MOLECULAR POLYMORPHISMS FOR PHYLOGENY, PEDIGREE AND POPULATION STRUCTURE STUDIES

by

YEAN WANG

UNIVERSITY OF SYDNEY

SCHOOL OF BIOLOGICAL SCIENCES

NSW, AUSTRALIA

A thesis submitted in the fulfillment of the requirements for the

degree of Doctor of Philosophy

December 2006

Declaration

This thesis is presented in fulfillment of the requirements for the degree of Doctor of Philosophy. I hereby declare that the work demonstrated in this thesis is my own unless indicated otherwise. No part of this work has been submitted for any degree or diploma. The microsatellite isolation in *Bactrocera tryoni* reported in Chapter 6 of this thesis was done jointly with MSc student Hong Yu as part of her project. This thesis may be made available for reference from the university library and may be photocopied or loaned to other libraries.

Yean Wang
December 2006

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Acknowledgements

This project proceeded with the support and care of my supervisors Professor Chris Gillies and Professor Chris Moran who have generously dedicated their time and expertise. I received substantial instruction and advice from Dr. Kathie Raphael, Dr. Marianne Frommer and Dr. John Sved of the Fruit Fly Research Centre. I have completed this thesis due to the patience and help from all of them. Especially when I was anxious, depressed, and impatient, they gave me enthusiastic and warm encouragement, and helped me to overcome those difficulties.

I am grateful to my external supervisor Professor Nelson Freimer of UCLA who has provided all the laboratory facilities and services for my research on green monkeys at UCLA.

Many thanks to Dr. Stuart Gilchrist for giving me strategic advice for the project. I would like to thank Dr. Cathy Turney and staff of the Fruit Fly Research Centre for technical advice and practical support; Dr. Deborah Shearman for providing the competent *Escherichia coli* DH5α cells used in transformation of recombinant plasmid DNA; and Sasha Curthoys for assistance with species morphological identification.

I thank Dr. Lynn Fairbanks of the Neuropsychiatric Institute at UCLA for giving me the opportunity for green monkey research; Matthew Jorgensen and other members of the UCLA-Sepulveda vervet monkey colony who have provided green monkey blood

samples; Dr. Roel Ophoff and his colleagues who has provided primer prescreening and Robot & MultiPROB II techniques; Dr. Charles Glatt and other members of Freimer lab who have lent facilities and technical support.

Abstract

A number of types of molecular polymorphisms can be used for studying genetic relationship and evolutionary history. Microsatellites are hypervariable and can be very useful tools to determine population structure, distinguish sibling species, as well as verifying parental relationships and pedigrees. However, while microsatellite polymorphisms are useful for solving relationships between populations within a species, relations among species or genera will probably be obscured due to a high degree of homoplasy —identity arising from evolutionary convergence not by descent. For long range evolutionary history, such as phylogeny from old world monkey to human, mtDNA markers may be better candidates. The aim of this thesis is to assess molecular polymorphisms of different types and their optimal use in different situations. Two widely separated taxa were used for testing –the green monkey *Chlorocebus sabaeus*, and the sibling dipteran flies *Bactrocera tryoni* and *B. neohumeralis*, known collectively as the Queensland fruit fly.

In the present study a complete 16,550 bp mtDNA sequence of the green monkey *Chlorocebus sabaeus* is reported for the fist time and has been annotated (Chapter 2). Knowledge of the mtDNA genome contributes not only to identification of large scale single nucleotide polymorphisms (SNPs) (Chapter 4) or other mtDNA polymorphisms development, but also to primate phylogenetic and evolutionary study (Chapter 3). Microsatellites used for the green monkey paternity and pedigree studies were developed by cross-amplification using human primers (Chapter 5). For studies of population

structure and species discrimination in Queensland fruit fly (Chapter 7), microsatellites were isolated from a genomic library of *Bactrocera tryoni* (Chapter 6)

The total length of 16550 bp of complete mtDNA of the green monkey *C. sabaeus*, which has been sequenced and annotated here, adds a new node to the primate phylogenetic tree, and creates great opportunity for SNP marker development. The heteroplasmic region was cloned and five different sequences from a single individual were obtained; the implication of this are discussed. The phylogenetic tree reconstructed using the complete mtDNA sequence of *C. sabaeus* and other primates was used to solve controversial taxonomic status of *C. sabaeus*. Phylogenies of primate evolution using different genes from mtDNA are discussed. Primate evolutionary trees using different substitution types are compared and the phylogenetic trees constructed using transversions for the complete mtDNA were found close to preconceived expectations than those with transversions + transitions. The sequence of *C. sabaeus* 12SrRNA reported here agrees with the one published by ven der Kuyl *et al*. (1996), but additional SNPs were identified. SNPs for other regions of mtDNA were explored using dHPLC. Twenty two PCR segments for 96 individuals were tested by dHPLC. Fifty five SNPs were found and 10 haplogroups were established.

Microsatellite markers were used to construct a genealogy for a colony of green monkeys (*C. sabaeus*) in the UCLA Vervet Monkey Research Colony. Sixteen microsatellites cross-amplified from human primers were used to conduct paternity analysis and pedigree construction. Seventy-eight out of 417 offspring were assigned paternity

successfully. The low success rate is attributed to a certain proportion of mismatches between mothers and offspring; the fact that not all candidate fathers were sampled, the limitations of microsatellite polymorphisms; and weakness of the exclusion method for paternity assessment. Due to the low success rate, the pedigree is split into a few small ones. In a complicated pedigree composed of 75 animals and up to four generations with multiple links a power male mated with 8 females and contributed 10 offspring to the pedigree. Close inbreeding was avoided.

Population structure within two species of Queensland fruit fly *Bactrocera tryoni* and *Bactrocera neohumeralis* (Tephritidae: Diptera) is examined using microsatellite polymorphisms. Queensland fruit flies *B. tryoni* and *B. neohumeralis* are sympatric sibling species that have similar morphological and ecological features. They even share polymorphism at the molecular level. Mating time difference is the main mechanism by which they maintain separate species. In the present study, 22 polymorphic and scorable microsatellites were isolated from *B. tryoni* and tested in the two species sampled from sympatric distribution areas. Pairwise genetic distance analysis showed explicit differentiation in allele frequencies between the two species, but very weak differences between conspecific populations. Gene flow is higher within *B. tryoni* than within *B. neohumeralis*, and gene exchange between the two species exists. An averaging linkage clustering tree constructed by UPGMA showed two major clusters distinguishing the two species, and it appears that population structure is highly correlated with geographic distance.

The relationship between molecular markers, evolution, and selection are discussed using comparative studies within two large taxa: primate and insect. The degree of conservation and polymorphism in microsatellites varies between taxa, over evolutionary time.

Abbreviations

A	adenine
aa	amino acid
ABC	ATP-binding cassette
ABCG1	gene product of *white*
AFLPs	amplified fragment length polymorphisms
Ala	alanine
AMOVA	**A**nalysis of **Mo**lecular **Va**riance
Arg	arginine
Asn	asparagine
Asp	aspartic acid
ATP	adenosine-5'triphosphate
ATPase6	gene encoding subunits 6 of ATP synthase
ATPase8	gene encoding subunits 8 of ATP synthase
bp	base pair
BSA	bovine serum albumin
C	cytosine
CEPH	Centre d'Études de Polymorphisme Human
CIP Calf	Intestinal alkaline Phosphatase
COI-III	cytochrome c oxidase subunit I-III
cry	circadian gene *cryptochrome*
CSIRO	Commonwealth Scientific and Industrial Research Organization
Cys	cystine
cytb	cytochrome b
dATP	2'-deoxyadenosine-5'triphosphate
DBY	A Y chromosome gene in NRY region
dCTP	2'-deoxycytidine-5'triphosphate
DFFRY	A Y chromosome gene in NRY region
dGTP	2'-deoxyguanosine-5'triphosphate
dHPLC	denaturing High Performance Liquid Chromatography
DMSO	Dimethyl Sulphoxide
DNA	deoxyribonucleic acid
dNTP	mixture of dATP, dCTP, dGTP and dTTP
DOP	degenerate oligonucleotide primer
dTTP	2'-deoxythymiding-5'triphosphate
EDTA	ethylene-diamine-tetra-acetic acid
EST	expressed sequence tag
F	gene diversity estimator in F-statistics
FAM	6-carboxy-fluorescein, a blue fluorescent dye
Fct	correlation of alleles between groups
FFRC	Fruit Fly Research Centre
Fsc	correlation of alleles of different population in the same group
Fst	correlation of alleles within populations over all groups
G	guanine
Gln	glutamine

Glu	glutamic acid
Gly	glycin
HEX	hexachloro-6-carboxy-fluorescein, a blue fluorescent dye
His	histidine
HLA-B	major histocompatibility complex, class I, B
HVS 1-2	high variable segment 1-2
HWE	Hardy-Weinberg Equilibrium
Ile	isoleucine
indel	insertion/deletion
IPTG	isopropylthio-β -D-galactoside
ITS1-2	ribosomal internal transcribed spacer 1-2
IVF	in-vitro fertilization
kb	kilobase
LB	Luria-Bertani (media)
LB-AXI	L-Broth plates containing ampicillin, X-gal & IPTG
Leu	leucine
Lys	lysine
Mb	mega base
MCAs	most common alleles
MELAS	mitochondrial myopathy, encephalopathy, lactacidosis, and stroke
Met	methionine
MHC	human major histocompatibility complex
MRCA	most recent common ancestor
mtDNA	mitochondrial DNA
MYBP	million years before present
NADH1-6	subunits 1–6 of nicotinamide adenine dinucleotide dehydrogenase
NADH4L	subunits 4 of NADH transcribed on light strand
NED	a yellow fluorescent dye
NJ	Neighbor-Joining
NRY	non-recombining region of the human Y-chromosome
PCR	polymerase chain reaction
per	circadian gene *period*
Phe	phenylalanine
PNK	T4 polynucleotide kinase
Pro	proline
PVP	Polyvinylpyrrolidone
RAPD	randomly amplified polymorphic DNA
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
RNase	ribonuclease
Rox	red size standard
rRNA	ribosomal RNA
SDS	sodium dodecyl sulfate
Ser	serine
SMCY	A Y chromosome gene in NRY region
SMM	single-step mutation model

SNPs	Single Nucleotide Polymorphisms
SOC	media contain tryptone, yeast extract, NaCl, $MgCl_2$, KCl, and glucose
SSCP	single stranded conformation polymorphism
SSLP	simple-sequence length polymorphism
SSPE	Sodium Phosphate EDTA
STR	short tandem repeats
STS	sequence tagged site
SUPAMAC	Sydney University and Prince Alfred Macromolecular Analysis Center
T	thymine
TAMRA	a fluorescent dye
TBHs	terminal branch haplotypes
TE	Tris-HCl and EDTA
Thr	threonine
TPM	two-phase model
Tris	tris (hydroxymethyl) aminomethane
tRNA	transfer RNA
Trp	tryptophan
Tyr	tyrosine
UCLA	University of California Los Angeles
UEPs	unique event polymorphisms
UPGMA	unweighted pair-group method with arithmetic mean
UTY	A Y chromosome gene in NRY region
Val	valine
VMRC	Vervet Monkey Research Colony
X-gal	5-bromo-4-chloro-3-indolyl-b-D-galactoside
YAP	Y Alu Polymorphism
YBP	years before present

Chapter 1

Introduction

The study of eukaryotic evolution has moved into the molecular biological era. Various polymorphisms of molecular markers and gene sequences have been used to investigate evolutionary scenarios, develop evolutionary theory, and construct evolutionary trees and phylogenetic affiliations within pedigrees, populations, species, and other taxonomic levels. For these purposes DNA polymorphisms in animals can be divided into two categories: mitochondrial DNA (mtDNA) and nuclear DNA. The latter can be further subdivided into autosomal (microsatellites are among the most important) and sex-linked classes.

Both the nuclear genome and the mitochondrial genome have coding and non-coding DNA sequences. The majority of the nuclear genome is composed of non-coding DNA sequences (Zane *et al*., 2002), while most of the mitochondrial genome consists of coding DNA sequences (Arnason *et al*., 1991, 1993, 1996a; Graf and Sparks, 2000). From comparison of the mutation rates in hypervariable microsatellites of the nuclear genome (mostly non-coding DNA) with highly variable control region (non-coding DNA) and coding region in mtDNA, and between genes in autosomal and sex-linked nuclear genome with mitochondrial DNA, it can be seen that the mitochondrial genome has a higher average mutation rate than the nuclear genome, and non-coding DNA sequences have higher average evolution rates than coding DNA sequences (Brown *et al*., 1979; Wallace, 1994). However, mutation rates are expected to vary according to specific lineage, effective population size, genomic region, type and size of polymorphisms (Loxdale and Lushai 1998; Ellegren, 2000; Pesole *et al*., 2001; Whittaker *et al*. 2003; Sainudiin *et al*. 2004).

When attempting to resolve evolutionary relationships using microsatellite polymorphisms, researchers find that the main difficulty is how to determine

unambiguously whether alleles observed to be identical in size are in fact identical by descent, and therefore identical in sequence. A high degree of homoplasy, identity arising from evolutionary convergence, not by descent, may exist in microsatellites. Size homoplasy causes an underestimate of genetic variability. Identical alleles can occur in different populations without shared ancestry. Population differences may be increasingly difficult to detect as the time of divergence between them increases. Because the shared alleles may not be identical by descent, but rather a product of the high mutation rate, or recurrent mutation.

Microsatellite markers are hypervariable and can be a very useful source of data for distinguishing sibling species, detecting population structure, determining parental relationship and pedigree construction (Brinkmann, 1998; Irvin, *et al*., 1998; Vaňková *et al*., 2001; Taylor *et al*., 2003; Wang *et al*., 2003). But for research of long-range evolutionary history, such as construction of a phylogeny from old world monkeys and humans, mtDNA polymorphism may be one of the better candidates (Graf and Sparks 2000; Arnason *et al*., 2002; Raaum *et al*., 2005). A review of different types of DNA polymorphisms will be a great help in deciding their importance in phylogenetic and evolutionary studies.

1.1 Review of DNA polymorphisms

1.1.1 Mitochondrial DNA

The mitochondrial genome of higher animals is a highly polymorphic, high-copy-number circular DNA, most of which is composed of coding sequences (Figure 1.1). Mitochondrial genes have been used extensively in evolutionary studies because of their uniparental mode of inheritance and high rate of mutation accumulation (Howell *et al*., 2003).

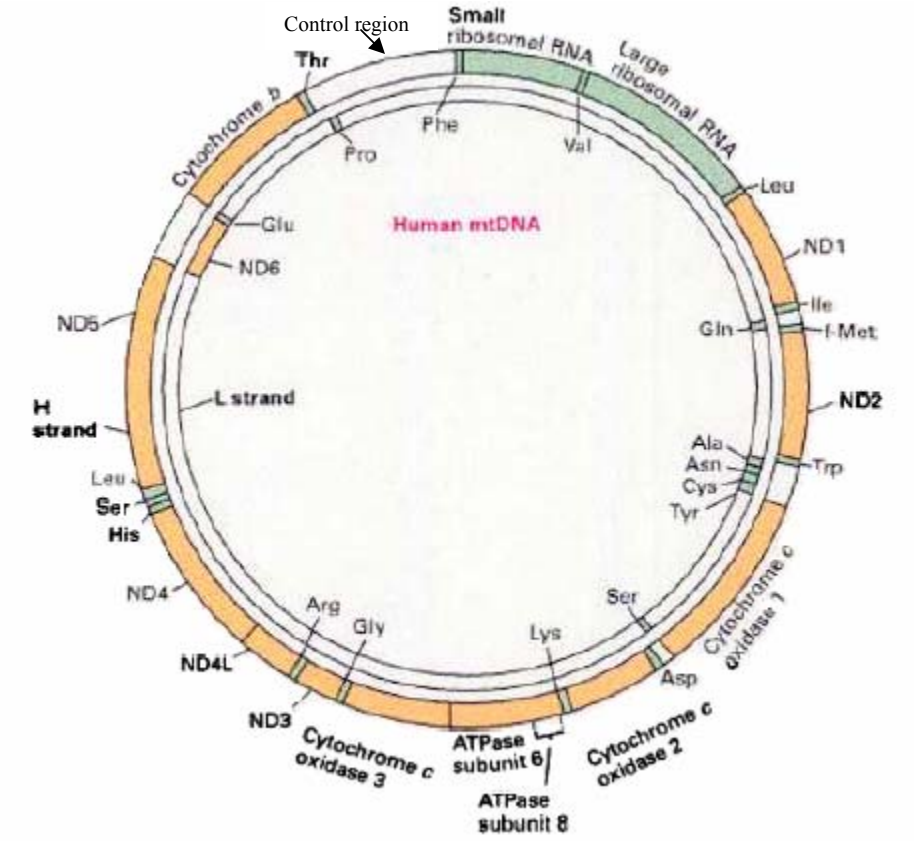


Figure 1.1 Human mtDNA showing heavy (outer circle) and light (inner circle) strands. For the gene designation please see Chapter 2. From Lodish *et al.* (1999)

Multiple copies of mtDNA are located in the within every mitochondrial organelle; several organelles are usually presented and inherited with the cytoplasm of the female gamete. New mutations in mtDNA segregate rapidly in the female germline under the effect of a genetic bottleneck in early oogenesis, even over a span of a single or a few generations (Howell *et al.*, 1996). Research has revealed that 2.8% of human mother–offspring pairs show single nucleotide differences in a 610 bp region of the control region (Parsons *et al.*, 1997). Sequence variants accumulate along maternal lineages without genetic recombination. When the transmission lineage was further traced by repeated backcrossing an F1 female from two strains with different mtDNA genotypes to males from the same strain, no paternal mtDNA was observed in moths (Lansman *et al.*, 1983), but a positive result for paternal mtDNA inheritance was seen in mice (Gyllensten *et al.*, 1991). Examples of paternal mtDNA inheritance are now known in *Drosophila* (Satta *et al.*, 1988; Kondo *et al.*, 1990), anchovy (Magoulas and Zouros, 1993) and mice

(Gyllensten *et al.,* 1991). It is assumed that a leaky contribution from the father (accounting for ~10^{-4} of an individual's mtDNA) may be the rule rather than an exception in animals (Zouros and Rand, 1998; Schwartz and Vissing, 2002; Jacobs, 2002; Williams, 2002). Selective proteolysis (hydrolytic break down of paternal mtDNA which is tagged with the universal proteolytic marker ubiquitin) usually ensures maternal inheritance. However, paternal inheritance and doubly uniparental-biparental inheritance (in which female offspring inherit the mtDNA from their mother, and male offspring inherit the mtDNA from both parents, leading to the phenomenon of heteroplasmy - the existence within an individual of two or more distinguishable orthologous mtDNA molecules), have been observed (Rokas *et al.,* 2003).

The mtDNA genome is a single completely linked entity, presenting in a circular molecule. Because mtDNA is usually believed not to recombine (but this is challenged by the suggestions of Eyre-Walker *et al.,*1999; Awadalla *et al.,*1999; Hagelberg, 2003 and Rokas *et al.,* 2003), genetic hitchhiking will affect all sites on the molecule. The frequency of an advantageous mutation increases, and its fixation may sweep all linked polymorphisms out of the population (Begun and Aquadro, 1992, 1993; Zouros and Rand, 1998). The population can only become polymorphic again by the accumulation of new mutations. Similarly, selection acting to maintain multiple alleles in the population will lead to their divergence, as happens in diversifying selection (Ballard and Kreitman, 1994). Most of the extant variation in mtDNA has been assumed to be neutral (Sigurðardóttir *et al*., 2000). However, comparisons between species of the ratio of replacement to silent nucleotide substitutions imply that the evolution of mtDNA is not strictly neutral. (Kilpatrick and Rand, 1995; Gerber *et al*., 2001; Rand, 2001).

In comparison to the nuclear genome, the mitochondrial genome is small (16,569 bp in humans), but has a very high mutation rate, up to 5-10 fold higher than nuclear DNA (Brown *et al.,* 1979, 1982). In the control region there is a mutation rate about 10 times higher than that of the coding region (Greenberg *et al.,* 1983), in a range of $9.5x10^{-1}$ to $1x10^{-6}$ nucleotides per million years (Howell *et al*., 2003). The substitution rate, estimated to be 0.19-3.2% per million years in the coding region (Giles *et al.,* 1980;

Wallace *et al.,* 1987; Wallace and Torroni, 1992), rises to 8.4% per million years in the control region (Greenberg *et al.,* 1983; Horai and Hayasaka, 1990; Vigilant *et al.,* 1989). The substitution rate is not uniform across the control region. The number of polymorphic sites is about double in the high variable segment 1 (HVS1) compared to the high variable segment 2 (HVS2), and between the two segments the substitution rate is site-dependent, with some extremely variable positions and some more constant sites (Figure 1.2) (Wakeley, 1993; Piercy *et al.,* 1993). Using the haplogroup method, it is possible to date human continental and ethnic expansions up to 30,000 years before present (YBP), and coalescence of human Paleolithic and Neolithic lineages up to 50,000 YBP can be detected (Kivisid *et al.,* 1999; Metspaly *et al.,* 1999; Richards *et al.,* 1996). Using the coding region, Arnason *et al.* (1998) dated divergences between primates in Cercopithecoidea and Hominoidea up to >50 MYBP. Merriwether *et al.* (1999) showed that the mtDNA variation (accounting for maternal lineages) is much more diverse than the Y chromosome variation (accounting for paternal lineages).

Figure 1.2 Structure of the mtDNA control region between tRNA-Pro and –Phe. HVS1 and HVS2 are high variable regions. Central domain and three conserved sequence boxes (black blocks) are conservative regions.

Nucleotide substitutions and indels (insertion/deletion) are common variations in a genome. Nucleotide substitutions in a coding region of a genome can be divided into non-synonymous substitutions (amino acid replacement), which are slowly occurring sequence changes; and synonymous substitution, which can be further divided into transversions (purine-pyrimidine changes), which are fair slowly occurring sequence changes; and transitions (purine-purine or pyrimidine-pyrimidine changes), which are rapidly occurring sequence changes (Mcdonald and Kreitmen, 1991; Wu *et al.,* 2000). In human, as well as in mouse and fly mtDNA genomes, a greater number of amino acid replacements were seen than expected based on interspecific comparisons (Nachman *et al.,* 1996). Wu *et al.* (2000) believed that the transversions and amino acid replacements contained lower levels of homoplasy and thus provided more accurate information on

cladistic relationships than transitions. According to their analysis, the phylogenetic tree based on transversions only in cytochrome *c* oxidase subunit I (COI) nucleotide sequences was less affected by homoplasy and therefore more informative than ones based on COI nucleotide sequences for all bases or for transitions only. A highly parsimonious tree of 266 amino acid (aa) replacements was created and it has a branching pattern that agrees with the previous molecular evidence on primate phylogeny (Wu *et al.,* 2000).

In addition to substitutions and indels, mtDNA heteroplasmy is common and has been reported in *Drosophila* (Stordeur, 1997; Kann *et al.,* 1998; Rand, 2001), fish (Broughton *et al*., 2001); birds (Moum and Bakke 2001); pigs (Ursing *et al*., 1998), mouse (Meirelles and Smith, 1997; Chinnery *et al.,* 2000), monkeys (Hayasaka *et al.,* 1991) and human (Grzybowski, 2000; Tully *et al.,* 1999) (see Chapter 2 for more details). A mouse model for heteroplasmic mtDNA segregation and transmission has been made (Jenuth *et al.,* 1996; Meirelles and Smith, 1997). Jenuth *et al.* (1997) investigated segregation of two different tissue-specific mtDNA genotypes in heteroplasmic mice and found that random genetic drift is in effect in some tissues, but strong, tissue-specific and age-related directional selection is very active in others. mtDNA heteroplasmy could be relevant to morphology and function heterogeneity of mitochondrial organelles within cells (Collins *et al*., 2002).

There are quite a few reports of mtDNA mutation caused diseases, including neuropsychoses, heart disease, respiratory dysfunction, and aging (Brown *et al.,* 1992; Lertrit *et al.,* 1994; Schwarze *et al.*, 1995; Lopez *et al.,* 2000).

1.1.2 Microsatellites

A microsatellite is a DNA motif of tandemly repeated units of one to six base pairs. Microsatellite markers, especially non-trinucleotide repeat motifs are predominantly non-coding loci (Hancock, 1995). Some studies indicate that microsatellites probably play a significant role in gene regulation (Kashi *et al.,* 1997, Harr, 2000).

Microsatellites are widely dispersed throughout eukaryotic genomes (Figure 1.3), occurring at 0.20-0.44 and 0.21-0.34 microsatellites per kilobase (kb) in human and *Drosophila melanogaster* genomes, respectively (Harr and Schlötterer, 2000; Katti *et al.,* 2001). Some researchers report that dinucleotide repeat loci have been characterized with such abundance that they occur once every 30 kb or less in human genome, an average of once every 20-30 kb in mammals (Stallings *et al*. 1991; Deka, 1995), and every 60 kb in the euchromatic *D. melanogaster* genome (Schug *et al.,* 1998a). In exons, trinucleotide repeats are abundant in all taxa surveyed by Tóth *et al.* (2000). In primates, mononucleotide repeats are the most common type; di- and tetra-nucleotide repeats are the next most abundant, especially in introns and intergenic regions. In arthropods, di-nucleotide repeats are the most abundant, especially in introns and intergenic regions (Tóth *et al.,* 2000). It has been shown that the abundance of specific dinucleotide microsatellite motifs differs between species. While in mammals and *Drosophila* CA is the most common dinucleotide (Beckmann and Weber, 1992; Schug *et al.,* 1998a), in plants, TA dinucleotides are the most abundant (Langercrantz *et al.,* 1993); the most frequent dinucleotide in *Caenorhabditis elegans* is GA (Schlötterer, 2000).

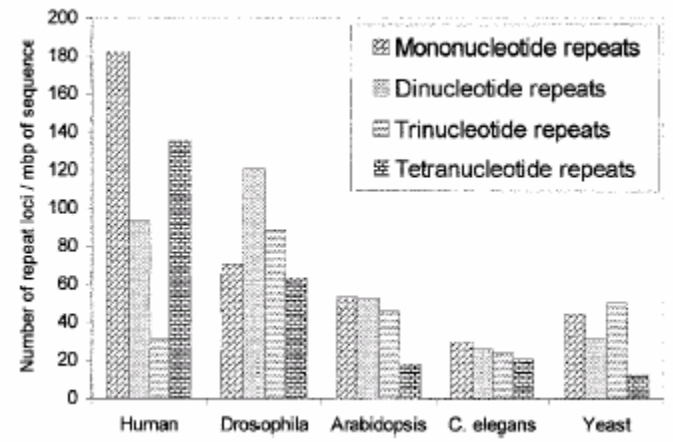


Figure 1.3 Frequencies of microsatellite repeat loci per million base pairs of chromosome sequences in different genomes. From Katti *et al.* (2001)

Microsatellites display relatively higher mutation rates compared to the eukaryotic substitutes average of around 10^{-9} mutation per nucleotide per generation (Ellegren, 2000). Mean microsatellite mean mutation rates may be as high as ~$2x10^{-3}$ per meiosis in human tetranuceleotide repeats (Ellegren, 2000), or as low as $4.87x10^{-5}$ to $3.76x10^{-5}$ per locus per generation (Sainudiin *et al*. 2004). The general microsatellite change rate is between the order of 10^{-3} and 10^{-4} per gamete per generation in humans and mice (Dallas, 1992; Jeffreys *et al.,* 1988; Weber and Wong, 1993; Zhivotosky and Feldman, 1995; Ellegren, 2000; Whittaker *et al*. 2003), $2.1x10^{-3}$ in human Y chromosomal microsatellites (Heyer *et al.,* 1997), and $6.3x10^{-6}$ to $9.3x10^{-6}$ for dinucleotides in *D. melanogaster* (Schug 1998b; Schlötterer *et al.,* 1998; Vazquez *et al.,* 2000). A growing amount of information on the microsatellite mutation rate is reported (Weber & Wong 1993, Di Rienzo *et al.,* 1994; Sainudiin *et al*. 2004). At most times, rates of around 10^{-4} changes per locus per meiosis in mammals are assumed (Weber & Wong 1993), but information about rate variation among microsatellites remains largely unknown (Bertranpetit and Calafell, 1996).

Replication slippage is the predominant mutational mechanism (Figure 1.4). However, certain phenomena of microsatellites, especially certain trinucleotide repeats associating with dramatic length expansion (Caskey *et al.,* 1992. Kunst & Warren 1994, Imbert *et al.,* 1993), could be explained by unequal recombination models in addition to slippage models (Tóth *et al.,* 2000). Individual loci probably start from regions of high cryptic simplicity (Tautz *et al.,* 1986; Messier *et al.,* 1996), and expand due to replication slippage, with a low initial rate. The mutation rate accelerates as heterozygosity increases, creating an interaction loop in which both allele length and heterozygosity hasten towards some upper stability threshold (Schlötterer and Tautz, 1992). *In vitro* experiments suggest that the initial slippage frequency must be much higher than the rate of observed microsatellite mutations (Schlötterer and Tautz, 1992). The DNA mismatch repair system recognizes and removes primary DNA slippage mutations effectively, and produces much lower observed mutation rates (Eisen, 1999). In yeast lacking a functional mismatch repair system up to a 6,000-fold increase in microsatellite mutation rates was found (Sia *et al.,* 1997). The single-step mutation model (SMM), where mutation

introduce a gain or loss of a single repeat, or the two-phase model (TPM), where mutation introduces a gain or loss of single and multiple repeat, are thought to be the main mutation model, but deviations from these models are reported (Palsbøll *et al.*, 1999).

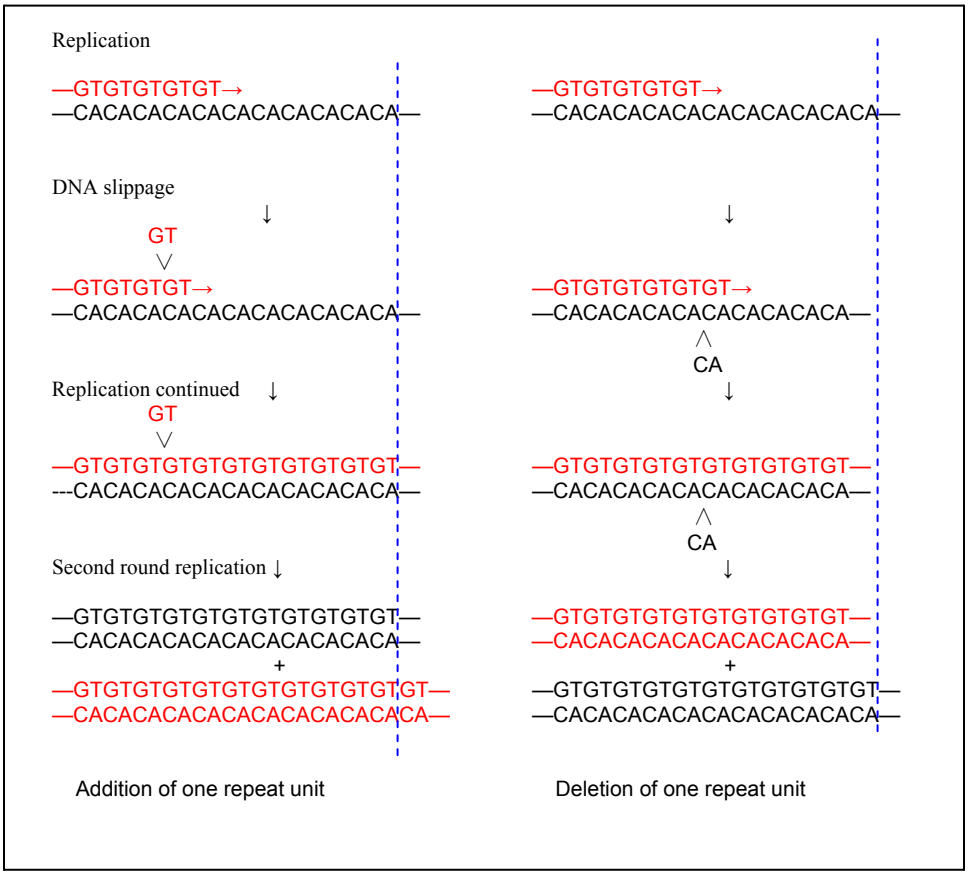


Figure 1.4 Model of DNA slippage adding (left part of the Figure) or removing (right part of the Figure) one repeat unit. Motifs in red are new replicons from original template. ⋁ and ⋀ indicate the position where a two nucleotide GT or CA loop formed.

The factors affecting mutation rate include repeat number, the type of repeat motif, and the nature of flanking genes and position in the chromosome. Several studies found a positive relationship between repeat number and microsatellite mutation rates (Schug *et al.*, 1998b; Kruglyak *et al.*, 1998) and an inverse relationship between repeat motif length and microsatellite mutation rates (Chakraborty *et al.*, 1997; Kruglyak *et al.*, 1998; Schug *et al.*, 1998b). Using a Markov chain model Lee *et al.* (1999) found that dinucleotide

repeats had the highest mutation rate, followed by tri- then tetranucleotide repeats. Their results are consistent with other determined mutation rates from experimental and natural data. Pupko and Graur (1999) compared the observed and expected frequencies of occurrence of microsatellites in the yeast genome. Their results indicated that the mutation rate of a microsatellite is more dependent on length of repeat array than number of repeats, or length of repeat unit. This means that a microsatellite with six repeats of trinucleotides has the same probability of mutation as a microsatellite with nine repeats of dinucleotides. Recently, Bachtrog *et al* (2000) provided evidence from *D. melanogaste*r for significant influence of base composition of repeat motif on the mutation rate of microsatellites. They found that CA repeats have the highest mutation rate, followed by AG. The lowest mutation rate was observed for AT microsatellites. Adding to the complication, selection acts differently at different chromosome positions, and the hitchhiking effect links a particular microsatellite allele tightly to a flanking gene.

Kruglyak *et al.* (1998) compared the length distribution of microsatellites in 1 mega bases (Mb) of nonredundant sequence in human, mice, yeast, and *Drosophila*. The length distribution differed, with long microsatellites being more frequent in humans. The insect *D. melanogaster* is a species with significantly shorter average microsatellite loci and a lower proportion of long microsatellites (Figure 1.5) (Kruglyak *et al.,* 1998; Harr and Schlötterer 2000). Accordingly, a comparison to mammals has shown that mutation rates of *D. melanogaster* microsatellites are substantially lower (Schug *et al.,* 1997, 1998a 1998b; Kruglyak *et al.,* 1998; Schlötterer *et al.,* 1998). Long microsatellites have high mutation rates in natural populations of *D. melanogaster* **(**Harr and Schlötterer 2000). This may be a reasonable explanation as to why the number of repeats decreases exponentially with repeat length (Katti *et al.,* 2001). Selection may act as an upper truncating mechanism, imposing a ceiling on alleles with large repeat counts (Feldman *et al.,* 1997; Samadi *et al.,* 1998). But microsatellite length can reach a stationary length distribution without an upper limit (Kruglyak *et al.,* 1998). The locus then either stabilizes or degenerates into the background sequence by acquisition of point mutations (Hancock, 1999). So mutation in extremely long alleles (more than 25 repeats in yeast and more than 15 repeats in *Drosophila*) is significantly biased toward a lower number of

repeats (Wierdl *et al.,* 1997; Harr and Schlötterer 2000). Inter- and intraspecific analysis for exceptionally long microsatellite alleles revealed a lack of old long alleles. Thus long microsatellite alleles arose in *D. melanogaster* only very recently, because long microsatellite alleles have short persistence times (Harr and Schlötterer 2000). Some researchers suggest that a simple difference in slippage rates, not selective constraints on length, is the cause for the different microsatellite length distributions among organisms and repeat motifs (Kruglyak *et al.,* 1998).

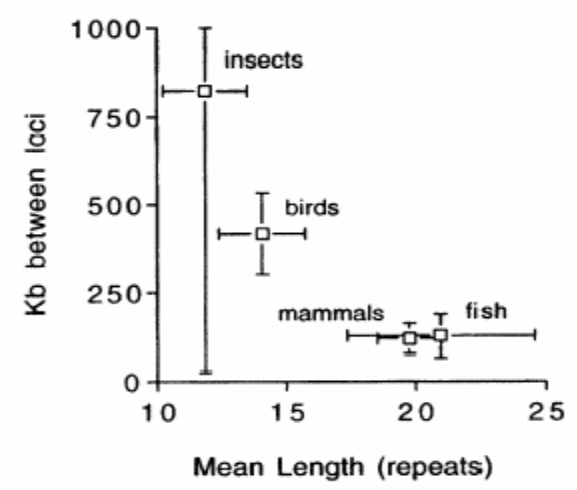


Figure 1.5 Range and mean repeat number plot for microsatellite loci across a wide range of organisms collected from 56 studies. From Amos (1999)

Microsatellites cloned from humans were, on average, longer than their chimpanzee homologues. Rubinstein *et al.* (1995a, b) suggested that this could be a result of population size, directional evolution, or a result of bias for longer repeat arrays during the isolation procedure. Ellegren *et al.* (1995) point out that most microsatellites are cloned from abundant species, and suggest that it is the cloning process, not population size differences and directional evolution that cause the observed length differences. Interspecific comparisons of microsatellite loci have frequently shown that the loci are longer and more variable in the species from which they are derived (the focal species), when compared to homologous loci in other (non-focal) species (Hutter *et al.,* 1998). Nevertheless, the arguments cannot explain all the size variation in the microsatellites cloned from the human or chimpanzee genomes (Cooper *et al.,* 1998). Two reciprocal

studies were previously performed using microsatellites derived from cattle (*Bos taurus*) and sheep (*Ovis aries*). Consistent with the ascertainment bias hypothesis, Ellegren *et al.* (1997) isolated 27 microsatellite markers from both cattle and sheep and found longer and more variable microsatellites in the focal species. In contrast, using an independent collection of loci, a survey of 472 microsatellites identified from both sheep and cattle (Crawford *et al.,* 1998) found that microsatellites were longer in sheep than in cattle regardless of whether they were originally identified from sheep or cattle. A reciprocal study in two species of *Drosophila* that diverged about 2.3-3.4 million years before present (MYBP) showed no PCR fragment length differences between focal and nonfocal species, but the average perfect repeat unit length, heterozygosity, and number of alleles were significantly higher in the focal species than in the nonfocal species (Hutter *et al.,* 1998). However, additional evidence showed that loci tend to be longer in humans even when they were cloned from chimpanzees (Amos and Rubinsztein 1996; Cooper *et al.,* 1998). These observations are consistent with the directional evolution hypothesis (Rubinstein *et al.,* 1995a).

A few authors observed that microsatellite loci with low allele variances were more appropriate for distinguishing between human ethnic populations than loci with high variances (Jorde *et al.,* 1997; Brinkmann *et al.,* 1998; Guarino *et al.,* 1999). In the former there is less homoplasy.

1.1.3 The Y chromosome

The Y chromosome is inherited from only one parent, but in contrast to mtDNA, it is paternally inherited. Therefore it changes by the accumulation of sequential mutations along paternal radiating lineages. It differs from mtDNA by the size, ~60 Mb in humans; a low average mutation rate, which remains undetermined; and the sex of the transmitting parent: male. The nonrecombining portion of the Y chromosome (NRY), which consists of ~30 Mb of linked DNA sequence in human, appears to be less variable than autosomal chromosomes. However, many polymorphisms have been discovered on the Y chromosome and it has rapidly emerged as a major source of information for

phylogeographic research (Richards and Macaulay, 2001). The most frequently used markers on the Y chromosome are biallelic loci, with the polymorphism limited to two alleles at a locus, and microsatellite markers. The former usually has a low mutation rate, although quite high mutation rates have also been reported (Jobling *et al.,* 1996). This type of polymorphism is suitable for detection of deep evolutionary branches. Markers such as single nucleotide polymorphisms (SNPs) and unique event polymorphisms (UEPs) are sometimes used to describe similar sets of biallelic Y chromosomal polymorphisms (Tyler-Smith, 1999).

Y chromosome microsatellites have higher mutation rates than biallelic markers and can be used to study recent evolutionary history, such as a human population migration history in Finland of around 2000 YBP (Kittles *et al.,* 1999; Nevanlinna, 1972), although some Y chromosome microsatellite genetic "signatures" can be detected up to 50,000 YBP (de Knijff *et al.,* 1997). Heyer *et al.* (1997) concluded that microsatellites on the Y chromosome have mutation frequencies comparable to those on the autosomes, after comparing variation in tetranucleotide microsatellites of the Y chromosome to that of autosomes. Their results supply additional evidence for the hypothesis that microsatellite variability is attributed to slippage-generated force. Using haplotypes deduced from these markers, and reasonable assumptions about the rate at which different types of mutations occur, the date for the most recent common ancestor (MRCA) of human Y chromosome can be estimated. Hammer (1995) sequenced a 2400-bp segment from the same Y chromosome region in 16 ethnically diverse humans plus four chimpanzees. He was able to calculate the date of the common ancestor of modern human Y chromosome as 188,000 YBP, with 95% confidence limits from 51,000 to 411,000 years. Shen *et al*. (2000) estimated continental and ethnic human nucleotide diversity of 2.45×10^{-5} to 8.54×10^{-5} for the coding regions of NRY genes *DBY, SMCY*, *DFFRY*, and *UTY1*, and an average of 9.16×10^{-5} to 14.2×10^{-5} for non-coding segments of these genes. In other research, Thomson *et al.* (2000) estimated a single nucleotide substitution rate of 1.24×10^{-9} per site per year from three NRY genes: *SMCY*, *DBY*, and *DFFRY*.

Both the Y chromosome and the mtDNA provide essential data, important in understanding the difference between human male and female migration and their contribution to present day distribution of people.

1.1.4 Other molecular polymorphic markers

Protein sequence variation reflects nucleotide replacement in DNA, mostly but not limited to coding DNA molecules. Some non-coding DNA molecules are functional in transcriptional regulation, modulation and recombination (Stallings *et al*., 1991; Meloni *et al*., 1998). DNA polymorphisms, such as single nucleotide polymorphisms (SNPs), restriction fragment length polymorphisms (RFLPs), simple-sequence length polymorphisms (SSLP), single stranded conformation polymorphisms (SSCP), amplified fragment length polymorphisms (AFLPs), random amplified polymorphic DNAs (RAPDs), sequence tagged sites (STS), and expressed sequence tags (EST) are extensions of markers for detection of classical sequence polymorphisms.

SSLP is another term for a microsatellite. SSCP (Orita *et al*. 1989) is a sensitive molecular analysis technique used for rapid detection of point mutation and DNA sequence polymorphisms. A STS is a unique, simple-copy segment of the genome whose DNA sequence is known and which can be amplified by specific PCR. When STS loci contain DNA length polymorphisms, such as SSLPs, they become valuable genetic markers (Olson *et al*., 1989). STSs can contain microsatellites, RFLPs and ESTs. ESTs are STSs developed from cDNA.

SNPs and RFLPs are two basic tools for detecting genomic variation in both mitochondrial and nuclear genomes. Both can be analyzed by their haplogroups (arrangement of groups according to haplotypes). SNPs are easier to characterize by their known nucleotide substitution pattern. SNPs have recently been recognized as the most widespread type of sequence variation (Brumfield *et al*. 2003) in either the nuclear or mtDNA genome, comparable with microsatellites that are used mostly in the nuclear genome. Both microsatellites and SNPs are valuable genetic markers for revealing

evolutionary history at different depths. The fact that SNPs occur throughout the genome makes them ideal for analyses of species differences, historical genealogies and all levels of evolution (Lewis, 2002; Brumfield *et al.* 2003). SNPs make the comparison of genomic diversities and histories of different species more straightforward than has been possible with microsatellites due to their lower variation and fewer alleles compared with microsatellites. The main complication in SNP analysis is a bias towards analyzing only the most variable loci. This bias is usually introduced by the limited number of individuals initially used to screen for polymorphisms, and it is correctable. The use of SNPs as markers in population analyses and phylogenetic studies is a new approach, but it is expected to become a widely used marker following the innovation of high throughput, accurate and cheap methods for SNP identification, inference, and statistical analysis. (Syvänen, 2001; Brumfield *et al.* 2003).

AFLPs and RAPDs typically are dominant markers, resulting in complicated allele distribution patterns, in which it is difficult to distinguish heterozygotes from homozygotes of the dominant allele, thus causing serious losing of information (Piepho and Koch, 2000; Ma *et al*. 2004). It has been reported that RAPDs were difficult to reproduce (Schierwater and Ender, 1993). Like microsatellites, these two types of markers are PCR based approaches. Unlike microsatellites, this group of loci involves various types of mutations such as insertions, deletions, SNPs, and others. They provide relatively little information for evolution rates and gene genealogies. Such loci are relevant to modern population geneticists only because of their vast number. They are random markers, and provide limited power for making inferences about molecular clock and kinship relationships.

All the animal genetic markers for detection of polymorphisms mentioned above are found in the two classes: mtDNA and nuclear DNA.

1.1.5 Sex specific molecular polymorphisms

Many population analyses use autosomal and Y chromosomal microsatellites, and mtDNA (Salem *et al.,* 1996; Seielstad *et al.,* 1998). Comparing microsatellites or SNPs between autosomes and the Y chromosomes in human and other primates, analysts tried to infer dynamic differences between male and female in the uses of specific modes of transmission (Seielstad, 2000). Binary polymorphisms associated with NRY and mtDNA preserve the paternal and maternal genetic legacies of extant species, allowing inference about human evolution, population affinity, and genealogical history. Differences in the reproductive output or migration rates of males and females will influence the geographic patterns and relative level of genetic diversity on the Y chromosome, autosomes and mtDNA (Seielstad, 2000). It was found that Y chromosome variants tend to be more localized geographically than those of mtDNA and autosomes (Seielstad *et al.,* 1994; Ruiz-Linares. *et al.,* 1996; Underhill *et al.,* 1996, 1997). In a recent study in Iceland the lineage-sorting process (descendants traced back to an individual ancestor) showed that the rate of evolution was markedly faster in matrilines (mtDNA) than in patrilines (Y chromosomes) (Helgason *et al.,* 2003). The fraction of variation within human populations for Y chromosome SNPs is 35.5%, versus 80–85% for the autosomes and mtDNA (Barbujani *et al.,* 1997; Excoffier *et al.,* 1992). The cause of the disparities could be transmission rate differences between male and female, shorter matrilineal generation interval, lower male transgeneratonal migration due to patrilocality, polygyny, or higher rate of mortality by natural selection in males (Seielstad *et al.,* 1998).

Only in cases of contemporaneous evolutionary histories, where mutation rates of different polymorphisms overlapped (Figure 1.6) can we contrast different migration or evolutionary rates between genders, by comparing sex-associated polymorphisms with others. There are a few good comparisons between paternally and maternally transmitted polymorphisms. Kittles *et al.* (1999) suggested a male-specific bottleneck in the Finnish human population signaled by a greater genetic distance among autosomal microsatellites and within the mtDNA control region, as compared to Y chromosome microsatellites. Shen *et al.* (2001) attribute the age discrepancy between NRY and mtDNA (estimated

92,000 and 230,000 years, respectively) in African humans to significant reduction in variability on NRY, or, possibly, demographic differences between males and females.

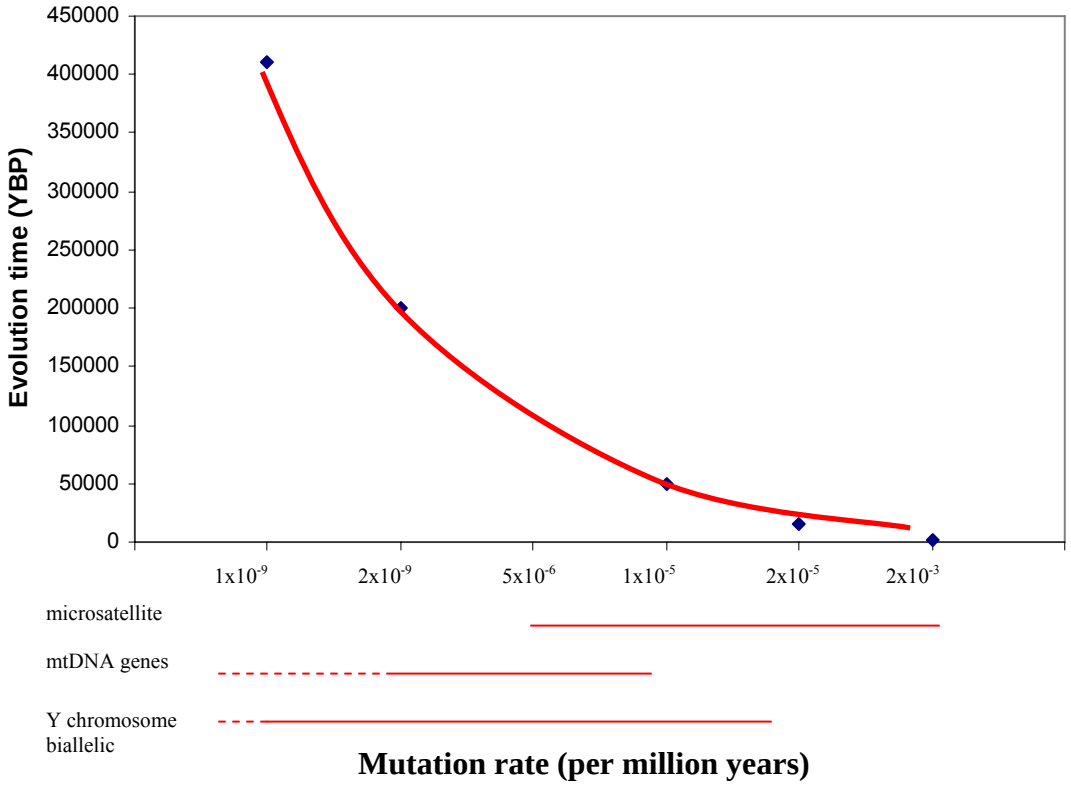


Figure 1.6 The relationship between mutation rate (per million years) (x-axis) and evolution time (YBP) (y-axis). The graph is drawn by best fit, according to direct and indirect data inferences from literatures mentioned throughout the chapter, and five deduced typical data sets for mutation rate (per million years) vs evolutionary time (YBP), [($2x10^{-3}$, 2000), ($1x10^{-5}$, 50000), ($2x10^{-5}$, 15000), ($1x10^{-9}$, 200000) and ($2x10^{-9}$, 411000)], were plotted as diamonds. Below the graph, the range of mutation rates for different polymorphisms is expressed by a solid line, with dashed line for possible extension of the rate range, according to reference literatures used for this graph.

Entomological population studies are quite different from human and primate population studies in dealing with sex-associated polymorphisms. For most insects, the discrepancy in evolutionary rate between different polymorphism types could be difficult to explain by transmission rate differences between male and female. In these cases, differences in mutation rates between markers and evolutionary depth could be used to explain the discrepancy, as discussed in Chapter 7 of the present study. However it will be worth

investigating the effect of some phenomena on sex-associated polymorphisms. Productivity of asexuality could be an interesting aspect. In insects, as well as other arthropods, some apomictic (asexually reproductive) parthenogenetic individuals appear in response to seasonal or environmental changes (Huigens *et al*., 2000). An example is the holocyclic aphid which has one sexual generation per year when the weather is getting cold, and numerous asexual ones throughout the rest of the year. Some species, such as *Drosophila*, parasitic *Trichogramma* and other insects become parthenogenetic or have their sex switched when they are infected by the microbe *Wolbachia pipientis* (Schilthuizen *et al.,* 1998; Stouthamer, 1997; Stouthamer *et al.,* 1999). Fruit fly *Drosophila melanogaster* geneticists should not overlook the effect on their studies for up to one third of their laboratory stock could be infected by this bacterium (Ainsworth, 2005). Sex-associated morphology can cause differentiation in migration rate, for example wingless aphids occur only as females in some species. Sex-associated polymorphisms may allow further insight into these phenomena.

1.1.6 Incongruencies between polymorphism types

Inferences and results derived from genetic analysis using molecular polymorphisms are affected by sampling features. The most important factors are sample size, sample number, and the number of polymorphisms tested. Inadequate sample sizes and/or polymorphism sizes are the main cause of discrepancy between and within polymorphism types (Rokas *et al*. 2003). Estimates from single genes may have large statistical errors, and use of multiple genes can allow a more reliable estimate of divergence time (Kumar and Hedges, 1998; Wilson *et al.,* 1977; Hedges *et al.,* 1996; Takezaki *et al.,* 1995). The use of different outgroups or calibrations can also cause inference discrepancies in analysis results. Arnason *et al.* (1996b), estimating from two non-primate calibration points using complete mtDNA and other genes with compatible evolutionary rates, proposed 6.1 MYBP for the date of the divergence between *Pan* (chimpanzee) and *Homo*. This estimate fits well with the age of the new hominid fossil *Sahelanthropus* (Brunet *et al.,* 2002) (6-7 MYBP). Previous estimates of the molecular timescales for divergence

between the human and chimpanzee lineages were 4.9 and 5.5 MYBP by Horai *et al.* (1995) and Kumar and Hedges (1998), respectively.

Examples of analysis discrepancy between nuclear and mtDNA polymorphisms are found in a range of organisms. Large genetic differences between populations of the American oyster was found using mtDNA (Reeb and Avise, 1990), but no significant geographic variation was detected using allozymes (Buroker, 1983) and some nuclear RFLPs (McDonald *et al.,* 1996). The harbor porpoise exhibits population subdivision with mtDNA, but not with microsatellite loci (Rosel *et al.,* 1999). Studies using nuclear polymorphisms showed low but statistically significant population structure in a seabird, the marbled murrelet, but no significant population structure was found in mtDNA (Friesen *et al.,* 1996, 1997). Slight but significant morphologic differences also were found between tree- and ground-nesting murrelets, but no variation was found when mtDNA was used (Pitocchelli *et al.,* 1995). A high degree of population structure in *Drosophila subobscura* exists between European populations for chromosomal arrangements (Prevosti *et al.,* 1988) and mtDNA but not for allozymes (Latorre *et al.,* 1992) or microsatellites (Pascual *et al.,* 2001). When a Y chromosome marker was used in human populations, it outperformed mitochondrial and other nuclear polymorphisms for detecting population structure (Semino *et al*., 1996; Richards *et al*., 1996). Y-linked short tandem repeats (STRs) make human African population structures more detectable than autosomal STRs (Seielstad, 2000). On a worldwide scale, Y-linked SNPs are more powerful than mitochondrial SNP for human population genetic analysis (Seielstad, 2000).

There have been very few comparisons of nuclear and mitochondrial polymorphisms in old world monkeys. Among them, Tosi *et al.* (2000) compared Y chromosome and mtDNA phylogenies of macaques and attributed the incongruence between "sex-biased" topologies to either Y chromosome introgression or mitochondrial differential lineage sorting.

Some inconsistencies exist between polymorphisms of the same type. This is because mutation rates may differ between specific loci. mtDNA, autosomal DNA, and the Y chromosome, coding and non-coding regions have different mutation rates. Even between regions of the same coding/non-coding type, selection imposes different effects on them and lineage sorting could lead to different inferences for different loci as well. There is a negative relationship between mutation rate and evolutionary depth (Figure 1.6).

1.2 The present study

1.2.1 Aims of the research

Nonhuman primates such as old world monkeys are closely related to humans, from an evolutionary point of view, and their colonies are extremely valuable in human disease related research. However, prerequisites for gene mapping and linkage studies are polymorphic molecular marker development and pedigree construction. A multigenerational old world monkey colony is a very valuable resource for these kinds of studies. However, when molecular polymorphisms are to be evaluated in a population and species structure study, naturally widely distributed taxa such as fruit flies are more helpful than captive or scattered animals. The present research aims to use appropriate molecular polymorphisms in studies on two distantly separated taxa (Insecta and Primates) to assay genetic and evolutionary issues in several biological levels: pedigree, populations, species and high level taxa. Similarities and differences in genetic relationship with several polymorphic markers of different inheritance types will be investigated in the present research.

1.2.2 mtDNA and phylogeny

Phylogenetic contributes to our understanding of genetic polymorphisms and population and species relationships. Medium to long range phylogenetic studies rely on polymorphisms with medium to slow mutation rates, such as mtDNA genes and binary polymorphisms on autosomal or sex linked chromosomes. Microsatellites and coding

regions of mtDNA are seldom used for long range historical studies due to the possibility of homoplasy.

The phylogenetic relationship of the old world monkey and modern humans has attracted many researchers. Complete mtDNA has already been sequenced for human and a few great apes, and their evolutionary relationships have been investigated by many authors (Horai *et al.,* 1995; Xu and Arnason, 1996a, b; Arnason *et al.,* 1996a, b). A phylogenetic study of the old world monkey tribe Papionini, composed of macaques, baboons, mandrills, drills, and mangabeys has been performed using the mitochondrial genes Cytb and COII (Disotell *et al.,* 1992). Raaum *et al* (2005) estimated Catarrhine primate divergence date from complete mtDNA sequence of representatives of two cercopithecoid subfamily, Cercopithecinae (*Chlorocebus aethiops*, *Macaca sylvanus* and *Papio hamadryas*) and Colobinae (*Colobus guereza* and *Trachypithecus obscurus*), together with that of hominoid species (humans and apes). *C. aethiops* is closely related to the green monkey *Chlorocebus sabaeus*, and sometimes both are called vervets.

The present aim is to determine the complete sequence of the mtDNA of the green monkeys, develop SNP markers on the mtDNA and conduct primate phylogenetic analysis for hominoid and Cercopithecinae species.

Three pairs of the sister species, two *Cercopithecus* species, two chimpanzee species and two *Macaca* species will be compared and that will add the most recent branches to the deeply developed tree in a primate phylogenetic study, and give an impression of primate genus divergence depth. As the genus *Cercopithecus* needs urgent taxonomic revision (Groves, 2001), the study of the two *Cercopithecus* species will be an incentive to get more complete mtDNA sequenced to facilitate taxonomic study in this genus.

1.2.3 Paternity testing and pedigree construction

An even shorter timescale than evolution among populations within species is that within a pedigree of a colony. Paternity analysis and pedigree reconstruction within a community is another application of microsatellites. Paternity research has mostly been done for higher animals, such as humans and other mammals (Pemberton *et al*., 1995; Foster *et al.,* 1998; Vaňková *et al.*, 2001; Blouin, 2003), because the parental and kinship relationships are easily traced.

African green monkeys (*Chlorocebus sabaeus*: Cercopithecoidea; an old world monkey) have been extensively studied in biology (Hall and Gartlan, 1965; Struhsaker, 1967; Isbell *et al.,* 1998a, b), molecular biology (Lucotte *et al.,* 1982; Castro *et al.,* 1996; Pero *et al.,* 2002), and medical sciences (Barnicot and Jewett-Emmett, 1971; Mc Dermid and Ananthakrishnan, 1972; Herbin *et al.,* 1997; Boire *et al.,* 2002). Paternity assessment and pedigree reconstruction using microsatellites is a preliminary step toward genetic mapping and further research in the green monkey. These researches would greatly facilitate research into the genetic basis of a number of important neurobehavioral and biological characteristics, offer a better understanding of the green monkey genome and provide basic knowledge for further investigation using the green monkey.

A colony of more than 500 green monkeys *C. sabaeus* in the Vervet Monkey Research Colony (VMRC) at the University of California Los Angeles (UCLA), USA, is an ideal subject for lineage research. The aim of this research is to develop parentage testing in the green monkey. Paternity assignment and pedigree reconstruction are carried out on these animals by using human primers to cross-amplify microsatellites.

1.2.4 Population analysis within species or between sibling species

Differentiation within species or between sibling species normally occurs over short evolutionary time scales. Hypervariable markers like microsatellites are commonly used for the study of species differences. Two species of Queensland fruit fly, *Bactrocera*

tryoni and *B. neohumeralis*, are closely related sympatric species. Species identity is maintained by a pre-mating isolation mechanism, i.e. their mating time difference. *B. tryoni* mates at dusk, like the majority of Tephritidae, but *B. neohumeralis* mates in the middle of the day (Smith, 1979). A study of the genetic difference of the two species provides an opportunity for an insight into issues related to sympatric speciation and provides necessary basic knowledge for pest control of the particular species. Information about adaptation of local populations to their environment is essential to our understanding of biodiversity and the processes of speciation; nevertheless, the genetic changes required for such local adaptations are poorly understood. Studies have shown that comparison of different populations for several loci can help to identify genomic regions carrying a mutation that results in a local adaptation. (Schlötterer, 2000).

The aim of the present study is to distinguish the sibling Queensland fruit flies *B. tryoni* and *B. neohumeralis*. Therefore, isolation of microsatellite markers and their use to analyze natural population structure of and genetic relationships between the two sympatric species will be studied.

1.3 An outline of the thesis

Different molecular polymorphisms can be used to study evolutionary history among taxa, kinship relationship within a high related colony, and genetic diversity between sibling species and populations. Proper analyses come from informed decisions about the polymorphism type and sample size to be used. This theme is addressed in the thesis by a review of case studies which use different strategies and organisms. Primates are favorite subjects for kinship and phylogenetic studies which use mtDNA or microsatellites cross-amplified using human primers. For studies which distinguish close related species over a wide range of sympatric zone, natural populations of fruit flies are better subjects for sampling and analysis using *de novo* developed microsatellite marker.

Chapter 2 reports the complete sequence of mtDNA of the green monkey *Chlorocebus sabaeus*. Comparisons of mtDNA within Cercopithecinae species, and among

Cercopithecinae and Hominidae species are made. Heteroplasmy in mitochondrial genome is discussed. Chapter 3 presents a general phylogenetic study for Hominidae and Cercopithecinae species using mtDNA, and evaluates different genes in the mtDNA for a phylogenetic study. Chapter 4 detects SNP sites along the green monkey mtDNA sequence, and gives a detailed explanation for an identification method for the St. Kitts green monkeys. In chapter 5, genealogical relationships within a colony of green monkey in UCLA Vervet Research Center is inferred using microsatellites developed by cross-amplification using human primers, and the possible cause of some errors is discussed. Chapter 6 shows the process of *de novo* isolation of the microsatellite from screening a partial genomic library of Queensland fruit fly *B. tryoni*, and analyses the possible linkage between these microsatellites and functional genes using blasting against sequences in GenBank. Chapter 7 describes the construction of structure within the two species of Queensland fruit fly *B. tryoni* and *B. neohumeralis* using these microsatellites. The effects of sample quantity and type of the polymorphisms are discussed. In Chapter 8, the relationship between molecular polymorphisms, evolution and selection, using comparative study for molecular conservation and the degree of molecular polymorphisms of the genomes of two large taxa: primates and insects are discussed.

Chapter 2

Complete mtDNA sequence of the green monkey

2.1 Introduction

Complete mitochondrial DNA (mtDNA) has been widely used in evolutionary and phylogenetic studies. mtDNA sequence length, base composition, gene order and start and stop codons vary among different taxa. Figure 2.1 illustrates the highly conserved gene arrangement of mammalian mtDNA relative to other vertebrates. These genes include ATP6 and ATP8, genes encoding subunits 6 and 8 of ATP synthase; COI-III, genes encoding subunits I–III of cytochrome c oxidase; Cytb, genes encoding cytochrome b; ND1-6, 4L, genes encoding subunits 1–6 and 4L of nicotinamide adenine dinucleotide dehydrogenase. Locations of two ribosomal RNA genes, srRNA which is 12S ribosomal RNA, and lrRNA which is 16S ribosomal RNA are also shown in Figure 2.1.

Figure 2.1 A comparison of mtDNA protein encoding genes and rRNA genes among human (*Homo sapiens*), gorilla (*Gorilla gorilla gorilla*), mouse (*Mus musculus molossinus*), donkey (*Equus asinus*), pig (*Sus scrofa*) wallaroo (*Macropus robustus*), quail (*Coturnix chinensis*), sea urchin (*Strongylocentrotus purpuratus*), ascidian (*Halocynthia roretzi*) and frog species (*Fejervarya limnocharis, Polypedatas magacaphalus* and *Rana nigromaculata*), shows highly conservative gene arrangement among mammalian mtDNA and various gene arrangements among frog species. For the gene designatum please see text above.

At the beginning of the present research on the green monkey *Chlorocebus sabaeus*, four genes from the mtDNA of a sibling subspecies *C. aethiops* had been sequenced: cytochrome c oxidase subunit II (COII) (Ruvolo *et al.*, 1991; Disotell *et al.*, 1992), cytochrome c oxidase subunit I (COI) (Wu *et al.*, 2000), 12S rRNA (van der Kuyl *et al.*, 1995, 1996), and cytochrome b (Cytb) (Kaminska *et al.*, 1997). However, it was clear that a complete mtDNA sequence of the green monkey *C. sabaeus* would be valuable, as it would fully characterize the mitochondrial genome of the species with precise

information on gene location, and could also be used to develop a large range of mtDNA markers for reliable evolutionary evaluation and other comparative studies. Therefore generating and annotating the complete mtDNA sequence of the green monkey, *C. sabaeus*, is the main aim of the work reported in this chapter.

2.2 Materials and Methods

2.2.1 DNA isolation

The genomic and mitochondrial DNA extraction procedure is described in detail in chapter 5. Briefly, all the green monkey DNA used in all chapters of the present research is from blood samples collected from the Vervet Monkey Research Colony (VMRC) at the University of California Los Angeles (UCLA), USA. The DNA was extracted either manually or automatically. The manual DNA isolation was done using a method modified from Puregene and Clotspin protocols provided by Gentra (Minneapolis, MN). Most of the automatic DNA extraction was done by the DNA Core (a molecular service lab) at UCLA.

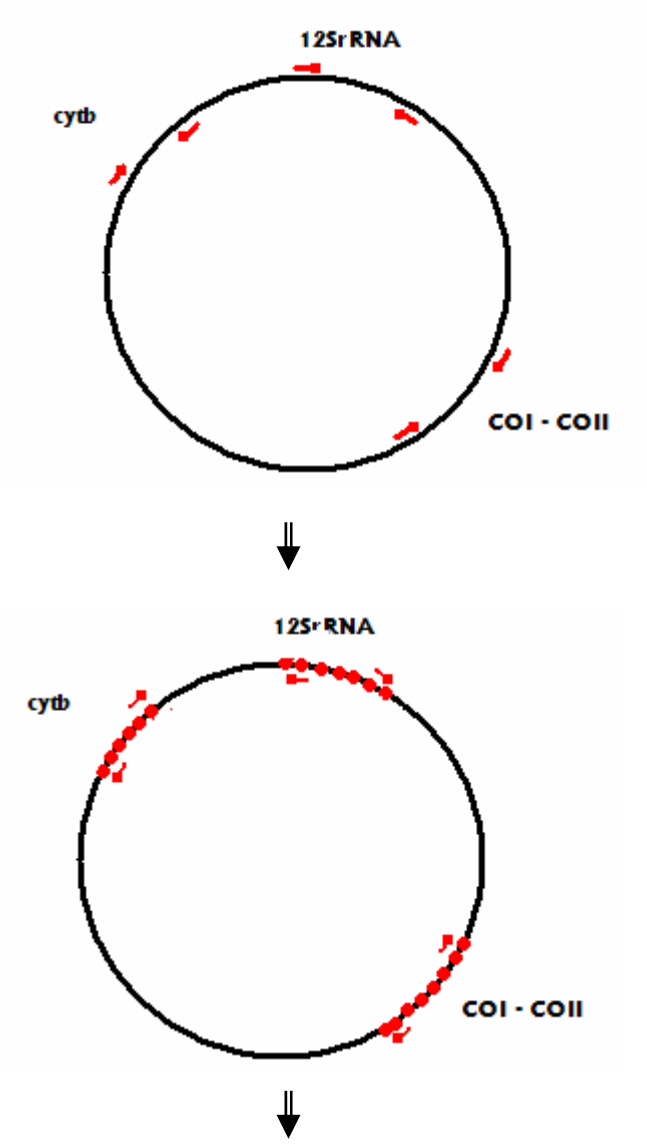
2.2.2 mtDNA sequencing

For direct PCR product sequencing, the animal 1993-072 was chosen as the reference sample. Each time when the reference sample was processed, several other samples were randomly chosen to be sequenced together with it. In the control region, there was a heteroplasmic region (several different sequences in the homologous region within the same individual), which was problematic when sequencing PCR products directly. This sequencing was done by cloning PCR products using sample 1992-086.

2.2.2.1 Normal and long PCR

mtDNA PCR products were sequenced mainly using primer walking (Figure 2.2). The first few primers were designed within the sequences of 12S rRNA, COI-COII and Cytb based on reference sequences of the sibling species *C. aethiops* (van der Kuyl *et al.* 1995,

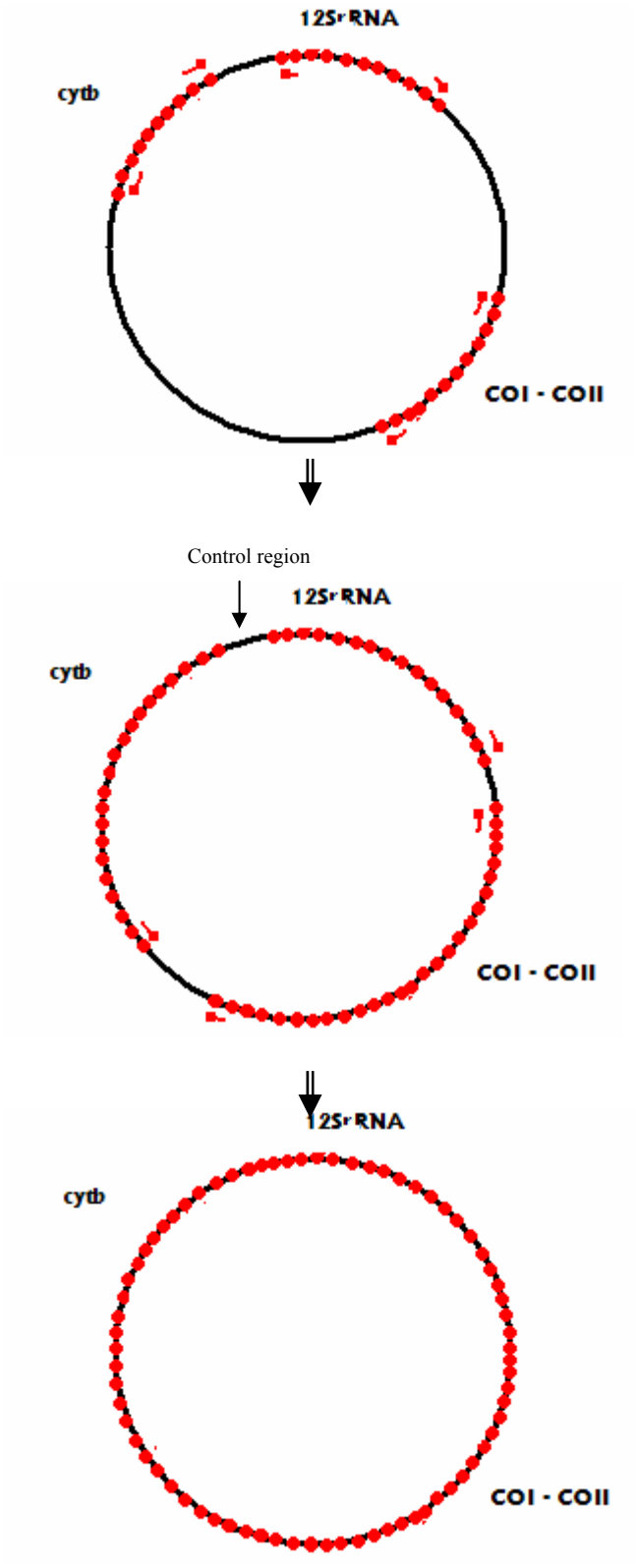
1996; Wu *et al.* 2000; Ruvolo *et al.*1991; Disotell *et al.* 1992). The PCR reaction mix of 20 µl usually contained PCR Master Mix (AmpliTaq Gold, 50U/ml) 12 µl, DNA (10ng/ µl) 6 µl and 1 µl of each primer (10 µM). The thermal cycle program comprised one cycle at 94°C for 1 min, followed by 30 cycles at 94°C for 30 sec, 50°C for 45 sec, 72°C for 50 sec. A final cycle was done at 72°C for 5 min.



Direct sequencing was started within three regions: 12S rRNA, COI-COII, and Cytb.

Long PCR products were made between these genes by designing appropriately oriented primers within them.

Figure 2.2 (continued over page)



The three sequences were extended by sequencing two ends of long PCR products. New primers were designed and the processes of PCR and sequencing were repeated.

Heteroplasmic segments were cloned in the control region while all gaps between the sequenced regions were filled.

A complete mtDNA was assembled

Figure 2.2 Flowchart illustrating method for generating the complete mtDNA sequence. Short lines with heads beside the DNA circle represent primers. Bead strand on the DNA circle represent sequenced portions.

PCR products longer than 5 kb were amplified using the protocol "Expand 20 kb PCR system" (Roche, Alameda CA) which includes a unique enzyme blend containing Taq and Pwo DNA Polymerase and Tgo, a thermostable enzyme with proofreading activity. Following the manufacturer's instructions, a 50 µl reaction mix included 2 solutions. Solution A contained 10 µl of buffer, 2 µl of each primer, 1.2 µl of DNA solution, 1 µl of DMSO, and 10.8 µl of ddH_2O. Solution B contained 0.7 µl of enzyme mix, 10 µl of dNTP mix, and 12.3 µl of ddH_2O. Solution B was added after solution A was heated at 92°C for 2 min. The thermal cycle program comprised one cycle at 92°C for 2 min; followed by 10 cycles at 92°C for 10 sec, 62°C for 30 sec, and 68°C for 18 min, then 30 cycles at 92°C for 10 sec, 62°C for 30 sec, and 68°C for from 18 min, extending by 10 sec for each cycle; a final extension was held at 68°C.

2.2.2.2 Cloning the heteroplasmic region

The heteroplasmic region by definition contains several different orthologous sequences in a single individual, and thus a heterogenous and unsequenceable PCR product will be amplified. It was therefore sequenced by cloning the PCR products (amplified by primers L34 and H560, see 2.3.1 and 2.3.2 of this chapter) using TOPO Reporter Kits (Invitrogen, Carlsbad CA) according to the manufacturer's instructions as follows.

MinElute Gel Extraction (QIAgen, Valencia CA): Each PCR product from the heteroplasmic region was loaded on a 1% agarose gel (1g agarose/100ml 1xTAE + 0.2 µg ethidium bromide) and run at 60v for 2 hours. The DNA fragment was excised from the agarose gel and weighed. Three volumes of Buffer QG were added to 1 volume of gel slice (100 mg ~ 100 µl). The tube was incubated at 50°C for 10 min (or until the gel slice was completely dissolved). The tube was vortexed every 2–3 min during the incubation. After the gel slice dissolved completely, the color of the mixture was checked to be sure it was yellow. If the color of the mixture was orange or violet, 10 µl of 3 M sodium acetate was added to adjust the color. Isopropanol equivalent to 1 gel volume was added and the tube was inverted several times. The sample was applied to the MinElute column,

in the provided 2 ml collection tube, and centrifuged at 13000 rpm for 1 min. The flow-through was discarded and the MinElute column was placed back in the same collection tube. 500 µl of Buffer QG was added to the spin column and centrifuged 13000 rpm for 1 min. The flow-through was discarded and the MinElute column was placed back in the same collection tube. 750 µl of Buffer PE, with pre-added ethanol as instructed, was added to the MinElute column and centrifuged for 1 min. The flow-through was discarded and the MinElute column was centrifuged for an additional 1 min at ≥10,000 x g (~13,000 rpm). The MinElute column was placed in a clean 1.5 ml microcentrifuge tube. 10 µl of Buffer EB (10 mM Tris·Cl, pH 8.5) or H_2O was added to the center of the membrane, the column was let to stand for 1 min, and then centrifuged for 1 min to elute the DNA into the microcentrifuge tube.

TOPO cloning (Invitrogen): DNA samples extracted from the gel were amplified using the same primer pair used before MinElute gel extraction. 4 µl fresh PCR product, 1 µl Salt Solution provided with the kit, and 1 µl TOPO vector were mixed gently and incubated for 5 minutes at room temperature (22-23°C) before being placed on ice. 2 µl of the TOPO cloning reaction was added to a vial of chemically competent *E. coli* provided. The vial was mixed gently, and incubated on ice for 5 to 30 minutes. The cells were heat-shocked for 30 seconds at 42°C without shaking, then immediately transferred to ice. 250 µl SOC medium at room temperature was added. The tube was capped tightly and shaken horizontally (200 rpm) at 37°C for 1 hour. From each transformation, 25-200 µl was taken to spread on a pre-warmed selective plate and incubated overnight at 37°C. ~10 colonies were taken for analysis.

QIAprep: 10 colonies were picked and cultured overnight in LB medium containing 50 µg/ml Carbenicillin (70 µl stock of 50 mg/ml for 70 ml LB). Cultures were centrifuged at 13,000 rpm at room temperature for 30 seconds to pellet the bacteria. Bacterial cells were resuspended in 250 µl Buffer P1, to which RNase A had been added as instructed, and transferred to a microcentrifuge tube. No cell clumps were visible after resuspension of the pellets. To each tube 250 µl Buffer P2 was added and the tube was gently inverted 4–6 times. If necessary, inversion was continued until the solution became viscous and

slightly clear, but this did not proceed more than 5 min. 350 µl Buffer N3 was added and the tube was gently inverted immediately and repeatedly, until the solution became cloudy. The tube was centrifuged for 10 min at maximum speed in a tabletop microcentrifuge. A compact white pellet formed in each tube. The supernatant was applied to the QIAprep columns. The flow-through was discarded after centrifugation for 30–60 sec. The QIAprep spin column was washed with 0.75 ml of Buffer PE, with pre-added ethanol as instructed, and centrifuged for 30–60 sec. The flow-through was discarded and the tube was centrifuged for an additional 1 min to remove residual wash buffer. The QIAprep column was placed in a clean 1.5 ml microcentrifuge tube. 50 µl Buffer EB (10 mM Tris·Cl, pH 8.5) or water was added to the center of each QIAprep column. The column was let to stand for 1 min, and then centrifuged for 1 min to elute QIAprep DNA.

Verification of presence of target DNA: The resulting QIAprep DNA was used as template in PCR (see 2.2.2.1 for normal PCR protocol) using primer L34 and H560 (see Table 2.1). The PCR product was loaded in 2% agarose gel (2 g agarose/100ml x1 TAE + 0.2 µg ethidium bromide) and run at 100 V for 1 hour. The fluorescent band was detected under UV illumination.

2.2.2.3 Sequencing

Purification of PCR products: 2 µl shrimp alkaline phosphatase (1 U/µl) was added to a 4 µl PCR product, (either from direct PCR products or cloned templates) to remove phosphate groups from the 5'-ends of DNA to prevent re-ligation of linearized PCR products. It also dephosphorylates dNTPs which allows for efficient removal of unincorporated nucleotides from PCR products prior to DNA sequencing. 0.2 µl exonuclease I was added to catalyze the removal of nucleotides from single-stranded DNA in the 3' to 5' direction. Finally 6 µl ddH_2O were added. The mixture was incubated at 37^oC for 60min, then 80^oC for 15min.

For sequencing, purified PCR products were submitted with both forward and reverse primers to the sequencing Core (a molecular service lab at the UCLA), which uses the

ABI3700 system. Each submission included a sample from the reference monkey 1993-072 and samples from several randomly selected animals. Each reaction mixture included 2 µl BigDye Terminator Mix, 2 µl HalfBigDye (Sigma, St. Louis), 1 µl Primer (3 µM), several µl Template depending on size (<200 bp ~20ng, 200-300 bp ~50ng, and >300 bp ~90ng), and ddH_2O to make up a 10 µl total volume. The Cycle Sequencing profile comprised an 8 min denaturing step at 95°C; followed by 25 cycles of 10 sec at 95°C, 5 sec at 50°C, and 4 min at 60°C; after the final cycle the sample was held at 4°C.

The resulting sequence chromatograms were aligned, assembled and analyzed using the Sequencher program (Gene Codes Corporation) or aligned manually if needed. All genes were identified by aligning the assembled sequence with human and chimpanzee sequences (Arnason *et al*. 1996a).

2.3 Results

2.3.1 Complete mtDNA sequence

The complete mtDNA of the green monkey *C. sabaeus* was sequenced and deposited into GenBank (Accession number: DQ069713). The primers used for sequencing are listed and described in Table 2.1, and their annealing temperatures (Tm) were checked using Amplify 2.5Br software (Engels 1993, 1994). The arrangement of the 13 mRNA, 2 rRNA and 22 tRNA genes and the control region is the same as in human mtDNA (Figure. 2.3). This arrangement is conserved in all mammals whose mtDNA has been sequenced so far (Akimoto *et al*., 2005; Anderson *et a*l., 1981; Janke *et al*., 1997; Ursing and Arnason, 1998; Xu *at al*., 1996; Xu and Arnason, 1996a) (Figure 2.1). The initial site, end site and size in nucleotides for each gene are listed in Table 2.2. The length of total mtDNA sequenced is 16550 bp. The L-strand (light strand) has a composition of A: 32%, C: 30%, G: 13% and T: 25%.

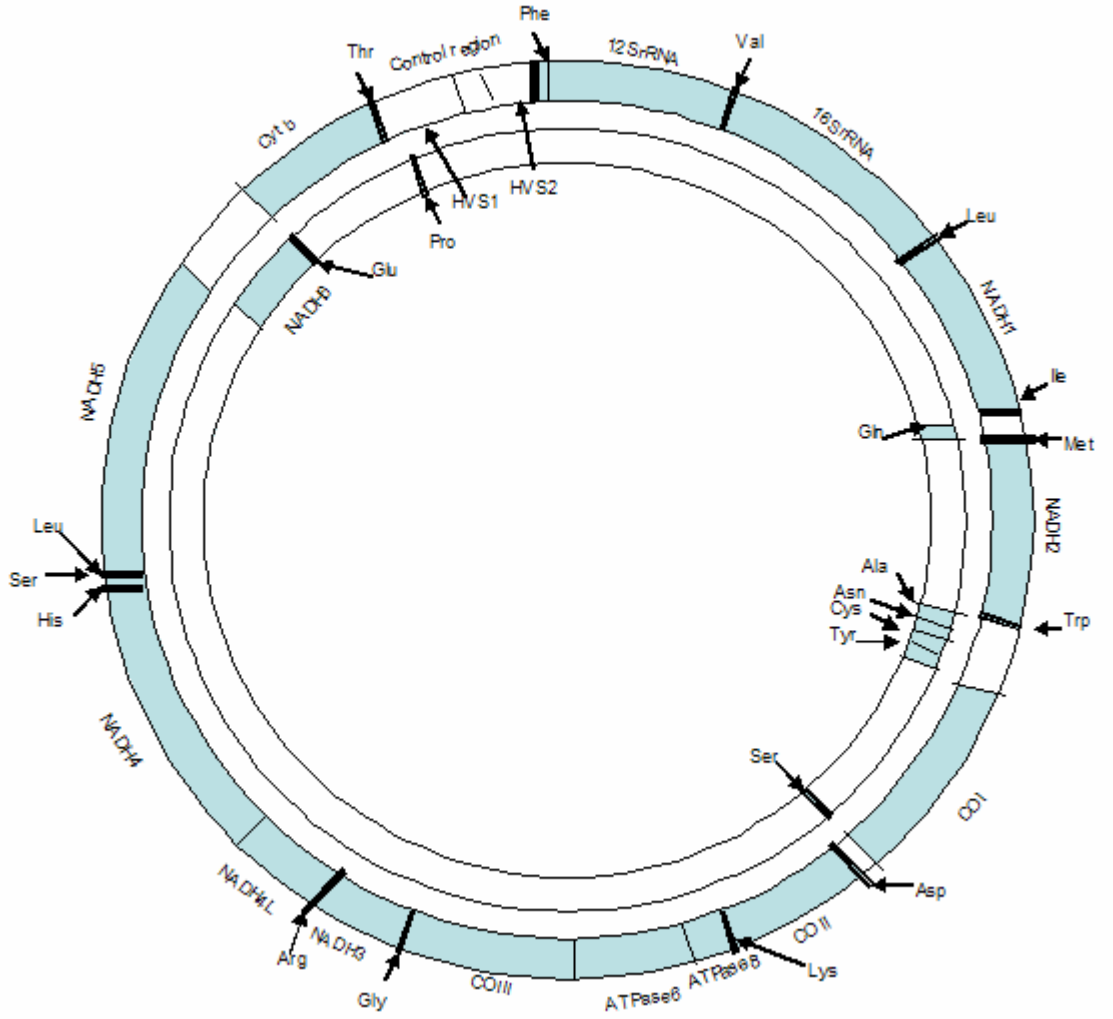


Figure 2.3 A schematic representation of the mitochondrial genome of the green monkey *Chlorocebus sabaeus*. Scaling is approximate. Genes on the inner circle are transcribed from the light (L) strand, and those on the outer circle are transcribed from the heavy (H) strand. Genes are designated with the same nomenclature and font used in humans and chimpanzees (Arnason *et al.* 1996a).

Table 2.1 Primers used for sequencing complete mtDNA of the green monkey

position	sequence	Tm (°C)	position	sequence	Tm (°C)
L8	GGTCTGTCACCCTATTAACC	57.5	H7821	GAGATGGTAGGGCAATCAGG	59.5
L34	TGGAGCTTCCATGCATTTGG	57.5	H8028	CGGGGGCTTCAATTGGGAGA	61.5
H36	CCCGTGACTGGTTAATAGGG	59.5	L8097	GATGCAGTACCCGGACGCTT	61.5
H560	GTGGCATCTTTGGGGTTGTG	59.5	L8128	CATTTACCGCCACACGACCA	59.5
L593	**CCTTAGTTGG**CCCCAAAGCA*	59.5	H8192	GTGGTTAGCGCCACAGATTT	57.5
H654	TCTATGGGATGTGTGAATCC	55.5	L8350	CTCTGCAGCGAAATGCCTCA	62.5
H679	GAAAGGCCAGGACCAAACCT	59.5	H8680	GCGAGTAGTACTAGGTTACG	57.5
L1093	GCTTAGCCCTAAAGGTCAGT	57.5	L8796	TTTACACCTACCACCCCAAC	57.5
H1183	GCACCGCCAAGTCCTTTGAG	61.5	H9242	TGTGATAAGGCGGGGGACAG	61.5
L1299	AGCGCAAGTACCCCTTTTCG	59.5	L9500	GAGCATTCTACCACTCAAGC	57.5
H1480	CGTGCTTTATGGCCTTGTTC	57.5	H11004	CATAGGTGGTATTGGCTAGC	57.5
H1904	GGGGTTTGGCTTTAGTTCTC	57.5	L11282	CTACTAACACTCACAGGACC	57.5
L1991	GACAAGCCTACCGAGCCTGA	61.5	H11358	AAGCCATTATGCATGCCAGT	55.5
H2485	GGTTAGTTTTGCCGAGTTCC	57.5	H11864	GGGGGTAAGGCTAGGTTGGT	61.5
L2813	GGTGACCTCGGAGCACAGTC	63.5	L12024	CAGCATTAAACCCCCTTTCA	55.5
L3273	ACCAGAGGTTCAACCCCTCT	59.5	L12217	CTGCTAACCCATGCCCCCAT	61.5
H3349	GCGGTTAGTGTAGATGCAAC	57.5	H12235	TGGGGGCATGGGTTAGCAGT	61.5
L3396	CAACTACGCAAAGGACCCAA	57.5	H12237	TATGGGGGCATGGGTTAGCA	59.5
L4036	CATTCTACCCCACCCACTCG	61.5	H12327	CTTTTATTTGGAGTTGCACC	53.5
L4084	CCAAAACGCTCCTCCTAACC	59.5	L12998	CAGCTCAACTTGGGCCTTCA	59.5
L4729	CCGCCAACCAACTTCCATCG	61.5	H13618	GGTGAGGAGAATGATTCGAG	57.5
H5386	GGGGTGGGAAATAGTGATAC	57.5	L13693	ACGCCTCACAATGGGCAGCA	61.5
H5656	TGGTCTAGTAGGGGCTTAGC	59.5	H14279	AGAGGGGTCAGGGTTAGTTC	59.5
L5959	TATTCGGCGCATGAGCTGGA	59.5	H14396	AGGGGGTCTTTTGGTGGGTT	59.5
H6030	CGGGTTGGCCTAATTCAGCTC	62	H14580	GGGGGCTAGCATTGATTGTT	57.5
L6482	ACTCCTCTCCCTGCCAGTCT	61.5	L14659	GTCATTGCTCTTGCATGGAC	57.5
L6532	ACCGCAACCTCAACACTACC	59.5	L15196	CCATACATCGGGACCGACCTC	64
H6578	GTCTCCCCCTCCAGTAGGAT	61.5	L15234	TGGGTACTCCATTGGCAACC	59.5
L6790	ATGTGGACACACGGGCCTAC	61.5	H15267	TCGTGAAAGGGTGGGGTTGC	61.5
H7035	CATAGTGGAAATGGGCGACA	57.5	H15367	TTCCGCAGGGGTTGTTTGAT	57.5
L7205	ACCCCGACGCTATTCTGACT	59.5	L15757	CAGCCAGTAAGCCAACCCTT	59.5
L7349	GGCCGAACAATCCCCCACCA	63.5	H15853	CGATTAGGGAGGTTAGTGGT	57.5
L7611	GACGCTACATCCCCTGTCAT	59.5	H16236	CGGGTTATGGTGTTTCGTGG	59.5
L7701	GCCCTATCCTCAACACTCACA	60	L16245	**CAACAACCA**ACCCACAACAC*	57.5
H7766	CTCTATCTCCTGGGCGTCTG	61.5	L16340	CCTCCCCACCACGGATGACC	65.5
L7807	TTGCCCTACCATCTCTACGC	59.5	L16409	CCGCACAAGCAGTGTTACCC	61.5

* nucleotides in bold mismatch to template DNA

Table 2.2 mtDNA gene map for the green monkey *Chlorocebus sabaeus*

gene*	from (nt)	to (nt)	length (nt)	5′ base	3′ base	intergenic nucleaotide**	strand	start codon	stop codon
HVS2	56	364	309						
tRNA-Phe	583	656	74	GTC	ACA		H		
12S rRNA	657	1604	948	AAT	AAT	0	H		
tRNA-Val	1605	1673	69	CAA	TGA	0	H		
16S rRNA	1674	3225	1552	GCC	TTT	0	H		
tRNA-Leu(UUR)	3226	3300	75	GTT	ACA	0	H		
NADH1	3303	4259	957	ATG	TAG	2	H	ATG	TAG
tRNA-Ile	4259	4326	68	GAA	CTA	-1	H		
tRNA-Gln	4324	4395	72	TAG	TAG	-3	L		
tRNA-Met	4397	4464	68	AGT	CTA	1	H		
NADH2	4465	5508	1044	ATT	TAG	0	H	ATT	TAG
tRNA-Trp	5507	5569	63	AGA	TTT	-2	H		
tRNA-Ala	5580	5648	69	AGG	TTA	0	L		
tRNA-Asn	5650	5722	73	TGG	TAG	1	L		
L-rep-origin	5724	5755	32	TTC	GAA	1	H		
tRNA-Cys	5754	5815	62	GAG	CTT	-2	L		
tRNA-Tyr	5816	5886	71	GGT	CCC	0	L		
COI	5899	7440	1542	ATG	AGC	13	H	ATG	AGC
tRNA-Ser(USN)	7439	7510	72	TTG	CGC	-2	L		
tRNA-Asp	7512	7580	69	GAG	TTA	1	H		
COII	7582	8265	684	ATG	TAA	1	H	ATG	TAA
tRNA-Lys	8296	8361	66	CAC	CGA	31	H		
ATPase8	8363	8569	207	ATG	TAA	1	H	ATG	TAA
ATPase6	8524	9204	681	ATG	TAA	-46	H	ATG	TAA
COIII	9204	9987	784	ATG	CCT	-1	H	ATG	T--
tRNA-Gly	9988	10055	68	ACT	GTA	0	H		
NADH3	10056	10395	340	ATC	AAT	0	H	ATC	T--
tRNA-Arg	10396	10460	65	TGG	CAA	0	H		
NADH4L	10461	10757	297	ATG	TAA	0	H	ATG	TAA
NADH4	10751	12128	1378	ATG	CCT	-7	H	ATG	T--
tRNA-His	12129	12197	69	GTA	ACC	0	H		
tRNA-Ser(AGY)	12198	12256	59	GAG	TCA	0	H		
tRNA-Leu(CUN)	12257	12327	71	ACT	GTA	0	H		
NADH5	12328	14136	1809	ATA	TAA	0	H	ATA	TAA
NADH6	14137	14658	522	ATG	AGG	0	L	ATG	AGG
tRNA-Glu	14659	14727	69	GCT	GCA	0	L		
Cytb	14732	15872	1141	ATG	CTT	3	H	ATG	T--
tRNA-Thr	15873	15939	67	GCC	ACA	0	H		
tRNA-Pro	15941	16009	69	CAG	GAG	1	L		
HVS1	16010	16361	352	TCC	AAG				

Genes COIII, NADH3, NADH4 and Cytb have incomplete stop codon presented by T-.

* All genes were identified by aligning the assembled sequence with human and chimpanzee sequences (Arnason *et al.* 1996a). Genes are designated with the same nomenclature and font used in humans and chimpanzees (Arnason *et al.* 1996a).
** Intergenic nucleotides are numbers of the nucleotides between the end of the last gene and beginning of the next gene. Overlap between two genes is counted as negative.

2.3.2 Heteroplasmic control region

Heteroplasmy (described in section 2.2.2 and 2.2.2.2 of this chapter, and Chapter 1) was observed in the control region around the highly variable segment 2 (HVS2). The sequence of PCR products around this region showed a consistent pattern of triplet or quartet wave forms in chromatographs even when different primers were used. For the control region between sites 58 and 549, one individual (1992-086) had at least 5 different sequences (GenBank Access number: DQ069714, DQ069715, DQ069716, DQ069717 and DQ069718) (Figure 2.4) identified among 10 clones from a heterogeneous PCR product using primer L34 and H560 (Table 2.1). Among the five sequences, 38.7% mismatches were found. However, there are few differences between sequences DQ069714 and DQ069716 (0.2%), DQ069717 and DQ069718 (0.6%), although these two groups differ in about 21-22%. DQ069715 has the greatest disagreement with the other four (27-28%) (Table 2.3). These five sequences represent three groups (Figure 2.5). Because sequence DQ069718 had the lowest mismatch with human sequence, it was assembled with the remaining mtDNA sequence obtained by primer walking, and designated as the canonical sequence of the green monkey *C. sabaeus*.

```
DQ069715        GATCACGGGT CTGTCACCCT ATTAACCAGT CACTGGAGCT TCCATGCATT
DQ069714        GATCACGGGT CTGTCACCCT ATTAACCAGT CACTGGAGCT TCCATGCATT
DQ069716        GATCACGGGT CTGTCACCCT ATTAACCAGT CACTGGAGCT TCCATGCATT
DQ069717        GATCACGGGT CTGTCACCCT ATTAACCAGT CACTGGAGCT TCCATGCATT
DQ069718        GATCACGGGT CTGTCACCCT ATTAACCAGT CACTGGAGCT TCCATGCATT
                .......... .......... .......... .......... ..........
        #1      GATCACGGGT CTGTCACCCT ATTAACCAGT CACTGGAGCT TCCATGCATT


DQ069715        TGGTATTTTT TATCTCTGGT CTGCACGCAA CACCATTGCA GAATGCTGAC
DQ069714        TGGTATTTTT TATTTCAAGT GTGCATGCGA CCCCATCGCA GAAAGCTGGC
DQ069716        TGGTATTTTT TATCTCAAGT GTGCATGCGA CCCCATCGCA GAAAGCTGGC
DQ069717        TGGTATCTTT TATCTCTGGT CTGCACGCGA CCCCATCGCA GAATGCTGGT
DQ069718        TGGTATCTTT TATCTCTGGT CCGCACGCAG CCCCATCGCA GAATGCTGGT
                .......... .......... .......... .......... ..........
        #51     TGGTATTTTT TATCTCTGGT CTGCACGCGA CCCCATCGCA GAATGCTGGC
                      *       *  **   **   *  **  *    *       *    **

DQ069715        TGCCACCACA TCCCGTCCTG AATGAGCCTG CCTTTGATTC CCAACCCATA
DQ069714        C:CCACCACA CCCTATGTTG CAGGA:TCTG TCTTTGATTC CTGGTTCATA
DQ069716        C:CCACCACA CCCTATGTTG CAGGA:TCTG TCTTTGATTC CTGGTTCATA
DQ069717        C:TCGCCATA ACACACGCCA CAGAA:CCTG TCTTTGATTG CTACTCCATA
DQ069718        C:TCGCCATA ACACACGCCA CAGAA:CCTG TCTTTGATTG CTACTCCATA
                .......... .......... .......... .......... ..........
        #101    C:CCACCACA MCCCATGCTG CAGGA:CCTG TCTTTGATTC CTASTCCATA
                *** *   *  * ******** * ** **    *        *  *****

DQ069715        TCATTATTGA TCGCACCTAC GTTCAATGTC CTAGCCCCAC ATGACTACCA
DQ069714        CCATTGTTGA TCGCACCTAC GTTCAATATT CCAGCCCCGC ATA:ATACCA
DQ069716        CCATTGTTGA TCGCACCTAC GTTCAATATT CCAGCCCCGC ATA:ATACCA
DQ069717        CCATTATTAA TCGCACCTAC ATTCAATATC CTAGTCCTGC ATAGACGTTA
DQ069718        CCATTATTAA TCGCACCTAC ATTCAATATC CTAGTCCTGC ATAGACGTTA
                .......... .......... .......... .......... ..........
        #151    CCATTATTGA TCGCACCTAC GTTCAATATC CTAGCCCCGC ATAGATACCA
                *    *  *              *      * *  *  *  **    *******

DQ069715        ATAAGGTGTT ATTTAATTCA TGCTTGTAAG ACATAC:AAC AACCAACCC:
DQ069714        :TAAGGTGTT ATTTAATTCA TGCTTGTAGG ACATACCAAT AACTGACTAT
DQ069716        :TAAGGTGTT ATTTAATTCA TGCTTGTAGG ACATACCAAT AACTGACTAT
DQ069717        :CAGGGTGTT ATTTAATTCA TGCTTGTAGG ACATACCAAT AACC:ACTAT
DQ069718        :CAGGGTGTT ATTTAATTCA TGCTTGTAGG ACATACCAAT AACC:ACTAT
                .......... .......... .......... .......... ..........
        #201    :TAAGGTGTT ATTTAATTCA TGCTTGTAGG ACATACCAAT AACCGACTAT
                ** *                           *        *  *    **  ***

DQ069715        ACAACACCAC CCCCACAACA CTACAAAAAA GCAAAAAACC TT:ATCAAAC
DQ069714        GCAGCATATC CCTCGCTACA CTATAAAATC ACAACAAATT TTTACCAAAC
DQ069716        GCAGCATATC CCTCGCTACA CTATAAAATC ACAACAAATT TTTACCAAAC
DQ069717        ACAGCCTATC TCTAACTACA CTATAAAAAC ACAACAAATT TTTAACAAAC
DQ069718        ACAGCCTATC TCTAACTACA CTATAAAAAC ACAACAAATT TTTAACAAAC
                .......... .......... .......... .......... ..........
        #251    ACAGCATATC CCTCACTACA CTATAAAAAC ACAACAAATT TTTAMCAAAC
                *  * ****  * *** *       *    ** *   *   **   * *
```

Figure 2.4 (continued over page)

```
DQ069715        CCCCC:CCAC :ATCTCTGAC ::CTC:CACA TAAACCCCCC TTT:GCCAAA
DQ069714        CCCC:TCCCC CATCTCTGAC TTTTTTCACC CAAAACTCAC TTTTGCCAAA
DQ069716        CCCC:TCCCC CATCTCTGAC TTTTTTCACC CAAAACTCAC TTTTGCCAAA
DQ069717        CCCCCTCCCC CATCTCTGAC TTCTCCCATA :AAAACTCGC TTTTGCCAAA
DQ069718        CCCCCTCCCC CATCTCTGAC TTCTCCCATA :AAAACTCGC TTTTGCCAAA
                .......... .......... .......... .......... ..........
       #301     CCCCCTCCCC CATCTCTGAC TTCTCYCACA CAAAACTCRC TTTTGCCAAA
                    **  *  *          *** **  ** *   * * *     *

DQ069715        CCCCAAAAAC AAAGTCCT:A TTCCTCCG:G CCAAAGCTTA CATTTTCATC
DQ069714        CCCCAAAAAC AAAAGCCTTA ATCCACTTAG CCAGAGCTTA TATTTTCATC
DQ069716        CCCCAAAAAC AAAAGCCTTA ATCCACTTAG CCAGAGCTTA TATTTTCATC
DQ069717        CCCCAAAGAC AAAAGCCTTG ATCCACTTAG CCAGAGCTTA CGTTCTCATC
DQ069718        CCCCAAAGAC AAAAGCCTTG ATCCACTTAG CCAGAGCTTA CGTTCTCATC
                .......... .......... .......... .......... ..........
       #351     CCCCAAAAAC AAAAGCCTTA ATCCACTTAG CCAGAGCTTA CATTTTCATC
                       *      **   ** *   * ***     *       **  *

DQ069715        TTTTAGGTGT GCAT:GCTTT AACTACCAAC CCCCCAACTA ATATTCACTT
DQ069714        TTTTAGCTAT GCATAACTTC AACTGCTA:C CCCTCAACT: ATCTACTCCA
DQ069716        TTTTAGCTAT GCATAACTTC AACTGCTA:C CCCTCAACT: ATCTACTCCA
DQ069717        TTTTAGCTAT GCACAGCTTC AACTGCTATC CCCTCAAATG ATATAAAATT
DQ069718        TTTTAGCTAT GCACAGCTTC AACTGCTATC CCCTCAAATG ATATAAAATT
                .......... .......... .......... .......... ..........
       #401     TTTTAGCTAT GCATAGCTTC AACTGCTATC CCCTCAACTG ATATACACTT
                      * *     ***   *     * * *     *   * *   * ******

DQ069715        ACCTCACAA: ACCACCCTTC ACACCAACTT CCCCCCCCCC ACA:ACCC:C
DQ069714        CCCT:ACC:: AC:::TCTCA ACTTCAATTT CTTC:CTTTT CTTTTCTTTC
DQ069716        CCCT:ACC:: AC:::TCTCA ACTTCAATTT CTTC:CTTTT CTTTTCTTTC
DQ069717        ACTTCACCTT ACCACTCTCA ACTTCAATTT CTTTTCCCTT CCATACTTTC
DQ069718        ACTTCACCTT ACCACTCTCA ACTTCAATTT CTTTTCCCTT CCATACTTTC
                .......... .......... .......... .......... ..........
       #451     ACCTCACCT: ACCACTCTCA ACTTCAATTT CTTCTCCCTT CCATACTTTC
                * * *  ***   ****  **   **   *    **** **** ***** ***

DQ069715        AAAGATG::: ::CC:ACACA CTTACCCACT ACCAGATCTA TACCCTTACT
DQ069714        TTTCCTTTTT TTCCTACACA CTTACCCACT ACCAGATCTA TACCCTTACT
DQ069716        TTTCCTTTTT TTCCTACACA CTTACCCACT ACCAGATCTA TACCCTTACT
DQ069717        TCAC:TC::: ::CC:ATGCA TTTACTCACT ACTGAATCTG TGCCCTTATT
DQ069718        TCAC:TC::: ::CC:ATGCA TTTACTCACT ACTGAATCTG TGCCCTTATT
                .......... .......... .......... .......... ..........
       #501     TYACCTY::: ::CC:ACACA CTTACCCACT ACCAGATCTA TACCCTTACT
                ***** **** **  * **   *    *       ***    *  *      *
```

Figure 2.4 Comparison of five heteroplasmic sequences of the green monkey *C. sabaeus*. * indicates a mismatch at the site among the five sequences, and the nucleotide appearing in the majority of the sequences was taken as the consensus (with position number on the left). If the consensus could not be determined, then an ambiguity code M (C & A), Y (T & C) and R (A & G) was used. : indicates a deletion.

Table 2.3 Pairwise comparison of the five heteroplasmic sequences

between sequence	total(bp)	mismatch	%mismatch
DQ069714 and DQ069716	489	1	0.2
DQ069717 and DQ069718	489	3	0.6
DQ069714 and DQ069718	489	109	22.3
DQ069716 and DQ069718	489	108	22.1
DQ069715 and DQ069718	489	139	28.4
DQ069714 and DQ069715	489	134	27.4
DQ069714 and DQ069717	489	106	21.7
DQ069715 and DQ069716	489	133	27.2
DQ069715 and DQ069717	489	138	28.2
DQ069716 and DQ069717	489	105	21.5

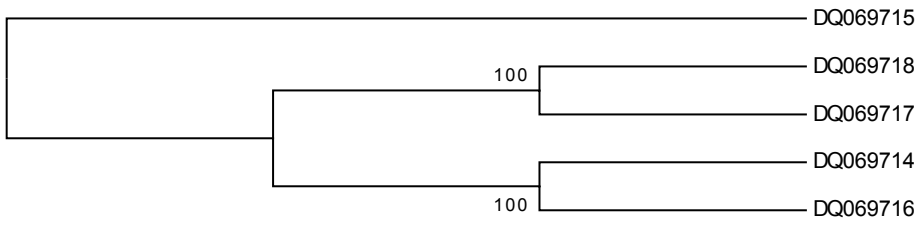


Figure 2.5 Clustering tree shows three groups of the five heteroplasmic sequences within the control region of green monkey mtDNA, based on Maximum Parsimony method. Numbers on branches are bootstrap values.

A 122 bp repeat unit was observed in the control region during direct PCR product sequencing. The sequence of three amplicons contained 3 or 4 repeats of this unit (Figure 2.6 & Figure 2.7). Comparisons among the five heteroplasmic sequence reported in the present study revealed that this 122 bp repeat unit has 24.2% mismatch to a region of DQ069714 and DQ069716; 15.6% to that of DQ069715, 31.5% to that of DQ069717 and DQ069718. However, heteroplasmy related to the number of copies of the repeat unit was not investigated in the 10 cloned segments because these segments were cloned from fragments of similar size excised from gels, thus excluding any longer segment that could contain additional repeat units.

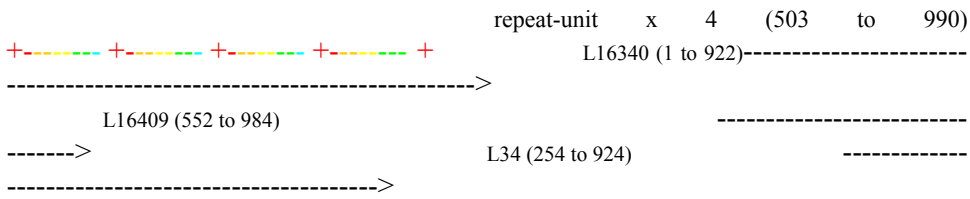


Figure 2.6 An overview of alignment of three amplicons with different primers oriented in the same direction. Top segmented line represents concatenated four repeat units. The other three arrowed lines represent amplicons with primer names above the line. The range number within each parenthesis is based on consensus length. The 122 bp repeat unit is: AACCACCCTT TGCCAAACCC CGAAAACAAA GTCCTGCTCC CCCGGCCAAA GCTTGTATTT CCATCTTTTA GGTATGCATG CCCTAACTAC TAACCCCTCA ACTAACCTTC GCTTACCCTA CC

```
>L16340             #1                CACTTAGAAC TCTCTTACTC ACCATCCTCC GTGAAATCAA

>L16340             #41               TATCCCGCAC AAGCAGTGTT ACCCTCCTCG CTCCGGGCCC

>L16340             #81               ACGACCCGTG GGGGTAGCTA AGAAATGAGC TGTATCCGGC

>L16340             #121              ATCTGGTTCT TACCTCAGGG CCATGCCAAC TAAAACCGTC

>L16340             #161              CACACGTTCC TCTTAAATAA GACATCTCGA TGGATCACGG

>L16340             #201              GTCTATCACC CTATTAACCA GTCACTGGAG CTTCCCATGC

>L16340             #241              ATTTGGTATT TTTTATCTCT GGTCTGCACG CAACACCATC
>L34                #1                   TATCTCT GGTCTGCACG CAACACCATC

>L16340             #281              GCAACATGCT GACTCCCACC ACATCCCGCC CTGAATGCGC
>L34                #28               GCAACATGCT GACTCCCACC ACATCCCGCC CTGAATGCGC

>L16340             #321              CTGCCTTTGA TTCCTAGTCC ATGCCATTAT TAATCGCACC
>L34                #68               CTGCCTTTGA TTCCTAGTCC ATGCCATTAT TAATCGCACC

>L16340             #361              TACGTTCAAT ATTCCAGCCC CGCATGACTA CTAGCGATGT
>L34                #108              TACGTTCAAT ATTCCAGCCC CGCATGACTA CTAGCGATGT

>L16340             #401              GTTATTTAAT TCATGCTTGT AAGACATACA ACAACCAACC
>L34                #148              GTTATTTAAT TCATGCTTGT AAGACATACA ACAACCAACC
```

Figure 2.7 (continued over page)

```
>L16340             #441              CACAACACTA CAAAAAAATT CAAAAACCTT AACCAAACCC
>L34                #188              CACAACACTA CAAAAAAATT CAAAAACCTT AACCAAACCC
```

```
>L16340          #481            CCCCCCCCTT TGACCTTCCA ATAACCACCC TTTGCCAAAC
>L34             #228            CCCCCCCCTT TGACCTTCCA ATAACCACCC TTTGCCAAAC
                                                       ──────►

>L16340          #521            CCCGAAAACA AGTCCTGCT CCCCCGGCCA AAGCTTGTAT
>L16409          #1                                              AGCTTGTAT
>L34             #268            CCCGAAAACA AGTCCTGCT CCCCCGGCCA AAGCTTGTAT

>L16340          #561            TTCCATCTTT TAGGTATGCA TGCCCTAACT ACTAACCCCT
>L16409          #10             TTCCATCTTT TAGGTATGCA TGCCCTAACT ACTAACCCCT
>L34             #308            TTCCATCTTT TAGGTATGCA TGCCCTAACT ACTAACCCCT

>L16340          #601            CAACTAACCT TCGCTTACCC TACCAACCAC CCTTTGCCAA
>L16409          #50             CAACTAACCT TCGCTTACCC TACCAACCAC CCTTTGCCAA
>L34             #348            CAACTAACCT TCGCTTACCC TACCAACCAC CCTTTGCCAA
                                                       ──────►

>L16340          #641            ACCCCGAAAA CAAAGTCCTG CTCCCCCGGC CAAAGCTTGT
>L16409          #90             ACCCCGAAAA CAAAGTCCTG CTCCCCCGGC CAAAGCTTGT
>L34             #388            ACCCCGAAAA CAAAGTCCTG CTCCCCCGGC CAAAGCTTGT

>L16340          #681            ATTTCCATCT TTTAGGTATG CATGCCCTAA CTACTAACCC
>L16409          #130            ATTTCCATCT TTTAGGTATG CATGCCCTAA CTACTAACCC
>L34             #428            ATTTCCATCT TTTAGGTATG CATGCCCTAA CTACTAACCC

>L16340          #721            CTCAACTAAC CTTCGCTTAC CCTACCAACC ACCCTTTGCC
>L16409          #170            CTCAACTAAC CTTCGCTTAC CCTACCAACC ACCCTTTGCC
>L34             #468            CTCAACTAAC CTTCGCTTAC CCTACCAACC ACCCTTTGCC
                                                       ──────►

>L16340          #761            AAACCCCGAA AACAAAGTCC TGCTCCCCCG GCCAAAGCTT
>L16409          #210            AAACCCCGAA AACAAAGTCC TGCTCCCCCG GCCAAAGCTT
>L34             #508            AAACCCCGAA AACAAAGTCC TGCTCCCCCG GCCAAAGCTT

>L16340          #801            GTATTTCCAT CTTTTAGGTA TGCATGCCCT AACTACTAAC
>L16409          #250            GTATTTCCAT CTTTTAGGTA TGCATGCCCT AACTACTAAC
>L34             #548            GTATTTCCAT CTTTTAGGTA TGCATGCCCT AACTACTAAC

>L16340          #841            CCCTCAACTA ACCTTCGCTT ACCCTACCAA CCACCCTTTG
>L16409          #290            CCCTCAACTA ACCTTCGCTT ACCCTACCAA CCACCCTTTG
>L34             #588            CCCTCAACTA ACCTTCGCTT ACCCTACCAA CCACCCTTTG
                                                          ──────►

>L16340          #881            CCAAACCCCG AAAACAAAGT CCTGCTCCCC CGGCCAAAGC
>L16409          #330            CCAAACCCCG AAAACAAAGT CCTGCTCCCC CGGCCAAAGC
>L34             #628            CCAAACCCCG AAAACAAAGT CCTGCTCCCC CGGCCAAAGC

>L16340          #921            TT
>L16409          #370            TTGTATTTCC ATCTTTTAGG TATGCATGCC CTAACTACTA
>L34             #668            TTGT

>L16409          #410            ACCCCTCAAC TAACCTTCGC TTAC
```

Figure 2.7 Sequence alignment summary of three amplicons with different primers of the same direction. Arrows under the sequences indicate beginning of the repeat unit. The amplicon names on the left are the same as in Figure 2.6.

2.4 Discussion

The complete mtDNA sequence of the green monkey *C. sabaeus* reported in the present study is within the primate range in sequence length (16380–16570 bp), base composition (A; 30.5-32%, C: 30-32.5%, G: ~13%, T: 24-25%), gene arrangement and basic codon features, although it uses the start codon ATC, which is not commonly used by primates other than *Chlorocebus* species, for the protein encoding gene NADH3.

2.4.1 Complete mtDNA comparison among primates

Among all primates whose mtDNA has been sequenced so far, the most closely related species to the one whose sequence is reported in this study are the vervet monkey *Chlorocebus aethiops* (GenBank Accession No. AY863426; Raaum *et al*., 2005), the Barbary ape *Macaca sylvanus* (GenBank Accession No. NC_002764), the rhesus monkey *Macaca mulatta* (GenBank accession No. AY612638) and the hamadryas baboon *Papio hamadryas* (GenBank accession No. Y18001, Arnason *et al*. 1998). All are in the family Cercopithecidae and the subfamily Cercopithecinae. However, a detailed comparison of these species was initially hindered by sequence position errors in the annotation of the complete mtDNA sequence of rhesus monkey. The most obvious mistake is for tRNA-Lys, defined from site 7761 to 7830 within the gene COII (7532-8215). It is high unlikely that the gene COII is interrupted by the gene tRNA-Lys in any mammal. By re-aligning the rhesus sequence with other primate sequences, errors in the annotation of tRNA-Ile, Trp, Cys, Tyr, Ser, Lys, Pro, mRNA COI and COIII in rhesus sequence were also corrected as indicated in Table 2.4. Some minor errors in the mtDNA annotation of *C. aethiops* and *P. hamadryas* were corrected as well (not shown).

Table 2.4 Correction to the annotation of mtDNA gene map for the rhesus monkey (*Macaca. mulatto*). Genes highlighted are corrected.

		Macaca mulatta (original)				*Macaca mulatta (corrected)*					
		from	to	5'	3'	*from*	to	5'	3'	length	intergenic nt
F	tRNA-Phe	536	607	GTT	ACA	536	607	GTT	ACA	72	
12S	12S rRNA	608	1554	TAT	AAT	608	1554	TAT	AAT	947	0
V	tRNA-Val	1555	1623	CAA	TGA	1555	1623	CAA	TGA	69	0
16S	16S rRNA	1624	3181	GCC	TTT	1624	3181	GCC	TTT	1558	0
L	tRNA-Leu(UUR)	3182	3256	GTT	ACA	3182	3256	GTT	ACA	75	0
ND1	NADH1	3259	4215	ATG	TAG	3259	4215	ATG	TAG	957	2
I	**tRNA-Ile**	**4214**	**4282**	**AGG**	**CTA**	**4215**	**4282**	**GAA**	**CTA**	68	-1
Q	tRNA-Gln	4280	4351	TAG	TAG	4280	4351	TAG	TAG	72	-3
M	tRNA-Met	4353	4420	AGT	CTA	4353	4420	AGT	CTA	68	1
ND2	NADH2	4421	5464	ATT	TAG	4421	5464	ATT	TAG	1044	0
W	**tRNA-Trp**	**5463**	**5529**	**AGA**	**CTG**	**5463**	**5526**	**AGA**	**TTT**	**64**	**-2**
A	tRNA-Ala	5537	5605	AAG	TTA	5537	5605	AAG	TTA	69	10
N	tRNA-Asn	5607	5679	TGG	TAG	5607	5679	TGG	TAG	73	1
	L-rep_origin	5681	5713	TTC	GAA	5681	5713	TTC	GAA	33	1
C	**tRNA-Cys**	**5712**	**5780**	**AGC**	**CTT**	**5712**	**5774**	**GAG**	**CTT**	63	-2
Y	**tRNA-Tyr**	**5781**	**5849**	**AAG**	**ACC**	**5775**	**5845**	**GGT**	**CCC**	71	0
COI	**COI**	**5850**	**7391**	**ATG**	**GTC**	**5850**	**7390**	**ATG**	**AGT**	1541	4
S	**tRNA-Ser(USN)**	**7391**	**7459**	**GAA**	**TCG**	**7390**	**7461**	**CCG**	**CGA**	72	-1
D	tRNA-Asp	7483	7530	AAG	TTA	7483	7530	AAG	TTA	48	21
COII	COII	7532	8215	ATG	TAA	7532	8215	ATG	TAA	684	1
K	**tRNA-Lys**	**7761**	**7830**	**CTC**	**AAC**	**8291**	**8355**	**CAC**	**TGA**	65	75
e8	ATPase8	8357	8563	ATG	TAA	8357	8563	ATG	TAA	207	1
e6	ATPase6	8518	9198	ATG	TAA	8518	9198	ATG	TAA	681	-46
COIII	**COIII**	**9198**	**9982**	**ATG**	**CTA**	**9198**	**9981**	**ATG**	**CCT**	**784**	**-1**
G	tRNA-Gly	9982	10049	ACT	GTA	9982	10049	ACT	GTA	68	0
ND3	NADH3	10050	10395	ATT	AAT	10050	10395	ATT	AAT	346	0
R	tRNA-Arg	10396	10460	TGG	CAA	10396	10460	TGG	CAA	65	0
ND4L	NADH4L	10461	10757	ATG	TAA	10461	10757	ATG	TAA	297	0
NADH4	NADH4	10751	12128	ATG	CCT	10751	12128	ATG	CCT	1378	-7
H	tRNA-His	12129	12197	GTA	ACC	12129	12197	GTA	ACC	69	0
S	tRNA-Ser(AGY)	12198	12256	GAG	TCA	12198	12256	GAG	TCA	59	0
L	tRNA-LeuCUN)	12257	12327	ACT	GTA	12257	12327	ACT	GTA	71	0
NADH5	NADH5	12328	14139	ATA	TAA	12328	14139	ATA	TAA	1812	0
NADH6	NADH6	14140	14667	ATG	AGG	14140	14667	ATG	AGG	528	0
E	tRNA-Glu	14668	14736	GCT	GCA	14668	14736	GCT	GCA	69	0
cyt b	cyt b	14741	15881	ATG	CTT	14741	15881	ATG	CTT	1141	4
T	tRNA-Thr	15882	15945	GCC	ACA	15882	15945	GCC	ACA	64	0
P	**tRNA-Pro**	**15947**	**16014**	**CAG**	**TGA**	**15946**	**16014**	**CAG**	**GAG**	**69**	**0**

Using the corrected annotations, the initial and terminal three nucleotides of all the genes have been compared among these five Cercopithecinae species and five Hominoid species [humans *Homo sapiens* (Accession No: X93334, Arnason *et al*. 1996a); common chimpanzees *Pan troglodytes verus* (Accession No: X93335, Arnason *et al*. 1996a); pygmy chimpanzees *P. paniscus* (Accession No: D38116); gorillas *Gorilla gorilla gorilla*

(Accession No: X93347, Xu and Arnason 1996a); and orangutans *Pongo pygmaeus* (Accession No: X97707, Xu and Arnason 1996b)] (Table 2.5). For most of the genes, the initial and terminal three nucleotides are identical. However, the initial three nucleotides for tRNA-Val, –Glu and NADH1 and the terminal nucleotides for 12S rRNA, NADH1, COI, tRNA-Ser(USN), and -Glu distinguish between Cercopithecinae species and Hominoid species. The initial nucleotides for tRNA-Phe and NADH3, and the terminal nucleotides for tRNA-Lys distinguish *Chlorocebus* species from all others. It was expected that the two *Chlorocebus* species would use the same initial and terminal nucleotides for most genes. However, it was unexpected that the green monkey *C. sabaeus* would share unique initial nucleotides with orangutan, which is different from all other species of both Cercopithecinae and Hominoid for tRNA-Asp. In addition to the differences mentioned above, the initial nucleotides differ across these species for tRNA-Ala, -Ser(USN), -Asn, -Thr, and both of the rRNA; and the nucleotides in the 3' termini differ in 16S rRNA, NADH2, NADH3, NADH4, COII, COIII, ATPase8, Cytb, tRNA-Tyr, -Thr, -Pro.

Table 2.5 Comparison of first and last three nucleotides of mtDNA genes among primate species *Chlorocebus sabaeus, C. aethiops, Macaca sylvanus, M. mulatta, Papio hamadryas, Homo sapiens, Pan troglodytes, P. paniscus, Gorilla gorilla gorilla,* and *Pongo pygmaeus*

	green monkey		vervet		rhesus		barbary		baboon		human		chimpanzee				gorilla		orangutan	
	Chlorocebus				*Macaca*				*Papio*		*Homo*		*Pan*				*Gorilla*		*Pongo*	
	sabaeus		aethiops		*mulatta*		*sylvanus*		*hamadryas*		*sapiens*		*troglodytes*		*paniscus*		*gorilla gorilla*		*pygmaeus*	
	5′	*3′*	*5′*	*3′*	*5′*	*3′*	*5′*	*3′*	*5′*	*3′*	*5′*	*3′*	*5′*	*3′*	*5′*	*3′*	*5′*	*3′*	*5′*	*3′*
tRNA-Phe	GTC	ACA	GTC	ACA	GTT	ACA	GTT	ACA	GTT	ACA	GTT	ACA	GTT	ACA	GTT	ACA	GTT	ACA	GTT	ACA
12S rRNA	AAT	AAT	AAT	AAT	TAT	AAT	AAC	AAT	AAT	AAT	AAT	AAC	AAC	AAC	AAC	AAC	AAT	AAC	AAT	AAC
tRNA-Val	CAA	TGA	CAA	TGA	CAA	TGA	CAA	TGA	CAA	TGA	CAG	TGA	CAG	TGA	CAG	TGA	CAG	TGA	CAG	TGA
16S rRNA	GCC	TTT	GCT	TTT	GCC	TTT	GCC	TTT	GCC	TTT	GCT	TTT	GCC	TTT	GCC	TTT	GCA	TTT	GCT	TAT
tRNA-Leu (UUR)	GTT	ACA	GTT	ACA	GTT	ACA	GTT	ACA	GTT	ACA	GTT	ACA	GTT	ACA	GTT	ACA	GTT	ACA	GTT	ACA
NADH1	ATG	TAG	ATG	TAG	ATG	TAG	ATG	TAG	ATG	TAG	ATA	TAA	ACA	TAA	ACA	TAA	ATA	TAA	ATG	TAA
tRNA-Ile	:GAA	CTA	GAA	CTA	GAA	CTA	GAA	CTA	GAA	CTA	GAA	CTA	GAA	CTA	GAA	CTA	GAA	CTA	GAA	CTA
tRNA-Gln	TAG	TAG	TAG	TAG	TAG	TAG	TAG	TAG	TAG	TAG	TAG	TAG	TAG	TAG	TAG	TAG	TAG	TAG	TAG	TAG
tRNA-Met	AGT	CTA	AGT	CTA	AGT	CTA	AGT	CTA	AGT	CTA	AGT	CTA	AGT	CTA	AGT	CTA	AGT	CTA	AGT	CTA
NADH2	ATT	TAG	ATT	TAG	ATT	TAG	ATT	TAG	ATT	TAG	ATT	TAG	ATT	TAG	ATT	TAG	ATT	TAG	ATT	TAG
tRNA-Trp	AGA	TTT	AGA	TTT	AGA	TTT	AGA	TTT	AGA	TTT	AGA	TTT	AGA	TTT	AGA	TTT	AGA	TTT	AGA	TTT
tRNA-Ala	AGG	TTA	AGG	TTA	AAG	TTA	AGG	TTA	AAG	TTA	AAG	TTA	AAG	TTA	AAG	TTA	AAG	TTA	GAG	TTA
tRNA-Asn	TGG	TAG	TGG	TAG	TGG	TAG	TGG	TAG	TGG	TAG	TAG	TAG	TAG	TAG	TAG	TAG	TAG	TAG	TGG	TAG
tRNA-Cys	GAG	CTT	GAG	CTT	GAG	CTT	GAG	CTT	GAG	CTT	GAG	CTT	GAG	CTT	GAG	CTT	GAG	CTT	GAG	CTT
tRNA-Tyr	GGT	CCC	GGT	CCC	GGT	CCC	GGT	CCC	GGT	CCC	GGT	TCC	GGT	TCT	GGT	TCT	GGT	TCC	GGT	CCC
COI	ATG	AGC	ATG	AAC	ATG	AGT	ATG	AAC	ATG	AGC	ATG	AGA	ATG	AGA	ATG	AGA	ATG	AAA	ATG	AGA
tRNA-Ser (USN)	TTG	CGC	TTG	CGT	CCG	CGA	TTG	CGT	TTG	CGC	TTG	TGT	TTG	TGT	TTG	TGT	TTG	TGT	TTG	TGT
tRNA-Asp	GAG	TTA	AAG	TTA	AAG	TTA	AAG	TTA	AAG	TTA	AAG	TTA	AAG	TTA	AAG	TTA	AAG	TTA	GAG	TTA
COII	ATG	TAA	ATG	TAG	ATG	TAA	ATG	TAA	ATG	TAA	ATG	TAG	ATG	TAG	ATG	TAA	ATG	TAA	ATG	TAA
tRNA-Lys	CAC	CGA	CAC	CGA	CAC	TGA	CAC	TGC	CAC	TGA	CAC	TGA	CAC	TGA	CAC	TGA	CAC	TGA	CAC	TGA
ATPase8	ATG	TAA	ATG	TAA	ATG	TAA	ATG	TAA	ATG	TAG	ATG	TAG	ATG	TAG	ATG	TAG	ATG	TAG	ATG	TAG
ATPase6	ATG	TAA	ATG	TAA	ATG	TAA	ATG	TAA	ATG	TAA	ATG	TAA	ATG	TAA	ATG	TAA	ATG	TAA	ATG	TAA
COIII	ATG	CCT	ATG	CCT	ATG	CCT	ATG	CTT	ATG	CTT	ATG	CTT	ATG	CTT	ATG	CTT	ATG	CCT	ATG	CCT
tRNA-Gly	ACT	GTA	ACT	GTA	ACT	GTA	ACT	GTA	ACT	GTA	ACT	GTA	ACT	GTA	ACT	GTA	ACT	GTA	ACT	GTA
NADH3	ATC	AAT	ATC	AAT	ATT	AAT	ATT	AAT	ATT	GAT	ATA	AAT	ATA	AAT	ATA	AAT	ATA	AAT	ATT	AAT
tRNA-Arg	TGG	CAA	TGG	CAA	TGG	CAA	TGG	CAA	TGG	CAA	TGG	CAA	TGG	CAA	TGG	CAA	TGG	CAA	TGG	CAA
NADH4L	ATG	TAA	ATG	TAA	ATG	TAA	ATG	TAA	ATG	TAA	ATG	TAA	ATG	TAA	ATG	TAA	ATG	TAA	ATG	TAA
NADH4	ATG	CCT	ATG	CCT	ATG	CCT	ATG	CCT	ATG	CCT	ATG	CTT	ATG	CCT	ATG	CCT	ATG	CCT	ATG	ACT
tRNA-His	GTA	ACC	GTA	ACC	GTA	ACC	GTA	ACC	GTA	ACC	GTA	ACC	GTA	ACC	GTA	ACC	GTA	ACC	GTA	ACC
tRNA-Ser (AGY)	GAG	TCA	GAG	TCA	GAG	TCA	GAG	TCA	GAG	TCA	GAG	TCA	GAG	TCA	GAG	TCA	GAG	TCA	GAG	TCA
tRNA-Leu (CUN)	ACT	GTA	ACT	GTA	ACT	GTA	ACT	GTA	ACT	GTA	ACT	GTA	ACT	GTA	ACT	GTA	ACT	GTA	ACT	GTA
NADH5	ATA	TAA	ATA	TAA	ATA	TAA	ATA	TAA	ATA	TAA	ATA	TAA	ATA	TAA	ATA	TAA	ATA	TAA	ACA	TAA
NADH6	ATG	AGG	ATG	AGG	ATG	AGG	ATG	AGG	ATG	AGG	ATG	AGG	ATG	AGG	ATG	AGG	ATG	AGG	ATG	AGG
tRNA-Glu	GCT	GCA	GCT	GTA	GCT	GCA	GCT	GCA	GCT	GCA	GTT	ATA	GTT	ATA	GTT	ATA	GTT	ATG	GTT	ACA
cyt b	ATG	CTT	ATG	CTT	ATG	CTT	ATG	CCT	ATG	CCT	ATG	CCT	ATG	CCT	ATG	CCT	ATG	CCT	ATG	CCT
tRNA-Thr	GCC	ACA	GCC	ACA	GCC	ACA	GCC	GCA	GCC	GTA	GTC	ACA	GCC	ACA	GCC	ACA	GCC	GCA	GCC	GCA
tRNA-Pro	CAG	GAG	CAG	GAG	CAG	GAG	CAG	GAG	CAG	GAT	CAG	GAT	CAG	GAT	CAG	GAT	CAG	GAA	CAG	GAT

For the 13 peptide-coding genes, the start codons normally used by mammals are ATG, ATT and ATA. However, both of the *Chlorocebus* monkeys use the unusual ATC for

NADH3, the same start codon used by the horse *Equus caballus* for NADH5 (Xu and Arnason, 1994; Xu *et al*., 1996). For 9 of the 13 protein encoding genes, COI, COII, ATPase8, ATPase6, COIII, NADH4L, NADH4, NADH6, and Cytb, the start codon is the most commonly used ATG, and is identical among all these primates. Like all other primates, the green monkey uses the same relatively unusual start codon ATT for NADH2. The green monkey employs the primate consensus stop codon TAA for ATPase6, NADH4L and NADH5; TAG for NADH2 and AGG for NADH6. Also it uses the consensus unusual incomplete stop codon T- for COIII, NADH3, NADH4, and Cytb, like all other primates (Arnason *et al*., 1996a; Xu and Arnason 1996a, b). The incomplete stop codon in COIII, NADH3 and NADH4 is a common feature in mammals. Pigs, donkeys, horses, harbor and grey seals, fin and blue whales all exhibit this feature (Arnason *et al*., 1991; Arnason and Johnsson 1992; Arnason *et al*., 1993; Arnason and Gullberg, 1993; Xu *at al*., 1996; Ursing and Arnason, 1998; Le *at al*.; 2002), although there is a complete stop codon in mouse and rats for NADH3 (Xu and Arnason 1994) and an incomplete stop codon in pigs (Ursing and Arnason 1998) for COII. These incomplete stop codons are completed by post-transcriptional polyadenylation (Ojala *et al*. 1981).

The nucleotide difference between the complete mtDNA sequence of human and that of the common chimpanzee is 8.9%. (Arnason *et al*., 1996a). Between gorilla and human, the difference outside the control region is 10.5%, counting each gap as a single mutation irrespective of its length (Xu and Arnason 1996a). The corresponding figure for gorilla vs. chimpanzee is 10.0%, and that for human vs. chimpanzee is 8.5% (Xu and Arnason 1996a). When all the mitochondrial gene sequences were compared between the green monkey and other primates, the coding sequence differences between the two *Chlorocebus* species is 5%, between the green monkey and the rhesus monkey 14%; between the green monkey and the baboon 16%; between the green monkey and human 21.6%; between the green monkey and the chimpanzee 19%; between the green monkey and gorilla 20.3%.

The nucleotide differences between the green monkey and the other primates vary among genes (Table 2.6). As a general rule, the nucleotide difference within the Cercopithecinae

species is smaller than between Cercopithecinae species and Hominoid species. However, for quite a few tRNA, such as -Met, -Asn, -Asp, -His, –Leu(CUN) and -Glu, the mismatch rate is mostly on a similar level for each primate-*sabaeus* comparison except for that within *Chlorocebus*. Pairwise comparisons between the green monkey and the other primates for tRNA-Trp, -Ala, -Thr and -Arg do not show the trend of sequence differences (within *Chlorocebus* < among Cercopithecinae < between Cercopithecinae and Hominoid). These scenarios could be due to the special secondary structure of the transfer RNA. In the cloverleaf structure of transfer RNA, certain stem, arms and loops are relatively stable to ensure their functions. Variable sites are limited and so it may easily reach a selective constraint and consequently homoplasy. These tRNAs then appear misinformative in evolutionary and phylogenetic studies (see next chapter). When comparing *sabaeus* tRNA sequence to pig *Sus scrofa* (accession number AY574048), it might be expected that the mismatch rate would be higher between pig and *sabaeus* than that between *sabaeus* and hominoid species. However, for tRNA-Phen, -Ile, -Tyr, -Gly and -Leu(CUN), the mismatch rates between *sabaeus* and pig are sometimes lower than mismatch rates between *sabaeus* and hominoid species. For other tRNA genes such as -Ser(AGY) and -Glu, the mismatch rates between *sabaeus* and pig are at a similar level to that between *sabaeus* and hominoid species. This provides additional evidence for strong selective constraint on the evolution of tRNA sequences.

Table 2.6 Pairwise comparisons of mitochondrial genes between the green monkey and other primates

	within Chlorocebus			sabaeus-rhesus			sabaeus-barbary			sabaeus-baboon			sabaeus-human			sabaeus-common chin			sabaeus-pygmy chim			sabaeus-gorilla			sabaeus-orangutan		
	mis-	length	mis%	mis-	length	mis%	mis-	length	mis%	mis-	length	mis%	mis-	length	mis%	mis-	length	mis%	mis-	length	mis%	mis-	length	mis%	mis-	length	mis%
tRNA-Phe	6	74	8.1	11	74	14.9	9	74	12.2	11	74	14.9	19	78	24.4	20	78	25.6	21	78	26.9	20	78	25.6	20	74	27.0
12S rRNA	25	951	2.6	114	950	12.0	97	950	10.2	103	950	10.8	151	962	15.7	147	957	15.4	151	960	15.7	150	957	15.7	150	957	15.7
tRNA-Val	4	70	5.7	13	70	18.6	14	69	20.3	18	69	26.1	22	69	31.9	21	69	30.4	21	69	30.4	21	69	30.4	19	69	27.5
16S rRNA	67	1561	4.3	181	1562	11.6	172	1564	11.0	200	1575	12.7	254	1566	16.2	239	1565	15.3	235	1567	15.0	264	1566	16.9	227	1567	14.5
tRNA-Leu (UUR)	0	75	0.0	7	75	9.3	4	75	5.3	5	75	6.7	9	75	12.0	8	75	10.7	8	75	10.7	9	75	12.0	8	75	10.7
NADH1	50	958	5.2	133	958	13.9	125	958	13.0	124	958	12.9	175	958	18.3	185	958	19.3	194	958	20.3	185	958	19.3	189	958	19.7
tRNA-Ile	2	68	2.9	4	68	5.9	1	68	1.5	3	68	4.4	8	68	11.8	10	68	14.7	11	68	16.2	7	68	10.3	7	68	10.3
tRNA-Gln	2	72	2.8	9	72	12.5	8	72	11.1	9	72	12.5	13	72	18.1	13	72	18.1	14	72	19.4	13	72	18.1	12	72	16.7
tRNA-Met	0	68	0.0	2	68	2.9	2	68	2.9	2	68	2.9	2	68	2.9	1	68	1.5	2	68	2.9	1	68	1.5	2	68	2.9
NADH2	58	1044	5.6	187	1044	17.9	170	1044	16.3	188	1047	18.0	263	1045	25.2	252	1044	24.1	251	1044	24.0	260	1047	24.8	252	1045	24.1
tRNA-Trp	3	63	4.8	0	63	0.0	6	64	9.4	7	64	10.9	6	65	9.2	9	65	13.8	8	65	12.3	5	65	7.7	8	64	12.5
tRNA-Ala	0	69	0.0	2	69	2.9	4	69	5.8	8	69	11.6	5	69	7.2	5	69	7.2	5	69	7.2	4	69	5.8	8	69	11.6
tRNA-Asn	0	73	0.0	8	73	11.0	5	73	5.8	9	73	11.6	8	73	7.2	8	73	11.0	9	73	12.3	7	73	9.6	8	73	11.0
tRNA-Cys	1	62	1.6	4	63	6.3	6	63	9.5	6	62	9.7	12	62	19.4	12	62	19.4	11	62	17.7	13	62	21.0	9	62	14.5
tRNA-Tyr	2	71	2.8	10	71	14.1	12	71	16.9	5	71	7.0	12	72	16.7	12	72	16.7	12	72	16.7	10	72	13.9	12	73	16.4
COI	83	1542	5.4	219	1542	14.2	247	1542	16.0	245	1542	15.9	298	1542	19.3	300	1542	19.5	302	1542	19.6	311	1542	20.2	302	1542	19.6
tRNA-Ser (USN)	4	72	5.6	7	72	9.7	9	72	12.5	7	72	9.7	14	72	19.4	12	72	16.7	11	72	15.3	14	72	19.4	13	72	18.1
tRNA-Asp	2	69	2.9	13	69	18.8	10	70	14.3	10	69	14.5	11	69	15.9	13	69	18.8	12	69	17.4	14	69	20.3	13	69	18.8
COII	35	684	5.1	111	684	16.2	96	684	14.0	89	684	13.0	131	684	19.2	114	684	16.7	120	684	17.5	127	684	18.6	128	694	18.4
tRNA-Lys	3	66	4.5	14	67	20.9	17	66	25.8	8	66	12.1	19	70	27.1	16	70	22.9	16	70	22.9	18	70	25.7	17	70	24.3
ATPase8	11	207	5.3	56	207	27.1	41	207	19.8	48	207	23.2	73	207	35.3	65	207	31.4	61	207	29.5	64	208	30.8	68	208	32.7
ATPase6	59	682	8.7	138	682	20.2	147	682	21.6	140	682	20.5	174	682	25.5	165	682	24.2	169	682	24.8	182	685	26.6	175	682	25.7
COIII	53	784	6.8	118	784	15.1	138	784	17.6	126	784	16.1	159	784	20.3	151	784	19.3	149	784	19.0	151	784	19.3	152	786	19.3
tRNA-Gly	5	68	7.4	7	68	10.3	6	68	8.8	6	68	8.8	12	68	17.6	12	68	17.6	11	68	16.2	17	68	25.0	12	68	17.6
NADH3	28	346	8.1	59	346	17.1	52	346	15.0	61	346	17.6	83	346	24.0	89	346	25.7	82	346	23.7	83	346	24.0	78	346	22.5
tRNA-Arg	4	65	6.2	5	65	7.7	5	66	7.6	2	65	3.1	6	65	9.2	7	65	10.8	7	65	10.8	6	65	9.2	7	65	10.8
NADH4L	23	297	7.7	58	297	19.5	48	297	16.2	48	297	16.2	60	297	20.2	58	297	19.5	63	297	21.2	60	297	20.2	55	297	18.5
NADH4	77	1378	5.6	216	1378	15.7	217	1378	15.7	219	1378	15.9	285	1380	20.7	296	1380	21.4	288	1380	20.9	321	1380	23.3	308	1379	22.3
tRNA-His	2	69	2.9	10	69	14.5	6	69	8.7	9	69	13.0	10	69	14.5	9	69	13.0	9	69	13.0	10	69	14.5	12	69	17.4
tRNA-Ser (AGY)	3	59	5.1	9	59	15.3	9	59	15.3	11	59	18.6	10	59	16.9	15	59	25.4	14	59	23.7	11	59	18.6	13	59	22.0
tRNA-Leu (CUN)	2	72	2.8	6	71	8.5	5	71	7.0	2	71	2.8	5	71	7.0	6	71	8.5	7	71	9.9	4	71	5.6	5	71	7.0
NADH5	127	1810	7.0	254	1812	14.0	275	1812	15.2	287	1812	15.8	405	1818	22.3	418	1815	23.0	408	1815	22.5	425	1813	23.4	414	1815	22.8
NADH6	28	522	5.4	89	529	16.8	96	528	18.2	85	525	16.2	110	527	20.9	112	525	21.3	103	525	19.6	108	526	20.5	108	526	20.5
tRNA-Glu	5	69	7.2	12	69	17.4	10	69	14.5	11	69	15.9	11	69	15.9	11	69	15.9	11	69	15.9	15	69	21.7	15	69	21.7
cyt b	70	1141	6.1	174	1141	15.2	173	1141	15.2	193	1142	16.9	243	1142	21.3	232	1141	20.3	229	1141	20.1	219	1141	19.2	234	1141	20.5
tRNA-Thr	4	67	6.0	15	68	22.1	13	67	19.4	13	68	19.1	14	67	20.9	15	68	22.1	15	67	22.4	17	68	25.0	20	68	29.4
tRNA-Pro	3	69	4.3	8	69	11.6	9	69	13.0	10	69	14.5	20	69	29.0	17	69	24.6	19	69	27.5	16	69	23.2	17	69	24.6

* mis-: mismatch; length: bp; within Chlorocebus: between *C. sabaeus* and *C. aethiops*; rhesus: *M. mulatta*; barbary: *M. sylvanus*; baboon: *P. hamadryas* common chim: *Pan troglodytes*; pygmy chim: *Pan paniscus*; gorilla: *Gorilla gorilla gorilla*; orangutan: *Pongo pygmaeus*

2.4.2 Heteroplasmy in the control region of primate mtDNA

Heteroplasmy in the control region of primate mtDNA is not unusual. Xu and Arnason (1996a) confronted the problem of heteroplasmy when attempting to use uncloned PCR products for sequencing the heteroplasmic mtDNA region. Heteroplasmy usually relates to variation in repeat copy number. In a study of the Japanese macaque *Macaca fuscata*, using DNA samples isolated from blood and liver tissue, Hayasaka *et al.* (1991) identified six types of heteroplasmic variation in length of sequence in the control region from two geographic localities (12 out of 18 and 29 out of 31 individuals respectively). These heteroplasmic length numbers showed a 158 or 167/168 bp motif repeat. Mammalian mtDNA heteroplasmy caused by point mutation rather than variation of the number of repeat sequences has been reported in or around the control region (Hauswirth *et al*., 1984; Hanna *et al.*, 1998; Schwarts and Vissing 2002, Bayona-Bafaluy 2003). Jae-Heup *et al*. (2001) determined both repetitive and point heteroplasmic region in both HVS1 and HVS2 in the great cat. Boursot *et al*. (1987) even found a large deletion (5 kb) of mtDNA coding region in wild heteroplasmic mice *Mus mus musculus*. For human mtDNA, most point mutation heteroplasmy is reported to be associated with disease and aging (DiMauro *et al.*,1998; Chinnery and Turnbull, 1999; Chinnery *et al.*, 2000; Diaz, 2002). Hanna *et al.* (1998) detected a heteroplasmic mitochondrial DNA mutation in tRNA-Phe gene in muscle tissue associated with the mitochondrial encephalomyopathy, lactic acidosis, and strokelike episodes (MELAS). This mutation was not detected in the patient's blood or in her mother's blood. Schwarts and Vissing (2002) found muscle mtDNA of paternal origin in a myopathy patient, but the paternal types were completely undetectable from all other tissues tested, including blood, hair roots, and fibroblasts (Jacob 2002). Nevertheless, high levels of point heteroplasmy have been reported in HVS1 of the control region of normal human brain and hair tissues (Grzybowski, 2000; Tully *et al.*, 1999) and the point heteroplasmy in HVS1 and HVS2 remain constant during life (Lagerström-Fermér *et al*. 2001).

The sequence heteroplasmy detected in the 10 cloned amplicons in the present study was in blood with no other tissue investigated. This heteroplasmy was found in an 11-year old

monkey, so it is possible that it is age related (The animals in this colony have an average age of 11.6, excluding infant and experimental death, with a range from 6 to 23).

While the mismatch between DQ069714 and DQ069716, or DQ069717 and DQ069718 of less than 1% is consistent with the level of variation in other studies, the mismatch of DQ069715 of 27-28% with the other sequences is quite unprecedented and the possibility that it is an amplification artifact cannot be excluded. There are three potential reasons for observing sequence heteroplasmy: a. The amplicons are co-amplification of mtDNA and nuclear DNA sequences; b. The amplified copies are from tandemly duplicated mtDNA (control region) sequences; c. Different homologous mtDNA sequences coexist in a single individual (Abbott *et al*., 2005). The later is true heteroplasmy and the former two are artifacts. Hirano *et al*. (1997) provided the evidence for the co-amplification of nucleus-embedded pseudogenes of mtDNA cytochrome C oxidase genes in human. Davis *et al*. (1997) reported this artificial type of "heteroplasmy" present in 10-15% of normal human population and 20-30% of patients with sporadic Alzheimer's disease. Abbott *et al*. (2005) investigated 350 bp amplicons from the control region of 5 species of *Thalassarche*, and found tandemly duplicated control regions in their mitochondrial genomes. In 102 bp amplicons, artificial "heteroplasmic sites" were observed at frequencies of 15-18% for these 5 species.

By Blasting, the 5 heteroplasmic sequences reported in the present study against GenBank, similarities were found with sequences from other primates. There is more than 80% identity between *Macaca fuscata* sequence and DQ069717 or DQ069718); 79% between *Macaca mulatto* sequence and DQ069714 or DQ069716; and 90% between *Chlorocebus aethiops* sequence and part of DQ069715. However, more work is needed to clarify whether the heteroplasmic sequences reported in this thesis are from true heteroplasmy or amplification artifact caused by control region tandem duplication or co-amplification of nucleus-embedded mtDNA pseudogenes.

Chapter 3

Analysis of phylogeny of Cercopithecinae using complete mitochondrial DNA sequence

3.1 Introduction

3.1.1 Nomenclature of the green monkey

The nomenclature of the green monkey is quite confusing and the entire genus is in urgent need of revision (Groves, 2001). The green monkey has been classified to subspecies status as *Cercopithecus aethiops sabaeus* by Dandelot (1959) and Ruvolo (1988). It was classified as a species *Cercopithecus sabaeus* by Kingdon (1997) or lumped as *Cercopithecus aethiops* (including what others have recognized as *sabaeus*, *aethiops* and other species as well) by Wolfheim (1983). Groves (2001), in reviewing the taxonomy of the family Cercopithecidae, resurrected the generic name *Chlorocebus*, originally established by Gray in 1870 and called the green monkey *Chlorocebus sabaeus*. This name is used systematically throughout this thesis, although Newman *et al*. (2002) used the subspecies name *Chlorocebus aethiops sabaeus* and Grubb *et al*. (2003) retained the old genus name *Cercopithecus* for *sabaeus* (Table 3.1). Grubb *et al*. (2003) predicted that more species and subspecies would be identified in this genus.

Table 3.1 Historical changes in nomenclature for green monkeys from 1766 to January 2006

Time	species/subspecies name	Author
1766	*Samia sabaea*	Linnaeus
1845	*Cercopithecus chrysurus*	Blyth.
1850	*Cercopithecus werneri*	Geoffroy
1851	*Cercopithecus calitrichus*	Geoffroy
1959	*Cercopithecus aethiops sabaeus*	Dandelot
1983	*Cercopithecus aethiops*	Wolfheim
1988	*Cercopithecus aethiops sabaeus*	Ruvolo
1997	*Cercopithecus sabaeus*	Kingdon
2001	*Chlorocebus sabaeus*	Groves
2002	*Chlorocebus aethiops sabaeus*	Newman *et al*.
2003	*Cercopithecus sabaeus*	Grubb *et al*.
2006	*Chlorocebus sabaeus*	Lang

The *Chlorocebus sabaeus* monkeys from the Vervet Monkey Research Colony (VMRC) at University of California Los Angeles (UCLA) were originally from the island of St. Kitts. The earliest reported identification of the Kittitian monkeys was as *Cercopithecus callitrichus* (synonym of *sabaeus*) by Sclater (1866), followed by Elliot (1907) who identified Barbadian as well as Kittitian monkeys as *Cercopithecus callitrichus*. The species name *sabaeus* was subsequently used by many authors to refer to the old world nonhuman primate species that successfully established itself in the new world on the islands of St. Kitts and Barbados in the Eastern Caribbean (Colyer, 1949; Ashton, 1960; Sade and Hildrech, 1965; Poirier, 1972; Horrocks, 1985; Herbin *et al.*, 1997). However, Denham (1987) pointed out that their use of *sabaeus* is tentative rather than definitive.

One of the reasons for the West Indies monkey to be nominated as *sabaeus* is that the west African countries Senegal and Gambia were identified by Sade and Hildrech (1965) and Ervin and McGuire (1974) as the most likely place of origin of the green monkeys. The monkeys were introduced to the West Indies from French ships involved in the slave trade in the 17th century, as described by Labat (1722). What we are now calling *Chlorocebus sabaeus* was abundant in these west African countries (Figure 3.1).

However, having noted the existence of the *sabaeus* species, Sade & Hidrech (1965) also reported an anomalous population which differed in pelage color, skull and dental

features from *sabaeus* on St. Kitts and Nevis. There had been rare reports of other species of *Chlorocebus* or *Cercopithecus*, including *mona*, on St. Kitts (Hollister 1912) and *tantalus* (Horrocks, 1985) in Barbados (Schomburgk, 1848).

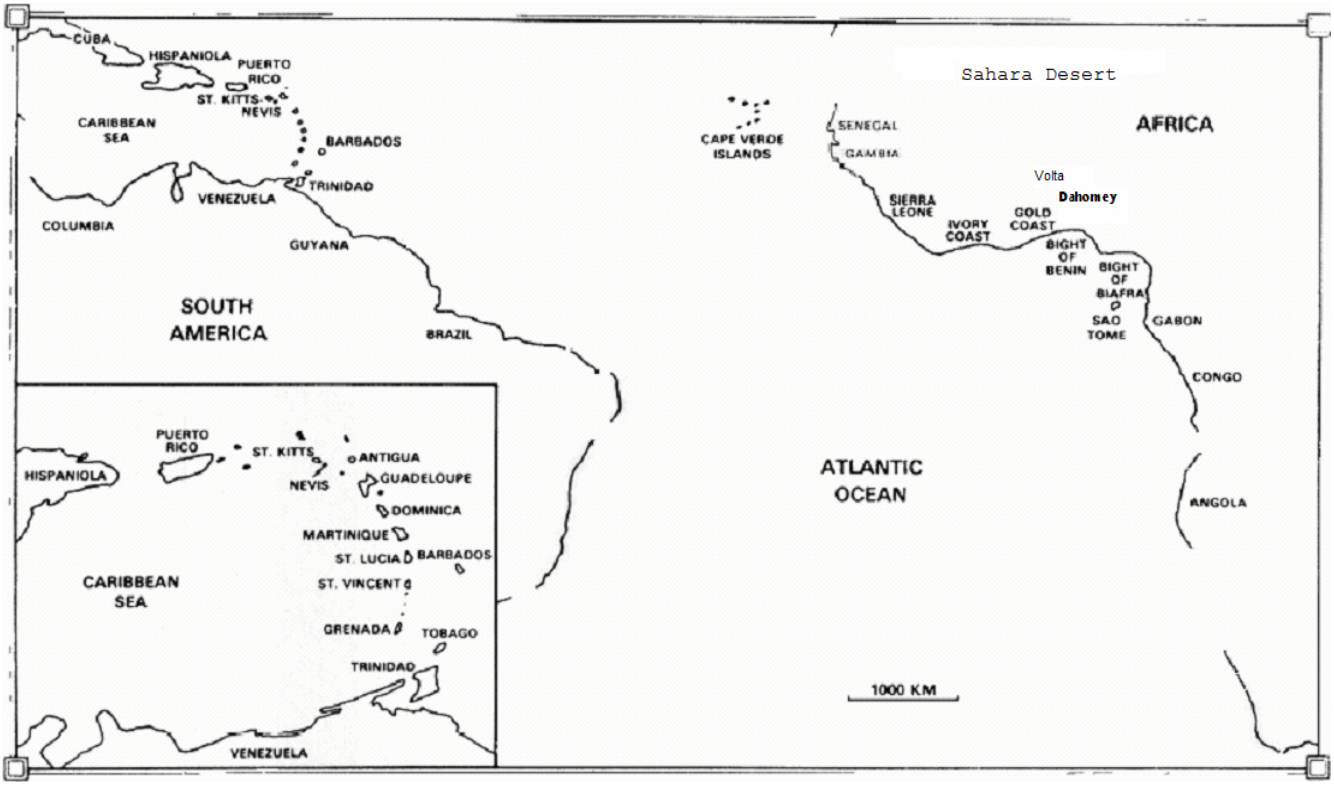


Figure 3.1 Map of West Africa and West Indies, showing their geographical proximity which supports the possibility of introduction of the green monkey to West Indies from Senegal and Gambia in west Africa

Booth (1956a, b) found the west African green monkey *sabaeus* occupied the part of Africa that begins in the east with the Volta and White Volta rivers and continues through the northern Ivory Coast west to Senegal (Figure 3.2). Dandelot (1959) agreed with Booth (1956a, b) on the distribution of the west African green monkey and pointed out that the definition of this animal was based on its distinguishing pelage. Tappen (1960) and Struhsaker (1967) have argued that the African green monkey is a species with varied pelage color. Wolfheim (1983) defined the species *C. aethiops* to include other species,

such as the west African green monkey described as *Chlorocebus sabaeus* in this thesis (Figure 3.2).

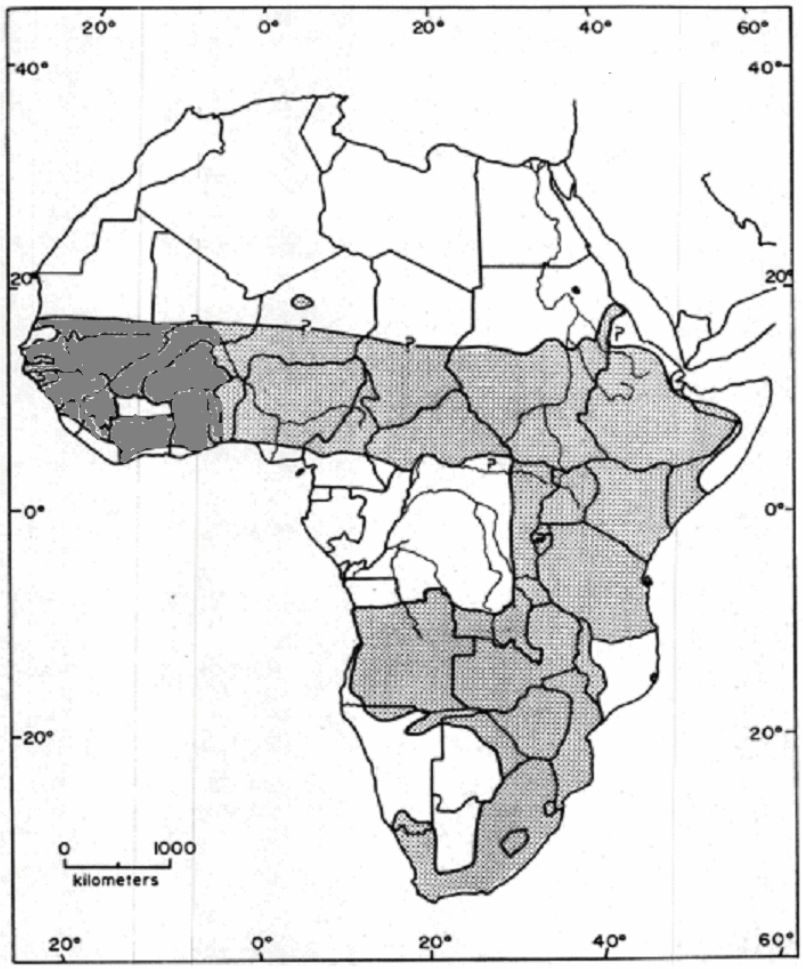


Figure 3.2 Distribution of *C. sabaeus* (dark shadowed area) and of *C. aethiops* as defined by Wolfheim (shadowed area) in Africa.

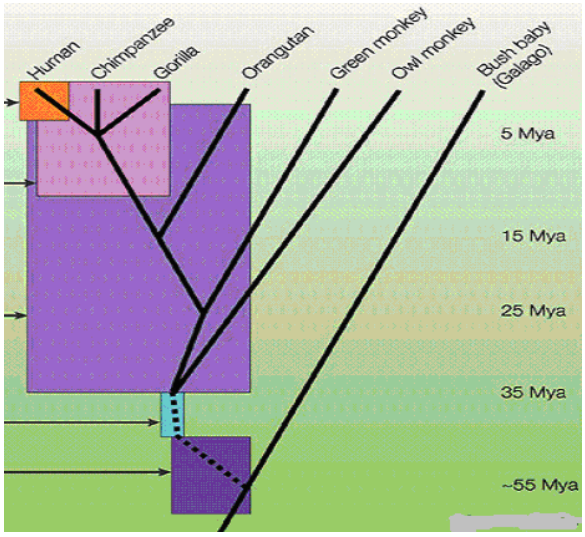
St. Kitts was a British colony from 1624, and occupied by both Britain and France between 1627 and 1665. Then the control of the island changed from one country to the other seven times until Britain finally acquired it permanently in 1713 (Dunn, 1973). From 1680 until 1807, British slavers supplied their colonies in the Caribbean with slaves from the Gold Coast and the Bights of Benin and Biafra, west Africa. During that period, less than ten percent of the British slaves originated in Senegal and Gambia (Handler and

Lange, 1978; Curtin, 1969). The historic data suggest that the Gold Coast, where both the *sabaeus* and *tantalus* species are present, and Angola and Mozambique, where both *pygerythrus* and *cynosurus* species occur, are the most likely sources of the Barbadian green monkeys (Handler and Lange, 1978; Curtin, 1969). The presence of green monkeys in the Cape Verde Islands raises the possibility that these islands may have been a stepping-stone to the West Indies, sometime between Portuguese settlement there in the 1460's and British settlement in Barbados, St. Kitts and Nevis in the 1620's (Jardine, 1833; Goldsmith, 1840; Duncan, 1972; and Denham 1987).

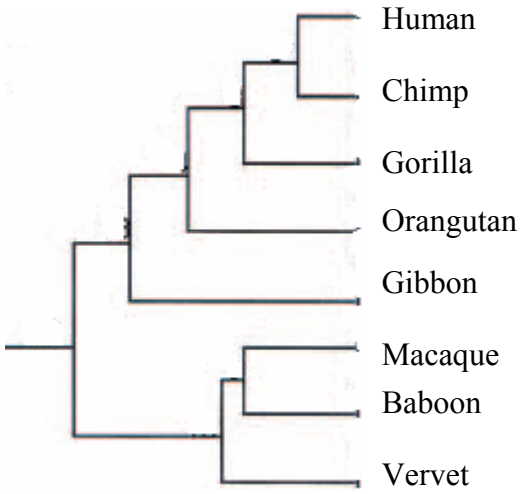
Denham (1987) investigated almost all possible historical documents, and concluded that it is conceivable that the St. Kitts green monkey originated elsewhere than the previously discussed west African countries. A Cape Verdian origin of the West Indian monkeys could not as yet be ruled out considering their geographical proximity to west Africa (Figure 3.1). The existence of a heterogeneous population in St. Kitts due to multiple introductions of monkeys is possible. The occurrence of infrequent hybridization between *sabaeus* and *tantalus* in Africa (Groves, 2001) could also explain the heterogeneous population of the green monkey in St. Kitts due to hybridized founders, or the population could be the product of interbreeding of animals introduced from several areas (McGure *et al*., 1974)

3.1.2 Phylogenetic analysis of the primate

The structure of the primate phylogenetic tree using the complete mtDNA gene sequence has been well established (Arnason *et al*. 1996b, 1998; Raaum *et al*. 2005) (Figure 3.3 c). This tree is consistent with the primate trees constructed with nuclear genomic markers (Figure 3.3a, b). The complete sequence of the mtDNA of the green monkey adds a new node to the tree and will be helpful in investigating more accurately the relationship between hominoids and old world monkeys, and the relationship within old world monkeys as well.

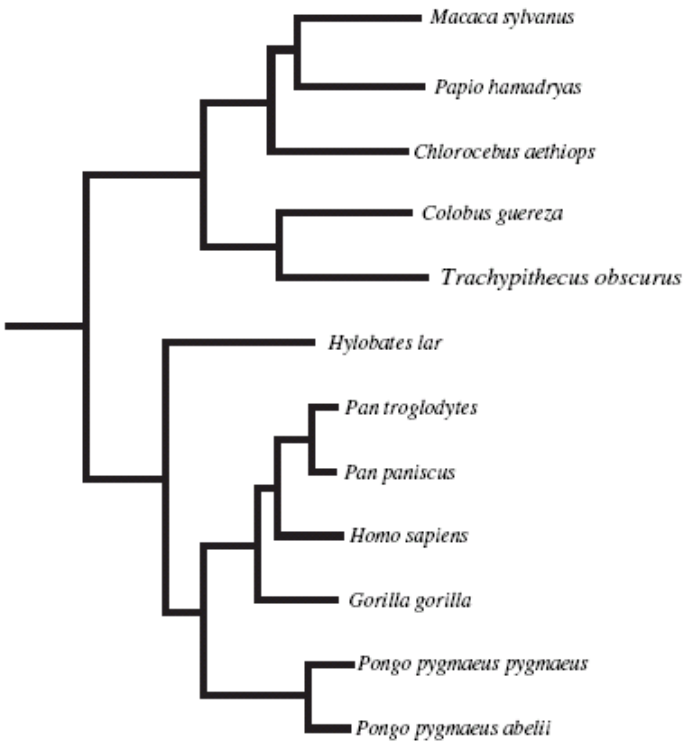


a. Primate divergences based on Alu repeat using bush baby (*Galago*) calibration, from Batzer and Deininger (2002).



b. Well established primate phylogeny based on targeted genomic region of nuclear genome, from Eichler and DeJong (1995).

Figure 3.3 (continued over page)



c. Mitochondrial DNA-based phylogeny of the primate species using an "all genes" alignment method, from Raaum *et al*. (2005)

Figure 3.3 (a-c) Consistency of primate phylogenies based on various data sets.

3.2 Materials and Methods

3.2.1 Data source

Complete mtDNA sequences for humans, common and pygmy chimpanzees, gorillas, orangutans (the family Hominidae), vervet monkeys, hamadryas baboons, rhesus monkeys and barbary apes (the subfamily Cercopithecinae of family Cercopithecidae) were obtained from GenBank. The new sequence reported in the present study was designated as *Chlorocebus sabaeus*. The trees established (see section 3.2.2 below) were compared to previously published well-know phylogenetic trees (Batzer and Deininger, 2002; Eichler and DeJong, 1995; Arnason *et al.*, 1998; Raaum *et al*., 2005) (Figure 3.3 b, c).

3.2.2 Phylogeny reconstruction among the species of primates

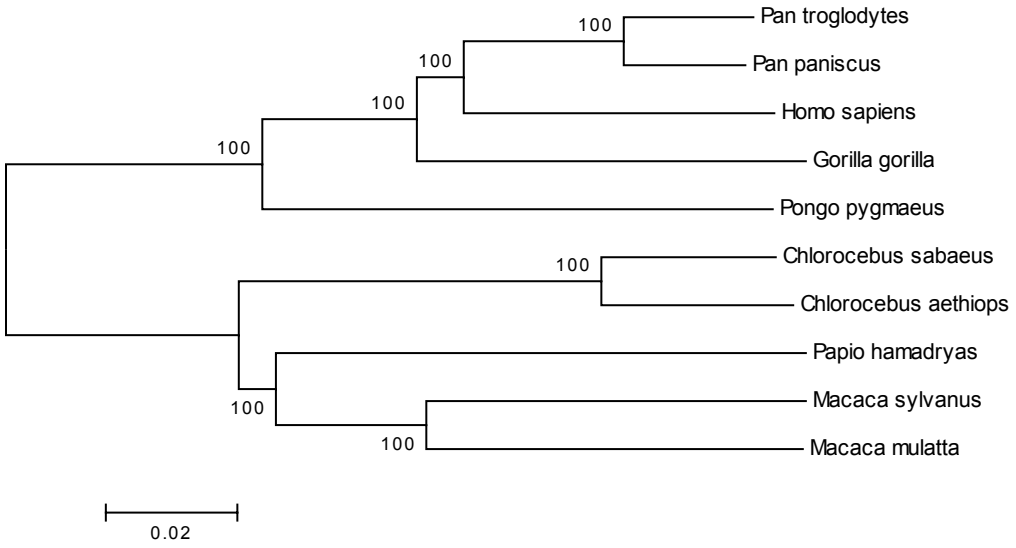
All phylogenetic trees were constructed using the neighbor-joining (NJ) method implemented in Mega software version 3 (Kumar *et al*. 2004). The NJ algorithm repeatedly combines closest taxonomic units based on the distance value, and eventually forms a tree with shortest total branch length (Saitou and Nei 1987). The Kimura 2-parameter model (Kimura 1980), which corrects for multiple hits, was used in the present analysis. A uniform evolution rate among the taxonomic units was assumed. The NJ method results in an unrooted tree, which has been displayed in a manner similar to the rooted tree here for easy inspection (Kumar *et al*. 2004). Bootstrap resamplings were performed 1050 times for all the phylogenetic trees. The transversion pairwise distance matrix used for the NJ tree was analyzed to investigate the taxonomic level of the two *Chlorocebus* species.

3.3 Results and discussion

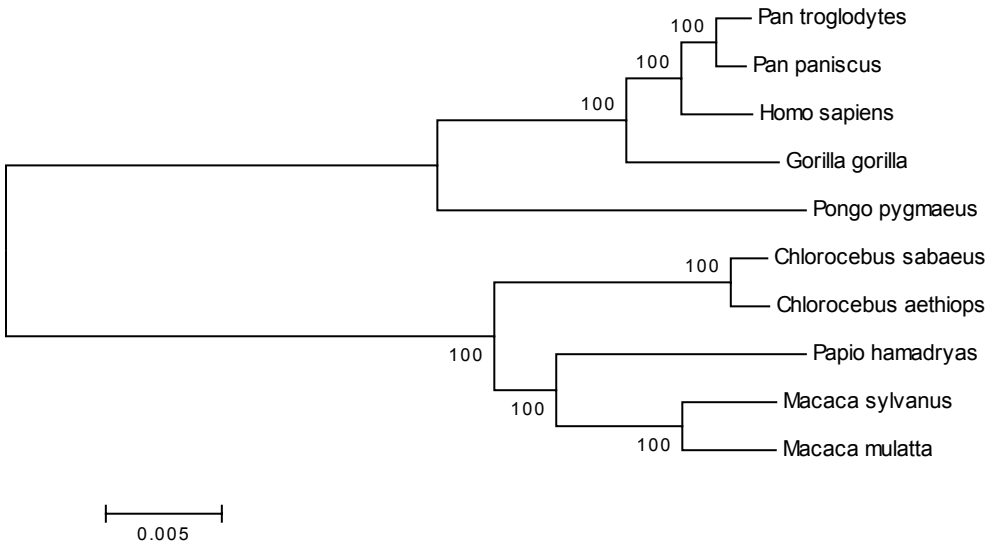
3.3.1 Phylogenetic analysis using complete mtDNA among primates

The topology of the phylogenetic tree (Figure 3.4) constructed with NJ using the concatenated sequence of mtDNA coding regions of these primates agrees with the common elements of the primate phylogeny based on nuclear and mitochondrial DNA sequence illustrated in Figure 3.3b, c. The newly constructed tree shows that two monophyletic lineages distinguish Cercopithecinae and Hominidae. At the most recent terminal branches there are three pairs of sister taxa, including the two species of chimpanzee (*Pan troglodytes* and *P. paniscus*), two species of *Macaca* (*mulatta* and *sylvanus*) and two species of *Chlorocebus* (*sabaeus* and *aethiops*). In the Hominidae lineage, the gorilla splits at a deeper branch than the human and chimpanzee branches, and in the Cercopithecinae lineage the *Chlorocebus* species radiated earlier than the baboon and the *Macaca* species. However, although supported with high bootstrap value, the tree constructed using transition + transversion substitutions (Figure 3.4a) showed a "leafy" phylogeny with much longer terminal branches than internal branches due to the

deep split on the tree (artifactually extending the divergence times for the terminal taxa) (Graf and Sparks 2000). Consequently for the "leafy" phylogeny, there is a problem in dating the divergence time between Cercopithecinae and Hominidae. If the date of the most recent common ancestor (MRCA) of human and chimpanzee is set at 5 MYBP (Adachi and Hasegawa, 1995; Gagneux *et al.*, 1999; Lewin, 1997), then the divergence time between Cercopithecinae and Hominidae will be underestimated at 13 MYBP, assuming that the root lies between cercopithecines and hominids. This is unreasonable. Using mtDNA coding sequences, Arnason *et al.* (1998) dated the divergence between Cercopithecoidae and Hominoidae at ≥50 MYBP, although previous authors calculated about 25 MYBP (Eichler and DeJong, 1995; Kumar and Hedges, 1998), and others computed up to 34.5 MYBP (Steiper *et al*. 2004). The misleading result produced by the "leafy" phylogenetic tree could be partially due to mtDNA evolutionary rate variation. The molecular clock can be affected by body size, metabolic rate, and generation time (Martin and Palumbi, 1993). The rate of nucleotide substitution for mtDNAs in hominines (human, chimpanzee, and gorilla) may have slowed down compared with that for old world monkeys (Hayasaka *et al.*, 1988), and adaptive evolution may cause accelerated evolution, even if the rate of transversion substitutions remain uniform among all lineages (Andrews *et al.*, 1998). When reconstructing the primate phylogenetic tree using transversions only, the branch topology remained the same, but the phylogenetic tree became much more stemmy (the internal branches become longer and the terminal branches relatively shorter) (Figure 3.4b). The divergence time between Cercopithecinae and Hominidae can then be calculated at 55 MYBP, assuming that the root lies between cercopithecines and hominids, which is close to that proposed by Arnason *et al.* (1998). The average transversion mutation rate for all coding genes would then be about 0.0006 per million years.



a "Leafy" phylogenetic tree constructed using transitions and transversions of the coding regions of the primates.



b. Stemmy phylogenetic tree constructed using transversions of the coding regions of the primates.

Figure 3.4 (a & b) Primate phylogenetic trees constructed using concatenated mtDNA gene sequence. Numbers on branching points are bootstrap values. The scale bar of units for each tree is to facilitate a comparison of branch length between different trees.

3.3.2 Taxonomic level of *Chlorocebus sabaeus* and *C. aethiops*

The Hominidae and Cercopithecinae phylogenetic tree constructed using data for transversions of the mtDNA coding regions (Figure 3.4b) shows that the most recent terminal branches consist of three sister taxa: two species of chimpanzee (*P. troglodytes* and *P. paniscus*), two species of *Macaca* (*mulatta* and *sylvanus*) and two species of *Chlorocebus* (*sabaeus* and *aethiops*). Among the three pairs, *C. sabaeus* and *C. aethiops* form a divergence branch of similar length to that of the two species of chimpanzee, which are inarguably distinct species. This result supports the distinct species status of the two *Chlorocebus* species, as proposed by Groves (2001). In a pairwise distance matrix (Table 3.2) generated during the phylogenetic tree construction procedure with transversion substitutions (Figure 3.4b), *C. sabaeus* and *C. aethiops* displayed a pairwise distance of 0.003, which is the same as that for the two chimpanzee species. Thus the new data support the species status of *Chlorocebus sabaeus* and *C. aethiops* (Groves, 2001).

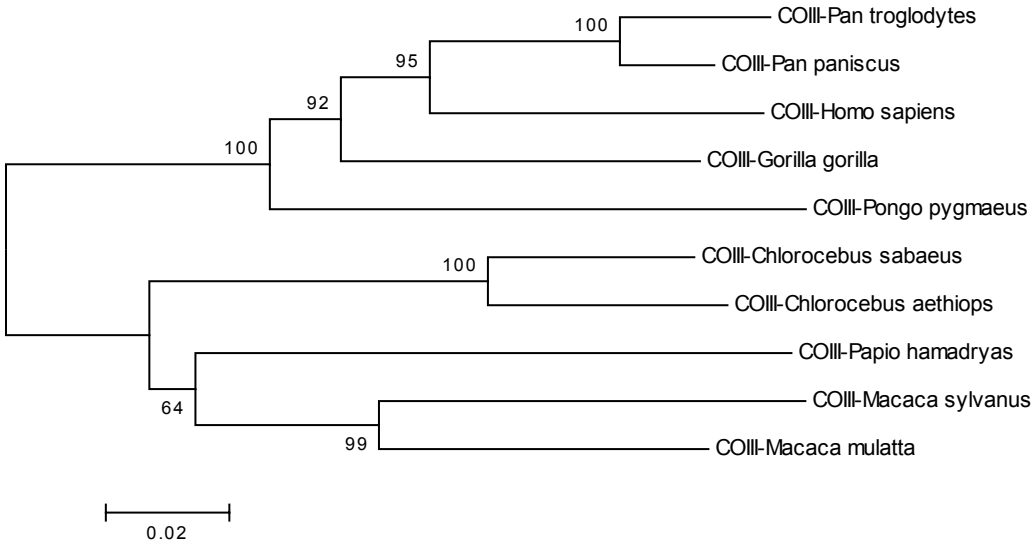
Table 3.2 Pairwise distance calculated using Kimura 2-parameter transversion model for phylogenetic tree construction among primate species in this study.

	H. sapiens	*P. troglodytes*	*P. paniscus*	*G. gorilla*	*P. pygmaeus*	*M. sylvanus*	*M. mulatta*	*C. sabaeus*	*C. aethiops*
Homo sapiens									
Pan troglodytes	0.006								
Pan paniscus	0.006	**0.003**							
Gorilla gorilla	0.012	0.012	0.012						
Pongo pygmaeus	0.029	0.029	0.029	0.031					
Macaca sylvanus	0.065	0.066	0.065	0.066	0.068				
Macaca mulatta	0.065	0.066	0.065	0.067	0.067	**0.008**			
Chlorocebus sabaeus	0.065	0.065	0.065	0.066	0.068	0.024	0.023		
Chlorocebus aethiops	0.065	0.065	0.065	0.066	0.068	0.024	0.024	**0.003**	
Papio hamadryas	0.067	0.067	0.066	0.068	0.068	0.02	0.02	0.025	0.026

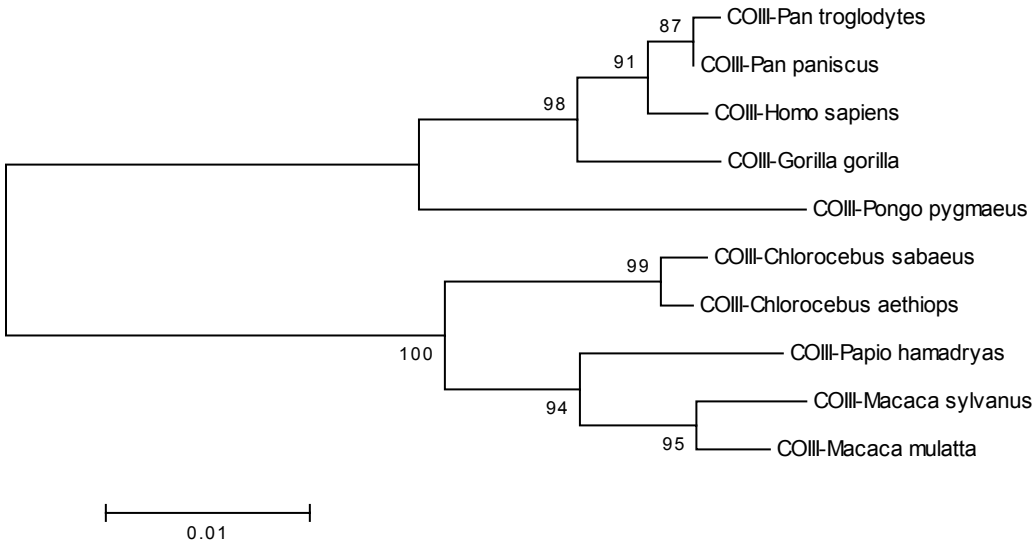
3.3.3 Phylogenetic analysis using single genes of mtDNA among primates

Mutation rate heterogeneity among genes can make a great impact on relationship inference for a single gene. Some biologically incorrect phylogenetic trees can be supported by a high bootstrap value (Takezaki and Gojobori, 1999), even when transversions that generally provided more accurate inference are used (Wu *et al.*, 2000).

For example, the primate phylogenetic tree based on transitions+transversions, and transversions only of COIII recovers the true topology (Figure 3.5a, b), but the 12S rRNA tree is misinformative for the Hominidae (Figure 3.5c, d), and NADH5 tree is misinformative for the Cercopithecinae lineage (Figure 3.5e, f). Like COIII, protein-encoding genes ATPase8, ATPase6, NADH2 and NADH4 recover the expected tree topology for both transitions+transversions, and transversions only. 16S rRNA, COI and Cytb recover the acceptable phylogenetic topology for transversions only, but not for transitions+transversions. In contrast, COII, ND1, ND4L and ND6 recover the proper tree topology for transitions+transversions, but not transversions only. As well as 12S rRNA, all the tRNA genes form an erroneous tree topology for both transversions+ transitions and transversions only (Table 3.3.).

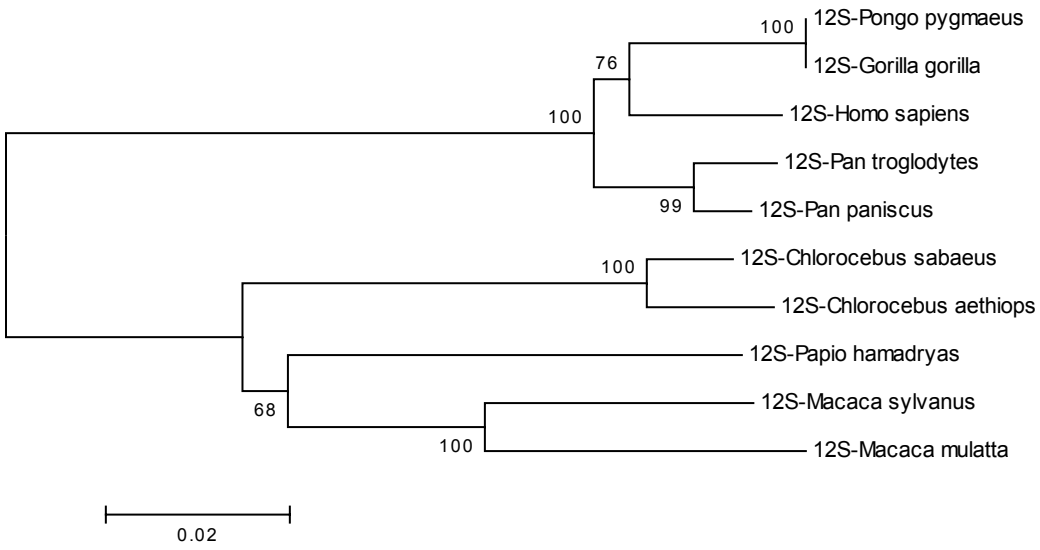


a. Informative phylogenetic topology using transitions and transversions of COIII.

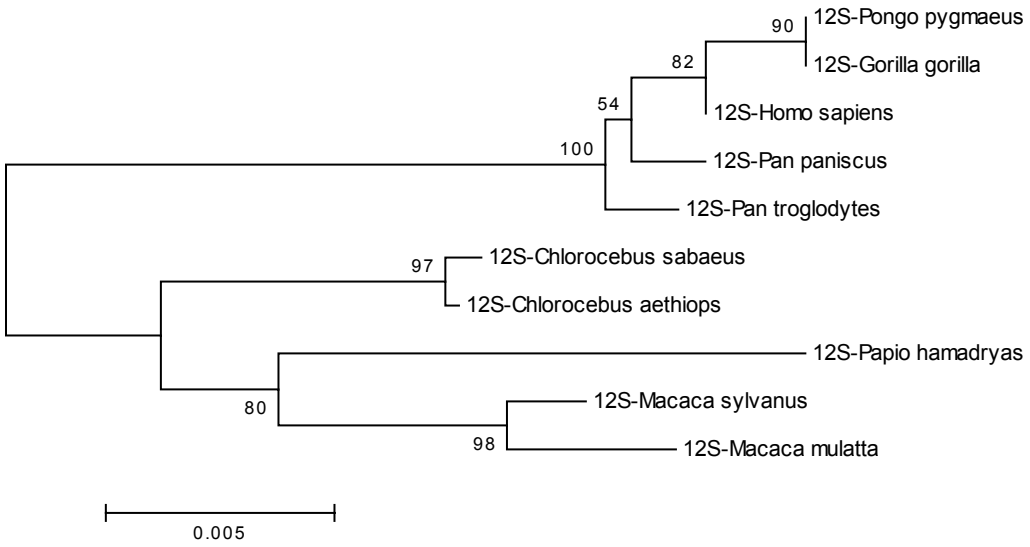


b. Informative phylogenetic topology using transversions only of COIII.

Figure 3.5 (continued over page)

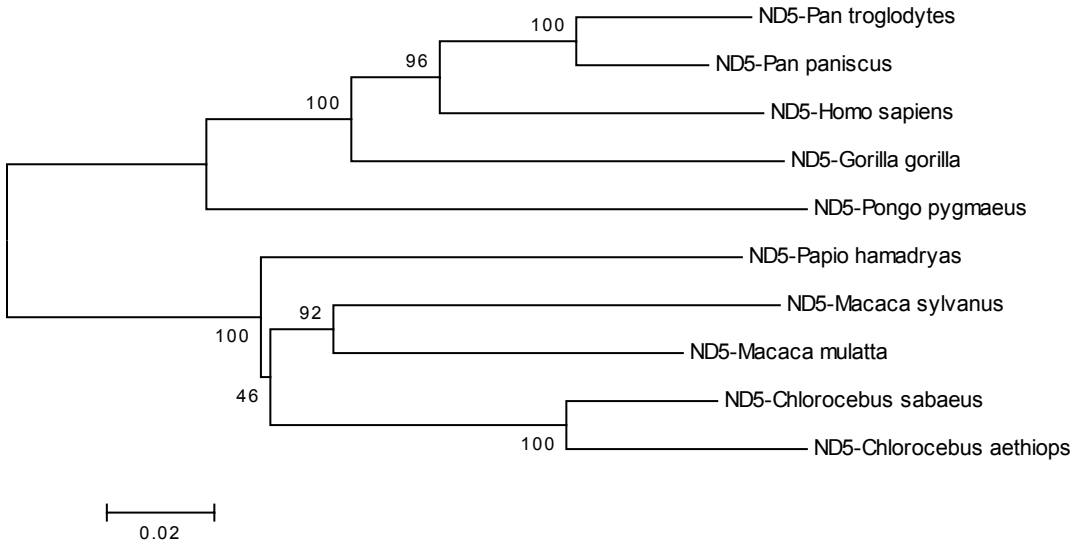


c. Misinformative (Hominidae lineage) phylogenetic topology using transitions and transversions of 12S rRNA.

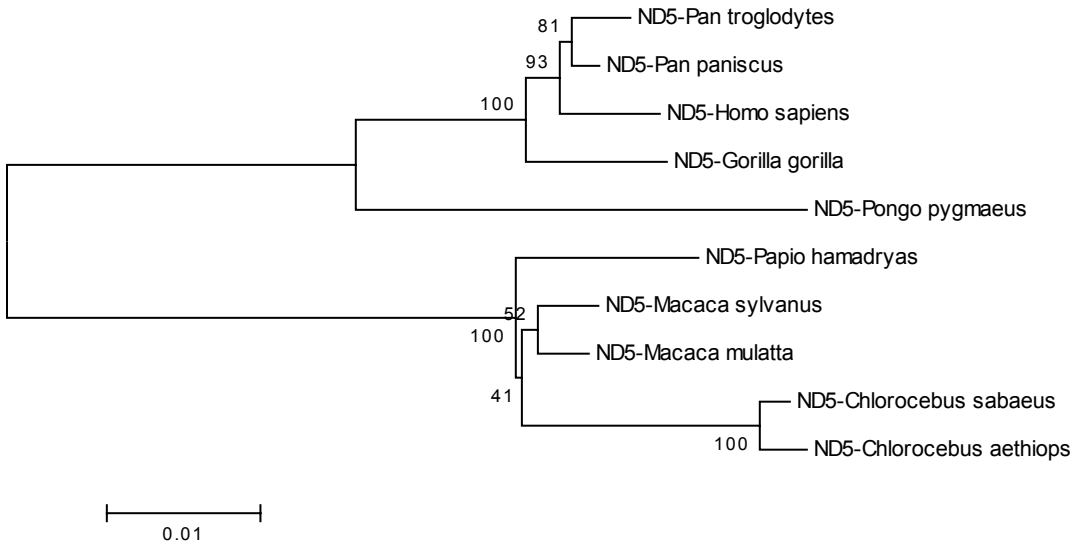


d. misinformative (Hominidae lineage) phylogenetic topology using transversions only of 12S rRNA.

Figure 3.5 (continued over page)



e. misinformative (Cercopithecinae lineage) phylogenetic topology using transitions and transversions of NADH5.



f. misinformative (Cercopithecinae lineage) phylogenetic topology using transversions only of NADH5.

Figure 3.5 (a-f) Phylogenetic trees of the primates reconstructed by single gene of mtDNA. Numbers on branching points are bootstrap values. The scale bar of units for each tree is to facilitate a comparison of brach length between different trees.

Table 3.3 Assessment of validity of phylogenetic trees recovered using single gene data. T--true (consistent with nuclear and mtDNA sequences illustrated in Figure 3.3), F--false (inconsistent with nuclear and mtDNA sequences illustrated in Figure 3.3)

	transition + transversion		transversion only	
	Hominoidae	Cercopithecine	Hominoidae	Cercopithecine
12S rRNA	F	T	F	T
16S rRNA	F	T	T	T
ATPase8	T	T	T	T
ATPase6	T	T	T	T
COI	T	F	T	T
COII	T	T	F	T
COIII	T	T	T	T
NADH1	T	T	T	F
NADH2	T	T	T	T
NADH3	T	F	F	T
NADH4L	T	T	F	T
NADH4	T	T	T	T
NADH5	T	F	T	F
NADH6	T	T	T	F
Cytb	T	F	T	T
tRNA-Phe	F	T	F	F
tRNA-Val	F	F	F	F
tRNA-Leu(UUR)	F	F	F	F
tRNA-Ile	T	F	F	F
tRNA-Gln	F	F	F	T
tRNA-Met	F	F	T	F
tRNA-Trp	F	F	F	F
tRNA-Ala	F	F	F	F
tRNA-Asn	T	F	F	F
tRNA-Cys	F	F	F	F
tRNA-Tyr	T	F	F	F
tRNA-Ser(USN)	F	F	F	F
tRNA-Asp	F	T	F	F
tRNA-Lys	F	F	F	F
tRNA-Gly	F	F	F	F
tRNA-Arg	F	T	F	F
tRNA-His	F	F	F	F
tRNA-Ser(AGY)	T	F	F	F
tRNA-Leu(CUN)	F	F	F	F
tRNA-Glu	F	F	F	F
tRNA-Thr	T	F	F	F
tRNA-Pro	F	F	F	F
HVS1	T	F	T	F

3.3.4 Conclusion

The mutation rate heterogeneity in the mtDNA of primates is a major problem for the construction of phylogenetic trees using a single or a few mtDNA genes (Hagelberg, 2003; Hasegawa *et al.*, 1993; Wakeley, 1993; Excoffier and Yang, 1999). Some genes produce anomalous branching patterns and exhibit discrepancies in the mtDNA clock between genes. All of the 22 tRNA genes showed a misinformative branching order, no matter whether transversions only or both transition and transversion substitutions were used. This could be caused by strong selective constraint in transfer RNA genes to ensure the functions of the stems, arms and loops in the cloverleaf structure (see discussion of previous chapter). All mRNA genes recover the true branching order when either transversions only, or transitions+transversions were used. For rRNA genes, 16S rRNA had fairly good performance, while 12S rRNA did not, when using either transversions only or transitions+transversions. For those misinformative genes, care must be taken when they are used to study long term evolution. However, misinformative genes cover only a small part of the complete mtDNA coding sequence, so the use of all coding sequence instead of just one or a few genes can avoid bias caused by rate heterogeneity between sites. As Russo *et al*. (1996) and Takezaki and Gojobori (1999) mentioned, a correct vertebrate phylogenetic tree can be obtained by using a large enough numbers of mitochondrial amino acid or nucleotide sequences, or the entire mtDNA sequence, even without considering rate variation across the sites in a unrooted tree.

If a single gene is to be used for phylogenetic study among distantly related taxa, such as from old world monkeys to human, good choices would be ATP6, ATP8, COIII, NADH2, and NADH4. These genes can provide a correct branch topology for both transitions+transversions and transversions only, with the latter giving better stemmy trees. This conclusion may change when a different range of target organisms are under study. When constructing trees for supposedly well known groups like whale, cow, rat, mouse, opossum, chicken, frog, and bony fish, Russo *et al.* (1996) believed that NADH5 produced the correct trees when either amino acid or nucleotide sequence data were used;

NADH4 and Cytb genes also produced correct trees in most tree-building algorithms using amino acid sequence data. By contrast, COII, NADH1, and NADH4L showed poor performance. Zardoya and Meyer (1996) suggested that the best genes to present a true picture of phylogenetic evolution among tetrapods in general or mammals in particular included NADH4, NADH5, NADH2, Cytb, and COI. Difference in phylogenetic performance cannot be completely explained by simple length differences and mutation rate differences between these genes. In general, larger genes produced better results; and the entire set of genes generated presumably true trees by all tree-building methods. (Russo *et al.*, 1996). Various factors determine the performance and the probability of discovering the correct phylogeny of these mitochondrial genes, such as density of lineage branching, the phylogenetic depth, lineage-specific rate heterogeneity, and the completeness of taxa representation (Zardoya and Meyer, 1996).

The major non-coding control region is hypervariable, and is considered to be more likely to show homoplasy than coding regions. This region is extensively used in human population or other intraspecies studies (Chen and Hebert 1999a; Helgason *et al*., 2003), but not for the study of long range evolution. The primate phylogenetic tree constructed using a sequence of the control region (around 950 bp, not shown) do not show any sibling relationship between the two *Chlorocebus* species, although the topology for hominoid species is reasonably distributed (Table 3.3). The performance of the control region is not as bad as some of the tRNA genes simply due to its length. All the tRNA genes of primate mtDNA are shorter than 80 bp.

The discrepancies in divergence dating in long range evolution are mostly attributed to variations of mutation rates. Using the wrong mutation rate for mtDNA data, often caused by study of an insufficient number of genes or insufficient length of sequence, could cause the dates for a presumed common ancestor to be either overestimated or underestimated (Kumar and Hedges, 1998). It is not surprising that different techniques, different external references, or different outgroups result in different divergence times (Arnason *et al.*, 1995; 1996b; Gagneux, 1999; Martin and Palumbi, 1993; Xu and

Arnason, 1996a; Adachi and Hasegawa, 1995). The use of transversion rates could be the most conservative approach to estimating timing in a long evolutionary history.

Chlorocebus aethiops and *C. sabaeus* are two closely related taxonomic groups whose taxonomic level has been controversial. A comparison of their relationship with that of the other two pairs of sibling species, i.e. chimpanzees and macaques, using the phylogenetic tree reconstructed in the present study and their consequent pairwise distance, showed that the two *Chlorocebus* species are as distinct as the well known pair of chimpanzee species. This strongly supports their species level as proposed by Groves (2001).

Since all mtDNA genes are on the one molecule and are maternal transmitted, lineage sorting and selective sweeps could have an impact on them (see Chapter 1 section 1.1.1). However, if the phylogenetic and evolutionary studies include mtDNA and other sex associated or autosomal polymorphisms, we can get more insight into factors such as different migration rate between genders, mutation susceptibility and rate differences between nuclear and mitochondrial genome, and selection pressure (Marvin *et al.*, 1993; Gissi *et al*., 2000; Battersby, and Shoubridge 2001; Helgason *et al*., 2003). However, ignoring these complexities, we know that primate mtDNA tree is consistent with the tree constructed with nuclear genomic markers (Figure 3.3 & 3.4). There is no doubt that mtDNA will remain an important source of data for evolution and phylogenetic studies in primates and other organisms in future.

Chapter 4

mtDNA SNPs detection and St Kitts monkey verification by 12S rRNA gene

4.1 Introduction

There have been disputes about single or multiple origins, and simple or mixed population structures of St. Kitts green monkeys (Denham, 1987; McGure *et al*., 1974) (see previous chapter). van der Kuyl *et al.* (1996) used 385 bp of mitochondrial 12S rRNA sequences to argue that the St. Kitts green monkeys originated from the adjacent ports of Senegal and Gambia, west Africa. They compared mtDNA sequence of a single monkey from St. Kitts to reference sequences from several African species. They also detected a single nucleotide difference between monkeys belonging to the same nominal species (*Chlorocebus sabaeus*) living in the new world and the old world. They suggested that speciation in West Indian *C. sabaeus* is in progress. However, their results are insufficient to exclude multiple origins for the green monkey populations on St. Kitts as the explanation for this difference.

The green monkeys in VMRC of UCLA are a resource for neurobehavioral and social genetic studies. However, for individual identification and pedigree verification in these animals, detailed information about their genetic characteristics, population structure and closely related species is required. The sequence and annotation of the complete mtDNA from a VMRC monkey has been reported in Chapter 2 and analyzed in Chapter 3. Here the detection and development of mtDNA molecular markers, specifically single nucleotide polymorphism (SNP), are reported.

Genotyping of SNPs requires high-throughput genotyping technologies. These technologies include allele-specific PCR (Liu *et al.*, 1997), DNA microarrays (Wang *et al.*, 1998), Taqman (Applied Biosystems. Foster City, CA), the use of fluorogenic probes

and the 5' nuclease assay (Lee *et al*., 1993), RFLP (Kan and Dozy, 1978; Gardner and Wagner, 2005; Zhang *et al*., 2005), oligonucleotide ligation assay (Grossman *et al.*, 1994; Iannone *et al.*, 2000), single base chain extension (Pastinen *et al.*, 1997; 2000), pyrosequencing (Alderborn *et al.*, 2000), multiplexed solid-phase amplification (Shapero *et al.*, 2001), degenerate oligonucleotide primer (DOP)-PCR (Jordan *et al.*, 2002), and denaturing high performance liquid chromatography (dHPLC). Among these, dHPLC is a relatively simple, robust and cost-effective way for genotyping large numbers of SNPs, and has been used in the present study.

The identification of SNPs and comparison of the UCLA green monkeys with those from St Kitts and with other African species using part of 12S rRNA gene are based on the new mtDNA sequence reported in Chapter 2. Although the data are limited, this genetic information will contribute to better biological and taxonomic understanding of these monkeys.

4.2 Materials and Methods

DNA was isolated from blood samples collected from the Vervet Monkey Research Colony (VMRC), at the University of California Los Angeles (UCLA), USA. See chapter 5 for detailed DNA isolation methods. mtDNA sequence from African and St Kitts green monkey mtDNA sequences reported by van der Kuyl *et al.* (1996) were used for comparison with the novel sequence studied in the present research.

4.2.1 PCR

Twenty two mtDNA segments to be used for denaturing high performance liquid chromatograph (dHPLC) detection were amplified by PCR from 96 animals. The PCR primers are listed in Table 4.1. The PCR reaction mix of 20 µl contained 12 µl PCR Master Mix (AmpliTaq Gold, 50U/ml, Applied Biosystems, Foster City, CA), 6 µl DNA (10ng/ µl) and 1 µl of each primer (10 µM). The thermal cycle program comprised one cycle at 94°C for 1 min, followed by 30 cycles at 94°C for 30 sec, 50°C for 45 sec, 72°C for 50 sec. A final cycle was done at 72°C for 5 min.

Table 4.1 mtDNA amplicons tested by dHPLC for SNPs. *The primer sequences are listed in Table 2.1 of Chapter 2. ** Reference samples are randomly selected after amplification (see text below).

mtDNA segment	start nt no.	end nt no.	forward primer*	reverse primer*	length (bp)	reference sample**
DE4	607	1174	L593	H1183	567	1991016
TU	1101	1470	L1093	H1480	369	1995047
DE1	1316	1892	L1299	H1904	576	1989086
DE5	2002	2473	L1991	H2485	471	1986017
DE3	2828	3339	L2813	H3349	511	1985005
COIab	5967	6570	L5959	H6578	603	1991030
AT516	6492	7019	L6482	H7035	874	1999031
AT824	6804	7393	L7349	H9242	527	1999031
AT1239	7223	7972	L7205	H8028	739	1993020
COII	7804	8180	L7611	H8192	376	1991030
COII1841	7609	7767	L7807	H8192	158	1994020
COII1735	7701	7821	L7701	H7821	120	1994020
COII1645	7804	8180	L7611	H7766	376	1994020
AT2	8142	8670	L8128	H8680	528	1980002
QQ	12055	12330	L12024	H12327	275	1997007
QP1	13014	13608	L12998	H13618	594	1989004
QP2	13707	14281	L13693	H14279	574	1993025
QP3	14676	15254	L14659	H15267	578	1991030
cyb	15243	15845	L15234	H15853	602	1998013
QP4	15770	16226	L15757	H16236	456	1991035
DE2	16265	16593	L16245	H36	328	1992051
QP5	16351	16837	L16340	H301	486	1992019

4.2.2 Single nucleotide polymorphisms (SNP) screening using dHPLC

Unpurified PCR product was mixed at an equimolar ratio with a randomly nominated reference PCR product and subjected to an 8 min 95°C denaturing step followed by gradual reannealing from 95°C to 25°C (70 steps of 1°C in total) with 19 sec per step. This allowed formation of less thermally stable heteroduplexes where the unknown test sample differed from the reference sample.

The hybridized PCR products were tested for presence of heteroduplexes using the dHPLC system, provided by the sequencing core at the UCLA. About 10 µl of each mixture was loaded onto a DNASep2 column (Transgenomic, San Jose, CA), and the

amplicon was eluted in 0.1 M triethylammonium acetate, pH 7, and 25% acetonitrile, with a linear acetonitrile gradient at a flow rate of 0.9 ml/min. The appropriate temperature conditions were optimized by computer simulation (http://insertion.stanford.edu/melt.html).

The elution profile was analyzed for evidence of a heteroduplex, recognized by the appearance of two or more peaks in the elution profiles (see result section).

4.2.3 Sequence and haplotype analysis of SNPs

The 22 segments covering 7439 bp of the mtDNA (Figure 4.1) were tested by dHPLC on 96 samples to recognize sequence variants from the elution profiles. For those elution profiles indicating presence of a heteroduplex in the reference/test mixture, the selected test PCR products were sequenced to identify the novel sequence variant and determine their haplotype. All the 96 haplotype were analyzed and summarized based on the elution profile to discern the haplogroups.

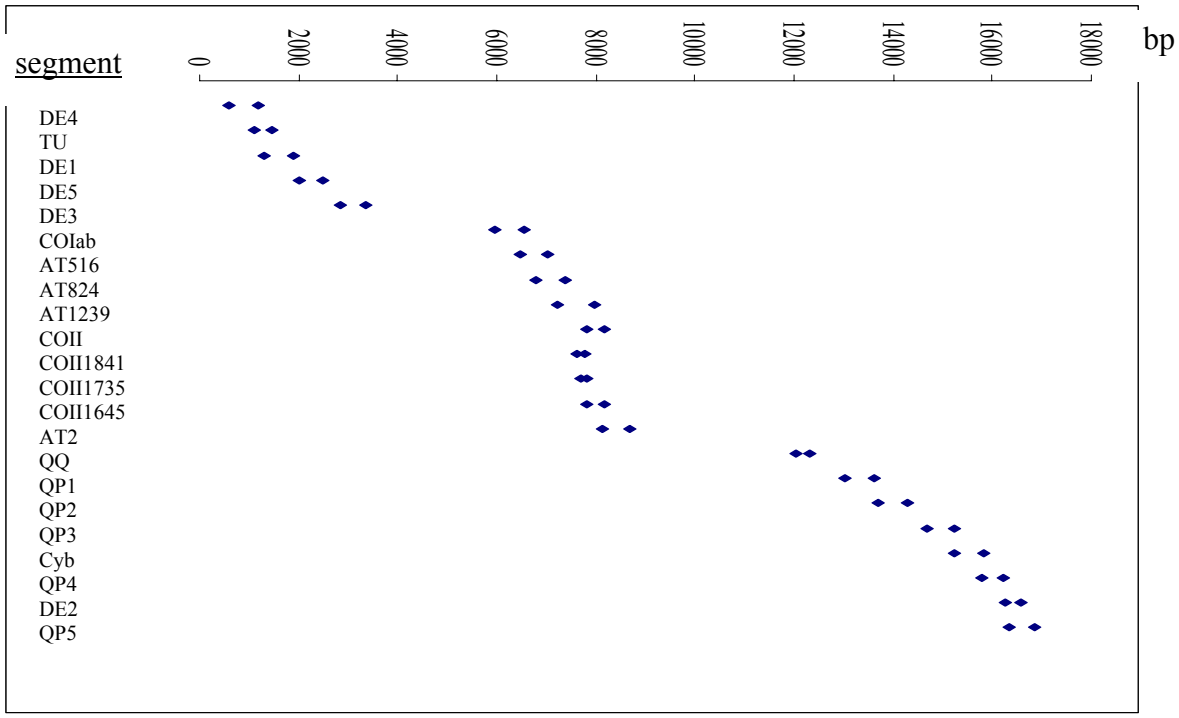


Figure 4.1 Start-End points of amplicons analysed by dHPLC in relation to the complete mtDNA

4.2.4 12S rRNA analysis of geographical origin of St Kitts monkeys

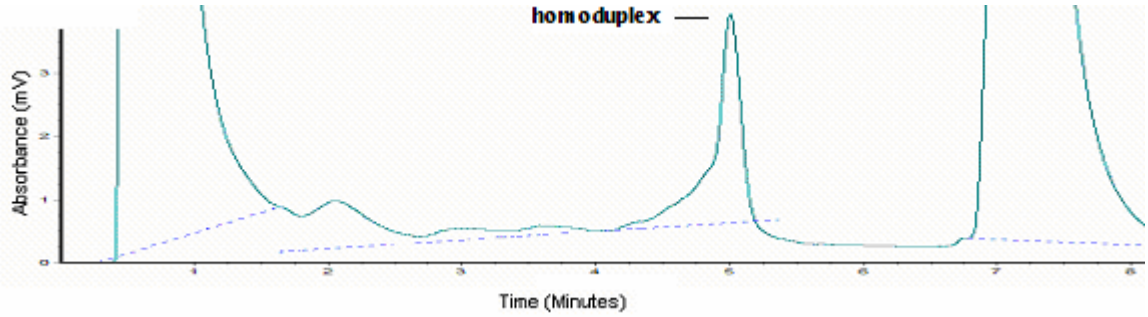
The new mtDNA sequence reported in previous chapters was aligned with 385 bp of 12S rRNA sequence produced by van der Kuyl *et al.* (1996). Two of the 385 bp sequences from UCLA green monkeys that had different SNP genotypes were compared with seven sequences produced by van der Kuyl *et al.* (1996), which included one St. Kitts green monkey, one African green monkey and five other closely related species from Africa.

4.3 Results

4.3.1 SNP identification by dHPLC

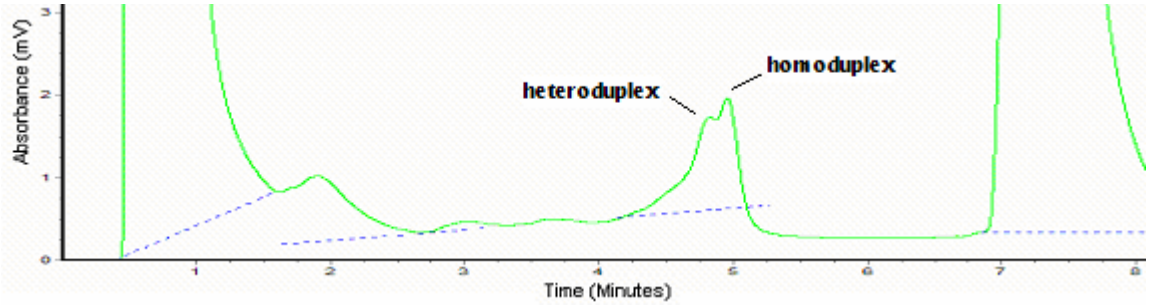
Whenever the dHPLC elution profile showed evidence of an early eluting heteroduplex indicative of a sequence difference between the test animal and the reference (Figure 4.2 as examples), the test sample was grouped to others with similar elution profiles. Selected

samples from each group were sequenced. Together with five SNP sites found during complete genome sequencing, a further 55 novel SNPs were found, excluding the heteroplasmic region (Table 4.2). These SNPs included one insertion/deletion (indel) and one di-nucleotide substitution. Most of the SNP variants were transitions, with only 9.26% being transversions (Table 4.2). Among the 22 segments tested by dHPLC, no SNPs were detected in DE5 and DE3. This could be due to insufficient sampling for sequencing or an inappropriate melting temperature. DE5 is within the 16S rRNA, while DE3 covers part of 16S rRNA and the whole Leu tRNA gene. As well as these two regions, DE4, AT2, QP2, QP4, DE2, and QP5 were not informative in the dHPLC elution profiles, although polymorphisms were detected by sequencing randomly selected samples. QP4 and DE2 (with 18 polymorphic sites) span the HVS1 region and QP5 (with 13 polymorphic sites) spans the HVS2 region. Hypervariability with multiple SNPs within a single segment was hard to detect by dHPLC due to the complicated elution profiles. Other uninformative segments may not have been tested at the optimal temperature.

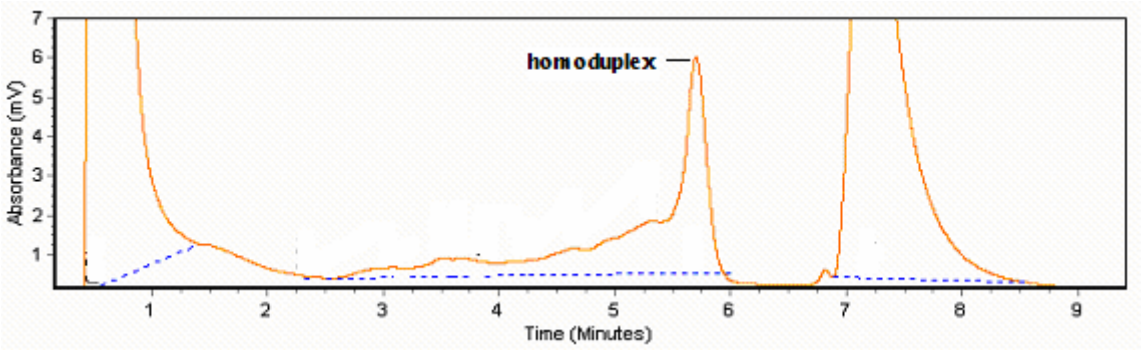


a. 1989087 + reference sample, position 1235 C

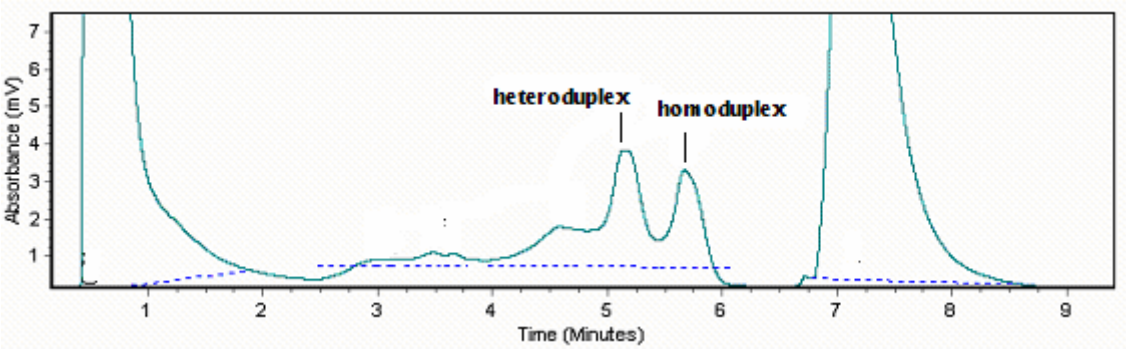
Figure 4.2 (continued over page)



b. 1992013 + reference sample, position 1235 T



c. 1988008 + reference sample, position 7838 G



d. 1990008 + reference sample, position 7838 A

Figure 4.2 (a-d) Examples of dHPLC profiles of mtDNA segments on the green monkeys, showing homoduplex only (a, c) and heteroduplex plus homoduplex (b, d). a and b are from fragment TU, and c and d are from segment COII1645.

Table 4.2 SNP sites in the green monkey *Chlorocebus sabaeus* mtDNA, excluding the heteroplasmic region, found both during routine sequencing (Chapter 2) and SNP sequencing. * :-A, insertion/deletion of nucleotide A.

SNP No.	position (NT No.)	base variation	SNP No.	position (NT No.)	base variation
1	1106	G-C	28	13812	T-C
2	1107	G-C	29	13905	G-A
3	1235	T-C	30	14058	T-C
4	1520	T-C	31	14082	T-A
5	1543	G-A	32	14214	G-A
6	1639	G-A	33	14980	T-C
7	1839	T-C	34	15157	G-A
8	5097	G-A	35	15442	T-C
9	5285	G-A	36	15703	G-A
10	5315	T-C	37	15958	T-C
11	5322	T-C	38	16150	T-C
12	6123	T-C	39	16164	G-A
13	6708	T-C	40	16175	G-A
14	6951	G-A	41	16181	T-C
15	7678	T-C	42	16190	G-A
16	7777	C-A	43	16202	T-C
17	7839	G-A	44	16259	G-A
18	8273	G-A	45-46	16276-7	CT-TC
19	8489	G-A	47	16286	G-A
20	8614	G-A	48	16296	T-C
21	12052	T-C	49	16299	T-C
22	12302	T-C	50	16309	:-A*
23	13089	T-C	51	16320	T-C
24	13120	T-C	52	16371	T-C
25	13332	G-A	53	16374	T-C
26	13740	T-C	54	16502	T-C
27	13773	T-C	55	16516	A-G

Haplotypes were constructed from all 96 animals studied by using their dHPLC elution profiles to infer the sequence variant at particular sites within those segments. Only 18 from the 55 polymorphisms reported in Table 4.2 could be scored in this way. These 18 polymorphic sites found in 10 segments were used for haplotype analysis. All the 96 haplotypes studied in dHPLC were used to summarize haplogroups, and 10 haplogroups were found (Table 4.3). These 96 samples are from 22 out of 24 matrilineal lines in the UCLA VMRC. Unlike sequence information that can be used to calculate a distance matrix, the haplogroup information is from limited sites and is best treated as character information. An unrooted haplogroup affiliation tree for these haplogroups was established (Figure 4.3) by Maximum Parsimony, which is based on shared and derived

characters (Nei and Kumar, 2000). Only those sites appearing at least twice were used by the Maximum Parsimony method.

Table 4.3 SNP haplogroup in the VMRC green monkey colony. Y=T-C; R=G-A.

segment	TU	DE1				COIab	AT516		AT123	COII	QQ	QP1			QP3		cyb		No.
Position	1235	1520	1543	1639	1839	6122	6708	6951	7678	7839	12302	13089	13120	13332	14980	15157	15442	15703	
SNP	**Y**	**Y**	**R**	**R**	**Y**	**Y**	**Y**	**R**	**Y**	**R**	**R**	**Y**	**Y**	**R**	**Y**	**R**	**Y**	**R**	
1	C	C	A	A	T	C	C	A	C	G	A	C	T	A	T	G	C	G	19
2	C	C	A	A	T	C	C	A	C	A	A	C	T	A	T	G	C	G	52
3	T	C	A	A	T	C	C	A	C	G	A	C	T	A	T	G	C	G	1
4	T	T	G	G	C	T	T	G	T	A	G	T	C	G	C	A	T	A	11
5	T	T	G	G	C	T	T	G	C	A	G	T	C	G	C	A	T	A	1
6	T	T	G	G	C	T	C	A	C	A	G	T	C	G	C	A	T	A	1
7	T	T	G	G	C	T	C	A	T	G	G	T	C	G	C	A	T	A	7
8	T	T	G	G	C	T	T	G	T	A	G	T	C	G	C	A	C	G	1
9	T	T	G	G	C	C	C	A	C	G	G	T	C	G	C	A	C	G	2
10	T	T	G	G	C	C	T	G	T	A	G	T	C	G	C	A	C	G	1

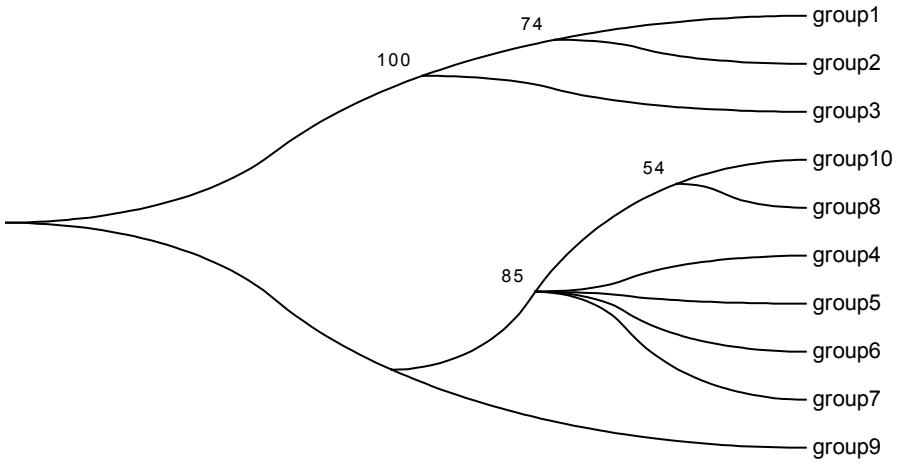


Figure 4.3 SNP Haplogroup affiliation tree (Maximum Parsimony) (displayed in a manner similar to a rooted tree). Numbers at branch points are bootstrap values.

4.3.2 Verification of species of the St Kitts green monkey by 12S rRNA gene

The UCLA VMRC was established using 57 green monkeys that were wild caught on the island of St. Kitts. Comparison of part of the new 12S rRNA sequence (385 bp) of the VMRC monkeys produced in this study with sequences published by van der Kuyl *et al.*

(1996) is consistent with their claim (Figure 4.4) that the green monkeys (from both St. Kitts and UCLA) have a very similar sequence to that of *Chlorocebus sabaeus* from west Africa. A minor difference, detected by van der Kuyl *et al.* (1996) between the St Kitts monkeys and African *C. sabaeus*, was interpreted by them as evidence of incipient speciation in the Caribbean. However in this study it was actually detected as a single nucleotide polymorphism within the VMRC colony and by inference at St Kitts as well.

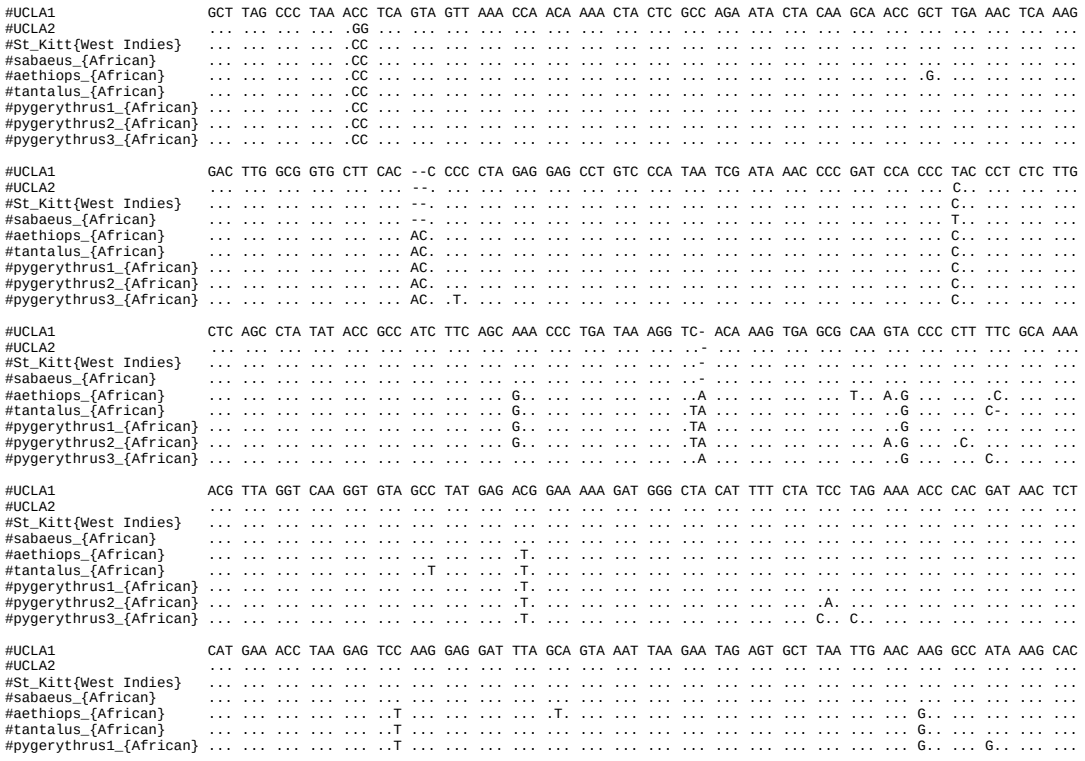


Figure 4.4 12S rRNA sequence comparison among geographic populations of *Chlorocebus sabaeus*, and other sibling species in Africa. UCLA1 and UCLA2 are new sequences produced by the present research. The other sequences were produced by van der Kuyl *et al.* (1996). Dot (.) = identical to reference sequence (UCLA1). Dash (-) = deletion.

4.4 Discussion

For dHPLC detection of single nucleotide polymorphisms, the ideal amplicon is around 500 bp in length, and should carry no more than one or two polymorphic sites. In some hypervariable region, such as the control region, the very high level of polymorphism makes the dHPLC elution profiles so complicated that SNP recognition is difficult. An example from segment DE2 is shown in Figure 4.5. This dHPLC profile showed multiple peaks even when tested at two different melting temperatures. When a few samples were randomly selected for sequencing, 14 SNPs were found within this segment, but none can be easily recognized from the dHPLC profile. Correct approaches to SNP detection generally involve large scale repeated sequencing and bioinformatics tools for SNP recognition as well as automated tools for genotyping (Stephens *et al.*, 2006). This approach is more comprehensive and accurate for SNP analysis.

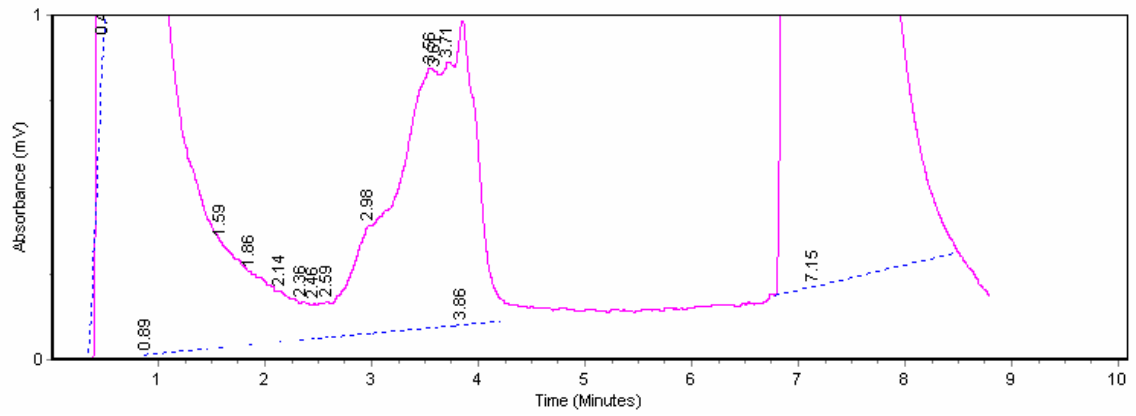


Figure 4.5 dHPLC profile for 1993029 + reference sample in segment DE2 of the control region, showing multiple peaks that are difficult to distinguish heteroduplex from homoduplex.

The 55 SNPs detected in the present study are predominantly transitions, which account for 90.74% of the total substitutions. This is very close to the reported frequency of transitions (around 92%) in mtDNA within humans, chimpanzees, and gorillas (Brown *et al.,* 1982). Although these 55 SNPs cover nearly half of the complete mtDNA genome of the green monkey, more work needs to be done to screen the whole length of mtDNA,

especially in the hypervariable control region where many more SNPs are likely to be found.

A comparison of 385 bp of the 12S rRNA gene among green monkey from different geographic localities and from other closely related monkey species showed that the UCLA green monkeys, and consequently the St. Kitts green monkey, are more similar to the West African green monkey *Chlorocebus sabaeus*. When more UCLA green monkeys were genotyped, it became clear that a SNP existed in the 385 bp sequence of 12S rRNA, contrary to the implication by van der Kuyl *et al.* (1996) of a fixed difference between St. Kitts and African green monkeys. van der Kuyl *et al.* (1996) used only one individual of the St. Kitts green monkey, so they had no chance to identify both variants. Their suggestion of incipient speciation after these animals colonized the Caribbean island of St. Kitts cannot be sustained.

SNPs are very useful tools in species/population comparison. When enough SNP markers are available in all target animals, an overall picture can be drawn about their relationships, taxonomic status and evolution. In the present study, only the 385 bp sequence of 12S rRNA was available for comparative studies. This covers only 2 of 55 SNPs detected in this study. In future, it will be informative to examine sequences containing these other sites in comparison between *Chlorocebus sabaeus* from different localities, and in comparisons between *Chlorocebus sabaeus* and other closely related species.

Chapter 5

Paternity and pedigree analysis using cross-amplified microsatellites in green moneys

5.1 Introduction

5.1.1 Vervet Monkey Research Colony

The Vervet Monkey Research Colony (VMRC), at the University of California Los Angeles (UCLA), USA was established in 1975 under the directorship of Dr. Michael McGuire (www.npi.ucla.edu/center/primate/index.html). From 1975 to 1987, a total of 57 green monkeys *Chlorocebus sabaeus*, which were wild caught from St. Kitts Island, West Indies, were introduced into the colony. Since then close to 1500 animals have been born into the colony. At the beginning of this project, the colony consisted of over 560 monkeys. Under the current directorship of Dr. Lynn Fairbanks, the main focus of the research at the colony is to study genomic mapping of the green monkey, behavior and social genetics, and other fundamental biological issues, in collaboration with the UCLA Center for Neurobehavioral Genetics and Department of Human Genetics. The large, extended pedigree is potentially useful for genetic linkage analyses aimed at finding genes contributing to specific behavioral and physiological traits.

The animals are housed in 22 social groups, mimicking the natural social composition of green monkey troops in the wild. Each outdoor animal enclosure has a unique name and the design of each enclosure is slightly different. However, all of the enclosures have the same components: a chain-link fenced outdoor area, a wooden night-room, an ante-room, and a squeeze tunnel. Animals can move to a new group from time to time, as occurs in nature.

The AnimalID name is a 7-digit number, the first 4 digits indicate the year of birth and the last 3 digits represent the order of birth in that year. Individual animal records

include AnimalID number, current enclosure name, enclosure history, putative mother's AnimalID number, possible fathers, sex, date of birth and date of death (if applicable), medical records, study records, weight/length records, and blood sample records.

5.1.2 Paternity assignment

Paternity testing and retrospective pedigree assignment are the purposes of the study of these green monkeys reported in this chapter, as accurate and complete pedigree records are essential for functional genomic studies such as quantitative trait locus mapping.

The exclusion principle is the fundamental basis for excluding parents of individuals, on the basis of incompatibilities of Mendelian inheritance rules between parents and offspring. Exclusion equations usually calculate the probability that a random individual from a population will be excluded as a parent, given a particular set of genetic markers. It is function of number of loci, number of alleles at each locus, and frequencies of alleles at these loci. Any parent which is excluded can be excluded with 100% certainty (assuming no genotyping errors). For the remaining non-excluded parents, it is possible to work out which is the most likely parent, but the efficiency and accuracy of distinguishing between these potential parents is not particularly high. Any non-excluded parents can be ranked, but hopefully only one non-excluded parent would remain if enough informative markers are used. The assumption is that all candidate parents are sampled. Usually several unlinked loci are used to reject candidate parents. The total exclusion probability is the complement of the product of the single-locus inclusion probabilities; i.e. the total exclusion probability is 1 minus the combined inclusion probability of the unlinked loci used (Evett and Weir, 1998; Marshall *et al.*, 1998). Sometimes excluding all but one potential parent is not possible, hence categorical and fractional likelihood techniques were developed to assign progeny to non-excluded parents based on likelihood scores derived from their genotypes, either categorically or fractionally. Vigilant *et al.* (2001) successfully assigned paternity to 34 of 41 offspring in wild chimpanzees after genotyping nine microsatellites on 108 individuals. Constable *et al.* (2001) successfully assigned fathers to all 14 offspring chimpanzees using 13-16

microsatellite loci. Newman *et al*. (2002) assigned a single sire to each of the 19 green monkey offspring of one of the social groups at VMRC.

5.1.3 Microsatellite cross-amplification

Markers with high allelic diversity associated with a high mutation rate are normally used for studies of recent evolutionary history. In addition to providing estimates of genetic variability for population analysis within species or between closely related species, hypervariable microsatellite loci allow assessment of parentage and other kinship relations among individuals (Luikart & England, 1999).

Generally, microsatellites can be cross-amplified successfully within a genus or a family of primates, but less efficiently across an entire order (Clisson *et al.*, 2000). Coote and Bruford (1996) failed to get amplification using human primers on Prosimians and New World monkeys. Takenaka *et al.* (1993) suggested that amplification using human primers in species distant from human would be so unsuccessful that designing species specific primers was required. Some microsatellites of non-human primates can be amplified through cross-amplification using human primers, but likely with low amplification success and polymorphism rate. It may be a general rule that a lower level of polymorphism is associated with the low cross amplification efficiency when orthologous loci are amplified in other species (Ellegren *et al.*, 1995; Zane *et al.*, 2002). Use of human microsatellite primers to cross-amplify non-human primate microsatellites results in about 20% polymorphic markers, with lower heterozygosity (65% in non-human compared to 80% in human) (Hadfield *et al.*, 2001). Clisson *et al.* (2000) demonstrated cross-amplification of eight microsatellites using human primers in all 18 species from the tested infraorder Catarrhini and also observed amplification for three of these microsatellites in a species of infraorder Platyrrhini (common marmoset). This conservation of microsatellite flanking sequences indicates their ancient origin and retention since before the separation of Catarrhini and Platyrrhini, more than 45 million years before present (MYBP).

Unsuccessful amplification in the marmoset of the other five microsatellites was found to be due either to the complete absence of the primer site due to deletion or to the excessive divergence of the flanking sequences (Clisson *et al.*, 2000). The success rate of 8% (6/76) in squirrel monkey *Saimiri (*Platyrrhini) when screening human primer pairs (Witte and Rogers, 1999) is lower than the 25% success rate obtained in baboons (*Papio hamadryas*) or macaques (Rogers *et al*., 1995; Perelygin *et al.,* 1996; Hadfield *et al.*, 2001). The lower success rate in *Saimiri* is probably due in part to the greater overall genetic divergence between human and platyrrhines compared with the divergence between human and cercopithecines (Catarrhini) (Schneider *et al.*, 1996). The green monkey diverged from the human lineage between 25 and >50 MYBP (Sarich and Wilson, 1967; Sibley and Ahlquist, 1987; Eichler and DeJong, 1995; Porter *et al.*, 1997; Arnason *et al.*, 1998), and is classified in the family Cercopithecidae and the subfamily Cercopithecinae, along with baboons and macaques. Cross-amplification of 55 microsatellites using human primers in the green monkey in VMRC, UCLA, has been reported (Newman *et al*., 2002). When tested on a social group of 51 monkeys (including 19 offspring, 18 mothers and 14 possible fathers), 43 produced PCR products. 14 of these PCR products proved polymorphic in the green monkeys. This success rate of around 25% is similar to that in baboon (*Papio*) or macaque monkey (*Macaca*) (20-25%) (Rogers *et al.,* 1995; Perelygin *et al.,* 1996; Hadfield *et al.*, 2001) of the same subfamily Cercopithecinae. Among other mammals the cross-amplification of microsatellite in peccaries using porcine primers was demonstrated by Gongora *et al*. (2002); in sheep using cattle primers by Ellegren *et al.* (1997) and Crawford *et al.* (1998); within mustelids by Davis and Strobeck (1998); between hare *Lepus* species by Kryger *et al*. (2002); among camelid species by Mariasegaram *et al*. (2002); and among cetaceans by Caldwell *et al*. (2002).

Being independent of the study carried by Newman *et al*. (2002), the current study evaluated 20 human primers initially (including 7 polymorphic markers tested by Newman *et al*., 2002) and 16 of them were ultimately used for large scale genotyping in the green monkey *C. sabaeus* of the VMRC (Table 5.1).

Table 5.1 Human microsatellites used in the green monkey for paternity assessment. In the present study, 20 microsatellites were screened, with 16 being used for paternity analysis. 4 unused for pedigree analysis are italicised. The set of 11 microsatellite used by Newman *et al*. (2002), which has 7 markers in common with the set reported here, is shown for comparison.

used for paternity test in the present study	used for paternity test by Newman *et al*., 2002
D1S102	
D1S207	
D1S244	
D1S518	D1S518
D2S144	D2S144
D3S3591	D3S3591
	D4S243
D4S2979	
	D5S1466
D6S1557	
D7S2446	
D8S1106	
D10S564	
D10S1483	
D11S928	
	D11S987
D11S1902	D11S1902
D12S84	D12S84
D12S314	
	D15S108
D18S42	D18S42
D18S72	D18S72
DXS1047	
DYS389	

5.2 Materials and methods

5.2.1 DNA isolation

Between 2000 and 2001 VMRC staff collected a blood sample from every living individual of the colony born in or before 2000, plus one born in 2001. However, not all animals of the large, extended pedigree could be sampled. For example, from a total of 52 animals of five generations in matrilineal line 1971-001, only 17 animals from three generations have been collected (Figure 5.1). A total of 564 samples (Table 5.2) were collected in duplicate batches. One batch of blood was frozen for several months before the start of the present study, and quite a few of these samples had become clotted. All these frozen samples were extracted for DNA using the manual protocol. Another batch of samples was collected freshly and DNA was extracted using the automatic protocol – AutoPure LS (Gentra, Minneapolis, MN).

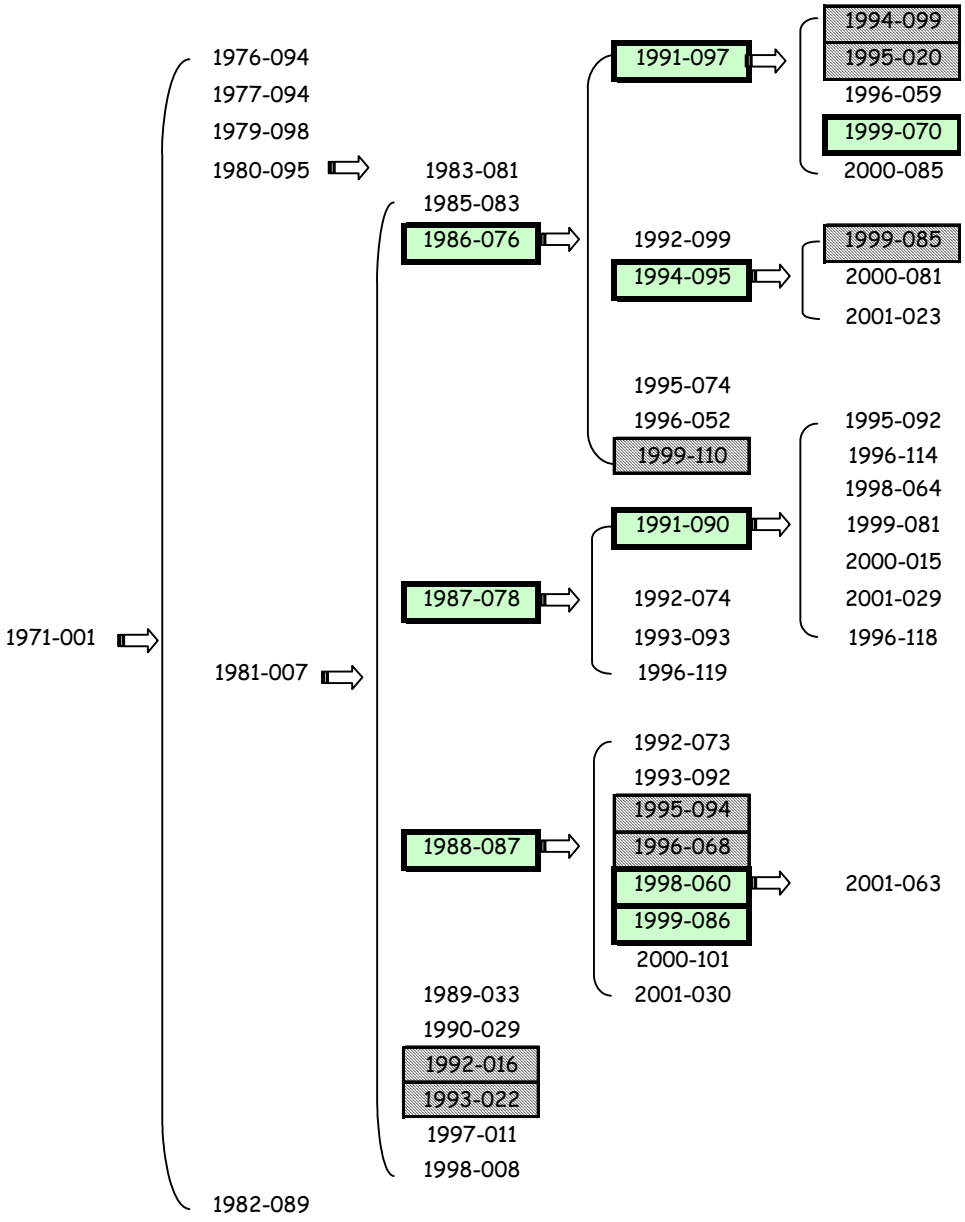


Figure 5.1 Putative structure of a matrilineal lineage of descent from 1971-001. Shadowed squares with bold border are sampled females. Crosshatching squares with lighter border are sampled males. Others were not available due to death.

Table 5.2 Samples collected from 6 founders and 558 descendents across several generations. The six sampled founders are in bold within cells without borders.

Sex	founder	sampled descendents	Sex	founder	sampled descendents
F	1970-001	3	F	1975-001	27
F	1971-001	17	M	1975-002	
F	1971-002	9	M	1975-003	
M	1971-003		F	1975-004	13
F	1971-004	8	F	1975-005	26
M	1971-092		M	1975-006	
M	1971-093		F	1975-007	1
M	1971-094		M	1975-008	
M	1971-095		F	1975-097	0
F	1971-096	35	M	1975-098	
M	1971-097		F	1975-099	15
M	1971-098		M	1976-096	
M	1971-099		M	1976-097	
F	1972-001	12	M	1976-098	
M	1972-002		M	1977-002	
M	1972-003		F	1977-003	23
M	1972-004		F	1977-005	1
F	1972-006	60	F	**1978-001**	12
F	1972-007	93	F	**1978-002**	7
F	1973-001	29	F	**1978-003**	11
F	1973-002	80	F	1978-004	5
M	1973-097		F	**1978-005**	10
M	1973-098		M	1978-006	
F	1973-099	0	F	**1978-007**	9
F	1974-001	16	F	**1979-006**	20
F	1974-002	16	M	1979-007	
F	1974-003		M	1979-008	
M	1974-004		M	1979-009	
M	1974-099				
		total		6	558

5.2.1.1 Autopure DNA extraction

The AutoPure LS (Gentra**)** is an automatic system specialized for DNA extraction from blood samples. During high throughput DNA extraction and purification of samples, the instrument uses bar-codes to track each sample. Whole blood samples 10 ml typically yield 350 μg of high molecular weight DNA that is 100-200 kb in size. Most of the Autopure DNA extraction was done by the DNA Core (a molecular service lab) at UCLA.

5.2.1.2 Manual DNA extraction

The related reagents for manual DNA isolation are from Gentra. The procedure for manual DNA isolation was modified from Puregene and Clotspin protocols provided by Gentra.

Whole blood (7 ml) was mixed with 21 ml RBC Lysis Solution (Gentra) and incubated for 10 minutes at room temperature. The tube was inverted at least once during the incubation. Supernatant was removed after 10 minutes centrifugation at 2,000 x g. The visible white cell pellet and about 200-400 µl of the residual liquid were left behind. 5 ml Cell Lysis Solution (Gentra) and 25 µl Proteinase K Solution (20 mg/ml, Gentra) were added and vortexed at high speed for 10 seconds. The tube was incubated at 37^{o}C overnight with occasional vortexing, until all particulates were dissolved completely. Protein Precipitation Solution (Gentra) 1.7 ml was added to the cell lysate after it was cooled on ice for 5 minutes, and followed by vigorous vortexing. The tube was centrifuged at 2,000 x g for 15 minutes, followed by 5 minutes incubation on ice. The supernatant containing the DNA was poured into a clean tube containing 5 ml 100% Isopropanol (2-propanol) and mixed by inversion. The sample was centrifuged at 2,000 x g for 10 minutes. The supernatant was poured off and the tube was drained briefly on clean absorbent paper. The DNA pellet was washed with 5 ml 70% ethanol and centrifuged at 2,000 x g for 10 minutes. The ethanol was poured off. The tube was inverted and drained on clean absorbent paper for 10 seconds. About 400 µl DNA Hydration Solution (Gentra) was added to rehydrate the DNA. The sample was shaken on a rotator overnight, before the concentration was measured by spectrophotometer. DNA was stored at 2-8^{o}C, or for long-term storage at –20^{o}C or –80^{o}C.

5.2.2 PCR amplification and genotyping

An aliquot of each DNA sample was diluted to a concentration of 10 ng/µl, and dispensed into each well of a total of 7 deep 96 well plates. A 384-well template microplate was then made from the DNA from each deep well plate using Hydra Microdispensers (Robbins), by dispensing each sample from the deep well plate into each well of a 384-

well microplate (Figure 5.2). Each sample was dispensed into 4 separate wells with 5 µl in each well, ready for 4 reactions with different primer pairs each. The operating system of the Hydra Dispensers uses the barrels of 96 (4x96=384) glass syringes which are held in a fixed array centered along a coordinate grid that corresponds to the exact center of each well in a microplate. The plungers within the syringes move up and down under computer control, dispensing or aspirating liquids to or from the microplate wells.

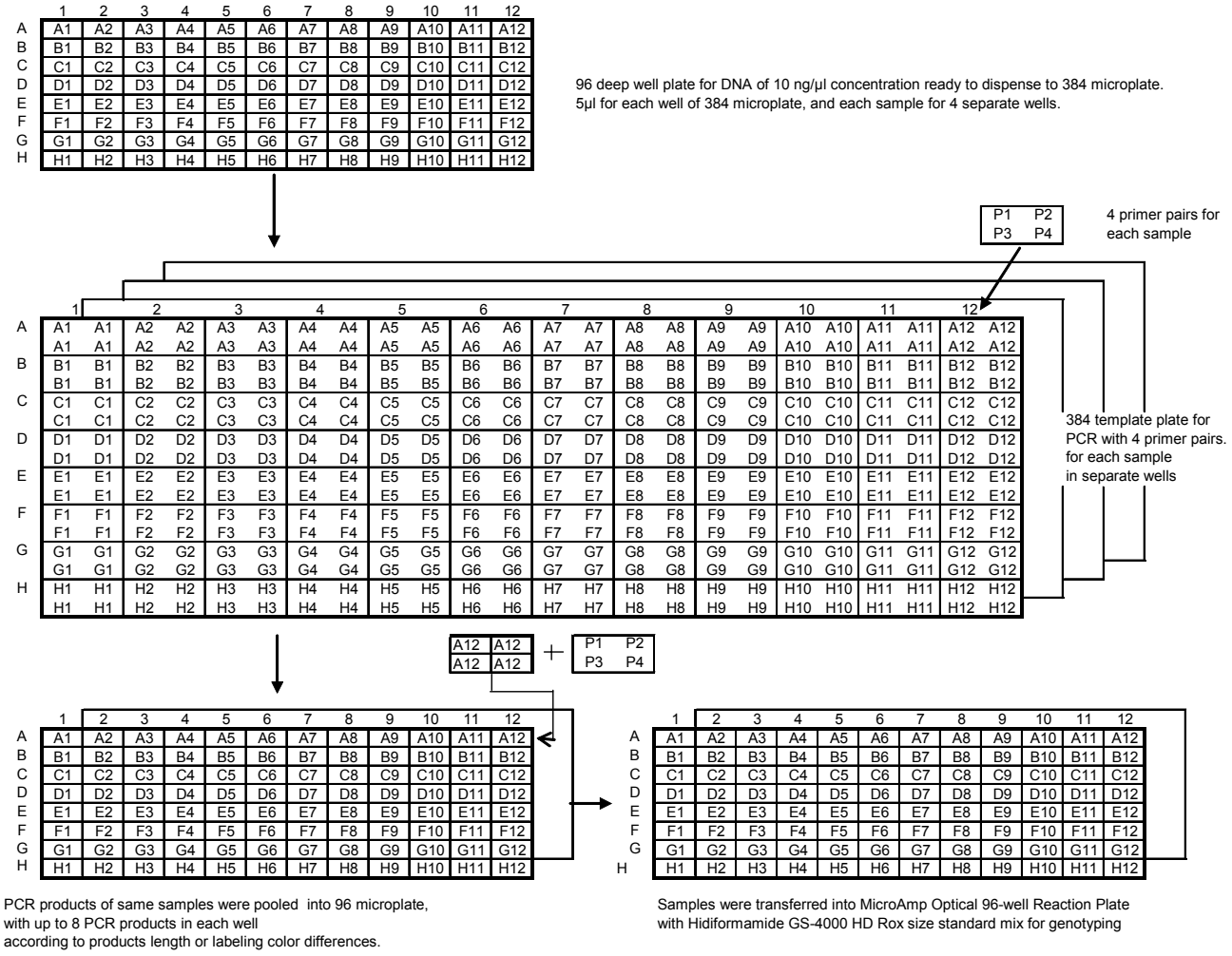


Figure 5.2 Flowchart for pre-genotyping preparation

PCR was carried out using the 384-well template microplate, with four primer pairs for each sample in separate wells (Figure 5.2). Each microsatellite marker was cross-

amplified using primers designed for the human genome. Each PCR amplification reaction was in a volume of 10 µl containing 0.5 µl of primer pairs (10 µM), one of which was fluorescently labeled (FAM, HEX or NED); 5 µl PCR Master Mix (AmpliTag Gold); and 4.5 µl ddH2O. The PCR mixture was dispensed into the 384-well template microplate bearing 50 ng dried DNA in each well previously dispensed by Hydra Microdispensers. The PCR mixture dispensing was done by MultiPROB II Robot (PerkinElmer. Wellesley, MA), a liquid handling and sample prep system, using software WinPREP (Promega, San Luis Obispo CA). PCR involved one cycle of 94^{o}C for 12 minutes; then 12 cycles of : 94^{o}C for 30 seconds, 56^{o}C, stepping down 0.5^{o}C each cycle, for 30 seconds, and 72^{o}C for 30 seconds; followed by 23 cycles of: 94^{o}C for 30 seconds, 53^{o}C for 30 seconds, and 72^{o}C for 30 seconds; finished by a final cycle of 72°C for 5 minutes.

PCR products were pooled in 96-well microplates using MultiPROB II Robot to combine up to eight reaction products with different primer pairs per sample into one. For confident scoring, the pooled products must be labeled with different colors and/or have different amplicon lengths which can be easily distinguished on a gel. 2 µl of pooled PCR product from each well was transferred to a MicroAmp Optical 96-well Reaction Plate with 10 µl 3.85% Hidiformamide-GS-400 HD Rox size standard mix (1000 µl Hidiformamide + 40 µl standard) in each well (Figure 5.2). The plates were denatured at 96^{o}C for 5 minutes.

Genotyping was done by the UCLA Sequencing and Genotyping Core (a molecular service lab) using an ABI3700 Capillary Sequencer, which uses a sheath flow fluorescence detection system.

Alleles were sized relative to internal size standard (HD400 labeled ROX). This appeared at defined molecular sizes as red peaks used by the Genotyper 3.6 software to calibrate the allele calling.

Multiple replicates are important for evaluation of cross species amplification, especially for lower quality DNA such as that isolated from faeces, hair, or other non-invasive samples (Taberlet *et al.,* 1997; Flagstad *et al.,* 1999; Smith *et al.*, 2000). In the present research, multiple independent repetitions for each genotyping were carried out in order to recover missing data and to get reliable scoring results, although the template DNA used was of high quality.

5.2.3 Data analysis

For comparison, human microsatellite information from NCBI and the Centre d'Études de Polymorphisme Human (CEPH) websites were used. Usually CEPH data were used, but other data were also occasionally used if relevant CEPH data were not available.

The monkeys studied in this project are founded by 57 captive animals. Unlike a natural population, these social groups could have some unusual features such as inbreeding or non-random mating. Therefore, Hardy-Weinberg equilibrium analysis (HWE) may be compromised since the animals were sampled across a multigenerational pedigree. The HWE was done using Arlequin software (Schneider *et al.*, 2000), using an exact test originally described by Guo and Thompson (1992). The *P*-value of the test is the proportion of the visited tables having a probability smaller or equal to the observed (initial) contingency table. More details about the exact test will be described in Chapter 7.

Paternal assessment was performed using Newpat.xls (Amos, 2000). In Newpat.xls, allele frequencies are calculated, and the frequency of putative null alleles are estimated by an iterative approach which assigns homozygote excess to the null allele category without considering whether non-amplification was due to sporadic PCR failure or to a genuine null allele. Resampling is used to estimate the null allele frequency confidence limits. The program Newpat.xls can be run without mother-offspring information. However, if maternal information is provided, mother-offspring analysis will be performed. Mother-offspring mismatches are identified, with homozygous mismatches (a mismatch of

homozygote from both mother and offspring) and heterozygous mismatches (a mismatch of heterozygote from one of the mother-offspring pair) tallied separately. Results of mother-offspring analysis are used for evaluation of paternity among available males. When a paternity match is found, the program then simulates large numbers of offspring-male relations to determine how frequently such a match will arise by chance. The relatedness value calculated by the program then represents the probability of the father most likely to be the father of the offspring.

Pedigree reconstruction was initially done using Pedigree Viewer program (Kinghorn and Kinghorn, 2005) and was finally redrawn manually.

5.3 Results

5.3.1 DNA extraction yield

The AutoPure LS (Gentra) for automated DNA extraction was found to produce very low yields from frozen samples stored for a long period. Several DNA extraction protocols were tried for these frozen samples. The use of the Clotspin manual protocol resulted in better yields than the Puregene manual protocol, but Clotspin costs more time and money for spin columns. In the present study, the Puregene and Clotspin protocols were combined and modified with extension of Proteinase K incubation time and omission of RNase, Glycogen solutions, and spin column, and other modifications. As shown in Table 5.3, the modified protocol had much better results than the recommended Puregene and Clotspin. This modified procedure was the best for frozen samples while the yield using the AutoPure system was highest for fresh blood.

Table 5.3 Summary of DNA yield from 7ml blood sample

extraction method	autopure for frozen samples		puregene for frozen samples		puregene+ proteinaseK for frozen samples		clotspin for frozen samples		modified for frozen samples		autopure for fresh samples	
yield level (ul)	No, of sample	%	No, of sample	%	No, of sample	%	No, of sample	%	No, of sample	%	No, of sample	%
<50	13	81.25	10	30.30	11	68.75	2	13.33	46	11.68	39	6.63
50-100	1	6.25	5	15.15	3	18.75	2	13.33	45	11.42	45	7.65
100-150			3	9.09	2	12.50			48	12.18	84	14.29
150-200	1	6.25	6	18.18			3	20.00	56	14.21	101	17.18
200-300			2	6.06			5	33.33	85	21.57	122	20.75
>300	1	6.25	7	21.21			3	20.00	114	28.93	197	33.50

5.3.2 Cross-amplification

Table 5.4 lists the human microsatellite primers genotyped on the green monkeys in the present research. The initial scoring rate for the 16 markers was around 80% on average. The final average rate of 91% was reached after the PCRs for each sample were repeated more than once and the results were consolidated. Bradley *et al.* (2000) achieved 62 to 75% initial scoring rate for cross-amplification of five successfully amplified human microsatellites on chimpanzee and gorilla. After these five primers were redesigned to produce shorter products using only the publicly available human sequence information for these loci, an average of 82% (62-95%) scoring rate was achieved with four of the five markers significantly improving the scoring rate.

Table 5.4 Primers used in paternity assignment

locus	primerF	primerR
D1S102	AAATCAGACAAGTACAGGTG	ATGAACTTGTTCTGGGAGGA
D1S207	CACTTCTCCTTGAATCGCTT	GCAAGTCCTGTTCCAAGTCT
D1S244	GAGCAGCACCGTACAAAT	AGCTCCGCTCCCTGTAAT
D1S518	TGCAGATCTTGGGACTTCTC	AAAAAGAGTGTGGGCAACTG
D2S144	TCTCCCTGACAGACTCTGCG	GCTGCATAGGCCGTACTGAG
D3S3591	CCAACTATGTTTTGGGTCTG	TGTGCCCAGTTAGATGATG
D6S1557	TCTCACCACTCTGAAGATAGTATG	CTAGGGACTCGCTGATTTG
D7S2446	TTTGAGTCTTCACAGCAGTTG	GGGAGGTTGATTTCCACAGT
D8S1106	TTGTTTACCCCTGCATCACT	TTCTCAGAATTGCTCATAGTGC
D10S564	TGGGAATGTGTCTTTATCCA	AGCTCTAACATAGAGGCCAGAT
D10S1483	CAATGCTATCCCGGCTATG	TCAAGACTGCAAGCGTGT
D11S928	AAGTGATCCACCTGCCTTG	GCCTCTGAGAATTAGTGTCTGTC
D11S1902	CTCTGGTTTCTGGCTCTTC	GAAGCATGAAATTGGGAAGG
D12S314	TTTGGGAACTGTCACTCAGAAAAG	AGCAGACCCTGTCTCTCATAATTG
D18S42	CCAGGGTCTTGGGGTAGCTT	GGTGATAGCGATTGAAATAGGAA
D18S72	GCTAGATGACCCAGTTCCC	CAAGAGAGCCCTTTGGTTT

5.3.3 Maternal mismatch pedigree errors

After genotyping the 564 samples from green monkeys (Table 5.2), an attempt was made to assign fathers to offspring. Initial analysis showed that quite a few genotypes of the mother and offspring did not match, and this was the main obstacle to full pedigree reconstruction. These mismatches were two types. First were mismatches when both the mother and the offspring are homozygous. The other is when at least one of the mother-offspring pairs is heterozygous (Table 5.5). Twenty two homozygote mismatches and 82 heterozygote mismatches were found among 50 female-offspring pairs (Table 5.6). All genotyping results were double checked and compared using control samples. As well as the three control samples in every plate, some of the samples (15) were duplicated either in the same plate or different plates. Moreover, all the samples were genotyped repeatedly in order to fill in missing data due to sporadic PCR failure. All these repeated data were compared to make sure the same samples in the multiple replications were identical. The only inconsistency was found for monkey 1985012 at locus D18S42, for which the PCR was repeated seven times. In only one case of the seven attempts was a different genotype recorded (one allele difference). The error rate due to sporadic PCR failure was negligible. To ensure that the samples were not mixed up, another plate (96

samples) amplified from another batch of the DNA template from the duplicate blood sample was genotyped for all the loci under the same conditions. All the genotyping results were the same as previously determined, although sometimes the peak position shifted slightly (Figure 5.3).

Table 5.5 Examples of three types of mismatch. In recorded scoring, offspring1 has a genotype consistent with mother's genotype, but offspring2 has not. Offspring 2 could be falsely excluded if nulls are present, as shown in the bottom part of the table. Homozygote exclusion is defined when both the mother and the offspring are homozygous. Heterozygote exclusion is defined when at least one of the mother-offspring pairs is heterozygous

recorded genotypes

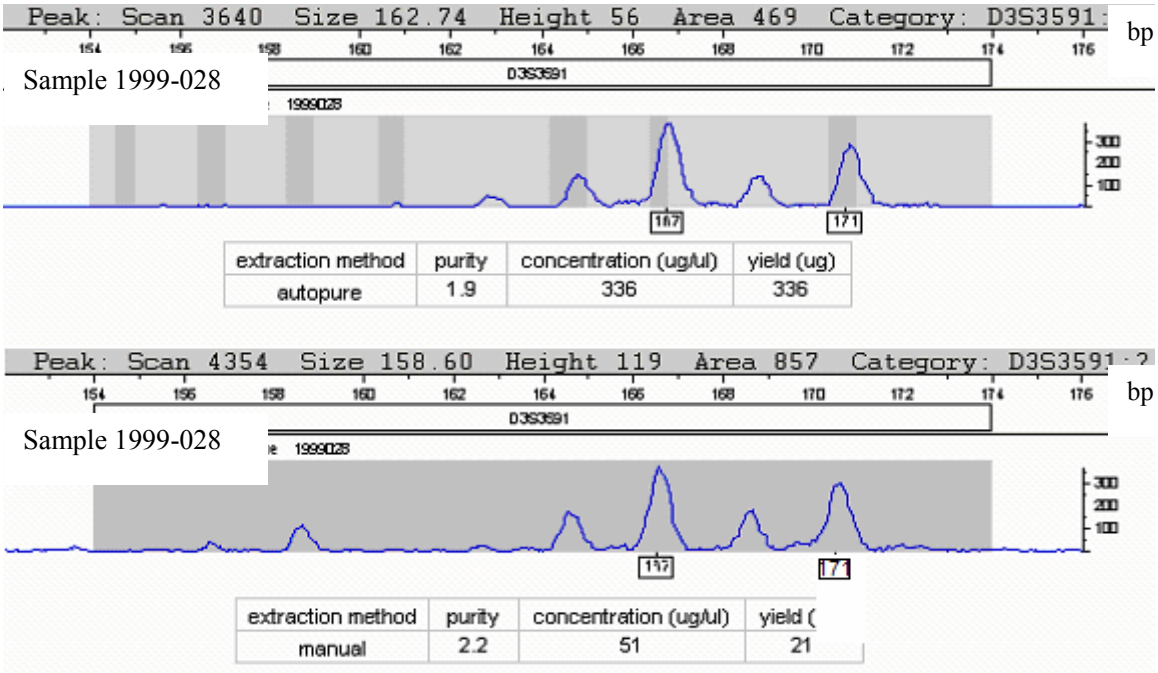
status	sample	D10S1483		D10S564		D1S102	
mother	**1992045**	130	130	268	276	197	201
offspring1	1996040	130	130	276	276	193	197
offspring2	1999028	140	140	278	278	193	203

posible genotypes if nulls are present

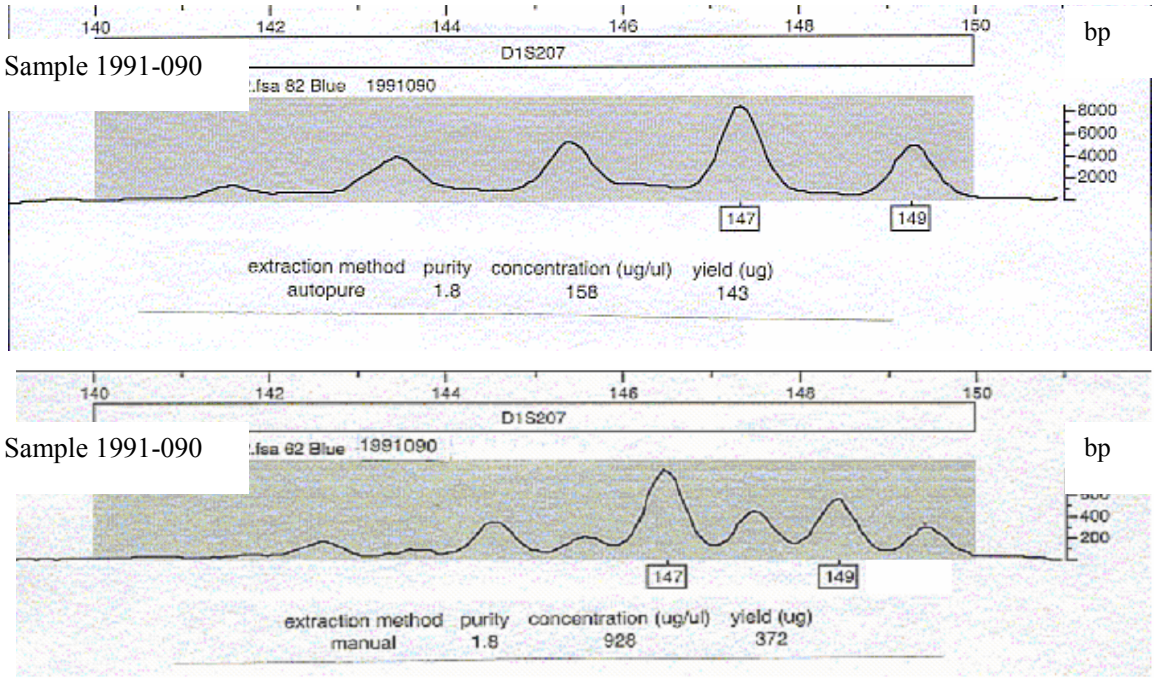
status	sample	D10S1483		D10S564		D1S102	
mother	**1992045**	130	null	268	276	197	201
offspring2	1999028	140	null	278	null	193	203
Possible analysis error from recorded data above		False homozygote exclusion if the two null are the same or the offspring null =130		False heterozygote exclusion only if null=268 or 276		True exclusion without analysis error	

Table 5.6 Homozygote mismatches (mis1) and heterozygote mismatches (mis2 and mis3) of the mother-offspring pairs. mis1: a mismatch when both mother and offspring are homozygous; mis2: a mismatch when one of the mother-offspring pair is heterozygous; mis3: a mismatch when both of the mother-offspring pair is heterozygous.

Female	Offspring	D10S1483	D10S564	D11S1902	D11S928	D12S314	D18S42	D18S72	D1S102	D1S207	D1S244	D2S144	D3S3591	D6S1557	D7S2446	D8S1106	D1S518	mis1	mis2	mis3	total
1978001	1991008		mis1			mis2				mis2				mis1				2	2	0	4
1982002	1987007										mis2							0	1	0	1
1982005	1990033														mis2			0	1	0	1
1982005	1994021														mis2			0	1	0	1
1982006	1991006												mis1					1	0	0	1
1982006	1993041												mis2					0	1	0	1
1982006	1995079												mis1					1	0	0	1
1983091	1992092								mis2						mis3	mis3	mis2	0	2	2	4
1985006	1994015			mis3				mis3										0	0	2	2
1985081	1993097			mis1		mis3											mis2	1	1	1	3
1985088	1989097																mis3	0	0	1	1
1985091	1993075	mis2				mis2		mis2			mis2				mis2		mis2	0	6	0	6
1985091	1995103														mis2			0	1	0	1
1985092	1993073										mis2						mis1	1	1	0	2
1985092	1994076																mis2	0	1	0	1
1985092	1997034																mis2	0	1	0	1
1986009	1998065								mis3				mis2		mis2			0	2	1	3
1986012	1999017		mis2							mis2								0	2	0	2
1986076	1999110													mis1				1	0	0	1
1986081	2000018								mis3					mis1				1	0	1	2
1987002	1991024		mis2	mis2			mis2									mis2		0	4	0	4
1987007	1993047										mis1							1	0	0	1
1987007	1996014										mis2							0	1	0	1
1987079	1993063									mis1		mis1			mis1	mis3		3	0	1	4
1987079	1999038							mis2			mis2	mis1						1	2	0	3
1987081	1994080								mis2									0	1	0	1
1987082	1999046								mis2						mis2		mis2	0	3	0	3
1988007	1992044														mis3			0	0	1	1
1988007	1996044						mis2								mis3			0	1	1	2
1988088	1993062				mis1				mis1									2	0	0	2
1990033	1996011											mis2						0	1	0	1
1990097	1994075		mis2					mis2		mis1						mis2		1	3	0	4
1991038	1998033						mis2											0	1	0	1
1991086	1996093											mis1						1	0	0	1
1991091	1995100		mis2					mis2					mis2					0	3	0	3
1991091	2000071							mis2					mis2				mis2	0	3	0	3
1991092	1998019						mis3									mis3		0	0	2	2
1991092	1999108		mis2		mis2								mis3					0	2	1	3
1992045	1999028	mis1	mis2			mis2			mis3	mis3			mis2			mis1		2	3	2	7
1992086	1999107													mis2				0	1	0	1
1992088	1998032													mis1				1	0	0	1
1992090	1999033															mis3		0	0	1	1
1992091	2000066					mis3		mis2			mis2					mis2		0	3	1	4
1992097	1999104							mis3					mis3				mis3	0	0	3	3
1992098	1999077					mis2												0	1	0	1
1993066	2000019															mis2		0	1	0	1
1994046	1999073							mis2								mis3		0	1	1	2
1994069	1999037								mis2				mis2					0	2	0	2
1995067	1999050				mis1													1	0	0	1
1995103	2000122														mis1			1	0	0	1
total mismatch		2	7	3	3	6	4	9	8	5	7	4	10	5	11	10	10	22	60	22	104



a. Consistent genotype pattern for marker D3S3591 for animal 1999028 of two DNA samples extracted from different batches of blood.



b, For animal 1991090, DNA extracted from two batches of blood have same score in genotype of marker D1S207, although the peak positions have been shifted slightly.

Figure 5.3 (a & b) Examples of genotyping profiles.

For the 412 pairs of mothers and offspring tested (from a total of 150 mothers), 50 pairs (12.1%) were mismatched for some loci. Among them, half are multilocus mismatch. Non-amplifying (null) alleles may have been present at some loci where apparent homozygous mismatches between parent and offspring were evident. Among the 6592 pair/locus combinations, 104 combinations were found to be mismatches, including those involving a homozygote in both the mother and offspring, and those involving a homozygote in either of the mother-offspring pair. Assuming that two of these types of mismatches can be explained by null alleles, only 22 mother-offspring combinations remain where the unequivocal heterozygous genotypes of both parents mismatch, setting a lower limit of 5.3% for maternal pedigree errors. However, given the consistency of replicated genotypes and the high quality of template DNA, the 12.1% error rate seems to be the more realistic figure for maternal pedigree errors. The number of mismatches, either considering or ignoring the putative null alleles, appears concordant with the total sample size (Figure 5.4).

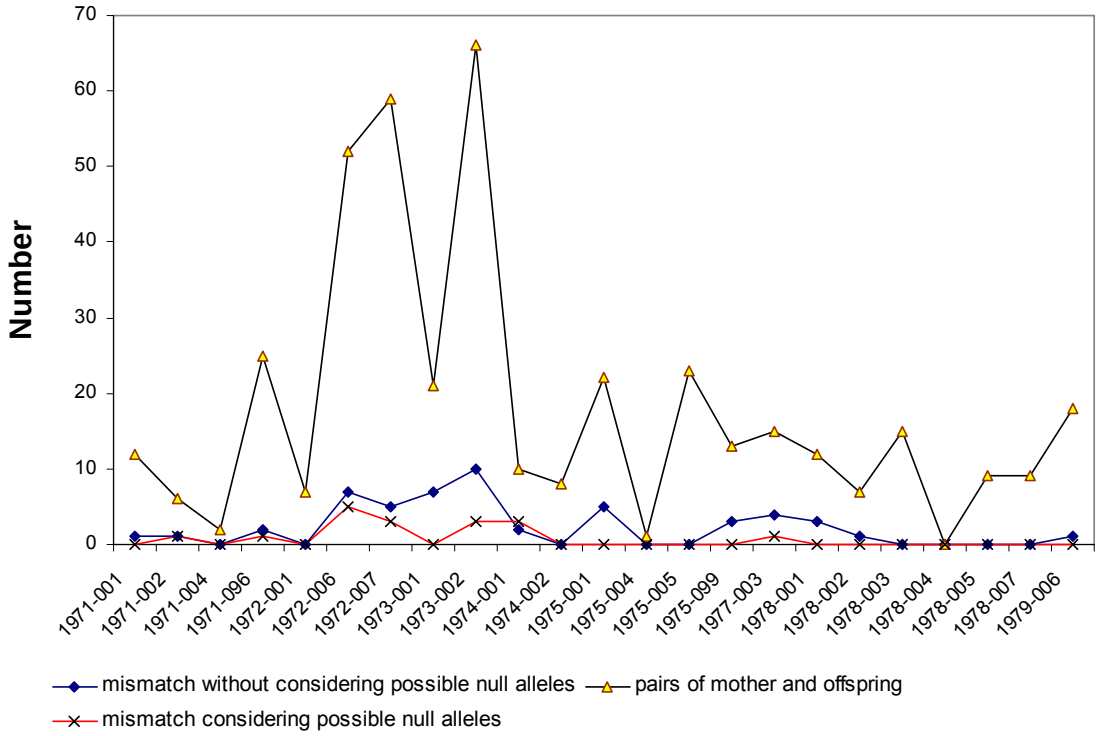


Figure 5.4 number of mismatches, either considering or ignoring the putative null alleles, appears concordant with the total pairs of mother and offspring

5.3.4 Hardy-Weinberg Equilibrium

An apparent homozygote excess could be explained by null alleles. Intermittently amplifying null alleles also contribute to both homozygote and heterozygote mismatch. A heterozygote excess could be explained by mis-scoring or artifacts of PCR amplification. The HWE evaluation computed by Arlequin (Schneider *et al*., 2000) using an exact test original described by Guo and Thompson (1992) for all the data showed that 3 of the 16 loci, D10S1483, D18S72 and D6S1557, deviated significantly from HWE at the 0.01 level (Table 5.7). This is 18.8% of the total markers. All these three loci showed slight homozygote excess. The cause of the departure could be null alleles or inbreeding. In any case, the monkeys genotyped in the colony are not a natural population but a multigenerational pedigree, and thus it is not strictly appropriate to apply the HWE test.

Table 5.7 Hardy-Weinberg equilibrium evaluation computed by Arlequin (Schneider *et al.*, 2000) using exact test original described by Guo and Thompson (1992) for all the data tested for microsatellites in monkeys at VMRC. * Divergence from HWE at 0.01 level.

Locus	No. tested	Obs.Heter.	Exp.Heter.	P. value
D10S1483	529	0.58601	0.64762	0.00747*
D10S564	534	0.70974	0.72051	0.51517
D11S1902	528	0.58902	0.61166	0.78184
D11S928	542	0.51107	0.49821	0.83269
D12S314	534	0.76217	0.7632	0.25303
D18S42	363	0.67769	0.68974	0.05106
D18S72	553	0.72694	0.76008	0.00099*
D1S102	474	0.63924	0.67862	0.32786
D1S207	542	0.66052	0.64952	0.34162
D1S244	505	0.68911	0.669	0.74952
D2S144	530	0.74528	0.73993	0.36677
D3S3591	499	0.76754	0.75494	0.10635
D6S1557	540	0.24259	0.2943	0*
D7S2446	514	0.65953	0.67309	0.56823
D8S1106	544	0.74632	0.7586	0.03037
D1S518	456	0.70833	0.68673	0.64587

5.3.5 Paternity analysis

Because of maternal mismatches and maternal unavailability, an attempt was made to assign 417 offspring to 225 potential male parents while ignoring maternal information. One hundred thirty six offspring were assigned, with half assigned to a single father (Table 5.8). For multiple possible fathers, the one with higher probability was taken as the assigned father. After eliminating age inconsistency (males too young to be fathers), 78 offspring were assigned paternally (Table 5.8, 5.9). When maternal genotypes were used, 51 animals were assigned to a single father. The majority was assigned to a single father, with only five having an alternative possible father (Table 5.8). Eliminating age inconsistency, 47 offspring were assigned paternity (Table 5.8, 5.9). The result was more reliable when the maternal information was used, but when the maternal information was not available, the result could still be used for the pedigree construction

Table 5.8 Summary of paternity assignment for the green monkeys at VMRC, UCLA

without maternal information			with maternal information		
no. of fathers	no. of assignment	no. of age acceptable	no. of fathers	no. of assignment	no. of age acceptable
1	68	33	1	51	46
2	26	22	2	5	1
3	18	10			
4	10	6			
5	3	1			
6	4	2			
7	6	3			
19	1	1			
Total	136	78		56	47

Table 5.9 Paternity assignment for the green monkeys at VMRC, UCLA. *assignment without maternal genotypes. **assignment with maternal genotypes.

offspring	father*	father**
1988085	1985002	
1990094	1985091	
1991024	1985090	
1991028	1988008	
1991030	1985002	1985002
1992028	1989002	1989002
1993018	1985002	1985002
1993035	1986014	1986014
1993075	1990008	
1993085	1986085	1986085
1994011	1986014	
1994019	1986014	1986014
1994029	1986014	1986014
1994033	1991044	
1994036	1989002	1989002
1994038	1989004	
1994039	1986014	
1994045	1986014	1986014
1994086	1986085	
1994090	1986085	1986085
1994099	1989097	1989097
1995020	1989097	1989097
1995031	1992041	
1995085	1989086	1989086
1995095	1985002	
1996014	1982002	
1996031	1990034	
1996086	1989086	1989086
1997007	1988008	1988008
1997018	1988008	1988008
1997026	1985090	
1997027	1991027	
1997030	1988008	1988008
1997033	1991027	1991027
1998002	1993025	1993025
1998010	1988008	1988008
1998013	1988008	1988008
1998015	1991027	1991027
1998016	1991028	
1998036	1988008	1988008
1998042	1993077	1993077
1998046	1994045	
1998053	1991027	1991027
1998059	1994045	
1998060	1992018	1992018
1998066	1988008	1988008
1998070	1990008	1990008
1998073	1986003	1986003
1999007	1994045	
1999008	1995086	
1999023	1991027	1991027
1999032	1988008	1988008
1999033	1995056	
1999048	1988008	1988008
1999064	1993077	
1999066	1990038	
1999067	1994086	1994086
1999069	1991024	1991024
1999082	1992018	1992018
1999087	1991027	1991027
1999102	1994021	1994021
1999103	1991024	
1999109	1989002	
1999110	1992028	
2000001	1997022	
2000006	1990008	1990008
2000022	1994057	1994057
2000025	1997019	
2000035	1995053	
2000043	1993020	1993020
2000046	1993020	1993020
2000062	1994041	1994041
2000064	1995113	1995113
2000087	1992018	1992018
2000099	1995031	1995031
2000111	1992028	1992028
2000115	1992028	1992028
2000117	1991027	

5.3.6 Pedigree construction

Besides some small isolated pedigrees, only one large pedigree could be reconstructed containing 75 animals (Figure 5.5). In this pedigree there is a dominant male 1988-008, who mated with 8 females and contributed 10 offspring to the pedigree. The other dominant male 1991027 mated with 7 females and contributed 7 offspring to the pedigree. Male 1986014 mated with 3 females and produced 5 offspring. Males 1985002, 1990008 and 1994045 each mated with 3 females and had 3 offspring. Close inbreeding was avoided within this pedigree.

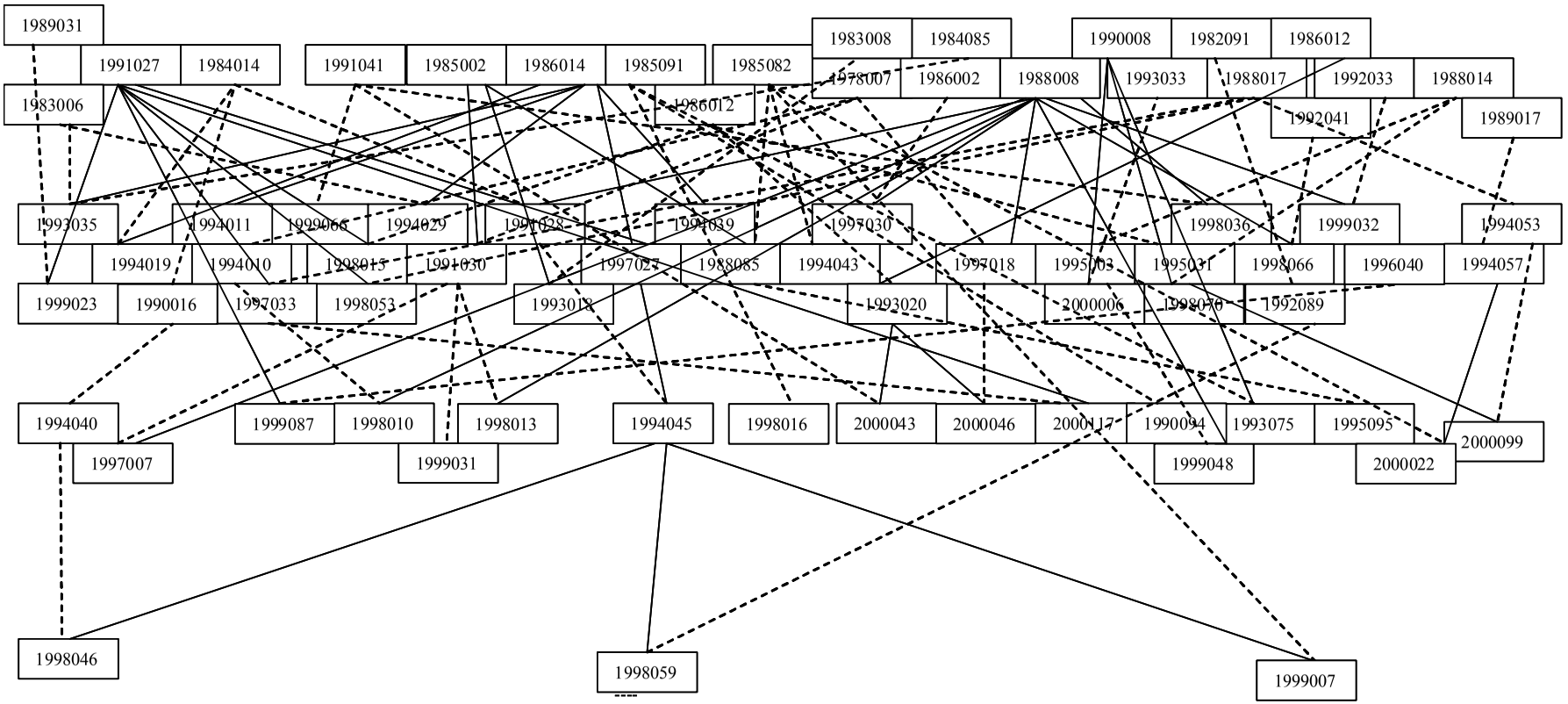


Fig 5.5 The major pedigrees constructed from paternal assignment results. Solid lines connect the father and the offspring. Dashed lines connect the mother and the offspring.

5.4 Discussion

Human microsatellites have been cross-genotyped to non-human primates for paternity and other assessment. The low efficiency of paternal assignment in the present study is partly due to mother-offspring mismatches, and partly because quite a number of blood samples were collected from individuals whose mother or father or offspring had died. The assumption for paternity assessment that all candidate parents were sampled was not fulfilled; this is an important cause of the low assignment rate, and could cause some of the currently favored paternity assignments to be incorrect. However, the limitation of the exclusion methodology could cause problems as well.

5.4.1 Microsatellite polymorphism comparison

When using human primers, the amplicon size differed between human (*Homo sapiens*) and the green monkey (Appendix 1), but the repeat sequence motif is most likely the same. Clisson *et al.* (2000) sequenced the non-human primate amplicons produced with human primers and confirmed that, in most cases, the repeat motif was identical to that of the human. Differences in the size of the PCR fragments between human and non-human amplified with the same primer were mainly caused by indels (insertion/deletion) in the sequence flanking the repeat, although base substitutions within the repeat motif sequences occurred at a high frequency (Clisson *et al.*, 2000). Nonetheless, compared to the CEPH reference families and other data from NCBI, the allele numbers of the green monkeys in the current study are much lower than in human (Table 5.10). For the green monkey, the largest number of alleles at one locus is seven, but for human it is 14. The total number of alleles in the green monkey is 85, and the average heterozygosity is 66.4%; compared to 133 total alleles and 72.8% average heterozygosity in human. The number of founders of green monkey (57) in VMRC is not much different from the number of individuals in the human data (around 60 in average) from CEPH reference families listed in Table 5.10. For the 16 loci studied here, the human is more polymorphic than the green monkey. This is consistent with the suggestion that non-human primates appear to have more mitochondrial genome diversity, whereas human have more nuclear

genome diversity (Ferris *et al*., 1981; Wise *et al.*, 1997; Rubinsztein *et al.*, 1995b; Cooper *et al.*, 1998). Rubinsztein *et al.* (1995a, b) proposed the directional-evolution hypothesis to explain why microsatellite loci in human are longer and more variable than their orthologues in other primates. Since it is possible that human microsatellites have more diversity than those of non-human primates, it may not be appropriate to attribute the lower heterozygosity of non-human primates to the low efficiency of cross-amplification. The high rate of successful genotyping and the consistency of genotypes across replications indicate that amplification problems or artifacts are unlikely to have interfered with the microsatellite diversity estimates in green monkeys.

Table 5.10 Allele number and heterozygosity for human (CEPH reference families) and green monkeys

	Human		Green monkey		amplification rate (%) in the green monkey
Locus	allele no.	%het	allele no.	%het	
D10S1483	12	83	4	56.96	94.15
D10S564	6	60	5	71.16	95.04
D11S1902	7	56	6	64.85	93.97
D11S928	6	70	4	53.33	96.45
D12S314	8	80	5	77.05	95.04
D18S42	9	78	6	71.14	64.36
D18S72	6	52	3	72.84	98.40
D1S102	5	56	6	59.39	84.40
D1S207	13	84	5	68.66	96.28
D1S244	11	81	6	68.68	89.89
D2S144	10	84	5	75.84	94.15
D3S3591	6	61	5	78.09	88.83
D6S1557	8	79	3	28.58	96.10
D7S2446	12	88	6	66.47	91.49
D8S1106	11	82.14	4	73.09	96.81
D1S518	9	78.57	5	70.73	81.38
total	139		85		
average		73.2944		66.053	91.21

5.4.2 Parent-offspring mismatch

The overall mother-offspring mismatch rate of 12.1% found in the present research ignoring possible null alleles is most likely correct. The cause of genotype mismatch between mother and offspring could be due to: 1) Colony record error. This problem can happen quite commonly in an uncaged animal study. The behavior of wild animal makes it hard to keep correct records of their kinship while mismothering or crossfostering is reasonably common. 2) A null allele mismatch. Non-

amplified alleles were detected in wild red deer (*Cervus elaphus*) through mismatches between undisputably known mother-offspring pairs and by significant deviations from Hardy-Weinberg equilibrium (Pemberton *et al*., 1995). However, in the case of the green monkey studied here, a half of the mismatch pairs are multilocus mismatches, which could not be simply explained by existence of null alleles. In addition, the consistence of replicated PCR result and the usage of quality template DNA confirm that the 12.1% mismatch rate was caused by colony record error.

Taberlet *et al.* (1997) and Smith *et al.* (2000) treated individuals that appeared "homozygous" for two different alleles in multiple replications as heterozygotes, even if the two alleles never appeared in the same amplification. The reason they did this is that they used a feces-derived DNA. About 48% of allele dropouts can be found in PCRs for feces-derived DNA; there was no allelic dropout observed in blood-derived DNA so far (Taberlet *et al.*, 1997).

Smith *et al.* (2000) tested 29 human microsatellite primers on baboons (*Papio cynocephalus*), six of which amplified and were polymorphic. Among the six loci, one exhibited departure from Mendelian inheritance, with three of six mother-daughter pairs having no shared alleles at this locus. These mother-daughter pairs could not be excluded at any other locus. Mislabeling of samples in the field was unlikely to be the cause of this problem. Furthermore, two of the three mother-daughters mismatches were confirmed on three or four gels, in three independent PCRs.

Unlike the case of Smith *et al.* (2000) mentioned above, the mismatches in the present research were distributed across all loci, most of which did not diverge from HWE. Since the genotyping for the same locus could be repeated without variation and with high score rates, and for most loci there was no obvious homozygote or heterozygote excess, null allele mismatches should not cause problems. Colony record error is almost certainly the main cause for the mismatches. Vaňková *et al.* (2001) tested the accuracy of visual observations of maternity assignment in farmed red deer by DNA analysis and confirmed that in 43 out of 48 cases (89.58%) the visual assignment through behavioral analysis were correct. They found 14 cases of adoption in their study. The percentage maternal assignment accuracy in the present study is close: to the deer figure of 87.4%. However, Launhardt *et al.* (2001) analyzed 67 pairs of observed mother-infant relationships in langurs (*Semnopithecus entellus*) using 5 microsatellite loci and found only one mismatched case (1.5% error rate).

Null alleles or PCR based artifact alleles can arise because of divergence in flanking sequences between species from which the primers originated and the species in which they are used. Lathuillière *et al.* (2001) found genotyping errors indicated by false alleles or false homozygotes when human primers were used to amplify Barbary macaque DNA. The cause of the errors was divergence between human and Barbary macaque DNA flanking sequences, which ranged from 2.8% to 25.5% depending on loci. They reduced the amplification error rate from 41% to 17% on average, when they redesigned one specific macaque primer of the primer pair in three loci. By redesigning both primers using macaque sequence, genotyping errors were eliminated completely. Similarly, Bradley *et al.* (2000) redesigned primers using human genome sequence data for some shorter amplicons and were able to reduce amplification artifacts when the primers were used for chimpanzee and gorilla.

In a small scale study of the green monkeys of VMRC by Newman *et al*. (2002), one of 11 loci showed nonsignificant homozygote excess. In the present study, 3 of the 16 loci showed homozygote excess. These homozygote excesses could be explained by null

alleles. However, the monkeys in VMRC are not a natural population. Given that the colony is established with 57 founders, the monkeys tested form a highly related group with chances of inbreeding. This could be the cause of homozygote excess instead of potential null alleles. Constable *et al.* (2001) found that most close inbreeding was avoided successfully in a pedigree construction of chimpanzees (*Pan troglodytes schweinfurthii*), although, one case of close inbreeding was detected in their study. Close inbreeding was found to be absent in the present study, but inbreeding can still not be ruled out due to the small population size.

There are reports of other microsatellite behavior which could contribute to artifactual genotypes. Harris and Crandall (2000) reported that a number of microsatellite loci within ribosomal internal transcribed spacer 1 and 2 (ITS1 and ITS2) regions of crayfish species show considerable variation in the number of repeats within individuals. Bénit *et al.* (2001) provided evidence of *de novo* deletion of paternal alleles in relation to human disease. Five microsatellites they tested flanking the NDUFS1 gene on chromosome 2q failed to be inherited from father to son. Cserhalmi-Friedman *et al.* (2002) presented the first documented case of paternal germline mosaicism for a recessive mutation in microsatellites flanking the NDUFS1 gene, which appeared as a 'new' mutation in the offspring of a clinically and/or genetically unaffected parent. Also a mutation of a maternally inherited microsatellite allele in the Olive Ridley sea turtle (*Lepidochelys olivacea*) has been reported and this kind of mutations are reasonably common (Hoekert *et al.*, 2002).

5.4.3 Shortcomings of approaches

Nurnberg *et al.* (1998) used 11 polymorphic loci from 51 markers cross-amplified using human microsatellite primers to test paternity for 200 male rhesus macaques (*Macaca mulatta*). They found some cases could not be solved unequivocally with the microsatellite markers, despite a combined exclusion rate of more than 99.9% in all three social groups tested. They resorted to minisatellite DNA fingerprinting to complete unresolved paternity decisions. However, microsatellites are still considered superior due

to their ease and high throughput procedure. Minisatellite DNA fingerprinting might be an additional tool for paternity assessment of the more than 500 green monkeys at VMRC, UCLA, but it cannot resolve the entire mismatch problem if there are observational and recording errors and large number of unsampled animals.

As Jones and Ardren (2003) pointed out, the exclusion technique is most powerful when there are few candidate parents that are preferably not related to each other, and highly polymorphic genetic markers are available. However, the large pool of candidate parents makes the exclusion approach impractical when a large number of loci are needed to find a single nonexcluded parent or parent-pair assignment. One of the potential weaknesses of a strict exclusion approach is that genotyping errors, null alleles, and mutations will impair complete exclusion or lead to false exclusions. The categorical and fractional likelihood techniques implemented in Newpat.xls (Amos, 2000) and other software alleviate the problems, but do not eliminate the complications. The more data to be analyzed, the more acute the problem becomes. The assay of additional loci or additional individuals increases the probability that a dataset will contain errors or mutations.

Discovery and understanding causes of experimental and analyses error are as important as construction of a pedigree. While the pedigree can be expanded and used for further research such as genetic or linkage mapping, the discovery of causes of error can be leant to improve future methodology and analysis.

Chapter 6

De novo isolation of microsatellite markers for the Queensland fruit fly *Bactrocera tryoni*

6.1 Introduction

In many cases microsatellite markers which have been isolated originally from a focal species, can often be used in other related species (see Chapter 5). However, it is normally required that microsatellites be isolated *de novo* by screening an established partial genomic library with simple sequence oligonucleotide probes. This chapter reports the isolation of novel microsatellite markers for the Queensland fruit fly *Bactrocera tryoni* (Diptera: Tephritidae).

A partial genomic library used for traditional microsatellite isolation is made by transforming bacteria with target DNA fragments ligated to vector DNA. Probes used to hybridize to the cloned DNA immobilized on a membrane or inert beads are concatenated simple sequences representing the target microsatellite. In Arthropoda, as well as in Mammalia, the most abundant short tandem repeats (STRs) are AC repeats (Tóth *et al*., 2000). AT repeats are the next most abundant STRs in human and *Drosophila* (Katti *et al*., 2001), but these repeats cannot be detected by specific probes due to their intrastrand-complementarity, which makes the probes self-annealing. DNA from positive clones recovered from screens is sequenced and characterized to enable primers to be designed based on the flanking unique sequences. Loci most likely to be polymorphic are those with higher numbers of repeats. PCR conditions are optimized to enable the amplification from every individual of a population.

The Queensland fruit fly *Bactrocera tryoni* is a major pest and a great threat to the Australian horticultural and fruit industry. This species infests almost all commercially grown fruit crops (Bateman, 1991), causing significant economic impact on agricultural

activities. *B. tryoni* is endemic to eastern Australia with potential to further spread both within and outside of Australia. Information on Queensland fruit fly population structure, gene flow and other population parameters estimated using microsatellites have proved to be very helpful in understanding fruit fly population history and invasion biology (Clarke *et al*., 2005). A sterile insect technique (SIT) population control project has been developed and implemented for the Queensland fruit fly in order to minimize the use of chemical insecticides. Research showed that the release of sterile male flies is a great help in reducing fly population density (McInnis *et al*., 1994). Molecular techniques have also proved to be very useful in managing the sterile male fly release program (Meats *et al*. 2002).

Bactrocera neohumeralis is a sympatrically distributed pest species sharing surprising similarity with *B. tryoni* (see next chapter for more details). Both species are serious pests for horticulture. *B. neohumeralis* is not considered to be as serious a pest as *B. tryoni* chiefly due to its more limited geographical distribution. Both feed on a wide range of native wild fruits and almost all cultivated fruits (Bateman, 1991). For effective control of these serious pests, the fruit fly laboratory at the University of Sydney has endeavored to characterize the two pest species at the biological and molecular level (Green and Frommer 2001; Kinnear *et al.*, 1998; Meats *et al.*, 2002, 2003; Morrow *et al.*, 2000; Wang *et al.*, 2003, Yu *et al.*, 2001; Zhao *et al.*, 1998, 2003).

Since microsatellites are useful for distinguishing between the two common *Bactrocera* species, and for developing further species-specific pest control protocols, Kinnear *et al.* (1998) isolated 16 microsatellite markers from *B. tryoni*. Six were polymorphic and informative in a population study of *B. tryoni*. A few, not only cross amplified in the sibling species *B. neohumeralis*, but also exhibited different allelic frequencies between the species (Frommer, personal communication). The microsatellite primers developed from *B. tryoni* (Kinnear *et al.*, 1998) work well on *B. neohumeralis* (Morrow, unpublished data). An analysis of population structures in *B. tryoni*, using these six microsatellite markers, showed remarkable homogeneity within regions between years and clear heterogeneity between regions (Yu *et al.,* 2001). To provide the best description

of the boundaries between populations and species, however, a larger number of loci than 6 is needed (Goldstein *et al.*, 1999). This chapter reports the development of additional new markers.

6.2 Materials and Methods

6.2.1 Genomic library and microsatellite isolation

6.2.1.1 Partial genomic library construction

Genomic DNA extraction: High molecular weight DNA was prepared from laboratory flies of *B. tryoni* maintained in cages under conditions of constant temperature (25°C) and natural light cycles over a period of 5-6 years. The extraction procedures are based on the method of Bender *et al.* (1983). Briefly, pooled flies were homogenized, lysed with extraction buffer, heat denatured, treated with proteinase K and RNase, extracted with phenol and chloroform-isoamyl alcohol, and precipitated with sodium acetate and ethanol. For details please consult Appendix 2.

DNA recovery by Gel purification: Genomic DNA was digested by either *Sau3*A1 (Boehringer, Ridgefield, CT) at 37°C or *Taq*1 (Biolabs, Arundel, Qld) at 65°C for three hours before heat inactivation at 80^{o}C for 20 minutes. Enzyme digested DNA fragments were separated on a 0.7 % agarose gel. DNA fragments sized from 360 bp to 1500 bp were excised from an ethidium bromide stained agarose gel. Target DNA was recovered using a Geneclean kit (Bresatec, Thebarton, SA). The principle of the protocol is to use the silica matrix ('glass milk') to bind double stranded DNA, the 'glass' is then pelleted from the DNA solution. The detailed procedure is described in Appendix 2.

Vector DNA preparation: Vector DNA *pBluescript KS+* (Stratagen, Miami, Florida) was extracted using BRESApure Plasmid kits (Bresatec Ltd.). A bacterial culture was alkaline lysed. During a gravity-flow anion exchange chromatographic procedure, DNA was bonded to the BRESApure anion exchange matrix while most contaminants passed through the column. A wash step removed all remaining contaminants. Ultra-pure DNA

was eluted in a high salt buffer and precipitated with isopropanol. Detailed procedures are available in Appendix 2. 5 μg vector DNA was digested with either *Bam*HI or *Cla*I at 37°C in a 50 μl reaction volume for three hours. Sample mixtures were heated at 65°C for 20 minutes to inactivate the restriction enzymes following the digestion. The digested mixtures were visualized by electrophoresis together with uncut controls.

Dephosphorylation of vector DNA: Both *Bam*HI and *Cla*I digested vector DNAs were dephosphorylated by calf intestinal alkaline phosphatase (CIP; New England Biolabs, Ipswich, MA) to remove 5'-phosphate groups from linear DNAs and prevent self-ligation. The use of CIP (10,000 U/ml) for dephosphorylation and phenol-chloroform for DNA recovery, modified from Sambrook *et al.* (1989) and manufacturer's instructions, is described in Appendix 2.

DNA ligation: *Sau3*A1 digested DNA fragments were ligated into dephosphorylated *Bam*HI digested vector DNA and *Taq*1 digested DNA fragments were ligated into dephosphorylated *Cla*I digested vector DNA, using a DNA ligation system provided by Amersham (Piscataway, NJ). Approximately 20 ng plasmid vector DNA was mixed with size-selected DNA fragments at a 3:1 molar ratio in 5 μl volume containing 100 mM Tris-HCl (pH7.6) and 5 mM $MgCl_2$. 20 μl reaction buffer solution A (Amersham) was added to each tube and the tube was mixed vigorously by vortex. 5 μl solution B (Amersham) containing T4 ligase was thawed on ice, then added to each tube and the tube was mixed gently by inversion. The tube was incubated at 16°C for 2 hours.

Bacterial transformation: Competent *Escherichia coli* DH5α cells were prepared for transformation using a heat shock method. 15 μl plasmid/insert ligation mixture was added to either 150 μl or 200 μl aliquots of ice-cold competent cells. 15 μl vector ligation mixture without insert was added to a 150 μl aliquot of ice-cold competent cells as one control, and 100 μl of ice-cold competent cells without ligation mixture were used as a second control. All the transformation tubes were left on ice for 30 minutes without mixing. Cells were then heat shocked at 42°C for 30 seconds, immediately returned to ice

and left for 1-2 minutes. 850 µl pre-warmed SOC media (2% tryptone, 0.5% yeast extract, 0.05% NaCl, 10mM $MgCl_2$, 2.5mM KCl, and 20mM glucose, pH 7.0) was added to each tube. The tubes were incubated at 37°C at ~210 rpm of banchtop microfuge for 1.5 hours to recover. 50-200µl aliquots were plated on 9 cm diameter LB-AXI agar plates (1.5% agar, 40 µg/ml IPTG, 100 µg/ml ampicillin, 50 µg/ml X-gal). Plates were inverted and grown overnight at 37°C until blue/white colonies were discernible.

6.2.1.2 Partial genomic library construction

Colony lifts: Single recombinant colonies (white) were transferred to a grid of 14cm diameter duplicated LB-AXI plates with a density of 554 colonies per plate (about 4704 recombinant clones totally). One of each duplicate LB-AXI plate was taken as a master plate to make plasmid DNA and another was used for colony lifting. The plates were inverted and grown at 37°C for 16-18 hours.

Colony lifts were made to nylon membranes (Hybond N+, Amersham). Two colony lifts were obtained for each plate in order to compare the positive signals. Orientation holes and lift tabs were made for accurate alignment of the master plate, membranes, and autoradiography. Plates were cooled at 4°C for at least 30 minutes before the procedure started. Gloves and forceps were used to avoid fingerprints and other contaminations. For detailed procedure please read Appendix 2.

Probe preparation: Oligonucleotides of short, simple-repeated sequences, $(GTT)_7G$, $(TG)_{10}$, $(TC)_{10}$, were end-labeled with [γ^{32}P]-ATP using T4 polynucleotide kinase (PNK) and purified. A 20 µl labeling mixture for each oligonucleotide probe consisting of 1 µl oligonucleotide (50 pmol/µl, GeneWorks, Thebarton, SA), 2 µl 10x One-Phor-All buffer (Pharmacia Biotech, Piscataway, NJ), 4 µl [γ^{32}P]-ATP (10 µCi/µl, GeneWorks), 1 µl PNK (9.5 U/µl, Pharmacia Biotech), and 12 µl MilliQ water was used for the labeling reaction in a 37°C water bath for 1 hour. 130 µl MilliQ water was added to increase the total volume to 150 µl. The labeled nucleotide probes were precipitated from the

unincorporated nucleotides with one volume (150 µl) 8M NH_4OAc and 2.5 volumes (750 µl) ice-cold 100% ethanol. End-labeled probes were recovered by centrifugation at 13,000 rpm of benchtop microfuge for 30 minutes. The supernatant was taken off, leaving 25-30 µl residue. The pellet was washed with 500 µl 70% ice-cold ethanol, spun down, air-dried, and re-dissolved in 20 µl $TE_{0.1}$ buffer. The level of incorporation was checked and the three labeled probes were pooled into one 1.5 microfuge tube. 400 µl $TE_{0.1}$ buffer was added to bring the volume up to ~460 µl. The tube was mixed well and set on ice until the probes were added to the hybridization bottle.

Prehybridization of nylon membrane: Five dry membranes, separated by Hybaid mesh with colony side up, were rolled up tightly and put into a large hybridization bottle (Hybaid). Membranes were spread out by tapping the bottle gently. 30 ml prehybridization solution consisting of 5x SSPE (0.75 M NaCl, 50 mM $NaH_2PO_4.2H_2O$ and 5 mM EDTA) and 5x Denhardts solution (0.1% Ficoll, 0.1% PVP, 0.1% BSA and 0.5% SDS) was added into each bottle carefully so that no visible bubbles formed between membranes. Caps were replaced and the weight balanced bottles were put in the Hybaid hybridization oven. The direction of bottle rotation was arranged to let the membranes unroll. The membrane was prehybridized at 46°C for one hour.

Hybridization: Hybridization was carried out in the same bottles and oven as used for prehybridization. About 115 µl (460 µl/4) an equal mix of [γ^{32}P]-ATP labeled $(GTT)_7G$, $(TG)_{10}$, $(TC)_{10}$ probes was added to each prehybridized bottle, after the mix was denatured at 95°C for 3 minutes. The bottle was gently agitated to ensure an even distribution of the probes in the hybridization. Overnight hybridization was at 46°C,

Membrane washing: After the hybridization fluid was removed, the membranes, still in place in each bottle, were washed twice with fresh 100 ml wash solution I (2x SSPE, 0.1% SDS) in the Hybaid hybridization oven at 42°C for 15 minutes. The wash solution was flushed down the sink with copious running water. The membranes were then washed with fresh 100 ml solution II (0.5x SSPE, 0.1% SDS) in the Hybaid oven at 42°C

for 1 hour. The bottles were arranged so that the rotation direction in the Hybaid oven rolled up the membrane and meshes more tightly, making them easier to take out of the bottles. The membranes were removed using forceps, separated from the meshes and rinsed in a container with 1L solution III (2x SSPE)]. More stringent washes were carried out in a plastic container with lid covered, using 1L washing solution II (0.5x SSPE, 0.1% SDS) in a water bath at 50°C to 52°C for another half hour with shaking.

Autoradiography: The wet nylon membrane was heat-sealed in a plastic sheet and placed in a cassette (Amersham) with intensifying screens (Dupont), superposed with a Kodak double emulsion film (Integrated Sciences, Willoughby, NSW). Cassettes were then stored at -70°C overnight or longer depending on signal intensity. Films were developed using CURIX 60 X-ray film processor (AGFA, Burwood, VIC).

Plasmid DNA preparation: Positive hybridization signals were located on the master plates, and plasmid DNA was prepared from the positive clones using a modification of the alkaline lysis and phenol-chloroform extraction method described by Sambrook *et al.* (1989) (Appendix 2).

Sequencing: 16 μl of each sample containing 1.5 μg mini-prep plasmid DNA, 4.8 pmol either T3 or T7 primers, and autoclaved MilliQ water was prepared for sequencing in both directions. Plasmid DNA was sequenced by Sydney University and Prince Alfred Macromolecular Analysis Center (SUPAMAC). The sequencing system used was either an ABI Prism 373 or 377 (version 2.1.1 or version 3.0, respectively) automated DNA sequencer. The sequencing data were then aligned and analyzed using the Sequencher 3.0 computer program (Gene Codes Corporation) for identifying the microsatellite sequences.

6.2.2 Microsatellite genotyping

6.2.2.1 DNA extraction

DNA used for PCR was extracted from single fly heads using a modified quick Chelex extraction method (Walsh *et al.*, 1991; Moritz *et al.*, 1992).

Each head was cut from a frozen fly using forceps, put in a 1.5ml microfuge tube and ground with a sterile wooden stick. 1 ml of near boiling (95-100°C) 5% Chelex solution [5 g Chelex® 100 resin in $TE_{0.1}$ (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) to a total volume of 100 ml] was added. The tubes were capped well and placed in a hot water bath (95-100°C) for 15 minutes with intermittent vortexing, and then cooled on ice.

DNA prepared using Chelex can be stored in a –20°C freezer for long term purposes. When needed, the tubes were thawed and vortexed, and centrifuged at 13,000 rpm of banchtop microfuge for 3-5 minutes to settle the chelating resin. The supernatant was used as a template in PCR reactions.

6.2.2.2 PCR and genotyping

Using Amplify 2.5Br software (Engels 1994) or sometimes manually, primers (18-24 nucleotides) were designed to amplify fragments from 80 to 200 bp in length. The PCR amplification reaction was performed in 1x Tth Plus DNA polymerase PCR reaction buffer, containing a final concentration of 0.15 µM of each primer (Life Technologies, Lino Lakes, Minnesota; one of which was 5'-HEX labeled), 2 mM $MgCl_2$, 0.2 mM dNTPs, 0.0165 unit/µl of Tth plus polymerase (Biotech International, Needville, Texas), and 0.3 µl/µl supernatant of Chelex extraction DNA. The thermal cycling program was fixed for all the reactions: 3 minutes for initial denaturation at 94°C; then 30 cycles of the following: 30 seconds at 94°C, 45 seconds at 50°C, and 50 seconds at 72°C; then 5 minutes of extension at 72°C as a final step.

All amplified fragments were detected on 6% polyacrylamide gels, using GS-2000 Automatic Fragment Analyser (Corbett Research, Sydney) which uses a solid state laser (green 530nm) to scan fluorescent signals; the gel images formed by these signals were manipulated and scored by the computer program One-Dscan (Scanlytics, Rockville, MD).

6.3 Results and Analysis

A total of 4704 clones were screened in the partial library of *B. tryoni* genomic DNA. One hundred fourteen clones showed positive signals, and these positive clones were stored as glycerol stocks. Forty-three strong positive clones were sequenced (Table 6.1), with four duplicate sequences being found. Excluding the approximately 10% of duplicated clones, the total length of screened clones was ~5080 kb, derived from the total clones with an average length of 1.2kb (Table 6.1). A total of 74 di- and tri-nucleotide microsatellites were identified from 43 clones (quite a few inserts had more than one microsatellite), with about 72% being dinucleotide repeat loci. The estimated dinucleotide repeat density was approximately one in every 96 kb. Because only two dinucleotide probes were used and only strong positive clones were sequenced, the actual density of dinucleotide repeats could be much higher, and could be close to that of *Drosophila melanogaster*, around one in every 60 kb (Schug *et al.*, 1998a). Among these dinucleotide repeats, CA repeats were most abundant, accounting for about 83%. AT repeats were found at a level of 14% of the total dinucleotide repeats, although no specific probes were used for it in this screen and indeed it is impossible to screen for AT repeats in this way because of the feature of self-annealing. For trinucleotide repeats, CAA repeats were most abundant, as expected. However, ATT or TAA repeats were occasionally found without using any probe for them.

Table 6.1 List of the sequenced positive clones, identified simple sequence repeats and approximate insert sizes. Those sequences of poor quality were not deposited in GenBank.

Sequenced clone	Simple sequence repeats	Approximate insert size (kb)	length submitted to GenBank (bp)	accession number
1.1	$(GT)_5GC(GT)_3$, $(CA)_2(TG)_2(CA)_5$	0.8	424	DQ462609
1.5	$(GGT)_4$	1.2		
1.6	$(GT)_{10}$, $(AGC)_4$	1	583	DQ462612
1.7	$(AC)_8$	1	553	DQ462610
2.2	$(GT)_3AT(GT)_5$, $(GTTTTT)_2(GTT)_2$, $(CACC)_3$	0.8		
2.4	$(CA)_7$	1.4		
2.6	$(GT)_8$, $(CAA)_{10}$	1.2	902	DQ462611
2.9	$(CT)_8$	0.6		
2.10	$(GT)_7$, $(AT)_2(GT)_4$	1	679	DQ462613
2.12	$(GT)_8$, $(ACAT)_3(AT)_2GT(AT)_4$	1.6	1096	DQ462614
3.2	$(GTT)_4GCT(GTT)_2$, $(GAGT)_4$	1	449	DQ462615
4.1	$(GT)_8$	0.8	322	DQ462616
4.3	$(TTG)_6TGG(TTG)_3$, $(A)_9(CAA)_3(A)_4$	1	580	DQ462617
4.6	$(GT)_8$	1	560	DQ462618
5.3	$(CA)_6$, $(CA)_7$	1	708	DQ462619
5.4	$(CA)_5$	1.2		
5.5	$(CA)_4$, $(AT)_2(TA)_3$	1.1	757	DQ462620
5.8	$(TG)_9$, $(CA)_8$	1.6		
5.10	$(TG)_{10}$, $(CAA)_4$…$(CAA)_3$	1.2	899	DQ462621
5.16	$(CAA)_6$	1.2	990	DQ462622
5.18	$(AC)_9$, $(GT)_6$	1.6	1278	DQ462623
6.1	$(GTT)_7$	1.2	999	DQ462624
6.5	$(CTT)_4$	1.2	779	DQ462625
6.6	$(GT)_6$, $(GT)_8$	1.2	885	DQ462626
*6.8=6.11	$(CAA)_5$....$(CAA)_5$…$(CAA)_4$, $(CAT)_5$	1.2	773	DQ462627
6.9	$(GT)_9$	1	569	DQ462628
*6.10=6.13	$(CA)_4TA(CA)_7$, $(CAA)_4CATTA(CAA)_4$	1.3	941	DQ462629
6.12	$(GT)_7$	1.2	893	DQ462630
6.19	$(TA)_5(TGTA)_3(TA)_2$, $(CAA)_3(CA)_2$	1	722	DQ462631
6.26	$(ATT)_4$	1.3	1027	DQ462632
7.2	$(CA)_6CG(CA)_3$	1	465	DQ462633
7.6	$(TG)_3$	1	548	DQ462634
7.9	$(TG)_{10}$	1.3	1048	DQ462635
*8.4=8.11	$(CA)_7$	1.6	1365	DQ462636
8.5	$(TA)_2(TG)_5$, $(GTT)_5G(GTT)_2$, $(CAA)_5$	1.6	1288	DQ462637
8.6	$(GT)_3G(T)_4(GT)_3$, $(GT)_3GC(GT)_4$, $(GTT)_7$…$(GTT)_3$	1	597	DQ462638
8.9	$(TA)_4C(TA)_3$	1	724	DQ462639
8.12	$(TG)_{11}$, $(CA)_6$, $(TG)_6$	1.6	1384	DQ462640
*8.15=10.3	$(GTT)_6$	1	703	DQ462641
8.17	$(GT)_4A(TG)_7$, $(ATT)_4ACT(ATT)_4$	1.6	1308	DQ462642
8.18	$(TAA)_3$	1.8		
8.19	$(GT)_{10}$, $(AC)_4TCGC(ATGT)_3$	1.1	776	DQ462643
9.1	$(AC)_{11}$	1	625	DQ462644

* indicates duplicated clones

Sequences from 36 clones providing sequence from both strands were deposited into GenBank (Table 6.1). Blasting these sequences against GenBank using nucleotide-nucleotide searching for short, nearly exact matches under default settings (with repeat

masker and without limit to organism and subset of database sequences) (Table 6.2) revealed that clone 6.26 and 7.6 are highly likely from the mitochondrial genome. Based on the bit score (using a specific formula that takes account of similarity and gap penalty) and e-value (the number of "matches" expected by chance) (see BLAST web pages in NCBI website for details of the bit score and e-value), clone 6.26 produced significant alignment with the NADH gene of several dipteran mtDNAs, including *B. tryoni*, *B. aquilonis*, *B. curvipennis*, *B. quadrisetosa*, *B. caryeae*, *B. occipitalis*, *B. musae*, *B. oleae*, *B. viscenda*, *B. cognate*, *B. albistrigata*, *B. papayae*, *B. correcta*, *B. carambolae*, *B. umbrosa*, *B. kirki*, *B. trilineola*, *B. antilocapra* and many *Drosophila* species. Clone 7.6 has significant alignment with mtDNA sequences of flies including *B. oleae*, *Ceratitis capitata* (Mediterranean fly), *Cochliomyia hominivorax* (primary screwworm) and quite a few *Drosophila* species. Sequence from clone 5.5 aligned significantly to the *mariner* family of transposable elements in *B. tryoni* genome (Green and Frommer 2001). For the remaining 33 clones, most blasted sequences did not significantly align with anything in GenBank, but there were still some interesting alignments. Clone 6.12 significantly aligned to one of the microsatellites markers isolated from *B. cacuminata* (Song *et al*. 2006). Clone 9.1 significantly aligned to a microsatellite marker from *Glossina palpalis palpalis*. Clone 2.6 best aligned to one of the microsatellite from *B. cacuminata*, but also aligned to the optomotor-blind gene (*omb*) of *Drosophila polymorpha* with lower bits (73.8) and e-value of $9 \times e^{-10}$. Clone 8.5 aligned to the membrane transporter gene, better known as *white* gene, of *B. tryoni* (Table 6.2). The *omb* gene product not only plays a role in early embryo development, but is also essential for the abdominal pigment pattern of *Drosophila* (Brisson *et al*. 2004). The *white* gene product, cooperating with other transporters, is responsible for transportation of eye pigment precursors to pigment cells in *Drosophila* (Sumitania *et al*. 2005).

Table 6.2 Blast data for the sequences submitted in GenBank, using bit score and e-value (see BLAST web pages in NCBI website) for the measure of the alignment.

Sequenced clone	Species with best alignment	Aligned accesion No	DNA source of the aligned sequence	Bit	e-value
1.1	*Mus musculus*	AC126032	chromosome 1 clone	60	6.0E-06
1.6	*Mus musculus*	AC101983	chromosome 6 clone	61.9	2.0E-06
1.7	*Mus musculus*	AC137749	chromosome 6 clone	52	0.002
2.6	*Bactrocera cacuminata*	DQ023215	microsatellite marker	135	3.0E-28
2.10	*Homo sapiens*	AC074349	chromosome 4 clone	48.1	0.036
2.12	Human	AL390741	chromosome 6 clone	48.2	0.057
3.2	*Drosophila melanogaster*	AY069770	SD01241 full length cDNA	48.1	0.024
4.1	Zebrafish	BX890576	clone DKEY-203O14	44.1	0.027
4.3	*Homo sapiens*	AC024023	chromosome 4 clone	44.1	0.5
4.6	*Drosophila melanogaster*	NM_132328	SD01241 full length cDNA	44.1	0.48
5.3	*Drosophila melanogaster*	AC104287	X chromosome clone	54	6.0E-04
5.5	*Bactrocera tryoni*	AF505880	transposon mariner transposase pseudogene	123	9.0E-25
5.10	*Dictyostelium discoideum*	XM_640881	hypothetical protein partial mRNA	52	0.003
5.16	*Vipera ammodytes*	AF332697	Bov-B LINE	46.1	0.22
5.18	*Mus musculus*	AC099612	chromosome 14 clone	44.1	1.1
6.1	*Mus musculus*	AC098726	BAC clone RP23-3E20	46.4	0.17
6.5	*Homo sapiens*	AC108713	chromosome 3q clone	50	0.11
6.6	*Bos taurus*	XM_606455	similar to eukaryotic translation initiation factor	44.1	0.77
*6.8=6.11	*Trypanosoma cruzi*	XM_806865	hypothetical protein mRNA, partial cds	46	0.17
6.9	*Schistosoma japonicum*	AY81003	protein mRNA, partial cds	50.1	0.008
*6.10=6.13	*Homo sapiens*	AC093891	chromosome 4 clone	46.1	0.21
6.12	*Bactrocera cacuminata*	DQ023218	microsatellite marker	113	1.0E-21
6.19	*Candida albicans*	XM_715579	cyclic nucleotide phosphodiesterase mRNA	46.1	0.61
6.26	*Bactrocera tryoni*	AY037474	mtDNA NADH	301	2.0E-78
7.2	*Dictyostelium discoideum*	XM_640523	hypothetical protein partial mRNA	52	0.002
7.6	*Bactrocera oleae*	AY210703	mtDNA complete genome	351	1.0E-93
7.9	*Drosophila melanogaster*	AC007891	cromosome 3R clone	52	0.004
*8.4=8.11	*Drosophila erecta*	AY190937	clone DERF01_40_C09 (D1409)	60	2.0E-05
8.5	*Bactrocera tryoni*	BTU9710	memberan transporter gene (*white*)	52	0.005
8.6	*Mus musculus*	AC127577	BAC clone	44.1	0.51
8.9	Zebrafish	BX08856	clone CH211-239H19 in linkage group 4	50.1	0.01
8.12	*Mus musculus*	AC165236	chromosome 5 clone	50.1	0.02
*8.15=10.3	*Homo sapiens*	AC026470	chromosome 16 clone	46.1	0.15
8.17	*Mus musculus*	AC125467	BAC clone	52	0.005
8.19	*Dictyostelium discoideum*	XM_631008	hypothetical protein partial mRNA	44.1	0.67
9.1	*Glossina palpalis palpalis*	AY008368	microsatellite marker	80.2	8.0E-12

* indicates duplicated clones

From the sequenced inserts, 38 microsatellite markers from 31 clones were chosen and primers were designed (Table 6.3). Except for locus 9.1a, which was generated from the *Sau3A*1/*Bam*HI combination site, all of the loci were generated from the *Taq*1/*Cla*I clones. Of these, 26 were polymorphic and amplified well in a preliminary test. A further test on both *B. tryoni* and *B. neohumeralis* showed 22 of them could be scored and used for population analysis (Table 6.3).

Table 6.3 Microsatellites isolated from *B. tryoni*. Clone names followed by a letter were used for the locus names. The first 22 were used for population analysis of Queensland fruit flies in the present study. Allele number was based on 60 flies of *B. tryoni*.

Locus	repeat sequence	legth (bp)	number of alleles	Primer name	Primer	Tm (°C)
1.1a	(GT)5GC(GT)3	131	8	1.1a5	ATTGCCAAGTTCCTTCAAACC	60.35
				1.1a3	ATTAGCCACAACCAAGCTCG	60.27
1.7a	(TA)4 .....(AC)8	165	27	1.7a5	CCAAAATATTACACTCAGCGG	57.36
				1.7a3	ATCATTGCTACTGCCTCTGC	58.08
2.6a	(GT)8	87	13	2.6a5	CTGAATGCTGCGCTTGTATG	60.57
				2.6a3	CGGCCATCATTAATCATTCG	61.18
2.6b	(CAA)10	113	10	2.6b5	GTGCGTGTGTCTTGGAATATG	59.06
				2.6b3	CCAGAAACCAACGACAATGC	61.47
2.9a	(CT)8	121	19	2.9a5	CCGAATTTCATTGGAAGCAC	60.45
				2.9a3	CACGAAGAGGTGCAGTAAATTG	59.81
3.2b	(GTT)4GCT(GTT)2	169	11	3.2b5	AACACGAGCACTTGTTGCAC	59.95
	(TA)3 .....(GAGT)4			3.2b3	AGGCAATATTCCCACCACAC	59.68
4.1a	(GT)8	80	10	4.1a5	ATGCACACCTATTTGCGGTG	61.88
				4.1a3	CACGCAACTCATTTCAATGC	60.27
4.3a	(TTG)6TGG(TTG)3	94	11	4.3a5	TATGCTGCGTCGCTTATACC	58.96
				4.3a3	CAAGTGCTTTAACTGCGTTCG	60.97
5.3b	(AC)7	144	8	5.3b5	CGTCAACATGCAGAGTTGCT	60.06
				5.3b3	CAACCAACAACAATATGTTTGC	58.01
5.8a	(TG)9	133	7	5.8a5	CACGACGTATGATTTCAATTGC	60.38
				5.8a3	GCTTACTGCAATTCCCACTTC	58.85
5.10a	(TG)10	141	8	5.10a5	GGCGGAATAAAGTGCGAAGA	62.47
				5.10a3	GCCAAACTGCTGTGGCATAG	61.77
6.6b	(GT)8	160	8	6.6b5	GTGTGTTGCTGTGGTAAATAAC	54.87
				6.6b3	CTACTTACAGGAAAAGCATGTG	54.47
6.8a	(CAT)5	147	4	6.8a5	AGCAACACTATAGGCTGGTC	54.55
				6.8a3	GTAACGCGTTACTGTCATTG	54.39
6.8b	(CAA)5 ..... (CAA)5	132	8	6.8b5	TAACGCGTTACATTAATGACG	56.51
				6.8b3	GCATTTACTTTCATCATTCAGC	56.59
6.10b	(CA)5 .....(CA)4	147	6	6.10b5	GGAAAATGTGCTGGACTAAGC	58.85
				6.10b3	ACTACGCCATTCTTCCTTCAG	58.47
6.12a	(GT)7	130	11	6.12a5	CTGGGAAGTGATTAATTAGCAC	55.67
				6.12a3	TCTGAAAGCTGCCAATATCC	57.91
7.2b	(CA)6CG(CA)3	147	6	7.2b5	GCGATCCAATTTCCGAATAAC	60.64
				7.2b3	CAATTGCCAGCAATAAGTGG	59.18
7.9a	(TG)10	132	8	7.9a5	AGTTCGCTTGTAGTTGCTGC	58.34
				7.9a3	CCGCTTGCATCCTACTTTAG	57.69
8.5a	(CAA)5	145	5	8.5a5	CAAGAAGAATAATAAAGCGCGG	60.23
				8.5a3	CGAACTATGCATGTATGCGTG	60.17
8.6a	(TTG)7	128	26	8.6a5	CCTTTGCCACTCTTGCCTAC	59.88
				8.6a3	ATTAGTGTCAGGGCAGGACAG	59.22
8.12a	(TG)11	141	24	8.12a5	ATGGAGTTGCAACGAACAGG	61.1
				8.12a3	ACTATGCGCCACTAACTTCCAC	60.56
9.1a	(AC)11	142	11	9.1a5	ACAAATGGCGTGGCATGTC	62.94
				9.1a3	TGTGCGCTAACAATCTAACGC	61.32
1.6a	(GT)10	137		1.6a5	TGCATGTCTCGTTCTAAGGC	59.03
				1.6a3	TGAAGTGTTTGCGATAGCACAG	61.37
2.12a	(TA)3A2(TA)4CA(TA)2(TGTA)3	120		2.12a5	GCTGATAATACATTTCGTTTTGG	57.8
				2.12a3	TGGTGCAGTGCGACTAAAAG	60.05
2.12b	(CA)8	142		2.12b5	GGCGAAATCACATCAATTTC	58.01
				2.12b3	TAATACTTTCGACTGGTATGCG	57.09
3.2a	(TTG)3CC(GTT)2A(TTG)2	109		3.2a5	TTATTAGCGCACAGCTGGTC	59.09
				3.2a3	CCCAAACTGCTCCGATTACC	61.75
4.6a	(GT)8	137		4.6a5	GCATGCATGTGACAAGGAGA	60.85
				4.6a3	CCATGTACAGCCGAGGTAAATG	61.62
5.3a	(CA)6 .....(GT)3	131		5.3a5	TTTATCTCGTCGGCTCAGTG	59.02
				5.3a3	AGTTGGAGGGAGTGATTTCG	59.14
5.16a	(CAA)6	132		5.16a5	ACGAGTGACCAACCCTTTCC	61.32
				5.16a3	GTTCATGTCCTTAGTCGTCGC	59.75
5.18a	(TG)9	97		5.18a5	CAGATTCATACGATTTGATGTGG	59.35
				5.18a3	ATGCTTTGCTTGAACACAGC	59.08
6.1a	(AAC)7	100		6.1a5	TGTGTATATGCATGGGAACG	57.86
				6.1a3	TCTGTGCAGCATTCAATACAG	57.52
6.6a	(GT)6	160		6.6a5	AGCTAGCTGAACTAGTAGCTGC	55.33
				6.6a3	CAGCTATGGCACTCAATCCT	57.92
6.9b	(GT)9	100		6.9b5	CTAAATGCATGTGAACGTCG	57.79
				6.9b3	ACGAAAATCCACATCTAATAATC	54.25
6.19a	(TA)5(TGTA)3(TA)2	117		6.19a5	AATATTCACCGGAATGATATG	54.11
				6.19a3	ATTTGGAATTACATAAAGGGC	54.99
8.6c	(GT)3GC(GT)4	129		8.6c5	TTTGCATGCATCTACTTATGC	56.59
				8.6c3	GAAATGGCATCAACAAACAG	56.61
8.15a	(GTT)6	159		8.15a5	CGGTCACACGGTAGAAACAG	58.97
				8.15a3	AACTAGGCTTGACTTCAGGAATG	59.21
8.17a	(TTA)4CTA(TTA)4	113		8.17a5	CCTTTAACGGTCGCTCAGTTC	61.15
				8.17a3	AGCGCGAAATACCAGAATGC	62.51
8.19a	(CA)7(TA)3 .....(GT)10	156		8.19a5	TGGAAATCCACATTGGCAAG	61.81
				8.19a3	TGAACTGTGAGTTTTGCTGGG	61.4

6.4 Discussion

6.4.1 Microsatellite isolation and cross-amplification

Microsatellite primers from one species can be used on another relatively closely related species, as has been shown already in the present research (Chapter 5), However, *de novo* isolation of microsatellites is always needed when no primers are available from a related species. In genomic screening, use of different restriction endonucleases and nucleotide probes could affect the efficiency of microsatellite isolation. Kinnear *et al.* (1998) isolated 6 polymorphic markers out of 16 microsatellites from *B. tryoni* when they used *Sau3A*1/*Bam*HI restriction endonuclease combination on genomic DNA segments and vector DNA for genomic library construction. In the present study, all 38 microsatellite markers of interest were generated from the *Taq*1/*Cla*I combination inserts except for one *Sau3A*1/*Bam*HI combination. However, the probes used in the two isolations were different, and that could affect the outcome as well.

In insects, cross-amplification is not easy, even within a genus. Cross-amplification from *Drosophila melanogaster* to *D. hibiscus* was tried for 17 markers, but none was successful (Wang, unpublished results). Cross-amplification for the closely related species *D. melanogaster* and *D. simulans* was successful for 15 microsatellites, but four loci displayed an almost completely different allele spectrum (Harr *et al*., 1998). Song *et al*. (2006) tried to cross-genotype *B. cacuminata* using 18 pairs of primers designed for *B. tryoni* and *B. dorsalis*, with half of them generating PCR products but with no or very little polymorphism. Fortunately, the present research showed that microsatellite primers developed in *B. tryoni* work well on *B. neohumeralis*, with a similar allele frequency distribution. Microsatellite primers developed from *B. tryoni* can be used to type *B. neohumeralis* with 100% success rate. Not only these 22 polymorphic markers, but also the monomorphic markers showed similar sized PCR products between the two species (data not shown). These primers can also be used to genotype several other *Bactrocera* species: *B. aquilonis, B. papayae, B. cucurbitae, B. opiliae* and *B. breviculeus* (Wang *et*

al., 2003; Shearman, unpublished), although the allele spectrum varies between these species.

6.4.2 Chromosome mapping of the microsatellites

Zhao *et al*. (2003) generated an integrated genetic and physical map of the five autosomes/linkage groups of *B. tryoni* using most of the microsatellite markers reported here and previously (Kinnear *et al.*, 1998), but also including two restriction fragment length polymorphisms (RFLP) and three visible morphological markers (*orange eyes*, *white*, and *bent wings*) (Meats *et al*. 2002). The 28 mapped microsatellites included 9 isolated by Kinnear *et al.* (1998) and 19 isolated in the present project. The chromosome map for the microsatellites (Figure 6.1) shows that they are not distributed evenly along the genome: 1.1a, 1.7a, 2.9a, 6.1a, 6.8b and 7.9a are on chromosome 2; 4.3a and 9.1a on chromosome 3; 6.12a, 8.12a and 8.19a on chromosome 4; 2.6a, 2.6b and 8.5a on chromosome 5; 3.2b, 5.8a, 6.10b, 7.2b and 8.6a on chromosome 6. Among these five autosomes, from a karyotype (Zhao *et al*., 1998), chromosome 3 is the shortest. The lengths of the remaining four chromosomes are more similar, especially when the left arms are compared, although the exact length of the five autosomes was not indicated in the original study of Zhao *et al*. (1998). However, there are insufficient data to test the uniformity or otherwise of microsatellite distribution across the genome of *B. tryoni*.

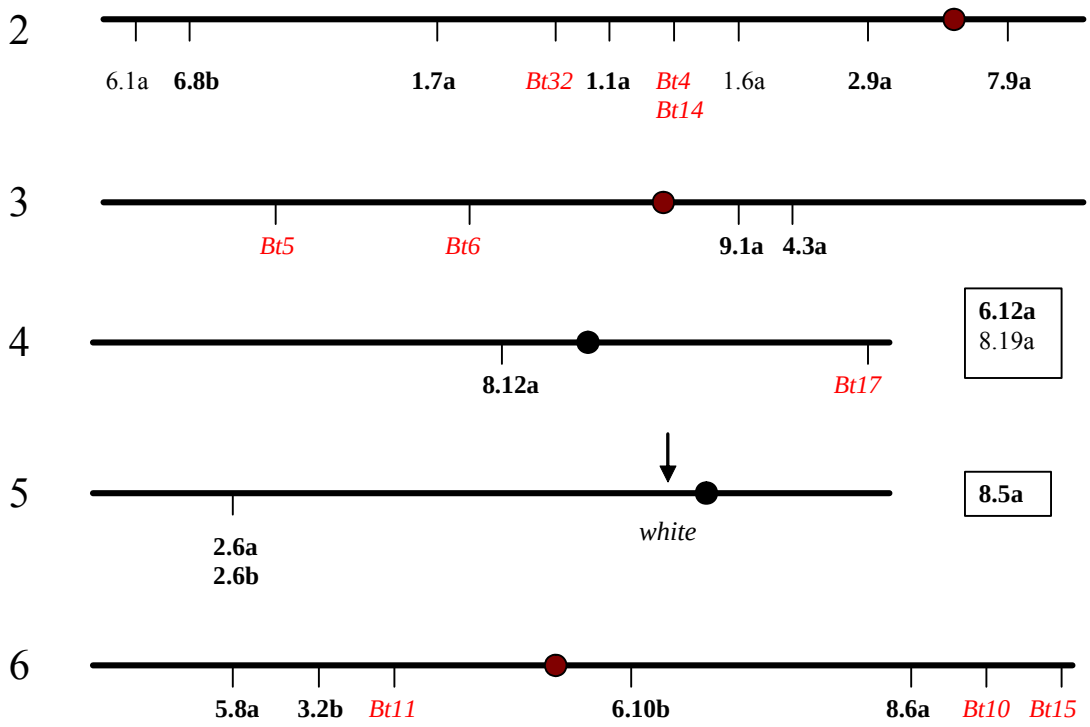


Figure 6.1 Location of microsatellites on the genetic map of *B. tryoni* as reported by Zhao *et al.* (2003). Italic loci (except *white*) are microsatellites previously isolated (Kinnear *et al.* 1998). Bold black loci are microsatellites isolated and used for population analysis in the present study. Three microsatellites in the boxes, 6.12A, 8.19A, and 8.5A, have been assigned to a chromosome, but not positioned on the chromosome. Three microsatellite (not bold), 6.1A, 1.6A, and 8.19A, were isolated in the present project, but have not been used in the consequent population analysis of this research. The position of the *white* gene, and by inference 8.5A, is shown with an arrow. Chromosome lengths are not proportionally drawn.

Microsatellite markers with the same number in their names are from the same clone and thus must map to the same position. Therefore the positions of some microsatellite markers not mapped in *B. tryoni* can be inferred from the marker name. These markers include 3.2a, 6.8a and 8.6c. Locus 6.8a will collocate with 6.8b in chromosome 2. Locus 3.2a and 8.6c will collocate with 3.2b and 8.6a respectively in chromosome 6.

Figure 6.1 shows three loci that were assigned to chromosomes but have not been mapped to sites. Among them, 8.5a was shown by blast to correspond to the *white* gene, which has previously been mapped on chromosome 5 close to the centromere about 41 cM away from 2.6a and 2.6b (Figure 6.1) (Zhao *et al*., 2003). Clearly microsatellite 8.5a must map to this same position.

Mutations in the *white* gene were first reported determining eye colour variation in *Drosophila* by Morgan in 1910 and provided the first ever example of sex linkage. The gene encodes a member of the ATP-binding cassette (ABC) transporter superfamily. It is interesting that the *white* gene maps to autosome 5 in *B. tryoni* (Zhao *et al*., 2003), while it is located on the X chromosome of *Drosophila* (Hazelrigg and Petersen 1992), indicating that sex chromosomes are not so highly conserved in insects as in mammals.

6.4.3 Abundances of different base repeats

In mammals, CA repeats are the most abundant microsatellites, being twice as frequent as AT repeats, and three times as common as GA repeats (Beckmann and Weber, 1992; Stallings, 1995). In the present research, the exact proportions of the three motif repeats could not be calculated because the AT specified probe could not be used. Even so, the AT repeat number was still three times higher than the GA repeat number. This could be due to A+T rich genome in insect (Powell 1997).

6.4.4 Microsatellite isolation approach

There are various methods for microsatellite isolation, including different proceedings for probe labeling and immobilizing hybridized DNA. The method used in the present research is a classical one. Many new protocols have been developed for fast, easy and environmentally friendly use. Probes labeled with non-radioactive tags (Engler-Blum *et al*., 1993; Toonen, 1997) are less hazardous (http://www.inapg.inra.fr/dsa/microsat

/microsat.htm), but are less sensitive than probes labeled with radioactive isotope (Zane *et al*., 2002). Alternatively the surface used for hybridizing genome DNA can be streptavidin-coated paramagnetic beads for 5' biotinylated probes (Kandpal *et al*. 1994; Kijas *et al*. 1994) instead of membranes (Karagyozov *et al*. 1993; Armour *et al*. 1994). The genomic DNA isolated this way for library construction is microsatellite enriched (Ostrander *et al*. 1992; Paetkau 1999; Lloyd *et al*. 2004; Song *et al*. 2006). These newly developed methods are much more efficient and more cost effective than traditional method (Zane *et al*., 2002). Future development of microsatellites from *Bactrocera* and other species will almost certainly use bead based technology for enriching libraries. However, in the longer run, SNP based methodologies may completely replace microsatellites.

Chapter 7

Species discrimination and population analysis using microsatellite polymorphisms

7.1 Introduction

7.1.1 Comparison of the sibling species *Bactrocera tryoni* and *B. neohumeralis*

Two closely related Queensland fruit flies, *Bactrocera tryoni* and *B. neohumeralis* (Diptera: Tephritidae), are sympatric in a narrow area of eastern Australia from northern NSW through Queensland, although *B. tryoni* extends much farther west and south (Figure 7.1). They are very similar morphologically and in their ecological requirements (Gibbs, 1965). There is very little evidence that interspecific competition occurs between the sympatric *Bactrocera*, even when the two species breed in the same fruits (Fitt, 1989). They can be distinguished by the color of the humeral calli: yellow in *B. tryoni* and brown in *B. neohumeralis*. Hardy (1951), in a taxonomic review, considered that flies with yellow calli and flies with brown calli should be regarded as color variants of one species, *B. tryoni*. In the laboratory, interspecific mating occurs at low frequency, and viable and fertile hybrids can be produced with intermediate callus coloration. Flies with intermediate coloration can be found in nature as well (Birch and Vogt, 1970; Gibbs, 1968; Wolda, 1967a, b). When extensive collections of naturally infested fruits were made in Rockhampton, Queensland, at various times of the year (Gibbs, 1967), only about 8% of emerging flies were intermediates (Gibbs, 1968). Wolda (1967a, b) regarded the wild intermediate color forms as not being hybrids but variants of either species; there was no problem in identifying the great majority of flies. Gibbs (1968), Birch and Vogt (1970) and Lewontin and Birch (1966) indicated that at least some of the intermediates must be true hybrids. However, in the field there is a clear difference in mating time: *B. tryoni* mates at dusk, whereas *B. neohumeralis* mates in the middle of the day (Gee, 1969a; Lewontin and Birch, 1966; Tychsen and Fletcher, 1971). Dusk mating is dominant in the F1 hybrids, but no simple segregation of behavior occurs in the F2

generation (Smith, 1979). Dusk- and day-mating lines can be selected from the F2 generation (Meats *et al.*, 2003). It has been shown that the biological rhythm which controls mating and pupal ecdysis is under the influence of light and temperature (Bateman, 1955, 1958; Barton-Browne, 1957; Smith, 1979; Tychsen, 1978 and Tychsen and Fletcher, 1971). It is likely that behavioral differences between the two species are controlled by multiple genes. Birch and Vogt (1970) suggested that pigmentation in the humeral callus and other parts of the body is controlled by multiple genes. Major changes in behavior caused by genetic factors are seldom the result of single gene polymorphisms, but more often the result of multifactorial genetic mechanisms (Ross and Keller, 1998).

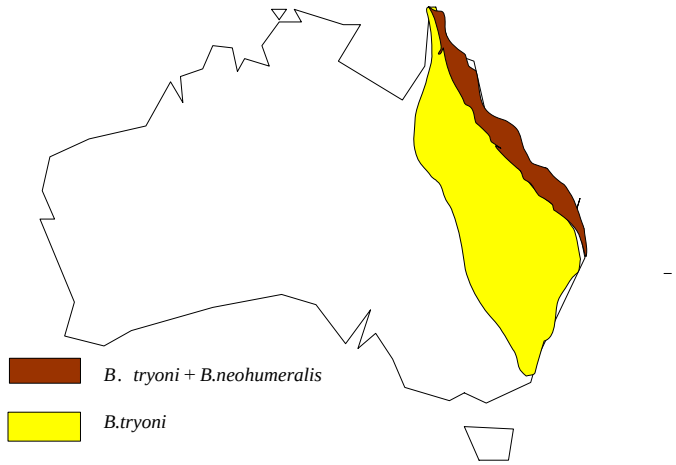


Figure 7.1 Distribution of *B. tryoni* and *B. neohumeralis* in eastern Australia

Previous researchers have attempted to use various polymorphic systems to distinguish the two species at the molecular level. McKechnie (1975) used isozymes and failed to find fixed differences between the two species. Morrow *et al.* (2000) used two nuclear DNA sequences [*white* gene and ribosomal internal transcribed spacer 2 (ITS2)] and two mtDNA sequences [cytochrome b (Cytb) gene and cytochrome oxidase subunit II (COII) gene]. A fixed difference in the ITS2 sequence represented the only molecular marker capable of distinguishing the two species in that study, but that could be an artifact of the small sample size. An *et al.* (2002) compared the gene sequence, expression level and Thr-Gly repeat number of the circadian rhythm gene *period* (*per*) of the two species, and found no species-specific differences between the two Queensland fruit fly species. Thr-

Gly repeats demonstrate species-specific differences between two drosophilid sibling species (Wheeler *et al.,* 1991; Nielsen *et al.,* 1994). A study on another circadian rhythm gene *Cryptochrome* (*cry*) showed no species specific difference in the CRY amino acid sequence, but the expression level of *cry* differed in antennae and brain between the two species (An, personal communication). All genomic and mtDNA sequences analyzed so far have shown that the two species exhibit a high level of shared polymorphic variation.

7.1.2 Use of microsatellites for population structure analysis

Microsatellite polymorphisms are broadly used for population analysis due to their hypervariability; they are richly informative; abundant in most genomes; and evolve rapidly (Goldstein and Schlötterer, 1999). Ross *et al.* (1999) assessed genetic structure of the fire ant *Solenopsis invicta* with molecular polymorphisms involving four classes of neutral nuclear markers: allozymes, codominant RAPDs, microsatellites, and dominant RAPDs. They found that microsatellites consistently provided the most effective markers for interpreting genetic structure and detected the greatest differentiation among higher level groups, although the four classes of neutral nuclear markers were largely concordant in describing population structure.

Another reason microsatellites are widely used for genetic and evolutionary studies of natural populations is their convenience in manipulation and assay. The alleles at microsatellite loci are usually scored by their molecular weight. Microsatellites are composed of either perfect repeats or imperfect repeats (interrupted or compound repeat). All these repeat types are involved in different mutation styles due to a complex pattern of single-strand slippage. Palsbøll *et al.* (1999) sequenced alleles at one locus that revealed a complex pattern of single-strand slippage on several levels. The mutation pattern involved not only single repeats, but also multiple and partial repeats. For microsatellites with single or multiple repeat mutation patterns, the step-wise mutation model can be relied on for subsequent analyses. Single-strand slippage involving partial repeats can cause less confident scoring due to the irregularity of allele sizes. When loci present a partial repeat mutation pattern, the nucleotide sequence of the microsatellite

array often provides additional evolutionary information, and precision will be reduced when microsatellites are analyzed based only on repeat number.

In the present study, the objective was to determine whether recently identified polymorphic microsatellites could be used to distinguish between populations of the two closely related species and to utilize them for analysis of population structure.

7.2 Materials and Methods

7.2.1 Microsatellite genotyping

Genomic DNA extraction and microsatellite genotyping are described in Chapter 6.

7.2.2 Population Sampling

Samples of both species were selected from six locations at different latitudes from 28^{o}S to 17^{o}S along a narrow area of the east coast of Queensland: Sunnybank, Nth Ipswich, Bundaberg, Mackay, Murray, and Freshwater (Figure 7.2). Fourteen individual flies from each location were tested for all the 22 loci (Table 6.2). However, three flies of the Freshwater population of *B. tryoni* were morphologically identified later as a species other than *B. tryoni*, so the data generated from these three flies were excluded from the analysis in this thesis. These three flies were identified using body remains after the heads were used for DNA extraction, so the exact species could not be identified. All the flies were collected during 1994 by the participants in the "National Fruit Fly Experiment," including members of CSIRO's Double Helix Club (Osborne *et al.,* 1997).

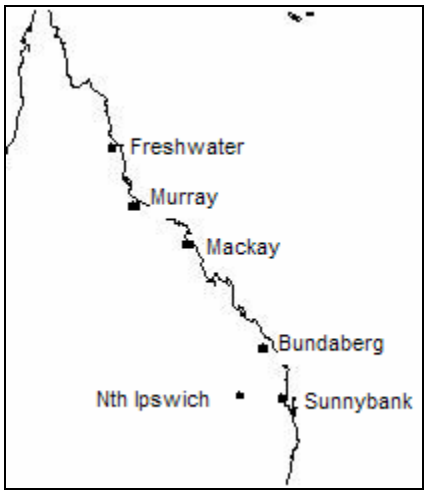


Figure 7.2 Map of Queensland showing geographic locations of the six populations sampled for the analysis of the two *Bactrocera* species in the present study

7.2.3 Statistical Analyses

The traditional chi-square (Pearson χ^2) test assumes that the dataset is large enough to approximate to the appropriate normal or chi-square distribution. The consequent *P* value is given by calculating the tail area of the limiting normal distribution. This so called asymptotic *P* value often brings problems when the sample is small or the sample has an abnormal distribution. Non-parametric statistics calculate *P* values based on true value distribution, by exact tests or permutation procedures, without the need for assumptions about the underlying distribution (Lehmann, 1975). This "exact" *P* value is most appropriate when the sample is not large. A standard error for the estimated *P* value can be obtained using a system of batches (Guo and Thompson, 1992). The computational burden of non-parametric tests has been overcome by development of suitable computer programs.

In the present study, computationally intensive non-parametric statistics have been used. Most analyses were performed using Arlequin 2.0 (Schneider *et al.*, 2000), including Hardy-Weinberg Equilibrium (HWE) test, population genetic structure inferred by the analysis of molecular variance (AMOVA), population pairwise genetic difference, and estimation of gene flow (Nm) derived from pairwise Fst values. Homogeneity and

differentiation tests were done using GenePop 7.3 (Raymond and Rousset, 1995a). The averaging linkage clustering trees were constructed using MEGA3 (Kumar *et al*, 2004).

7.2.3.1 Hardy-Weinberg equilibrium test

The exact test for HWE used a modified version of the Markov-chain random walk algorithm described by Guo and Thompson (1992). The modified version used in Arlequin leads to the same results as the original, but is more computationally efficient.

First, a k x k contingency table was built; k is the number of alleles. Under the null-hypothesis of no association (each allele is independent), the probability of obtaining the observed results is given by Levene (1949). All states of the contingency tables were explored by random walk between states of the Markov-chain. The switch to a new table with the same marginal allele counts would be accepted when the new table had a higher probability.

The *P*-value of the test is the proportion of the visited tables having a probability smaller or equal to the observed (initial) contingency table.

7.2.3.2 Homogeneity and differentiation test

The homogeneity and differentiation test was performed using Genepop, which uses an exact test described by Raymond and Rousset (1995b). This exact test used a contingency table built using genotype data, and explores the potential tables bearing the same marginal totals but with a probability equal or less than the observed sample configuration under null hypothesis of panmixia, with a Markov-chain procedure similar to that used in HWE test (Goudet *et al*., 1996). During the random walking between the Markov-chain states the probability of observing a table less or equally likely to the observed sample was estimated to give the *P* value for each batch. The error of the estimated *P* was calculated by partitioning the number of steps into the number of batches (Raymond and Rousset, 1995b).

7.2.3.3 Population genetic structure inferred by Analysis of Molecular Variance

A hierarchical analysis of molecular variance (AMOVA) partitions the total allele frequency variance into covariance components due to differences between species, between populations within species, and within populations respectively. The total molecular variance is the sum of the covariance components. The action of evolutionary forces results in intraspecific and interspecific differentiation, and the differentiation is conveniently quantified based on these parameters using fixation indices (Wright, 1951; Cockerham 1969, 1973).

The analysis of molecular variance used in Arlequin has two approaches. One is essentially similar to the analysis of variance framework defined by Cockerham (1969, 1973) and extended by Weir and Cockerham (1984), Long (1986) and Weir (1996); and the other is derived from Excoffier *et al*. (1992), but adapted to both diploid and haploid data.

In typical analysis of population structure such as in Cockerham (1969, 1973) and Weir and Cockerham (1984), the fixation index Fst is used to measure differentiation (reduction in heterozygosity) between populations, Fis measures the proportional reduction in heterozygotes within subpopulations due to inbreeding, and Fit measures the reduction in heterozygosity due to both.

The analyses implemented in Arlequin and used in this study do not include within population analysis analogous to Fis, but include a higher hierarchical level, namely species. Here the fixation index Fst measures differentiation between populations within species, Fsc differentiation between species, and Fct the effect of both

The significance of the fixation indices was tested using a non-parametric permutation approach described by Excoffier *et al.* (1992), consisting of permuting individuals among populations within and between species, and populations within species. Under this

distribution-free approach, the usual normality assumption in analysis of variance tests is no longer necessary, nor is equal variance premises among populations or groups of populations (Schneider *et al*. 2000).

The input file in the present study used genotypic data with an unknown gametic phase, as microsatellites are codominant markers. No within-individual level was chosen, and "species" is the "group" level.

7.2.3.4 Population pairwise genetic comparison

In random populations, the conventional pairwise genetic estimator Fst, obtained in Arlequin by permuting haplotypes between populations, compares allele frequencies between populations and may serve as a measure of genetic isolation by distance (Raymond and Rousset, 1995a; Weir, 1996). Under the hypothesis of no population differentiation between the populations, all the individuals were permuted to obtain a null distribution. The pairwise Fst values are weighted F-statistics averaged over loci (Weir and Cockerham, 1984; Michalakis and Excoffier, 1996), and given in a matrix. The *P* value of the test is the proportion of permutation leading to a pairwise Fst value higher than or equal to the one observed. The *P* value is also given in the form of a matrix.

7.2.3.5 Gene flow measurement

Slatkin and Barton (1989) compared three indirect methods for estimating average levels of gene flow. The conclusion was that maximum-likelihood methods tend to yield biased estimates when relatively small numbers of locations are sampled, and the Fst method is likely to be more useful under realistic conditions although Fst and rare-allele methods are expected to be equally effective in analyzing ideal data.

In Arlequin, the average level of gene flow across all the loci was calculated from the Fst of Weir and Cockerham (1984) averaged over populations within species. In addition, a matrix of Nm was obtained from the Slatkin linearized Fst, as Fst/(1-Fst), under the assumption that gene flow is the main source for gene variation and the mutation rate is

negligible as compared to gene flow. The absolute number of gene exchanges is estimated under the equilibrium of drift and mutation.

7.2.3.6 Averaging linkage clustering tree construction

An averaging linkage clustering tree was constructed with MEGA3 (Kumar *et al*, 2004). The most widely used sequential agglomerative, hierarchical and nonoverlapping approach to find clusters is the unweighted pair-group method using an arithmetic average (UPGMA) (Weir 1996). Assuming the rate of evolution is constant throughout the evolutionary history of the taxa, the UPGMA tree-making method produces a rooted tree based on the pairwise distance data generated in section 7.3.4 and input as a distance matrix (Nei and Kumar, 2000).

7.3 Results and Analysis

7.3.1 Microsatellite genotyping and comparison of two sibling species

Fourteen flies from each population for each species were genotyped. In the Freshwater population, three putative *tryoni* flies were found having identical homozygote genotype, which were obviously distinct from other individuals and consequently departed from HWE. This was consistent with a morphological identification that these three flies were a species other than *B. tryoni* (see 7.2.2). So the data from these three flies were excluded from all the present analyses. Allele counts are presented in Table 7.1.

Table 7.1 Microsatellite allele numbers genotyped in 6 populations of two *Bactrocera* species

		B. tryoni						*B. neohumeralis*					
locus	allele	Sunnybank	NthIpswitch	Bundabarg	Mackay	Murray	Freshwater	Sunnybank	NthIpswitch	Bundabarg	Mackay	Murray	Freshwater
1.1a	105	0	0	0	0	0	0	0	**1**	0	0	0	0
	107	0	0	2	0	1	0	5	1	5	8	5	6
	121	0	0	0	0	0	0	0	0	0	0	**1**	0
	123	9	4	9	10	8	5	2	3	5	5	1	3
	125	**1**	0	0	0	0	0	0	0	0	0	0	0
	129	0	0	0	0	0	**4**	0	0	0	0	0	0
	131	12	21	14	9	13	12	19	16	15	14	16	18
	133	2	1	3	3	0	1	2	5	3	1	3	1
	total	24	26	28	22	22	22	28	26	28	28	26	28
1.7a	154	0	0	0	0	0	0	0	0	0	0	**2**	0
	159	0	0	0	0	0	0	0	1	0	0	1	0
	160	0	0	0	0	0	0	0	0	0	0	**1**	0
	162	0	0	0	0	0	0	0	0	**2**	0	0	0
	163	0	0	1	0	0	0	0	0	1	0	0	1
	164	0	0	0	0	0	0	0	0	0	0	0	**1**
	165	5	5	2	5	3	2	4	3	3	4	2	0
	166	0	**1**	0	0	0	0	0	0	0	0	0	0
	167	6	7	16	8	11	2	17	9	7	12	13	13
	168	1	1	0	2	2	3	0	0	1	0	1	3
	169	5	5	2	3	2	5	1	7	11	5	4	5
	170	0	0	0	0	0	2	0	2	0	0	1	1
	171	1	1	0	0	0	0	2	1	2	1	1	0
	172	0	0	1	0	0	0	0	0	0	0	0	1
	174	0	2	0	0	0	0	2	2	0	1	0	0
	175	0	0	1	0	0	1	0	0	0	1	1	2
	176	1	0	0	0	0	0	0	1	1	2	0	1
	177	0	0	0	0	**1**	0	0	0	0	0	0	0
	179	0	0	0	0	1	0	0	0	0	1	0	0
	181	0	0	2	0	0	0	0	0	0	0	1	0
	182	0	0	0	**1**	0	0	0	0	0	0	0	0
	183	1	0	2	0	1	0	0	0	0	1	0	0
	184	0	1	0	1	0	0	0	0	0	0	0	0
	185	0	**1**	0	0	0	0	0	0	0	0	0	0
	186	0	0	1	2	1	0	0	0	0	0	0	0
	187	0	0	0	0	0	**1**	0	0	0	0	0	0
	207	0	0	0	0	0	0	0	**2**	0	0	0	0
	total	20	24	28	22	22	16	26	28	28	28	28	28

Table 7.1 continued

2.6a	56	0	0	0	0	0	0	0	0	0	0	0	**1**
	76	0	1	1	0	0	0	0	0	0	0	0	0
	78	0	0	1	0	1	0	2	0	3	0	1	2
	80	0	2	0	1	0	0	1	0	0	0	1	2
	82	4	2	2	1	0	1	2	5	2	1	2	1
	84	1	1	3	2	0	1	2	0	0	1	1	0
	86	1	2	4	4	2	6	0	3	2	3	1	1
	88	14	9	12	17	10	5	7	10	6	10	7	5
	90	5	5	3	1	8	5	2	4	4	6	4	2
	92	3	1	1	1	2	1	0	4	1	2	2	3
	94	0	3	1	1	0	1	6	2	6	3	5	5
	96	0	1	0	0	1	0	0	0	4	2	2	2
	105	0	**1**	0	0	0	0	0	0	0	0	0	0
	total	28	28	28	28	24	20	22	28	28	28	26	24
2.6b	99	3	5	7	2	7	4	1	0	0	2	1	3
	102	8	8	8	11	8	9	4	6	6	7	8	7
	105	7	3	4	4	5	4	9	11	10	1	10	7
	108	5	4	3	5	3	0	2	1	1	1	0	1
	111	1	4	5	0	2	3	7	3	4	3	4	2
	114	2	1	0	3	3	0	5	3	3	4	5	5
	117	0	0	0	0	0	0	0	0	0	3	0	1
	120	0	1	0	1	0	0	0	0	0	0	0	0
	123	0	0	1	0	0	0	0	0	0	0	0	0
	129	0	0	0	0	0	0	0	0	0	1	0	0
	total	26	26	28	26	28	20	28	24	24	22	28	26
2.9a	99	0	**1**	0	0	0	0	0	0	0	0	0	0
	100	0	0	0	0	0	0	0	0	2	1	1	1
	101	4	7	9	5	5	6	0	3	0	2	2	2
	102	1	0	2	3	2	0	0	1	0	1	0	1
	103	2	1	0	0	0	1	2	0	2	0	3	1
	104	1	0	0	0	0	0	0	0	1	1	1	1
	105	0	0	0	3	0	0	0	0	0	1	0	0
	106	0	0	0	0	0	0	0	0	0	**2**	0	0
	107	0	1	3	1	1	1	0	2	1	0	0	2
	108	0	0	1	1	0	0	0	0	2	2	0	1
	109	0	0	3	1	0	0	3	0	1	0	4	4
	110	1	2	2	2	3	0	4	1	3	1	4	0
	111	2	2	2	1	2	0	0	0	3	0	2	1
	113	0	0	0	**2**	0	0	0	0	0	0	0	0
	114	0	0	0	0	0	0	**1**	0	0	0	0	0
	115	3	2	2	9	7	4	14	9	13	12	9	8
	117	0	4	2	0	0	0	0	0	0	2	0	0
	121	0	0	0	0	2	0	0	0	0	1	0	0
	122	0	**4**	0	0	0	0	0	0	0	0	0	0
	total	14	24	26	28	22	12	24	16	28	26	26	22

Table 7.1 continued

3.2b	154	0	**1**	0	0	0	0	0	0	0	0	0	0
	160	0	0	**1**	0	0	0	0	0	0	0	0	0
	162	0	0	0	0	0	0	0	2	0	1	0	0
	164	7	3	3	2	6	4	3	5	4	2	2	0
	166	0	0	0	0	0	0	1	0	0	1	1	0
	167	0	0	0	0	1	0	0	0	0	0	2	0
	168	8	7	11	7	4	5	2	4	4	2	3	3
	169	9	17	13	15	10	10	22	14	16	16	8	20
	170	0	0	0	0	0	**1**	0	0	0	0	0	0
	172	0	0	0	2	1	0	0	0	0	0	0	0
	178	0	0	0	0	0	0	0	1	0	2	0	1
	total	24	28	28	26	22	20	28	26	24	24	16	24
4.1a	72	0	0	0	0	0	**3**	0	0	0	0	0	0
	74	0	0	0	0	2	0	0	0	0	1	0	0
	76	0	0	0	**2**	0	0	0	0	0	0	0	0
	78	0	0	0	1	0	0	0	0	1	0	1	0
	79	1	0	1	1	1	1	5	2	3	4	2	2
	80	9	12	9	8	7	6	5	9	7	6	10	5
	82	11	15	12	15	11	9	14	16	14	12	11	17
	84	5	1	2	1	4	1	1	1	1	3	4	4
	86	0	0	0	0	1	0	1	0	0	0	0	0
	88	0	0	**2**	0	0	0	0	0	0	0	0	0
	total	26	28	26	28	26	20	26	28	26	26	28	28
4.3a	78	0	**1**	0	0	0	0	0	0	0	0	0	0
	84	0	0	1	0	0	0	3	2	3	2	3	2
	87	0	0	3	0	1	0	0	0	0	2	1	2
	90	4	3	1	4	0	2	0	3	0	1	2	2
	93	22	19	20	21	18	12	2	4	1	1	4	1
	96	2	4	3	1	3	2	14	15	18	16	9	11
	99	0	1	0	0	0	0	1	0	3	1	1	0
	102	0	0	0	0	1	0	1	0	0	0	2	1
	105	0	0	0	0	0	0	0	0	0	1	0	2
	111	0	0	0	0	1	0	5	0	1	1	2	5
	132	0	0	0	0	0	0	0	0	0	**1**	0	0
	total	28	28	28	26	24	16	26	24	26	26	24	26
5.3b	132	0	0	0	0	0	0	0	0	0	**1**	0	0
	138	0	0	0	0	0	0	0	**1**	0	0	0	0
	140	0	0	0	0	**1**	0	0	0	0	0	0	0
	142	0	0	**1**	0	0	0	0	0	0	0	0	0
	144	0	0	9	1	1	0	5	2	2	2	2	0
	146	24	27	14	27	22	14	21	24	26	25	25	25
	148	2	1	4	0	2	5	0	1	0	0	1	1
	152	0	0	0	0	0	**1**	0	0	0	0	0	0
	total	26	28	28	28	26	20	26	28	28	28	28	26
5.8a	110	0	0	0	0	**2**	0	0	0	0	0	0	0
	124	0	0	0	0	0	0	0	0	0	0	0	**1**
	128	0	0	0	0	0	3	0	0	0	0	0	1
	130	0	0	0	0	0	0	0	**1**	0	0	0	0
	132	25	21	20	21	18	13	22	19	24	25	24	20
	134	3	7	5	6	3	4	4	6	0	2	4	6
	136	0	0	1	1	1	0	2	0	4	1	0	0
	total	28	28	26	28	24	20	28	26	28	28	28	28

Table 7.1 continued

5.10a	118	1	2	0	1	1	2	6	4	3	6	5	1
	126	0	**1**	0	0	0	0	0	0	0	0	0	0
	138	0	0	0	0	0	0	2	0	0	0	1	1
	140	9	7	6	6	6	3	7	9	7	12	10	13
	142	5	5	3	5	3	5	1	3	7	2	5	2
	144	12	12	19	14	16	11	11	8	4	6	3	4
	146	1	1	0	2	0	1	0	2	1	0	0	3
	148	0	0	0	0	0	0	1	0	0	2	0	0
	total	28	28	28	28	26	22	28	26	22	28	24	24
6.6b	131	0	7	4	4	7	3	1	6	1	3	3	2
	133	**6**	0	0	0	0	0	0	0	0	0	0	0
	155	0	0	0	0	**2**	0	0	0	0	0	0	0
	157	0	0	0	0	0	0	0	0	0	0	**1**	0
	159	0	1	1	2	2	0	0	0	1	0	0	1
	161	21	17	18	17	13	11	10	12	21	19	18	14
	163	1	3	3	3	1	1	13	7	4	6	5	5
	165	0	0	0	0	1	1	0	1	1	0	1	0
	total	28	28	26	26	26	16	24	26	28	28	28	22
6.8a	145	3	2	4	1	2	1	1	0	0	0	0	0
	151	21	24	19	24	19	17	25	28	27	24	26	27
	157	2	2	5	3	6	2	2	0	1	2	2	1
	169	0	0	0	0	**1**	0	0	0	0	0	0	0
	total	26	28	28	28	28	20	28	28	28	26	28	28
6.8b	114	0	0	0	**1**	0	0	0	0	0	0	0	0
	117	0	**1**	0	0	0	0	0	0	0	0	0	0
	120	1	0	0	1	2	2	2	2	4	2	4	3
	123	0	0	0	0	0	0	0	0	0	**1**	0	0
	129	0	0	1	0	1	0	0	0	0	0	0	2
	132	22	19	23	20	19	17	22	26	22	23	24	22
	135	4	8	3	5	6	3	4	0	2	2	0	1
	138	1	0	1	1	0	0	0	0	0	0	0	0
	total	28	28	28	28	28	22	28	28	28	28	28	28
6.10b	125	0	0	0	0	0	0	0	0	0	0	**1**	0
	137	1	1	1	1	4	2	1	4	2	1	1	1
	141	0	0	0	0	0	0	0	1	1	0	1	0
	143	0	0	0	**1**	0	0	0	0	0	0	0	0
	149	4	2	6	1	4	1	1	0	0	0	0	2
	151	21	23	17	25	16	17	26	21	25	27	21	25
	total	26	26	24	28	24	20	28	26	28	28	24	28
6.12a	121	0	0	0	0	0	0	1	0	0	0	1	1
	124	0	0	0	1	0	0	0	1	2	1	2	1
	125	1	5	4	1	1	0	21	13	17	8	10	9
	127	4	3	2	2	1	5	2	1	1	0	0	1
	129	19	18	18	21	19	10	1	10	2	9	4	7
	131	1	2	4	3	5	1	2	1	4	1	2	1
	132	0	0	0	0	0	0	1	0	0	1	0	2
	133	0	0	0	0	0	0	0	2	2	0	1	0
	135	0	0	0	0	0	**2**	0	0	0	0	0	0
	137	0	0	0	0	**2**	0	0	0	0	0	0	0
	139	**1**	0	0	0	0	0	0	0	0	0	0	0
	total	26	28	28	28	28	18	28	28	28	20	20	22

Table 7.1 continued

7.2b	139	0	0	0	0	0	0	0	0	0	0	0	**1**
	145	0	4	3	0	2	2	3	1	5	2	2	1
	147	1	0	1	2	1	0	0	0	0	0	0	0
	149	23	22	18	21	22	13	21	26	18	24	25	23
	151	2	2	2	3	1	3	1	1	5	2	1	1
	153	0	0	0	0	0	0	1	0	0	0	0	2
	total	26	28	24	26	26	18	26	28	28	28	28	28
7.9a	125	0	0	0	0	0	**1**	0	0	0	0	0	0
	127	0	1	0	0	1	0	1	4	5	5	0	2
	129	2	5	1	0	7	2	9	7	11	14	18	15
	131	3	3	5	2	1	1	0	2	0	0	0	0
	133	7	8	14	10	10	8	14	4	5	7	6	5
	135	5	5	6	13	6	3	0	1	1	0	1	0
	137	9	4	2	2	3	5	0	2	0	0	1	2
	139	0	0	0	**1**	0	0	0	0	0	0	0	0
	total	26	26	28	28	28	20	24	20	22	26	26	24
8.5a	148	8	6	11	11	15	7	12	8	12	11	9	13
	151	0	0	0	0	0	0	0	**1**	0	0	0	0
	154	20	22	17	17	13	11	12	19	14	15	17	15
	172	0	0	0	0	0	**1**	0	0	0	0	0	0
	199	0	0	0	0	0	**1**	0	0	0	0	0	0
	total	28	28	28	28	28	20	24	28	26	26	26	28
8.6a	98	0	**1**	0	0	0	0	0	0	0	0	0	0
	104	0	0	0	0	**1**	0	0	0	0	0	0	0
	108	1	0	1	1	0	0	3	2	1	1	2	0
	110	0	0	0	0	0	0	0	0	1	0	1	0
	112	0	2	0	0	0	0	0	0	0	0	1	0
	116	0	1	0	0	0	1	0	0	0	1	0	0
	117	0	1	1	0	0	0	0	0	0	0	0	0
	118	1	0	0	1	0	0	1	1	0	1	0	0
	120	3	1	6	3	0	1	2	0	3	7	0	4
	121	0	0	0	0	1	0	2	1	3	3	3	2
	122	0	1	0	0	0	0	5	1	0	0	0	1
	124	7	10	10	8	11	9	3	4	3	3	4	6
	125	0	**1**	0	0	0	0	0	0	0	0	0	0
	126	3	0	2	2	3	2	1	4	8	5	1	6
	127	0	0	0	1	0	1	3	2	1	2	1	0
	128	6	5	5	3	1	2	1	5	5	1	0	1
	129	0	0	0	0	**2**	0	0	0	0	0	0	0
	130	6	4	2	4	2	2	1	1	3	1	4	2
	133	0	0	0	0	0	0	0	0	0	0	0	**1**
	134	1	0	0	0	2	0	0	0	0	0	0	0
	136	0	0	**1**	0	0	0	0	0	0	0	0	0
	139	0	0	0	0	0	0	0	0	0	0	**1**	0
	140	0	0	0	0	0	0	0	**1**	0	0	0	0
	144	0	1	0	3	2	0	6	2	0	3	5	3
	153	0	0	0	0	0	0	0	0	0	0	**1**	0
	156	0	0	0	0	**1**	0	0	0	0	0	0	0
	total	28	28	28	26	26	18	28	24	28	28	24	26

Table 7.1 continued

8.12a	106	0	0	0	0	0	**1**	0	0	0	0	0	0
	112	0	0	0	0	0	0	0	0	0	**1**	0	0
	131	0	0	0	0	**2**	0	0	0	0	0	0	0
	132	0	1	3	0	0	2	0	0	0	0	0	0
	133	2	1	0	1	1	1	14	6	11	9	8	10
	134	0	1	0	3	0	0	0	1	0	1	3	0
	135	0	0	2	0	1	1	4	1	2	2	5	0
	136	0	1	0	0	0	0	0	3	2	0	1	1
	137	2	0	0	0	2	0	0	0	0	0	1	0
	138	0	0	**2**	0	0	0	0	0	0	0	0	0
	139	0	0	1	1	1	2	2	1	2	0	2	2
	140	0	1	4	1	0	0	1	3	0	4	1	5
	141	0	1	0	1	2	2	4	3	6	6	4	4
	142	3	1	1	0	0	0	1	2	0	0	0	1
	143	1	1	0	3	1	5	0	2	0	1	1	1
	144	0	0	1	0	2	0	1	1	0	2	0	0
	145	11	6	10	9	8	3	0	1	1	0	0	0
	146	0	3	0	0	0	2	0	1	0	0	1	1
	147	7	2	1	9	3	0	1	1	4	2	1	1
	148	0	1	1	0	1	0	0	1	0	0	0	0
	149	0	0	0	0	0	0	0	**1**	0	0	0	0
	150	0	0	0	0	0	**1**	0	0	0	0	0	0
	157	**1**	0	0	0	0	0	0	0	0	0	0	0
	168	**1**	0	0	0	0	0	0	0	0	0	0	0
	total	28	20	26	28	24	20	28	28	28	28	28	26
9.1a	129	0	0	0	0	0	0	0	0	**1**	0	0	0
	135	21	19	21	16	20	11	9	9	11	10	16	11
	137	0	0	0	0	0	1	4	0	3	3	1	0
	139	0	0	0	0	0	0	1	0	0	0	0	2
	141	5	4	4	4	2	4	6	9	3	9	5	7
	143	1	1	1	5	4	0	5	1	3	4	2	6
	145	0	2	2	0	0	0	0	3	2	2	2	0
	147	0	1	0	0	0	0	1	0	1	0	0	0
	149	0	**1**	0	0	0	0	0	0	0	0	0	0
	151	0	0	0	**1**	0	0	0	0	0	0	0	0
	155	**1**	0	0	0	0	0	0	0	0	0	0	0
	total	28	28	28	26	26	16	26	22	24	28	26	26

All of the loci listed in Table 7.1 showed polymorphisms for all populations tested except that locus 6.8a for the Nth Ipswich population of *B. neohumeralis* was fixed for a single allele (Table 7.2). The data from Table 7.2 show slightly more alleles for all loci in *B. neohumeralis* (687) than in *B. tryoni* (651), but, in contrast, fewer rare alleles in *B. neohumeralis* (24) than in *B. tryoni* (38). The rare alleles are those occurring only once in the two species in the present study. The average number of alleles per locus is 5 (4.9 for *B. tryoni* and 5.2 for *B. neohumeralis*).

Table 7.2 Total number of alleles and number of rare alleles of the microsatellites for populations of the two species

	B.tryoni												*B.neohumeralis*											
	Sunnybank		Nthlpswitch		Bundaberg		Mackay		Murray		Freshwater		Sunnybank		Nthlpswitch		Bundaberg		Mackay		Murray		Freshwater	
Locus	allele	rare	allele	rare	allele	rare	allele	rare	allele	rare	allele	rare	allele	rare	allele	rare	allele	rare	allele	rare	allele	rare	allele	rare
1.1a	4	1	3		4		3		3		4		4		5	1	4		4		5	1	4	
1.7a	7		9	2	9		7	1	8	1	7	1	5		9		8		9		11		9	1
2.6a	6		11	1	9		8		6		7		7		6		8		8		10		10	1
2.6b	7		6		6	1	6		6		4		7		8		5		6	1	5		5	
2.9a	7		9	1	9		10		7		4		5	1	5		9		11		8		10	
3.2b	3		4	1	4	1	4		5		4	1	4		5		3		6		5		3	
4.1a	4		3		5		6		6		5		5		4		5		5		5		4	
4.3a	3		5	1	5		3		5		3		6		4		5		9	1	8		8	
5.3b	2		2		4	1	2		4	1	3	1	2		4	1	2		3	2	3		2	
5.8a	2		2		3		3		4		3		3		3	1	2		3		2		4	1
5.10a	6		3	1	5		5		4		5		5		6		5		5		5		6	
6.6b	3		4		4		4		6		4		3		4		5		3		5	1	4	
6.8a	3		3		3		3		4	1	3		3		1		2		2		2		2	
6.8b	4		3	1	4		5	1	4		3		3		2		3		4	1	2		4	
6.10b	3		3		3		4	1	3		3		3		3		3		2		4	1	3	
6.12a	5	1	4		4		5		5		4		6		6		6		5		6		7	
7.2b	3		3		4		3		4		3		4		3		3		3		3		5	1
7.9a	5		6		5		5	1	6		6	1	3		6		4		3		4		4	
8.5a	2		2		2		2		2		4	2	2		3	1	2		2		2		2	
8.6a	8		11	2	8	1	9		10	2	7		11		11	1	9		11		11	2	9	1
8.12a	8	2	12		10		8		11		10	2	8		15	1	7		9	1	11		9	
9.1a	4	1	6	1	4		4	1	3		3		6		4		7	1	5		5		4	

The allele frequency distributions for most of the loci in the two species are surprisingly similar (Appendix 3). Only eight of the 22 loci showed differentiated allele frequencies between the two species. Among them, two trinucleotide loci, 2.6b and 4.3a, and one dinucleotide, 2.9a, show that the most common alleles (MCAs) have longer bias in *B. neohumeralis*. The biases are not sharp in 2.6b and 2.9a because the frequency peaks are weakened by a sub-peak in *B. neohumeralis* and *B. tryoni*, respectively. There is one repeat difference between the MCAs of the two species in 4.3a, which is the only imperfect repeat among the eight loci that showed differentiated allele frequency distribution patterns. The interruption of the repeat in this locus could have interfered with the evolutionary trend of this locus. Four dinucleotides, 5.10a, 6.12a, 7.9a, and 8.12a exhibit a strong longer bias in *B. tryoni*. For loci 5.10a, 6.12a and 7.9a, *B. neohumeralis* has MCAs (140, 125, and 129, respectively) two repeats shorter than those of *B. tryoni* (144, 129, and 133, respectively). For locus 8.12a, the MCA in *B. neohumeralis* (133) is

six repeats shorter than that of *B. tryoni* (145) (Appendix 3). For the remaining locus of the eight, 8.6a, there is one MCA peak in *B. tryoni* (124) and three peaks in *B. neohumeralis* (124, 126 and 144). There are some allele-specific features, such as 1.1a-129 that occurred 6 times in the Freshwater population of *B. tryoni* but was undetected in all the other 11 populations of the two species; and 6.6b-133, which occurred 4 times in the Sunnybank population of *B. tryoni* but was undetected in all the other 11 populations of the two species (Table 7.1). Moreover, some alleles show species-specific frequency differentiation. Among them, 2.9a-100, 3.2b-162, 5.10a-138 and 6.12a-133 were not detected in *B. tryoni* but appear in at least two of the six populations of *B. neohumeralis*. Similarly, 6.8b-138 and 7.2b-147 were not detected in *B. neohumeralis* but appeared in at least three of the six populations of *B. tryoni* (Table 7.1).

7.3.2 Hardy-Weinberg equilibrium test

The observed and expected heterozygosities for the 12 populations and 22 loci are given in Table 7.3, and the significant deviations from HWE at 0.01 level are given in Table 7.4. All but one (8.5a) of the significant deviations were caused by a heterozygosity deficit. Of the 264 population-locus combinations, 24 (9%) departed from HWE.

Table 7.3 A comparison between observed and expected heterozygosity by each locus and each population for the two fly species

	B.tryoni												*B.neohumeralis*											
	Sunnybank		Nth Ipswich		Bundaberg		Mackay		Murray		Freshwater		Sunnybank		Nth Ipswich		Bundaberg		Mackay		Murray		Freshwater	
Locus	Obs.H	Exp.H	Obs.H	Exp.H	Obs.H	Exp.H	Obs.H	Exp.H	Obs.H	Exp.H	Obs.H	Exp.H	Obs.H	Exp.H	Obs.H	Exp.H	Obs.H	Exp.H	Obs.H	Exp.H	Obs.H	Exp.H	Obs.H	Exp.H
1.1a	0.25	0.66	0.31	0.4	0.79	0.64	0.55	0.71	0.82	0.54	0.44	0.5	0.5	0.52	0.54	0.64	0.79	0.7	0.79	0.66	0.54	0.64	0.5	0.55
1.7a	0.7	0.84	0.75	0.87	0.64	0.71	1	0.81	0.64	0.79	0.5	0.93	0.54	0.56	0.64	0.84	0.57	0.82	0.79	0.78	0.57	0.81	0.57	0.79
2.6a	0.43	0.75	0.79	0.88	0.71	0.8	0.64	0.66	0.5	0.76	0.7	0.82	0.73	0.86	0.93	0.81	0.86	0.9	0.71	0.83	0.85	0.88	0.92	0.9
2.6b	0.92	0.84	0.57	0.86	0.23	0.82	0.54	0.81	0.86	0.84	0.4	0.79	0.31	0.86	0.73	0.87	0.07	0.81	0.5	0.81	0.5	0.73	0.5	0.77
2.9a	0.43	0.92	0.17	0.92	0.46	0.9	0.57	0.88	0.55	0.87	0.33	0.73	0.17	0.7	0.25	0.76	0.23	0.78	0.69	0.82	0.23	0.89	0.64	0.89
3.2b	0.42	0.71	0.5	0.59	0.64	0.64	0.69	0.61	0.82	0.71	0.5	0.72	0.43	0.43	0.62	0.67	0.33	0.58	0.5	0.55	0.38	0.73	0.33	0.37
4.1a	0.62	0.7	0.36	0.58	0.85	0.68	0.64	0.65	0.46	0.78	0.5	0.77	0.38	0.67	0.43	0.58	0.62	0.69	0.77	0.72	0.64	0.77	0.64	0.6
4.3a	0.43	0.43	0.5	0.53	0.43	0.53	0.38	0.34	0.25	0.5	0.25	0.53	0.62	0.72	0.58	0.63	0.62	0.51	0.69	0.62	0.75	0.83	0.62	0.83
5.3b	0.15	0.22	0.07	0.14	0.36	0.69	0.07	0.14	0.15	0.35	0.2	0.54	0.38	0.38	0.21	0.33	0.14	0.2	0.21	0.2	0.21	0.27	0.08	0.15
5.8a	0.21	0.2	0.5	0.44	0.31	0.46	0.5	0.4	0.25	0.5	0.5	0.62	0.29	0.43	0.54	0.43	0.29	0.31	0.21	0.27	0.29	0.31	0.57	0.51
5.10a	0.79	0.75	0.36	0.51	0.79	0.73	0.43	0.74	0.46	0.62	0.82	0.7	0.64	0.81	0.57	0.79	0.46	0.78	0.29	0.75	0.58	0.8	0.58	0.68
6.6b	0.29	0.46	0.5	0.63	0.62	0.55	0.54	0.61	0.62	0.69	0.63	0.52	0.33	0.55	0.62	0.7	0.36	0.43	0.57	0.5	0.57	0.61	0.36	0.63
6.8a	0.31	0.4	0.29	0.33	0.57	0.51	0.29	0.26	0.43	0.55	0.3	0.36	0.21	0.2	0	0	0.07	0.14	0	0.22	0.14	0.2	0.07	0.07
6.8b	0.36	0.37	0.43	0.47	0.36	0.32	0.5	0.47	0.5	0.56	0.09	0.47	0.21	0.37	0.14	0.2	0.36	0.43	0.29	0.38	0.14	0.31	0.36	0.43
6.10b	0.38	0.4	0.23	0.22	0.42	0.53	0.21	0.27	0.67	0.58	0.3	0.36	0.14	0.21	0.38	0.4	0.21	0.27	0.07	0.14	0.25	0.31	0.14	0.27
6.12a	0.38	0.46	0.5	0.6	0.43	0.56	0.43	0.43	0.5	0.57	0.67	0.64	0.36	0.49	0.57	0.73	0.64	0.62	0.54	0.72	0.54	0.74	0.57	0.68
7.2b	0.23	0.22	0.36	0.43	0.5	0.49	0.23	0.41	0.31	0.35	0.44	0.55	0.38	0.34	0.14	0.21	0.57	0.59	0.29	0.33	0.21	0.27	0.36	0.39
7.9a	0.54	0.78	0.54	0.84	0.64	0.7	0.5	0.69	0.86	0.78	0.8	0.78	0.25	0.62	0.6	0.83	0.64	0.72	0.46	0.68	0.23	0.54	0.42	0.63
8.5a	0.43	0.47	0.29	0.42	0.64	0.52	0.64	0.54	0.21	0.58	0.6	0.65	0.5	0.56	0.64	0.52	0.46	0.52	0.69	0.51	0.23	0.47	0.93	0.52
8.6a	0.64	0.85	0.86	0.84	0.79	0.81	0.92	0.87	0.85	0.81	0.78	0.75	0.93	0.9	0.75	0.93	0.86	0.89	0.93	0.89	0.75	0.91	0.85	0.88
8.12a	0.79	0.79	0.6	0.93	0.85	0.83	0.86	0.8	0.58	0.91	0.6	0.92	0.71	0.72	0.93	0.93	0.71	0.79	0.79	0.85	0.79	0.88	0.62	0.82
9.1a	0.43	0.47	0.5	0.58	0.5	0.43	0.38	0.59	0.31	0.41	0.5	0.58	0.54	0.83	0.45	0.72	0.75	0.76	0.64	0.77	0.69	0.59	0.23	0.74

Table 7.4 Loci and populations showing significant departure from HWE. - heterozygosity deficit; + heterozygosity excess.

	B.tryoni						*B.neohumeralis*						
Locus	Sunnybank	Nth Ipswitch	Bundaberg	Mackay	Murray	Freshwater	Sunnybank	Nth Ipswitch	Bundaberg	Mackay	Murray	Freshwater	number of populations
1.1a	-												1
1.7a						-							1
2.6a													0
2.6b			-			-	-		-	-			5
2.9a	-	-	-	-			-	-	-		-		8
3.2b													0
4.1a													0
4.3a													0
5.3b			-										1
5.8a													0
5.10a									-	-			2
6.6b													0
6.8a													0
6.8b						-							1
6.10b													0
6.12a													0
7.2b													0
7.9a											-		1
8.5a												+	1
8.6a													0
8.12a		-			-								2
9.1a												-	1
loci	2	2	3	1	1	3	2	1	3	2	2	2	24

There are 12 population-locus combinations that deviated from HWE for each species; an average of 2 (1-3) loci deviated from HWE for each population. Comparing different loci, however, 0-2 populations deviated from HWE for each locus except 2.9a and 2.6b (Table 7.4). For locus 2.9a, two thirds of either *B. neohumeralis* or *B. tryoni* populations deviated from HWE. For locus 2.6b, two and three populations of *B. neohumeralis* and *B. tryoni*, respectively, deviated from HWE. Locus 2.9a is a partial repeat slippage mutation style microsatellite (see introduction and discussion: Microsatellite markers) with high allelic richness. The fact that most of the significant deviations are heterozygosity deficits implies the existence of null alleles. Linkage to loci which are under strong selection

pressure could also be likely. Two loci, 2.6b and 2.9a, contributed most of the departures (54%). For the rest of the loci, about 4.5% population-locus combinations deviated from HWE. As in the case described by Ross *et al.* (1999), exclusion of the two loci from the analyses of structure using hierarchical F-statistics had little influence on the results (data not shown), as the significance of the fixation indices was tested using a non-parametric permutation approach (Excoffier *et al.,* 1992). Therefore these loci were not excluded when inferring genetic population structure.

7.3.3 Population homogeneity and differentiation test

In an allele frequency based exact test of the 22 markers, according to the method described by Raymond and Rousset (1995b), four markers, 5.3b, 6.6b, 7.9a, and 8.12a, demonstrate population differentiation among the 6 populations of *B. tryoni* (Table 7.5). Five markers, 1.7a, 5.8a, 6.12a, 7.9a, and 8.6a, show population differentiation among the 6 populations of *B. neohumeralis*. Most of the markers show distinctions between the two species. The four markers which do not show significant difference between the two species are 4.1a, 6.8a, 7.2b, and 8.5a (Table 7.5). This exact test of Raymond and Rousset (1995b) showed relatively homogenous genotype frequencies within the species and differentiation between the species. However, as the two species share a great deal of similarity, even with similar allele frequency patterns (Appendix 3), only when the variances are partitioned into different levels to account for the structure of species, population and individuals, can the differences between species be recognized more clearly.

Table 7.5 *P* value of homogeneity test using GenePop exact test (Raymond and Rousset, 1995b) based on allele frequencies

Locus	Between species	Between populations of *tryoni*	Between populations of *neohumeralis*
1.1a	**0.001**	0.246	0.433
1.7a	**0.000**	0.086	**0.011**
2.6a	**0.020**	0.221	0.473
2.6b	**0.001**	0.474	0.330
2.9a	**0.000**	0.055	0.061
3.2b	**0.008**	0.365	0.193
4.1a	0.417	0.401	0.746
4.3a	**0.000**	0.440	0.186
5.3b	**0.000**	**0.000**	0.427
5.8a	**0.025**	0.134	**0.022**
5.10a	**0.000**	0.900	0.094
6.6b	**0.000**	**0.010**	0.098
6.8a	0.053	0.760	0.738
6.8b	**0.046**	0.760	0.403
6.10b	**0.022**	0.263	0.514
6.12a	**0.000**	0.123	**0.012**
7.2b	0.266	0.482	0.271
7.9a	**0.000**	**0.032**	**0.013**
8.5a	0.393	0.126	0.689
8.6a	**0.000**	0.280	**0.014**
8.12a	**0.000**	**0.000**	0.356
9.1a	**0.002**	0.432	0.212

7.3.4 Population structure analysis by AMOVA

The AMOVA results computed by Arlequin (Excoffier *et al*. 1992) are listed in Table 7.6, giving a hierarchical picture of structure between species and populations. For most of the loci, there are significant differences between species, but not between populations among species. These loci include 1.1a, 2.6a, 2.6b, 2.9a, 3.2b, 4.3a, 5.10a, 6.8a, 6.10b, 8.6a, and 9.1a. Nevertheless, five loci, 1.7a, 5.3b, 5.8a, 7.2b, and 8.5a, which show almost identical allele frequency distribution patterns for the two species (Appendix 3), show no significant differences between the two species and have the fixation index between species smaller than the one within species) (Table 7.6). They were therefore considered species-uninformative loci, in contrast to the rest of the 17 loci that defined as species-informative loci and can be used to distinguish the two species.

Table 7.6 AMOVA results obtained using Excoffier *et al.* (1992) permutation approach with Arlequin software, showing fixation index values of F-statistics for each locus and over all loci

	Between species		between populations whithin species		combined of the formers	
Locus	Fsc	P-value	Fst	P-value	Fct	P-value
1.1a	0.05	0.002**	0.01	0.285	0.05	0.006**
1.7a	-0	0.417	0.02	0.008**	0.02	0.004**
2.6a	0.02	0.011*	0.01	0.204	0.03	0.012*
2.6b	0.04	0.003**	-0	0.561	0.03	0.029*
2.9a	0.06	0***	0.01	0.085	0.07	0***
3.2b	0.04	0.0123*	0	0.326	0.05	0.015*
4.1a	0.01	0.049*	-0.01	0.901	-0.01	0.799
4.3a	0.37	0.003**	0	0.371	0.37	0***
5.3b	-0	0.439	0.09	0***	0.09	0***
5.8a	-0	0.334	0.01	0.276	0.01	0.281
5.10a	0.07	0***	0	0.359	0.08	0.001***
6.6b	0.02	0.009**	0.03	0.026*	0.05	0.005**
6.8a	0.07	0.002**	-0	0.548	0.07	0.009**
6.8b	0.03	0.003**	-0.01	0.833	0.02	0.222
6.10b	0.03	0.016*	0.01	0.237	0.04	0.034*
6.12a	0.25	0.002**	0.03	0.015*	0.27	0***
7.2b	-0	0.593	0.01	0.282	0	0.330
7.9a	0.15	0.001***	0.03	0.009**	0.17	0***
8.5a	-0	0.441	0	0.393	-0	0.417
8.6a	0.04	0.001***	0.01	0.080	0.04	0***
8.12a	0.11	0.002**	0.01	0.033*	0.12	0***
9.1a	0.07	0.003**	-0	0.477	0.07	0.001***
all loci	0.08	0.002**	0.01	0.004**	0.08	0***
all informative loci	0.09	0.003**	0	0.112	0.1	0***

* $P<0.05$
** $P<0.01$
*** $P<0.001$

The loci seen as species-uninformative in this study were a small proportion (22.7%) of all loci considered. The AMOVA indicates significant differentiation between the two species (Table 7.6). Excluding the five species-uninformative loci makes the overall Fsc value increases slightly (Table 7.6). Therefore the following pairwise distance analysis was based on the species informative loci only, similarly to Hedges *et al*. (1996) and Kumar and Hedges (1998). However, the inclusion of species-uninformative loci (not shown) does not change the basic picture in the pairwise distance analysis. It was considered that there is no general tendency for locus-specific effects to artificially mask

real population structure, whether all loci or only some are considered (Baer 1999), as long as a moderate number of loci are tested (Goldstein *et al.,* 1999).

7.3.5 Population structure analysis by pairwise comparison

Conventional pairwise Fst can be used as a measure of genetic isolation by distance (Raymond and Rousset, 1995a; Weir, 1996). The present research used the pairwise Fst, obtained by permutation procedure with Arlequin, as a basic data for population structure analysis. The matrix of pairwise Fst is shown in Table 7.7. A matrix of *P* values, including a correction for multiple comparisons, is shown in Table 7.8. Note that the between species pairwise Fst are more related to the Fsc values in Table 7.6, whereas the within species pairwise Fst here are more related to the Fst values in Table 7.6.

Table 7.7 Pairwise Fst data generated by averaging values across loci in Arlequin. Negative Fst values are computational artifacts and the true value is close to zero.

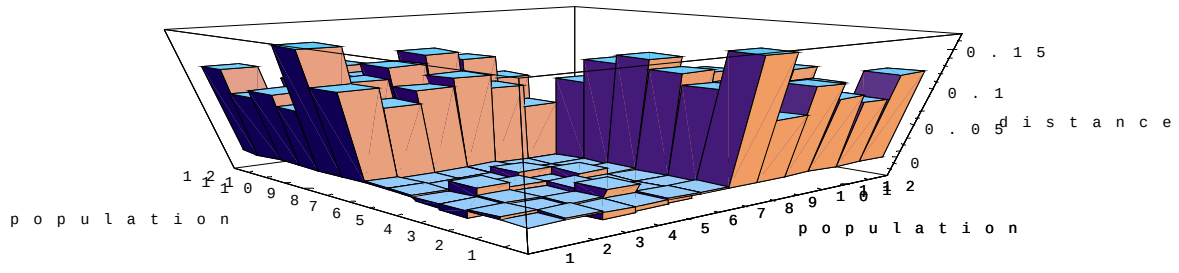
		B. tryoni						*B. neohumeralis*					
		Sunnybank	Nth Ipswitc	Bundaberg	Mackay	Murray	Freshwater	Sunnybank	Nth Ipswitc	Bundaberg	Mackay	Murray	Freshwater
B. tryoni	Sunnybank												
	Nth Ipswitch	0.03											
	Bundaberg	-0.09	0.02										
	Mackay	-0.06	0.13	0.06									
	Murray	0.01	0.00	-0.05	0.09								
	Freshwater	0.02	-0.21	-0.12	0.02	-0.07							
B. neohumeralis	Sunnybank	1.24	1.01	1.23	1.31	1.10	0.73						
	Nth Ipswitch	1.69	1.23	1.46	1.53	1.41	1.01	0.12					
	Bundaberg	0.81	0.42	0.66	0.71	0.61	0.39	0.07	0.26				
	Mackay	1.05	0.86	1.00	1.05	0.85	0.57	0.00	0.23	-0.02			
	Murray	0.89	0.69	0.81	1.07	0.76	0.49	0.01	0.20	0.07	0.01		
	Freshwater	1.21	0.82	1.02	1.16	1.00	0.71	-0.01	0.15	0.12	-0.09	-0.02	

Table 7.8 *P* values of pairwise differences

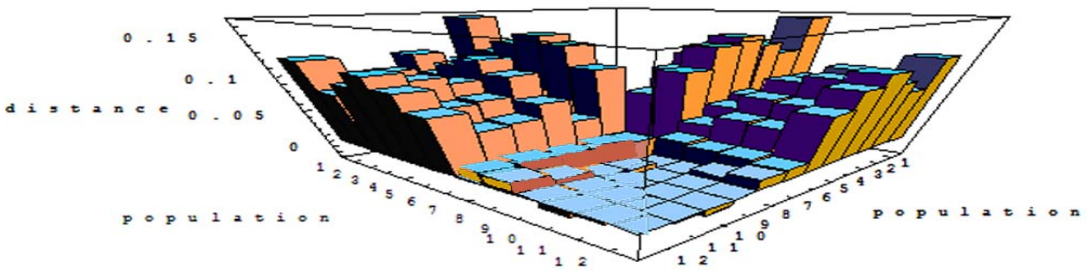
for conventional pairwise Fst from Table 7.7

		B. tryoni						*B. neohumeralis*					
		Sunnybank	Nth Ipswich	Bundaberg	Mackay	Murray	Freshwater	Sunnybank	Nth Ipswich	Bundaberg	Mackay	Murray	Freshwater
B. tryoni	Sunnybank												
	Nth Ipswich	0.309											
	Bundaberg	0.900	0.345										
	Mackay	0.773	0.073	0.218									
	Murray	0.400	0.509	0.682	0.127								
	Freshwater	0.791	1.000	0.973	0.927	0.936							
B. neohumeralis	Sunnybank	0.000	0.000	0.000	0.000	0.000	0.000						
	Nth Ipswich	0.000	0.000	0.000	0.000	0.000	0.018	0.000					
	Bundaberg	0.000	0.000	0.000	0.000	0.000	0.000	0.082	0.191				
	Mackay	0.000	0.000	0.000	0.000	0.000	0.018	0.018	0.600	0.500			
	Murray	0.000	0.000	0.000	0.000	0.000	0.027	0.009	0.173	0.436	0.455		
	Freshwater	0.000	0.000	0.000	0.000	0.000	0.000	0.036	0.082	0.500	0.927	0.645	

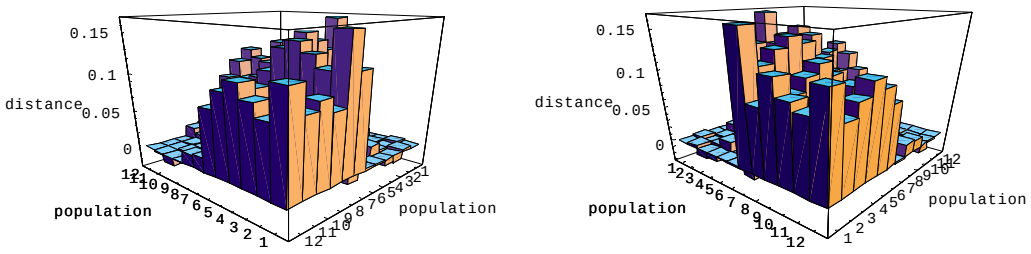
A bar chart established from Table 7.7 shows four three dimensional views of the genetic differentiation within and between species (Figure 7.3) with two lowlands and two plateaus. This illustrates explicitly that the pairwise genetic differences between populations between the two species are greater than those within species. The two lowlands reflect the pairwise distance within species and the two plateaus reflect the pairwise distance between the two species.



a. A lowland showing pairwise distance within *B. tryoni.*



b. A lowland showing pairwise difference within *B. neohumeralis.*



c. A plateau showing pairwise difference between *B.tryoni* and *B.neohumerallis*

	Population number	
	B.tryoni	*B.neohumeralis*
Sunnybank	1	7
Nt Ipswich	2	8
Bundaberg	3	9
Macay	4	10
Muttay	5	11
Freshwater	6	12

Figure 7.3 Four views of three dimensional pairwise differences using pairwise Fst from Table 7.7

7.3.6 Gene flow between the two species

An estimated gene flow (Nm) of 2.068 between the two species over all loci was obtained, derived from average Fst of Weir and Cockerham (1984) using Arlequin. This is much lower than that within species estimates (Table 7.9). High gene flow across loci within *B. tryoni* (31.125) indicates a relatively panmictic Queensland population. For *B. neohumeralis*, lower gene flow of 13.764 across loci indicates either population differentiation, or smaller effective population size. Table 7.10 lists detailed Nm values between populations of the two species derived from Slatkin linearized Fst, as Fst/(1-Fst) (Slatkin, 1995). We see that quite a few Nm values within species are infinite, and the lower limit is 25.19 (Mackay-Murray) for *B. tryoni* and 8.30 (Sunnybank-Nth Ipswich) for *B. neohumeralis*. Those values between species vary from 1.27 (Sunnybank) to 5.35 (Nt Ipswich).

Table 7.9 Gene flow (Nm) across all loci calculated from the Fst of Weir and Cockerham (1984) averaged over populations

	Fst	Nm
B. tryoni	0.004	31.125
B. neohumeralis	0.009	13.764
Between species	0.057	2.068

Table 7.10 Gene flow between populations of the two species obtained from Slatkin linearized Fst. *Inf = infinity, resulting from the zero Fst.

		B. tryoni						*B. neohumeralis*					
		Sunnybank	Nth Ipswitch	Bundaberg	Mackay	Murray	Freshwater	Sunnybank	Nth Ipswitch	Bundaberg	Mackay	Murray	Freshwater
B. tryoni	Sunnybank												
	Nth Ipswitch	79.80											
	Bundaberg	inf*	108.97										
	Mackay	inf	18.01	41.05									
	Murray	154.21	3476.00	inf	25.19								
	Freshwater	118.95	0.50	inf	176.29	inf							
B. neohumeralis	Sunnybank	1.27	1.80	1.59	1.44	1.56	2.01						
	Nth Ipswitch	2.67	5.35	3.50	3.10	3.62	5.25	8.30					
	Bundaberg	1.82	2.32	1.97	1.77	2.09	2.95	18.29	33.17				
	Mackay	2.23	2.78	2.51	2.29	2.79	4.00	9.89	inf	inf			
	Murray	2.48	3.30	2.90	2.10	2.94	4.29	10.83	30.47	187.93	226.01		
	Freshwater	1.78	2.71	2.26	1.90	2.19	2.86	14.43	17.09	inf	inf	inf	

7.3.7 Averaging linkage clustering

An averaging linkage clustering tree constructed by UPGMA based on the pairwise difference data in section 7.3.4 (Table 7.7) and input as a distance matrix (Nei and Kumar, 2000) showed two major clusters distinguishing the two species (Figure 7.4). The six populations can be grouped further within each species: Sunnybank and Nth Ipswich, Bundaberg and Mackay, Murray and Freshwater. Population structure appears highly correlated with geographic proximity, although the deeper branching is different between the two species. For *tryoni*, the most northern pair of populations, Murray and Freshwater, is distant from other populations; for *neohumeralis*, the most southern pair of populations, Sunnybank and Nth Ipswich, is distant from all other populations. The divergence levels are very low between populations within each species (not significant for most loci, Table 7.8), but quite high between species. The assumption, required by the UPGMA tree making method, that the rate of evolution is constant, is applicable given that the two species are very close related. The UPGMA tree is suitable for short evolutionary history where rate heterogeneity has less effect.

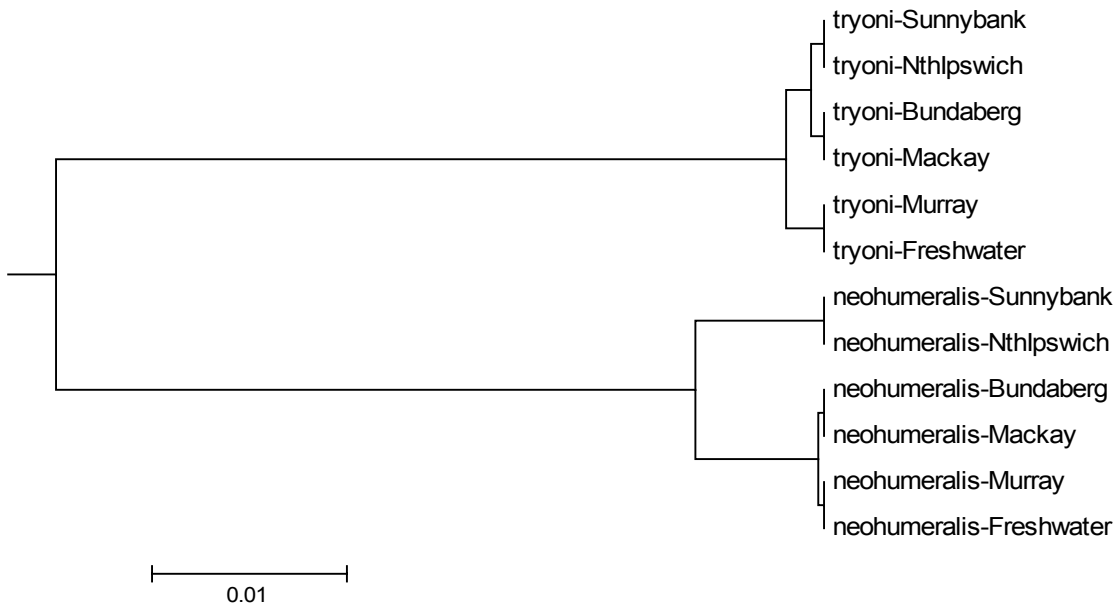


Figure 7.4 UPGMA clustering tree constructed from pairwise genetic differences of populations of *B. tryoni* and *B. neohumeralis* listed in Table 7.7

7.4 Discussion

7.4.1 Differentiation of *B. neohumeralis* and *B. tryoni*.

The results from the AMOVA test and the homogeneity & differentiation test described in the previous section are largely consistent for most of the loci except for loci 1.7a, 4.1a, 5.3b, 5.8a and 6.8a at the species level. The AMOVA test partitioned variances into three different levels of the species-population structure, and took account of the differences between population variances at two levels when computing differences between the two species at the same time. The homogeneity & differential test was run three times separately without taking account of the different structure levels each time. The differences between populations within a particular species could have caused the differences between species to be overestimated in the homogeneity & differentiation test, such as for loci 1.7, 5.3 and 5.8, which showed no significance between the two species in the AMOVA test. When the differences between populations within both species are not significant, differences between the two species could be underestimated as not significant in the homogeneity & differentiation test while they showed significant differences in the AMOVA test, such as for loci 4.1a and 6.8a. That is why the AMOVA hierarchical test is necessary after the polymorphisms are proved basically homogenized within species and differentiated between species.

B. tryoni and *B. neohumeralis* are very close related species, which share a high level of polymorphisms (McKechnie, 1975; Morrow *et al.,* 2000; An *et al.,* 2002). The population analysis using microsatellites detected not only differentiation between the two species, but also subtle substructure within the species, although very weak. As inferred from six populations in a narrow zone of sympatry, there is very weak geographic structure within each species. That the population differentiation is more profound in *B. neohumeralis* than in *B. tryoni* is consistent with higher gene flow within *B. tryoni* than within *B. neohumeralis*.

Wolda (1967a, b) produced data consistent with the hypothesis that the two species of Queensland fruit flies, *B. neohumeralis* and *B. tryoni*, are completely reproductively isolated. Gibbs (1968) investigated the issue based on the hypothesis that hybrids of *B. neohumeralis* and *B. tryoni* are produced in nature, and provided some evidence of gene exchange. Our study supports Gibbs's proposal by measuring gene flow using an indirect method. The gene flow between the two species is comparatively weak, but it overcomes genetic drift, indicated by an Nm greater than one. Gibbs (1968) suggested the possibility that hybridization between the two species was more frequent in the past (about 100 years before) than now.

7.4.2 A sufficient number of markers is required to differentiate the two species

Although differentiation levels are very low between the two species, there are still quite a few loci that show the differentiation. Of the 22 newly isolated microsatellites in the present study, 17 are species-informative in a population structure study of the two species. Four of the six microsatellites (Kinnear *et al.*, 1998) used in a *B. tryoni* population study (Yu *et al.* 2001) were also shown to be species-informative (not shown), but were insufficient to reveal clearly the weakly exhibited hierarchical population structure, as found using the 22 loci in the present study. Goldstein *et al.,* (1999) analyzed 19 microsatellite loci to infer population structure and demographic history in a natural model system of the gray fox *Urocyon cinereoargentatus*. They found that, whatever the complexities of microsatellite mutations and evolution, a moderate number of loci enabled the estimation of subtle aspects of population structure and demographic history. All the 22 loci isolated and analyzed in this study, plus the 6 loci cloned by Kinnear *et al.* (1998) will provide effective genetic markers in further studies of populations of Queensland fruit flies. Seven of the 28 markers are species-uninformative in the present study, but they may have biological meaning in other studies.

Different genomic regions may convey different answers to evolutionary questions. However, if enough markers are used, they may reveal on average an unbiased

evolutionary history. Jorde *et al.* (1997) demonstrated that a larger number of tetranucleotide loci (a minimum of 36) were needed to detect a significant increase in the level of diversity in African humans compared to Asians and Caucasians.

A sufficient number of markers can overcome the disadvantage of small sample size. Yu *et al.* (2001) tested 6 microsatellite markers on a large sample of *B. tryoni* and concluded that a panmictic population occurs in Queensland. The result of the present study, using more microsatellite markers, is consistent with Yu *et al.* (2001), although the sample sizes are smaller. The important point is that a sufficient number of markers compensated for the smaller sample size.

7.4.3 Choice of appropriate marker type is necessary to explain different evolutionary events

Another species of fruit fly *B. aquilonis* is morphologically very similar to *B. tryoni*, and has a narrower and different endemic wild host range. The distribution of *B. aquilonis* is restricted to the northern part of the Northern Territory and Western Australia and, for many years, did not overlap with *B. tryoni*. *B. aquilonis* mates at dusk and is able to mate freely with *B. tryoni* to produce fertile offspring in the laboratory (Drew and Lambert, 1986). It has been suggested that these two "species" could be two geographically separated populations of one species (Wang *et al.* 2003), perhaps commencing allopatric divergence. Normally allopatric speciation is quicker than sympatric speciation due to a physical barrier which interrupts gene flow. Gene flow slows down speciation. Two nuclear markers (*white* and ITS2) and two mitochondrial genes (Cytb and COII) were tested in the three *Bactrocera* complex species (Morrow *et al.*, 2000). For the two nuclear loci, no fixed species-specific differences between *B. tryoni* and *B. aquilonis* were found, and three minor differences were found between *B. tryoni* and *B. neohumeralis* for one of them (ITS2) (Morrow *et al.*, 2000). Since it is non-coding, ITS2 is a rapidly evolving marker, and has proven useful for comparing closely related insect species, subspecies, and populations (Porter and Collins, 1991; Collins and Paskewitz, 1996; Fenton *et al.*, 1998; Silva *et al.*, 1999; Zhu and Greenstone, 1999).The microsatellite results for *B.*

tryoni and *B. neohumeralis* from the present study support the results described by Morrow *et al.* (2000).

Comparisons among these three species using mainly the same microsatellites used in the present study (Wang *et al.* 2003) showed that the genetic distance was lower between *B. tryoni* and *B. aquilonis* than between *B. tryoni* and *B. neohumeralis*. For the two mitochondrial loci, COII distinguishes *B. aquilonis* from *B. tryoni* and *B. neohumeralis*, but does not differentiate *B. tryoni* from *B. neohumeralis*. Thus for hypervariable nuclear markers *B. neohumeralis* appears more evolutionarily distant from *B. tryoni* and *B. aquilonis*, while for mtDNA markers *B. aquilonis* appears more distant from *B. tryoni* and *B. neohumeralis*. A further investigation of mtDNA and nuclear DNA could be critical to clarify the cause of the different expressions with different polymorphisms: whether it is due to a deeper divergence between *B. tryoni* and *B. aquilonis* (which might be masked by hypervariable marker ITS2 and microsatellites) than between *B. tryoni* and *B. neohumeralis* (which can be detected by hypervariable marker ITS2 and microsatellites).

7.4.4 Properties of microsatellite markers related to their utility

7.4.4.1 Selection and microsatellites

Several studies of microsatellite variation in natural populations of *Drosophia* show that non-coding microsatellites can be influenced by the selection acting on linked sites (Irvin *et al.,* 1998; Schlötterer *et al.,* 1997). Furthermore, some microsatellites, especially trinucleotide repeat motifs, have coding functions, and others could act as regulatory elements (Meloni *et al.*, 1998). These microsatellites would be directly under selection instead of hitchhiking. In our study, locus 6.8A is completely fixed for one allele in the Nth Ipswich population of *B. neohumeralis*, while it remains variable in other populations. Similarly, locus bt14 isolated by Kinnear *et al.* (1998) is completely fixed for one allele in the Edge Hill population of *B. neohumeralis* and remains variable in other populations (unpublished data). Both of these microsatellites are trinucleotide repeats. Several reports have shown that expanded alleles at disease-associated or non-disease-associated

trinucleotide repeats have intergenerational instability (Deka and Chakraborty, 1999; Nakamoto *et al.*, 1997; Ikeuchi *et al.*, 1998 Breschel *et al.*, 1997), and such trinucleotide repeats are also most likely under selection. The phenomenon of locus-specific selection requires further investigation of the regions flanking the microsatellites to determine whether adaptive fixation (Begun and Aquadro 1991) or background selection (Charlesworth *et al.,* 1993) is responsible for the phenomenon (Irvin *et al.,* 1998). Of course the *Bactrocera* microsatellite fixation discussed here could be caused by chance from founder effect and/or small sample size.

7.4.4.2 Null alleles and othere factors causing HWE deviations

Deviations from HWE could be caused by null alleles or selective force. Ross *et al.* (1999) used multiple classes of molecular markers to assess genetic structure in the fire ant *Solenopsis invicta* and found that, except for allozymes and codominant RAPDs, the results of the exact test of microsatellite loci showed about twice as many significant departures from HWE as were expected by chance. They suggested a few sources of departure from HWE for microsatellites: (a). high allelic richness caused sensitivity to subtle deviations from HWE; (b). null alleles generated by mutations in the primer site (Pemberton *et al.,* 1995), expressed by homozygosity excess for all the populations; (c). particular susceptibility to scoring errors; (d). artifact bands, especially for those alleles differing greatly in length (Wattier *et al.*, 1998); and (e). linkage to loci which are under strong selection (Slatkin, 1995, Schlötterer *et al.,* 1997).

Direct sequencing also reveals that microsatellite alleles may be identical in length but not identical in sequence. Direct sequencing may be useful for collecting more information from a single locus. Substantially different mutation rates between different alleles at a single locus may be found by direct sequencing (Freimer and Slatkin, 1996).

7.4.4.3.Correlation between microsatellite attributes

Microsatellite replication slippage involves not only single repeats, but also multiple and partial repeats (Palsbøll *et al.*, 1999). The repeat type can be perfect or imperfect. Most of the microsatellite studied here do not depart from HWE and are species-informative, no matter what the repeat unit sizes are and what type of the repeat. To obtain a clear view of all these attributes of microsatellites, a summary of all microsatellite loci under study here, plus the six loci cloned by Kinnear *et al.* (1998), is provided in Table 7.11.

Table 7.11 A summary of microsatellite loci attributes, including 22 microsatellites isolated and used in the present analysis and 6 previously isolated by Kinnear *et al.* (1998)

Locus	base no. per repeat unit	mutation style	repeat type	HWE significant departure*	Species-informative
1.1A	2	single repeat	imperfect		+
1.7A	2	partial repeat	imperfect		-
2.6A	2	single repeat	perfect		+
2.6B	3	single repeat	perfect	*	+
2.9A	2	partial repeat	perfect	*	+
3.2B	2, 3	partial repeat	imperfect		+
4.1A	2	partial repeat	perfect		+
4.3A	3	single repeat	imperfect		+
5.10A	2	single/multiple	perfect		+
5.3B	2	single repeat	perfect		-
5.8A	2	single repeat	perfect		-
6.10B	2	single repeat	imperfect		+
6.12A	2	partial repeat	perfect		+
6.6B	2	single repeat	perfect		+
6.8A	3	multiple repeat	perfect		+
6.8B	3	single repeat	imperfect		+
7.2B	2	single repeat	imperfect		-
7.9A	2	single repeat	perfect		+
8.12A	2	partial repeat	perfect		+
8.5A	3	single repeat	perfect		-
8.6A	3	partial repeat	perfect		+
9.1A	2	single repeat	perfect		+
bt32	2	single repeat	perfect		-
bt10	2	partial repeat	imperfect		+
bt11	2	partial repeat	imperfect		+
bt14	3	partial repeat	perfect		+
bt17	2	single repeat	imperfect		-
bt15	2	partial repeat	perfect		+

* more than two out of the 12 populations depart from HWE.

There seems to be no correlation between the four categories: mutation styles, repeat types, species-informativity, and compliance with expectation of HWE. Looking at the categories including repeat types, single-slippage mutation styles, and HWE conformities (Table 7.11), the species-informative loci account for about 70% or more for each category (Table 7.12), including all the microsatellite markers tested on the two species of the Queensland fruit fly so far.

Table 7.12 Number and percentage of species-informative loci derived form all microsatellites listed in Table 7.11

	Total microsatellits	Species-informative	
Perfect array type	18	14	75%
Imperfect array type	10	7	75%
Single or multi mutation	17	11	69%
Partial mutation	11	10	83%
HWE conformed	26	19	73%
HWE departure	2	2	100%

Further study may be required into the implications and influences of the proportions of various mutation styles, repeat types, species-informativity, and HWE conformities. What is the correlation between them, and how can we get more information from them? Resolution of these questions may require an extensive review of this research area.

Chapter 8

General discussion and conclusion

The aim of the present study was to develop the use of microsatellite and mitochondrial DNA polymorphisms as tools to investigate genetic relationships at levels from pedigrees to populations to species. Two taxa widely separated phylogenetically –the green monkey (Primates) and the Queensland fruit fly (Insecta: Diptera) were used as model organisms to generate the data.

Insects generally have large populations and short life spans whereas primate species have small number of family members who do not reach sexual maturity for several years with a subsequent relatively long generation span. This may have large or small effects on the rate of genetically change depending on the organism. For example, loci in larger populations are more diversify and may evolve faster than those in small populations (Amos and Rubinsztein, 1996; Amos *et al*. 1996; Amos 1999). The much shorter generation times of flies should make their evolution even faster.

Two categories of polymorphisms, nuclear and mitochondrial, may be distinguished. The mutation rate of polymorphisms is distinctly different in each. The mitochondrial genome is maternally inherited. Because of its higher mutation rate, it evolves more rapidly. For nuclear microsatellites, the degree of conservation and polymorphism varies between taxa, and may respond quickly to selection pressure. Bazin *et al*. (2006) reported that there is an expectation that genetic variation increases as population size increases, but not for mtDNA diversity. Their argument trigged a debate on whether the positive relationship is sustained on mtDNA or not (Berry 2006; Mulligan *et al*. 2006). However, these arguments certainly do not affect the distinguishable feature of the two categories of polymorphisms.

8.1 Summary of main findings

A complete 16,550 bp mtDNA sequence of the green monkey *Chlorocebus sabaeus* is reported and annotated for the fist time in the present study (Chapter 2). The primate phylogenetic and evolutionary study using complete mtDNA sequences was implemented. Assessment of validity of phylogenetic trees recovered using single gene and different mutation types were investigated (Chapter 3). Transversion substitutions were considered as more informative for phylogenetic studies. Fifty five SNPs were found and 10 haplogroups were established. The SNPs were helpful in explanation of 12S rRNA differences between the same species *C. sabaeus* from different geographical origins (Chapter 4). Microsatellites used for the green monkey paternity and pedigree studies were developed by cross-amplification using human primers (Chapter 5). Sixteen microsatellites cross-amplified from human primers were used for paternity assignment, 78 progenies were assigned paternally; and a pedigree involved 75 animals was constructed. 38 microsatellites were isolated from 31 sequenced inserts of a partial genomic library of Queensland fruit fly *Bactrocera tryoni* (Chapter 6). Of these microsatellite markers 22 were used for Queensland fruit fly population structure and species discrimination studies (Chapter 7). Two microsatellites showed significant deviation from HWE. Most loci displayed similar allele frequency distribution.. Genetic differences proved to be significant between species of *Bactrocera tryoni* and *B. neohumeralis*, although weak gene flow was detected between the two species.

8.2 Molecular markers, evolutionary depth, and selection

Typically, different mutation rates and evolution depth (comparative branch length in a phylogenetic tree) can explain disparities in population distribution and evolutionary history between types of polymorphisms, including sex-associated markers (Fig 1.6). In the case of overlapping depths expressed for different polymorphisms, the geographic patterns derived from autosomal, mtDNA, and Y-chromosome markers showed remarkably similarities in human variation (Lell and Wallace, 2000; Seielstad, 2000; Cann, 2001). Allendorf and Seeb (2000) compared gene flow estimates among

microsatellites, mtDNA, allozymes, and RAPD markers and concluded that these Fst-type estimates show little difference as long as they are corrected for differing numbers of alleles and heterozygosity. It is believed that for a study of a shallow evolutionary history of a set of geographically proximate populations microsatellites are most appropriate for the application (Takezaki and Nei, 1996; Angers and Bernatchez, 1998). This means that microsatellites are ideal markers for intraspecies analyses, including paternity assessment and pedigree reconstruction within a small pedigree, as described in chapter 5 of this research, and for population analysis of intraspecific variation or for distinguishing sibling species, as described in chapter 7. However, for a deep range evolution history, differentiation is often underestimated when using hypervariable markers like microsatellites, due to size or sequence homoplasy. Queney *et al.* (2001) described unchanged allele distributions of microsatellite loci between well-defined divergent populations of the European rabbit (*Oryctolagus cuniculus*) due to allele-size homoplasy of the markers. Actually some of the size homoplasy could involve undetected differences in sequences. In a long range phylogenetic study, as described in chapter 3 and 4, mtDNA can provide satisfactory results; in particular when the complete sequence or more informative gene sequences are used. In some situations utilizing mtDNA application, however, the lineage sorting and single inherited unit effect should be considered (see Chapter 3 discussion and Chapter 1 section 1.1.1).

There are a number of criteria that can be used in choosing markers for population and species studies. First, one would choose a special functional gene that is subject to selection (Noor, 2003). Actually this is not easy, as most evolutionary processes involve multiple loci. However, Krieger and Ross (2002) identified a single gene that encoded a pheromone binding protein in fire ants *Solenopsis invicta*. Using starch gel electrophoreses of proteins they showed this protein had amino acid variation in different populations. The protein has major effects on the regulation of complex social behavior and played a diagnostic role in population analysis.

Second, one may try to assess evolutionary depth. Between flies *Drosophila simulans* and *D. yakuba*, with about 10 generation per year, an adaptive amino acid replacement takes

about 45 years (Smith and Walker, 2002). A sequence comparison of the important circadian gene *per*, which was considered as at least one of the functional genes in the mating timing regulation between the fruit flies *Bactrocera tryoni* and *B. neohumeralis*, showed that there was no amino acid sequence divergence between the two species (An *et al.*, 2002), although it was supposed that their divergence has been occurring for at least 100 years (Lewontin and Birch, 1966). This implies not only strong, positive, stabilizing selection (Fay *et al.*, 2002), but may also indicate very shallow evolutionary depth in the two *Bactrocera* flies, compared with that in the two *Drosophila* flies. Sympatric divergence may be a long-term process that eventually overcomes the existence of gene flow. It will be appropriate to use highly variable markers such as microsatellites, as used in this study, to distinguish *B. tryoni* and *B. neohumeralis*.

Third, in some special circumstances sex associated markers are needed, such as mitochondrial or Y chromosome loci. These special circumstances include biased sex ratios, populations with different migration rates between sexes, polyandry, polygyny, and patrilocality or matrilocality populations. Uni-parental markers can also be used to detect differences in male/female gene flow in short histories (Merriwether *et al.*, 1997; Green *et al.*, 2000; Mesa *et al.*, 2000; Carvajal-Carmona *et al.*, 2000).

In addition to effective evolution depth, active selection can affect different markers or genomes, in both insect and primate. Diversifying selection acts to increase variation and consequently maintains polymorphisms in the nuclear genome, and directional selection acts to reduce variation in the mitochondrial genome (Nachman *et al.*, 1996; Wise *et al.*, 1997). Mitochondrial markers could be a good choice for detecting differences in geographic population structure between allopatric *B. tryoni* and *B. aquilonis* under diversifying selection; Hypervariable nuclear polymorphisms might be better for analysis of sympatric species *B. tryoni* and *B. neohumeralis*, which reflect both diversifying and directional selection. Morrow *et al.* (2000) used two nuclear sequences (*white* gene and ITS2) and two mitochondrial genes (Cytb and COII) to find differences between the three *Bactrocera* species. One of the two mitochondrial genes, COII, showed differentiation

between *B. aquilonis* and the other two species, but not between *B. tryoni* and *B. neohumeralis*, which were distinguished with microsatellites showed in the present study.

McKechnie (1975) used three polymorphic enzymes to discriminate between the two species of Queensland fruit fly, sampled from four populations of *B. tryoni* and two populations of *B. neohumeralis*. He found allele $Adh^{1.20}$ was abundant in *B. neohumeralis* but not present in *B. tryoni*, whereas allele $Odh^{0.88}$ could not be detected in *B. neohumeralis* but was present in a few *B. tryoni*. A similar scenario occurred in the present study. Some alleles of microsatellites were detected in one of the Queensland fruit flies, *B. tryoni* or *B. neohumeralis*, but not in the other at the same locus. Microsatellites and enzymes are both nuclear markers, and could have similar manifestation in intra- or interspecies analysis. Another example of the similar demonstration for microsatellites and enzymes is that Estoup *et al.* (1998) found similar genetic distances for microsatellites and allozymes in brown trout *Salmo trutta*.

In many studies (Reeb and Avise, 1990; Buroker, 1983; McDonald *et al.,* 1996; Friesen *et al.,* 1996, 1997; Pitocchelli *et al.,* 1995), allozymes, microsatellites, multilocus fingerprinting, and other nuclear markers normally did not produce contradictive conclusions (although hypervariable markers like microsatellites or minisatellites are more sensitive to instant environment changes and evolutionary processes); but different result can be obtained when mtDNA is used. Environmental contamination, such as air pollution, is a potent source of mutation for both human and wildlife populations. When exposed to environmental contamination (such as Chernobyl accident and polynuclear aromatic hydrocarbons in Hamilton Harbour), both humans and wildlife showed about two-fold increase in mutation rate determined by microsatellites and minisatellites; but mtDNA exposed to environmental contamination was less affected, even in hypervariable control region (Chen and Hebert, 1999b; Dubrova *et al.*, 1996, 1997; Marvin *et al.*, 1993; Poulton, 1987; Yauk and Quinn, 1996). A notable European population structure in *Drosophila subobscura* is detected by mtDNA but not by allozymes or microsatellites (Prevosti *et al.,* 1988; Latorre *et al.,* 1992; Pascual *et al.,* 2001). The incongruent results

between nuclear and mitochondrial DNA could be overcome by using appropriate polymorphic markers according to the evolutional depth and the selection type.

8.3 Comparison of genomic conservation

The degree of genomic conservation is related to molecular clocks that vary across taxa. The order Primates arose in the Cenozoic about 65 MYBP. Recent studies have revealed paucity of variation in human population evolution showing that the human species is young and genetically quite uniform (Lewis, 2002). The human population is small and grew fast, from ~10,000 founders in Africa ~3,000 generations ago. Most of the variation in humans was present when they migrated out of Africa and modern humans are all closely related (Lewis, 2002). In contrast, the class Insecta is a relatively old Mesozoic radiation going back about 205 MYBP. The pace of the clock may be faster in insects than in primates. The former have small body size, large population, and short generation spans. So their divergence is faster and deeper, and the degree of conservation is lower even in recent lineages.

Wu *et al.* (2000) proposed that transversions are more conserved than transitions. Actually this is not universal although the results of the present study support it. In some insects, such as parasitic Hymenoptera, conversions between A and T (AT transversions) occur much more frequently in the mitochondrial genome than any other type of change (Dowton and Austin, 1997). This could be due to the much higher AT content in mtDNA in these wasps (Dowton and Austin, 1995).

For the order Diptera (flies, etc.), microsatellite loci cannot be cross-amplified between different families within the fly group Acalyptratae, such as between Drosophilidae (*Drosophila*) and Tephritidae (*Bactrocera*). It is also very hard to cross-amplify between some species even within the same genus, e.g. between *D. melanogaster* and *D. hibiscus* (unpublished results). Nevertheless, cross-amplification can be successful among some very closely related species within a genus. Song *et al*. (2006) tried to use 18 pairs of primers designed for *B. tryoni* and *B. dorsalis* to genotype *B. cacuminata*. Nine produced

amplicons but with significant low- or non-polymorphisms. Microsatelllites isolated from *B. tryoni* can be typed to *B. neohumeralis* and *B. aquilonis* with high success, and to several other *Bactrocera* species with lower success rates, e.g. *B. papayae, B. cucurbitae, B. opiliae* and *B. breviculeus* (Wang *et al.*, 2003; Shearman, unpublished).

Among Anthropoidea, microsatellites can be easily cross-amplified within the superfamily Hominoidea, between all the four nominal families (Hominidae, Panidae, Pongidae, and Hylobatidae) (Rubinsztein *et al.*, 1995a, b; Kanthaswamy *et al.*, 1998, Bradley *et al.*, 2000; Constable *et al.*, 2001). They can even be cross-amplified between superfamilies: human of Hominoidea and old world monkeys of Cercopithecoidea (including the baboon, langurs, macaque and green monkey) (Rogers *et al.,* 1995; Perelygin *et al.,* 1996; Hadfield *et al.*, 2001; Launhardt *et al.*, 2001) Microsatellite flanking regions are conserved within the infraorder Catarrhini (old world monkeys). There are a few new world monkey species, such as *Alouatta palliata* and *Saimiri sp.*, in which microsatellites were successfully cross-amplified between humans and new world monkeys and within new world monkeys (infraorder Platyrrhini), (Takenaka *et al.*, 1993; Ellsworth and Hoelzer 1998; Witte and Rogers 1999).

The presence of highly conserved flanking regions has been reported for some microsatellite loci in turtles (FitzSimmons *et al.,* 1995) and fishes (Rico *et al.,* 1996), allowing cross-amplification for species that diverged as long as 470 MYBP (Zane *et al.*, 2002).

8.4 Degree of polymorphism varies across taxa

For Queensland fruit flies the average number of alleles per locus for the 22 microsatellites is about 5, screened from six populations of each of the two species covering a large area of Queensland. As to the green monkey in a captive population founded with 57 animals, for the 17 loci the average number of alleles per locus is 5 and could be higher in the wild. The average number of alleles per locus for the 17 microsatellites in humans is about 8, analyzed from a CEPH reference family with

limited individuals. It could be higher in a large population. Although the organisms used in the present study are too few to draw a conclusion, the comparison of the two taxa is consistant with Katti *et al.* (2001): insects generally have lower number of repeats (Fig 1.3), and consequently lower allele numbers and polymorphisms.

Clisson *et al*. (2000) confirmed the existence of a repeat in non-human primates whose motif was identical to that of humans, although base substitutions within some motifs occur at high frequency. Base substitutions within repeats reduce repeat length and interfere with the polymorphisms, resulting in lower allele numbers. Hutter *et al.* (1998) isolated 55 microsatellites from *D. simulans* including 25 perfect repeats, 24 imperfect repeats and 6 compound repeats. The low rate of perfect repeats could be explained by a high frequency of base substitutions within the repeats. For the 22 microsatellites isolated from *B. tryoni* in the present study, two thirds are perfect repeats, a much higher proportion than found in *D. simulans* by Hutter *et al.* (1998). The one third of the imperfect repeats could still play a major role in interfering with repeat numbers and repeat abundance.

In primates in the regions flanking repeat stretches, indel events involving multiple nucleotides up to five bases often occur (Clisson, 2000); this is also true in some flies (Hutter *et al.*, 1998). These provide useful information for null alleles, size and sequence homoplasy. Size-defined alleles, as seen in microsatellites, may have multiple molecular structures (Clisson, 2000). Thus size homoplasy leads to an underestimate of genetic variability, and caution should be used when employing microsatellites in comparisons between species which have separated by considerable evolutionary time. For example, the relationships between populations within a species may be resolved with microsatellite polymorphisms, but relations among species or genera will probably be obscured when microsatellites are used. The degree of size homoplasy and divergence at a microsatellite locus between two species increases with time, although the pace of the increase declines (Clisson, 2000).

8.5 Future work

mtDNA markers are widely used for genetic, biological, and medical studies. The use of the complete mtDNA sequence or all informative sequences can minimize biases caused by mutation rate heterogeneity within regions of the whole genome when single or a few genes are used. The mtDNA sequence and SNPs of the green monkey reported in the present study will contribute to further studies on the taxonomic status of the green monkey, while the whole genus is in urgent need of revision (Groves, 2001). In addition species/subspecies verification and phylogenetic analyses will benefit from mtDNA sequence and SNP identification, along with the development of independent nuclear markers. The complete mtDNA sequence will help to stimulate the worldwide population analysis and evolutionary historical study of this Old World monkey.

The pedigrees of the green monkeys in the VMRC UCLA constructed in the present research, although not covering all the animals in the colony, will still be very helpful for further sibling linkage studies and genetic mapping, which is one of the basic goals of the VMRC. However, a larger pedigree and parental analysis can be developed when more generations and information are accumulated.

The microsatellites isolated from *Bactrocera tryoni* in the present project have been used in other closely related species, as well as in genetic mapping of this species (Wang *et al*., 2003; Zhao *et al.,* 2003). For further study in *Bactrocera*, mtDNA or other SNP markers may be used and compared with the results derived from microsatellites. The comparison will reveal a much clearer picture on population, speciation and related natural selection of the flies.

For development of molecular polymorphisms, such as SNPs and microsatellites, a growing list of improved approaches and instrumentation have been created and improved. While dHPLC is still attractive for large-scale screening due to its cost-effectiveness (Wuytsa *et al*., 2005; Yeh *et al*., 2005), direct sequencing methods have been largely used in SNP detection recently (Vignal *et al*., 2002; Stephens *et al*., 2006), and will replace dHPLC eventually. Microsatellite isolation from a conventional partial

genomic library is time consuming and tedious, and has been increasingly replaced by new approaches such as streptavidin-coated paramagnetic beads for 5' biotinylated probes (Kandpal *et al*. 1994; Kijas *et al*. 1994), and the cost are expected to decrease greatly following recent development. When genomic sequences of many species increasingly become available, microsatellite isolation will be easily performed by comprehensive computer programs (Tanaka *et al*., 2005). Similarly, SNP detection will be facilitated by availability of sequences in a related species or populations. In addition, Microarrays provide most efficient method for SNP genotyping. Molecular polymorphisms will be even more widely used in a range of taxa on aspects including evolutionary and ecological research, gene mapping, medical studies, and monitoring of biological pest control.

Bibliography

Abbott, C. L., M. C. Double, J. W. H. Trueman, A. Robinson and A. Cockburn 2005 An unusual source of apparent mitochondrial heteroplasmy: duplicate mitochondrial control regions in *Thalassarche* Albatrosses. Molecular Ecology 14: 3605-3613

Adachi, J. and M. Hasegawa 1995 Improved dating of the human/chimpanzee separation in the mitochondrial DNA tree: heterogeneity among amino acid sites. Journal of Molecular Evolution 40: 622-8

Ainsworth, C. 2005 Parasite infiltrates fruitfly research. Nature 436: 8

Akimoto, M., M. Niikura, M. Ichikawa, H. Yonekawa, K. Nakada, Y. Honma, J-I Hayashi 2005 Nuclear DNA but not mtDNA controls tumor phenotypes in mouse cells. Biochemical and Biophysical Research Communications 327: 1028-1035

Alderborn, A., A. Kristofferson and U. Hammerling 2000 Determination of single-nucleotide polymorphisms by real-time pyrophosphate DNA sequencing. Genome Research 10: 1249-1258

Allendorf, F. and L. Seeb 2000 Concordance of genetic divergence among sockeye salmon populations at allozyme, nuclear DNA, and mitochondrial DNA markers. Evolution 54: 640-651

Amos, B. 2000 NEWPATXL, a general paternity program. XL macro version. www.zoo.cam.ac.uk/ZOOSTAFF/amos/newpat.htm Dept. of Zoology, Cambridge, CB2 3EJ, UK

Amos, W. 1999 A comparative approach to the study of microsatellite evolution. p. 66-79 in Microsatellites: Evolution and Applications, Ed: D. B. Goldstein and C. Schlötterer. Oxford University Press, Oxford

Amos, W. and D. C. Rubinsztein 1996 Microsatellites are subject to directional evolution. Nature Genetics 12: 13-14

Amos, W., S. J. Sawcer, R. Feakes and D. C. Rubinsztein 1996 Microsatellites show mutational bias and heterozygote instability. Nature Genetics 13: 390–391.

An, X., K. Wilkes, Y. Bastian, J. L. Morrow, M. Frommer and K. A. Raphael 2002 The *period* gene in two species of tephritid fruit fly differentiated by mating behaviour. Insect Molecular Biology 11: 419-430

Anderson, S., A. T. Bankier, B. G. Barrell, M. H. L. de Bruijn, A. R. Coulson, J. Drouin, I. C. Eperon, D. P. Nierlich, B. A. Roe, F. Sanger, P. H. Schreier, A. J. H. Smith, R. Staden and I. G. Young 1981 Sequence and organization of the human mitochondrial genome. Nature 290: 457-465

Andrews, T. D, L. S. Jermiin and S. Easteal 1998 Accelerated evolution of cytochrome b in simian primates: adaptive evolution in concert with other mitochondrial proteins? Journal of Molecular Evolution 47: 249-257

Angers, B. and L. Bernatchez 1998 Combined use of SMM and non-SMM methods to infer fine structure and evolutionary history of closely related brook charr populations from microsatellites. Molecular Biology and Evolution 15: 143-159

Armour, J. A., R. Neumann, S. Gobert and A. J. Jeffreys 1994 Isolation of human simple repeat loci by hybridization selection. Human Molecular Genetics 3: 599-565

Arnason, U. and A. Gullberg 1993 Comparison between the complete mtDNA sequences of the blue and the fin whale, two species that can hybridize in nature. Journal of Molecular Evolution 37: 312-322

Arnason, U. and E. Johnsson 1992 The complete mitochondrial DNA sequence of the harbor seal, *Phoca vitulina*. Journal of Molecular Evolution 34: 493-505

Arnason,U., A. Gullberg, E. Johnsson and C. Ledje 1993 The nucleotide sequence of the mitochondrial DNA molecule of the grey seal, *Halichoerus grypus*, and a comparison with mitochondrial sequences of other true seals. Journal of Molecular Evolution 37: 323-330

Arnason, U, A. Gullberg, C. Ledje and S. Mouchaty 1995 A molecular view of pinniped relationships with particular emphasis on the true seals. Journal of Molecular Evolution 40: 78-85

Arnason, U., A. Gullberg and B. Widegren 1991 The complete nucleotide sequence of the mitochondrial DNA of the fin whale, *Balaenoptera physalus*. Journal of Molecular Evolution 33: 556-568

Arnason, U., X. Xu and A. Gullberg 1996a Comparison between the complete mitochondrial DNA sequences of Homo and the common chimpanzee based on nonchimeric sequences. Journal of Molecular Evolution 42:145-152

Arnason, U., X. Xu, A. Gullberg and D. Graur 1996b The ‘‘*Phoca* Standard’’: An external molecular reference for calibrating recent evolutionary divergences. Journal of Molecular Evolution 43: 41-45

Arnason, U., A. Gullberg and A. Janke 1998 Molecular timing of primate divergences as estimated by two nonprimate calibration points. Journal of Molecular Evolution 47: 718-727

Arnason, U., J. A. Adegoke, K. Bodin, E. W. Born, Y. B. Esa, A. Gullberg, M. Nilsson, R. V. Short, X. Xu and A. Janke 2002 Mammalian mitogenomic relationships and the

root of the eutherian tree. Proceedings of the National Academy of Sciences 99: 8151-8156

Ashton, E. H. 1960 The influence of geographic isolation on the skull of the green monkey (*Cercopithecus aethiops sabaeus*) Part V. Proceedings of the Royal Society of London B: Biological Sciences 151: 538

Awadalla, P., A. Eyre-Walker and J. M. Smith 1999 Linkage disequilibrium and recombination in hominoid mitochondrial DNA. Science 286: 2524-2525

Bachtrog, D., M. Agis, M. Imhof, and C. Schlötterer 2000 Microsatellite variability differs between dinucleotide repeat motifs—evidence from *Drosophila melanogaster* Molecular Biology and Evolution 17: 1277-1285

Ballard, J. W. O. and M. Kreitman 1994 Unraveling selection in the mitochondrial genome of *Drosophila.* Genetics 138: 757-772

Baer, C. F. 1999 Among-locus variation in Fst: fish, allozymes and the Lewontin–Krakauer test revisited. Genetics 152: 653-659

Barbujani, G., A. Magagni, E. Minch and L. L. Cavalli-Sforza 1997 An apportionment of human DNA diversity. Proceedings of the National Academy of Sciences USA 94: 4516-4519

Barnicot, N. A. and D. Jewett-Emmett 1971 Ted cell and serum proteins of talapoin, patas and vervet monkeys. Folia Primatology 15: 65-76

Barton-Browne, L. 1957 The effect of light on the mating behavior of the Queensland fruit fly, *Strumeta tryoni* (Frogg.) Australian Journal of Zoology 5: 145-158

Bateman, M. A. 1955 The effect of light and temperature on the rhythm of pupal ecdysis in the Queensland fruit-fly, *Dacus* (Strumera) *tryoni* (Frogg.) Australian Journal of Zoology 3: 22-33

Bateman, M. A. 1958 Ecological adaptations in geographic races of the Queensland fruit fly *Dacus* (Strumeta) *tryoni* Frogg. PhD thesis, University of Sydney

Bateman, M. A. 1991 The impact of fruit flies on Australian horticulture. Horticultural Policy Council Report No. 3. Department of Primary Industries, Canberra, Australia

Battersby, B. J. and E. A. Shoubridge 2001 Selection of a mtDNA sequence variant in hepatocytes of heteroplasmic mice is not due to differences in respiratory chain function or efficiency of replication. Human Molecular Genetics 10: 2469-2479

Batzer, M. A. and P. L.Deininger 2002 Alu repeats and human genomic diversity. Nature Reviews Genetics 3: 370-378

Bayona-Bafaluy, M. P., R. Acin-Pérez, J. C. Mullikin, J. S. Park, R. Moreno-Loshuertos, P. Hu, A. P Pérez-Martos, P. Fernández-Silva, Y. Bai and J. A, Enrĩquez 2003 Revisiting the mouse mitochondrial DNA sequence. Nucleic Acids Research 31: 5349-5355

Bazin, E., S. Gle´min and N. Galtier 2006 Population Size Does Not Influence Mitochondrial Genetic Diversity in Animals Science 312: 570-572

Beckmann, J. S. and J. L.Weber 1992 Survey of human and rat microsatellites. Genomics 12: 627-631

Begun, D. J. and C. F. Aquadro 1992 Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. Nature 356: 519-520

Begun, D. J. and C. F. Aquadro 1993 African and North American populations of *Drosophila melanogaster* are very different at the DNA level. Nature 365: 548-550

Bender, W., P. Spierer and D. S. Hogness 1983 Chromosomal walking and jumping to isolate DNA from the *Ace* and *rosy* loci and the *bithorax* complex in *Drosophila melanogaster*. Journal of Molecular Biology 168: 17-33

Berry, N, C. Davis, A. Jenkins, D. Wood, P. Minor, G. Schild, M. Bottiger, H. Holmes and N. Almond 2001 Vaccine safety -Analysis of oral polio vaccine CHAT stocks. Nature 410: 1046-7

Berry, O. F. 2006 Mitochondrial DNA and Population Size. Science 314: 1388-1389

Bertranpetit, J. and F. Calafell 1996 Genetic and geographical variability in cystic fibrosis: evolutionary considerations. p: 97-118 In: Variation in the Human Genome - Ciba Foundation Symposium 197. Ed: Chadwick, D. and G. Cardew. John Wiley & Sons Ltd.

Birch, L. C. and W. G. Vogt 1970 Plasticity of taxonomic characters of the Queensland fruit flies *Dacus tryoni* and *Dacus neohumeralis* (Tephritidae). Evolution 24: 320-343

Blouin, M. S. 2003 DNA-based methods for pedigree reconstruction and kinship analysis in natural populations. Trends in Ecology and Evolution18: 503-511

Boire, D., H. Theoret and M. Ptito 2002 Stereological evaluation of neurons and glia in the monkey dorsal lateral geniculate nucleus following an early cerebral hemispherectomy. Experimental Brain Research 142: 208-220

Booth, A. H. 1956a The Cercopithecidae of the Gold and Ivory Coasts: Geographic and systematic observations. Annals and Magazine of Natural History 9: 476-480

Booth, A. H. 1956b The distribution of primates in the Gold Coast. Journal of the West African Science Association 2: 122-139

Boursot, P., H. Yonekawa and F. Bonhomme 1987 Heteroplasmy in mice with deletion of a large coding region of mitochondrial DNA. Molecular Biology and Evolution 4: 46-55

Bradley, B. J., C. Boesch and L. Vigilant 2000 Identification and redesign of human microsatellite markers for genotyping wild chimpanzee (*Pan troglodytes verus*) and gorilla (*Gorilla gorilla gorilla*) DNA from faeces. Conservation Genetics 1: 289-292

Breschel, T. S., M. G. McInnis, R. L. Margolis, G. Sirugo, B. Corneliussen, S. G. Simpson, F. J. McMahon, D. F. MacKinnon, J. F. Xu, N. Pleasant, Y. Huo, R. G. Ashworth, C. Grundstrom, T. Grundstrom, K. K. Kidd, J. R. DePaulo and C. A. Ross. 1997 A novel, heritable expanding CTG repeat in an intron of the ser2-1 gene on chromosome 18q21.1. Human Molecular Genetics 11: 1855-1863

Brinkmann, B., A. Junge, E. Meyer and P. Wiegand 1998 Population genetic diversity in relation to microsatellite heterogeneity. Human Mutation 11: 135-144

Broughton, R. E., J. E. Milam and B. A. Roe 2001 The Complete sequence of the Zebrafish (*Danio rerio*) mitochondrial genome and evolutionary patterns in vertebrate mitochondrial DNA. Genome Research 11: 1958-1967

Brown, W. J., M. M. George and A. C. Wilson 1979 Rapid evolution of animal mitochondrial DNA. Proceedings of the National Academy of Sciences USA 76: 1967-1971

Brown, W. M., E. M. Prager, A. Wang and A. C. Wilson. 1982 Mitochondrial DNA sequences of primates: the tempo and mode of evolution. Journal of Molecular Evolution 18: 225-239

Brown, M. D., A. S. Voljavec, M, T. Lott, I. MacDonald, D. C. Wallace. 1992 Leber's hereditary optic neuropathy: a model for mitochondrial neurodegenerative diseases. The FASEB Journal 6: 2791-9

Brumfield, R. T., P. Beerli, D. A. Nickerson and S. V. Edwards 2003 The utility of single nucleotide polymorphisms in inferences of population history. Trends in Ecology and Evolution 18: 249-256

Brunet, M., F. Guy, D. Pilbeam, H. T. Mackaye, A. Likius, D Ahounta, A Beauvilain, C. Blondel, H. Bocherensk , J-R. Boisserie, L. De Bonis, Y. Coppens, J. Dejax, C. Denys, P. Duringerq,V. Eisenmann, G. Fanone, P. Fronty, D. Geraads, T. Lehmann, F. Lihoreau, A. Louchart, A. Mahamat, G. Merceron, G. Mouchelin, O. Otero, P. P. Campomanes, M. P. De Leon, J-C. Rage, M. Sapanetkk , M. Schusterq, J. Sudrek , P. Tassy, X. Valentin, P. Vignaud, L. Viriot, A. Zazzo and C. Zollikofer 2002 A new hominid from the Upper Miocene of Chad, Central Africa. Nature 418: 145-151

Brisson, J. A., A. R. Templeton and I. Duncan 2004 Population Genetics of the Developmental Gene *optomotor-blind* (*omb*) in *Drosophila polymorpha*: Evidence for a Role in Abdominal Pigmentation Variation. Genetics 168: 1999-2010

Buroker, H. E. 1983 Population genetics of the American oyster *Crassostrea virginica* along the Atlantic coast and the Gulf of Mexico. Marine Biology 75: 99-112

Caldwell, M., M. S. Gaines and C. R. Hughes 2002 Eight polymorphic microsatellite loci for bottlenose dolphin and other cetacean species. Molecular Ecology Notes 2: 393-395

Cann, R. L. 2001 Genetic Clues to Dispersal in Human Populations: Retracing the Past from the Present. Science 291:1742-1748

Carvajal-Carmona, L. G., I. D. Soto, N. Pineda, D. Ortiz-Barrientos, C. Duque, J. Ospina-Duque, M. McCarthy, P. Montoya, V. M. Alvarez, G. Bedoya and A. Ruiz-Linares 2000 Strong Amerind/White sex bias and a possible sephardic contribution among the founder of a population in Northwest Colombia. American Journal of Human Genetics 67: 1287-1295

Caskey, C. T., A. Pizzuti, Y-H. Fu, R. G. Fenwick and D. L. Nelson 1992 Triple repeat mutations in human disease. Science 256: 784-789

Castro, M. J., P. Morales, B. Fernandex-Soria, B. Suarez, M. J. Recio, M. Alvarez, M. Martiin-Villa and A. Arnaiz-Villena 1996 Allelic diversity at the primate Mhc-G locus: exon 3 bears stop codons in all Cercopithecinae sequences. Immunogenetics 43: 327-336

Chakraborty, R., M. Kimmel, D. N. Stivers, L. J. Davison and R. Deka 1997 Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. Proceedings of the National Academy of Sciences USA 94: 1041-1046

Charlesworth, B., M. T. Morgan and D. Charlesworth 1993 The effect of deleterious mutations on neutral molecular variation. Genetics 134: 1289-1303

Chen, J. Z. and P. D. N. Hebert 1999a Intraindividual sequence diversity and a hierarchical approach to the study of mitochondrial DNA mutations. Mutation Research 434: 205-217

Chen, J. Z. and P. D. N. Hebert 1999b Terminal branch haplotype analysis: a novel approach to investigate newly arisen variants of mitochondrial DNA in natural populations. Mutation Research 434: 219-231

Chinnery, P. F. and D. M. Turnbull 1999 Mitochondrial DNA and disease. Lancet (Suppl. I) 354: 17-21

Chinnery, P. F., D. R. Thorburn, D. C. Samuels, S. L. White, H-H. M. Dahl, D. M. Turnbull, R. N. Lightowlers and N. Howell 2000 The inheritance of mitochondrial DNA heteroplasmy: random drift, selection or both? Trends in Genetics 16: 500-505

Clarke, A. R., K. F. Armstrong, A. E. Carmichael, J. R. Milne, S. Raghu, G. K. Roderick and D. K. Yeates 2005 Invasive phytophagous pests arising through a recent tropical evolutionary raditation: the *Bactrocera Dorsalis* complex of fruit flies. Annual Review of Entomology 50: 293-319

Clisson, I, M. Lathuilliere and B. Crouau-Boy 2000 Conservation and evolution of microsatellite loci in primate taxa. American Journal of Primatology 50: 205-214

Cockerham, C. C. 1969 Variance of gene frequencies. Evolution 23: 72-83

Cockerham, C. C. 1973 Analysis of gene frequencies. Genetics 74: 679-700

Collins, F. H. and S. M. Paskewitz 1996 A review of the use of ribosomal DNA (rDNA) to differentiate among cryptic *Anopheles* species. Insect Molecular Biology 5: 1-9

Collin, T. J., M. J. Berridge, P. Lipp and M. D. Bootman 2002 Mitochondria are morphologically and functionally heterogeneous within cells. The EMBO Journal 21: 1616-1627

Colyer, F. 1949 Variations of the teeth of the green monkey in St. Kitts. Proceedings of the Royal Society of Medicine 41: 845-848

Constable, J. L, M. V. Ashley, J. Goodall and A. E. Pusey 2001 Noninvasive paternity assignment in Gombe chimpanzees. Molecular Ecology 10: 1279-300

Cooper, G., D. C. Rubinsztein and W. Amos 1998 Ascertainment bias cannot entirely account for human microsatellites being longer than their chimpanzee homologues. Human Molecular Genetics 7: 1425-1429

Coote, T. and Bruford M. W. 1996 Human microsatellites applicable for analysis of genetic variation in apes and Old World monkeys. Journal of Heredity 87: 406-410

Crawford, A. M., S. M. Kappes, K. A. Paterson, M. J. deGotari, K. G. Dodds, B. A. Freking and R. T. Stone, C. W. Beattie 1998 Microsatellite evolution: testing the ascertainment bias hypotheses. Journal of Molecular Evolution 46: 256-260
Curtin, P. 1969 The Atlantic slave trade: a census. University of Wisconsin Press, Madison

Dallas, J. F. 1992 Estimation of microsatellite mutation rates in recombinant inbred strains of mouse. Mammalian Genome 3: 452-456

Dandelot, P. 1959 Note sur la classification des Cercopitheques du groupe *aethiops*. Mammalia 23: 357-368

Davis, R. E., S. Miller, C. Herrnstadt, S. S. Ghosh, E. Fahy, L. A. Shinobu, D. Galasko, L. J. Thal, M. F. Beal, N. Howell and W. D. Parker 1997 Mutations in mitochondrial cytochrome *c* oxidase genes segregate with late-onset Alzheimer disease Proceedings of the National Academy of Sciences USA 94: 4526-4531

Davis, C. S. and C. Strobeck 1998 Isolation, variability, and cross-species amplification of polymorphic microsatellite loci in the family Mustilidae. Molecular Ecology Notes 7: 1776-1778

Deka, R, L. Jin, M. D. Shriver, L. M. Yu, S. DeCroo, J. Hundrieser, C. H. Bunker, R. E. Ferrell and R. Chakraborty 1995 Population genetics of dinucleotide $(dC\text{-}dA)_n\cdot(dG\text{-}dT)_n$ polymorphisms in world populations. American Journal of Human Genetics 56: 461-374

Deka, R. and R. Chakraborty 1999 Trinucleotide repeats, genetic instability and variation in the human genome. p. 53-64 In: Genomic diversity: Applications in human population genetics. Ed: Papiha, S. S., R. Deka and R. Chakraborty, Kluwer Academic/Plenum Publishers

de Knijff, P., M. Kayser, A. Caglia, D. Corach, N. Fretwell, C. Gehrig, G. Graziosi, F. Heidorn, S. Herrmann, B. Herzog, M. Hidding, K. Honda, M. Jobling, M. Krawczak, K. Leim, S. Meuser, E. Meyer, W. Oesterreich, A. Pandya, W. Parson, G. Penacino, A. Perez-Lezaun, A. Piccinini, M. Prinz, C. Schmitt, P. M. Schneider, R. Szibor, J. Teifel-Greding, G. Weichhold and L. Roewer 1997 Chromosome Y microsatellites: population genetic and evolutionary aspects. International Journal of Legal Medicine 110: 134-140

Denham, W. W. 1987 West Indian Green Monkeys: Problems in historical biogeography. Contributions to Primatology Vol. 24 Ed: F.S. Szalay, Karger Publishers, New York, N.Y

Diaz, F., M. P. Bayona-Bafaluy, M. Rana, M. Mora, H. Hao and C. T. Moraes 2002 Human mitochondrial DNA with large deletions repopulates organelles faster than full-length genomes under relaxed copy number control. Nucleic Acids Research 30: 4626-4633

DiMauro, S., E. Bonilla, M. Davidson, M. Hirano and E. A. Schon 1998 Mitochondria in neuromuscular disorders. Biochimica et Biophysica Acta 1366: 199-210

Di Rienzo, A., A. C. Peterson, J. C. Garza, A. M. Valdes, M. Slatkin and N. B. Freimer 1994 Mutational processes of simple sequence repeat loci in human populations. Proceedings of the National Academy of Sciences USA 91: 3166-3170

Disotell, T. R., R. L. Honeycutt and M. Ruvolo 1992 Mitochondrial DNA phylogeny of the old-world monkey tribe Papionini. Molecular Biology and Evolution 9: 1-13

Dowton, M and A. D. Austin 1995 Increased genetic diversity in mitochondrial genes is correlated with the evolution of parasitism in the Hymenoptera. Journal of Molecular Evolution 41: 958-65

Dowton, M. and A. D. Austin 1997 Evidence for AT-transversion bias in wasp (Hymenoptera: Symphyta) mitochondrial genes and its implications for the origin of parasitism. Journal of Molecular Evolution 44: 398-405

Drew, R. A. I. and D. M. Lambert 1986 On the specific status of *Dacus (Bactrocera) aquilonis* and *D. (Bactrocera) tryoni* (Diptera: Tephritidae). Annals of the Entomological Society of America 79: 870-878

Dubrova, Y. E. , V. N. Nesterov, N. G. Krouchinsk, V. A. Ostapenko, R. Neumann, D. L. Neil and A. J. Jeffreys 1996 Human minisatellite mutation rate after the Chernobyl accident. Nature 380: 683-686

Dubrova, Y. E. , V. N. Nesterov, N. G. Krouchinsk, V. A. Ostapenko, G. Vergnaud, F. Giraudeau, J. Buard and A. J. Jeffreys 1997 Further evidence for elevated human minisatellite mutation rate in Belarus eight years after the Chernobyl accident, Mutation Research 381: 267-278

Duncan, T. B. 1972 Atlantic islands: Madiera, the Azores, and the Cape Verdes in seventeenth century commerce and navigation. University of Chicago Press, Chicago

Dunn, R. S. 1973 Sugar and slaves: rise of the planter class in the English West Indies 1624-1713. Published for the Institute of Early AME. W. Norton, New York 379 pp

Eichler, E. E. and P. J. DeJong 1995 Biomedical applications and studies of molecular evolution: a proposal for a primate genomic library resource. Genome Research 12: 673-678

Eisen, J. A. 1999 Mechanistic basis for microsatellite instability. p: 34-48 In: Microsatellites: evolution and applications. Ed: Godstein, D. B. and C. Schlötterer, Oxford University Press, Oxford

Ellegren, H. C., R. Primmer and B. Sheldon 1995 Microsatellite evolution: directionality or bias in locus selection? Nature Genetics 11: 60-62

Ellegren, H., S. Moore, N. Robinson, K. Byrne, W. Ward and B. C. Sheldon 1997 Microsatellite evolution: a reciprocal study of repeat lengths at homologous loci in cattle and sheep. Molecular Biology and Evolution 14: 854-860

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

Elliot, D. G. 1907 A checklist of mammals of the North American continent, the West Indies, and the neighboring seas. Field Columbia Museum, Zoology Series 6: 536-537

Ellsworth J. and G. Hoelzer 1998 Characterization of microsatellite loci in a New World primate, the mantled howling monkeys (*Alouatta palliata*). Molecular Ecology 7: 657-658

Engels, W. R. 1993 Contributing software to the Internet: The Amplify program. Trends Biochemical Science 18: 448-450

Engles, B. 1994 Amplify program. WREngels@macc.wisc.edu

Ervin, P. M. and M. T. McGuire 1974 The St. Kitts vervet (*Cercopithecus aethiops*). Caribbean Quarterly 20: 36-52

Estoup, A., F. Rousset, Y. Michalakis, J. M. Cornuet, M. Adriamanga and R. Guyomard 1998 Comparative analysis of microsatellite and allozyme markers: a case study investigating microgeographic differentiation in brown trout (*Salmo trutta*). Molecular Ecology 7: 339-353

Evett, I. W. and B. S. Weir 1998 Interpreting DNA evidence: statistical genetics for forensic scientists. Sinauer Press, Sunderland 278p

Excoffier, L., P. Smouse and J. Quattro 1992 Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. Genetics 131: 479-491

Excoffier, L. and Z. Yang 1999 Substitution rate variation among sites in mitochondrial hypervariable region I of humans and chimpanzees. Molecular Biology and Evolution 16: 1357-1368

Eyre-Walker, A., N. H. Smith and J. M. Smith 1999 How clonal are human mitochondria? Proceedings of the Royal Society of London B: Biological Sciences 266: 477-483

Fay, J. C., G. J. Wyckoff and C. Wu 2002 Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. Nature 415: 1024-1026

Feldman, M. W., A. Bergman, D. D. Pollock and D. B. Goldstein 1997 Microsatellite genetic distances with range constraints: analytic description and problems of estimation. Genetics 145: 207-216

Fenton, B., G. Malloch and F. Germa 1998 A study of variation in rDNA ITS regions shows that two haplotypes coexist within a single aphid genome. Genome 41: 337-345

Ferris, S. D., W. M. Brown, W. S. Davidson and A. C. Wilson 1981 Extensive polymorphism in the mitochondrial DNA of apes. Proceedings of the National Academy

of Sciences USA 78: 6319-6323

Fitt, G. P. 1989 The role of interspecific interactions in the dynamics of tephritid populations. In: Fruit Flies, Their Biology, Natural Enemies and Control. World Crop Pests. Ed: A. S. Robinson and G. Hooper, pp. 281–300. Elsevier, Amsterdam

FitzSimmons, N. N., C. Moritz and S. S. Moore 1995 Conservation and dynamics of microsatellite loci over 300 million years of marine turtle evolution. Molecular Biology and Evolution 12: 432-440

Foster, E. A., M. A. Jobling, P. G. Taylor, P. Donnelly, P. De Knijff, R. Mieremet, T. Zerjal and C. Tylersmith 1998 Jefferson fathered slave's last child. Nature 396: 27-28.

Freimer, N. B. and M. Slatkin 1996 Microsatellites: evolution and mutational processes. p: 51-72 In: Variation in the human genome – Ciba Foundation Symposium 197. Ed: Chadwick, D. and G. Cardew, John Wiley & Sons Ltd, Hoboken, NJ

Friesen, V. L., J. F. Piatt, and A. J. Baker 1996. Evidence from cytochrome *b* sequences and allozymes for a ''new'' species of alcid: the long-billed murrelet (*Brachyramphus peridix*). Condor 98: 681-690

Friesen, V. L., B. C. Congdon, H. E. Walsh and T. P. Birt 1997 Intron variation in marbled murrelets detected using analysis of single-stranded conformational polymorphisms. Molecular Ecology 6: 1047-1058

Gagneux, P., C. Wills, U. Gerloff, D. Tautz, P. A. Morin, C. Boesch, B. Fruthi, G. Hohmanni, O. A. Ryder and D. S. Woodruff 1999 Mitochondrial sequences show diverse evolutionary histories of African hominoids. Proceedings of the National Academy of Sciences USA. 96: 5077-5082

Gardner, S. N and M. C Wagner 2005 Software for optimization of SNP and PCR-RFLP genotyping to discriminate many genomes with the fewest assays. BMC Genomics 6: 73

Gee, J. H. 1969a Effect of daily synchronization of sexual activity on mating success in laboratory populations of two species of *Dacus* (Diptera: Tephritidae). Australian Journal of Zoology 17: 619-624

Gee, J. H. 1969b An analysis of natural selection in laboratory populations of *Dacus* (Tephritidae: Diptera). Evolution 23: 212-224

Gerber, A. S., R. Loggins, S. Kumar, T. E. Dowling 2001 Does nonneutral evolution shape observed patterns of DNA variation in animal mitochondrial genomes? Annual Review of Genetics 35: 539-566

Gibbs, G. W. 1965 Comparative ecology and sexual isolation of two species of *Dacus*. PhD Thesis, University of Sydney

Gibbs, G. W. 1967 The comparative ecology of two closely related sympatric species of *Dacus* (Diptera) in Queensland. Australian Journal of Zoology 15: 1123-1139

Gibbs, G. W. 1968 The frequency of interbreeding between two sibling species of *Dacus* (Diptera) in wild populations. Evolution 22: 667-683

Gissi, C., A. Reyes, G. Pesole and C. Saccone 2000 Lineage-specific evolutionary rate in mammalian mtDNA. Molecular Biology and Evolution 17: 1022-1031

Giles, R. E., J. Blanc, J. M. Cann and D. C. Wallace 1980 Maternal inheritance of human mitochondrial DNA. Proceedings of the National Academy of Sciences USA 77: 6715-6719

Gill, P., P. L. Ivanov, C .Kimpton, R. Piercy, N. Benson, I. Evett, E. Hagelberg, K. Sullivan 1994 Identification of the remains of the Romanov family by DNA analysis. Nature Genetics 6: 130-135

Gokey, N. G., Z. Cao, J. W. Pak, D. Lee, S. H. McKiernan, D. McKenzie, R. Weindruch and J. M. Aiken 2004 Molecular analyses of mtDNA deletion mutations in microdissected skeletal muscle fibers from aged rhesus monkeys. Aging Cell 3: 319-326

Goldsmith, O. 1840 A history of the earth and animated nature. Blackie and Son, Glasgow

Goldstein, D. B. and C. Schlötterer 1999 Microsatellites: evolution and applications. Oxford University Press, Oxford

Goldstein, D. B., G. W. Roemer, D. A. Smith, D. E. Reich, A. Bergmen and R. K. Rayne 1999 The use of microsatellite variation to infer population structure and demographic history in a natural model system. Genetics 151: 797-801

Gongora, J., Y. Chen, J. E. Bernal, F. W. Nicholas and C. Moran 2002 Interspecific amplification of peccary microsatellite markers using porcine primers. Animal Genetics 33: 312-327

Goudet, J., M. Raymond, T. De Meeüs and F. Rousset 1996 Testing differentiation in diploid populations. Genetics 144: 1933-1940

Graf, D. L. and J. S. Sparks 2000 Sometimes too much is not enough: a comment on the use of mtDNA to recover ancient divergences. Naïve Verificationist 1: 1-13,

Green, L. D., J. N. Derr and A. Knight 2000 mtDNA affinities of the peoples of north-Central Mexico. American Journal of Human Genetics 66: 989-998

Green, C. L. and M. Frommer 2001 The genome of the Queensland fruit fly *Bactrocera tryoni* contains multiple representatives of the Mariner family of transposable elements. Insect Molecular Biology 10: 371–386

Greenberg, B. D., J. E. Newbold and A. Sugino 1983 Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. Gene 21: 33-49

Grossman, P. D., W. Bloch, E. Brinson, C. C. Chang, F. A. Eggerding, S. Fung, D. M. Iovannisci, S. Woo, E. S. Winn-Deen and D. A. Iovannisci 1994 High-density multiplex detection of nucleic acid sequences: Oligonucleotide ligation assay and sequence-coded separation. Nucleic Acids Research 22: 4527-4534

Groves, C. 2001 Primate Taxonomy. Smithsonian Institution Press, Washington and London

Grubb, P. Thomas M. Butynski, John F. Oates, Simon K. Bearder, Todd R. Disotell, Colin P. Groves, and Thomas T. Struhsaker 2003 An assessment of the diversity of African primates. International Journal of Primatology 24, 1301-1357

Grzybowski, T. 2000 Extremely high level of human mitochondrial DNA heteroplasmy in single hair roots. Electrophoresis 21: 548-553

Guarino, F. D., L. Federle, R. A. H. van Oorschot, I. Briceno, J. E. Bernal, S. S. Papiham M. S. Schanfield and R. J. Mitchell 1999 Genetic diversity among five Native American tribes of Colombia: Evidence from nine autosomal microsatellites. p. 33-51 In: Genomic diversity: Applications in human population genetics. Ed: Papiha, S. S., R. Deka and R. Chakraborty. Kluwer Academic/Plenum Publishers, New York

Guo, S. and E. Thompson 1992 Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 48: 361-372

Gyllensten, U., D. Wharton, A. Joseffson and A. C. Wilson 1991 Paternal inheritance of mitochondrial DNA in mice. Nature 352: 255-257

Hadfield, R. M., J. G. Pullen, K. F. Davies, S. E. Wolfensohn, J. W. Kemnitz, D. E. Weeks, S. T. Bennett and S. H. Kennedy 2001 Toward developing a genome-wide microsatellite marker set for linkage analysis in the rhesus macaque (*Macaca mulatta*): Identification of 76 polymorphic markers. American Journal of Primatology 54: 223-231

Hagelberg, E. 2003 Recombination or mutation rate heterogeneity? Implications for Mitochondrial Eve. Trends in Genetics 19: 84-90

Hall, K. R. L. and J. S. Gartlan 1965 Ecology and behavior of the vervet monkey (*Cercopithecus aethiops*), Lolui Islands, Lake Victoria. Proceedings of the Zoological Society of London 145: 37-56

Hamblin, M. and M. Veuille1999. Population structure among African and derived populations of *Drosophila simulans*: evidence for ancient subdivision and recent admixture. Genetics 153: 305–317

Hammer, M. F. 1995 A recent common ancestry for human Y chromosomes. Nature 378: 376-378

Hancock, J. M. 1995 The contribution of slippage-like processes to genome evolution. Journal of Molecular Evolution 41: 1038-1047

Hancock, J. M. 1999 Microsatellites and other simple sequences: genomic context and mutational mechanisms. In Microsatellites: evolution and applications. Ed: Godstein, D. B. and C. Schlötterer, Oxford University Press, Oxford

Handler, J. S. and F. W. Lange 1978 Plantation slavery in Barbados: an archeological and historical investigation. Harvard University Press, Cambridge MA

Hanna, M. G., I. P. Nelson, J. A. Morgan-Hughes and N. W. Wood 1998 MELAS: a new disease associated mitochondrial DNA mutation and evidence for further genetic heterogeneity. Journal of Neurology, Neurosurgery and Psychiatry 65**:** 512-517

Hardy, D. E. 1951 The Krauss collection of Australian fruit flies (Tephritidae-Diptera). Pacific Science 5: 115-189

Harr, B. 2000 Evolution of Microsatellite in *Drosophila*. PhD Dissertation, der Fakultät für Biologie der Universität Konstanz

Harr, B. and C. Schlötterer 2000 Long microsatellites in *D. melanogaster* have a downward mutation bias and short persistence times, which cause their genome wide under-representation. Genetics 155: 1213-1220

Harr, B., B. Zangerl, G. Brem and C. Schlötterer 1998 Conservation of locus-specific microsatellite variability across species: a comparison of two *Drosophila* sibling species, *D. melanogaster* and *D. simulans*. Molecular Biology and Evolution 15:176-184

Harris, D. J. and K. A. Crandall 2000 Intragenomic variation within ITS1 and ITS2 of freshwater crayfishes (Decapoda: Cambaridae): implications for phylogenetic and microsatellite studies. Molecular Biology and Evolution 17: 284-291

Hasegawa, M., A. Di Rienzo, T. D. Kocher and A. C. Wilson 1993 Toward a more accurate time scale for the human mitochondrial DNA tree. Journal of Molecular Evolution 37: 347-354

Hauswirth, W. W., M. J. Van De Walle, P. J. Laipis and P. D. Olivo 1984 Heterogeneous mitochondrial DNA D-loop sequences in bovine tissue. Cell 37: 1001-1007

Hayasaka, K. and T. Gojobori, S. Horai 1988 Molecular phylogeny and evolution of primate mitochondrial DNA. Molecular Biology and Evolution 5: 626-44

Hayasaka, K., T. Ishida and S. Horai 1991 Heteroplasmy and polymorphism in the major noncoding region of mitochondrial DNA in Japanese monkeys: association with tandemly repeated sequences. Molecular Biology and Evolution 8: 399-415

Hazelrigg, T. and S. Petersen 1992 An unusual genomic position effect on *Drosophila white* gene expression: pairing dependence, interactions with *xeste,* and molecular analysis of revertants. Genetics 130: 125-138

Hedges, S. B., P. H. Parker, C. G. Sibley and S. Kumar 1996 Continental breakup and the ordinal diversification of bird and mammals. Nature 381: 226-229

Helgason, A., B. Hrafnkelsson, J. R. Gulcher, R. Ward and K. Stefansson 2003 A population wide coalescent analysis of Icelandic matrilineal and patrilineal genealogies: evidence for a faster evolutionary rate of mtDNA lineages than Y chromosomes. American Journal of Human Genetics 72: 1370-1388

Herbin, M., D. Boire and M. Ptito 1997 Size and distribution of retinal ganglion cells in the St. Kitts Green Monkey (*Cercopithecus aethiops sabaeus*). The Journal of Comparative Neurology 383: 459-472

Heyer, E., J. Puymirat, P. Dieltjes, E. Bakker and P. de Knijff 1997 Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. Human Molecular Genetics, 6: 799-803

Hirano, M., A. Shtilbans, R. Mayeux, M. M. Davidson, S. Dimauro, J. A. Knowles and E. A. Schon 1997 Apparent mtDNA heteroplasmy in Alzheimer's disease patients and in normals due to PCR amplification of nucleus-embedded mtDNA pseudogenes. Proceedings of the National Academy of Sciences USA 94: 14894-14899

Hoekert W. E., H. Neufeglise, A. D. Schouten and S. B. Menken 2002 Multiple paternity and female-biased mutation at a microsatellite locus in the olive ridley sea turtle (*Lepidochelys olivacea*). Heredity 89: 107-13

Hollister, N. 1912 The mona monkey of the island of St. Kitts. Proceedings of the Biological Society of Washington 25: 93

Horai, S. and K. Hayasaka 1990 Intraspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA. American Journal of Human Genetics 462: 828-842

Horai, S., K. Hayasaka, R. Kondd, K. Tsugane and N. Takanata 1995 Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. Proceedings of the National Academy of Sciences USA 92: 532-536

Horrocks, J. A. 1985 Life-history characteristics of a wild population of vervets (*Cercopithecus aethiops sabaeus*) in Barbados, West Indies. International Journal of Primatology 7: 31-47

Howell, N., I. Kubacka and D. A. Mackey 1996 How rapidly does the human mitochondrial genome evolve? American Journal of Human Genetics 59: 501-509

Howell, N., C. B. Smejkal, D. A. Mackey, P. F. Chinnery, D. M. Turnbull and C. Herrnstadt 2003 The pedigree rate of sequence divergence in the human mitochondrial genome: there is a difference between phylogenetic and pedigree rates. American Journal of Human Genetics 72: 659-670

Huigens, M. E., R. F. Luck, R. H. G. Klaassen, M. F. P. M. Maas, M. J. T. N. Timmermans and R. Stouthamer 2000 Infectious parthenogenesis. Nature 504: 178-179

Hutter, C. M., M. D. Schug and C. F. Aquadro 1998 Microsatellite variation in *Drosophila melanogaster* and *Drosophila simulans*: a reciprocal test of the ascertainment bias hypothesis. Molecular Biology and Evolution 15: 1620-1636

Iannone, M. A., J. D. Taylor, J. Chen, M. S. Li, P. Rivers, K. A. Slentz-Kesler and M. P. Weiner 2000 Multiplexed single nucleotide polymorphism genotyping by oligonucleotide ligation and flow cytometry. Cytometry 39: 131-140

Ikeuchi, T., K. Sanpei, H. Takano, H. Sasaki, K. Tashiro, G. Cancel, A. Brice, T. D. Bird, G. D. Schellenberg, M. A. Pericak-Vance, K. A. Welsh-Bohmer, L. N. Clark, K. Wilhelmsen and S. Tsuji 1998 A novel long and unstable CAG/CTG trinucleotide repeat on chromosome 17q. Genomics 49: 321-326

Imbert, G., C. Kretz, K. Jognson and J-L Mandel 1993 Origin of the expansion mutation in myotonic dystrophy. Nature Genetics 4: 72-76

Irvin, S. D., K. A. Wetterstrand, C. M. Hutter and C. F. Aquadro 1998 Genetic variation and differentiation at microsatellite loci in *Drosophila simulans*: evidence for founder effects in New World populations. Genetics 150: 777-790

Isbell, L. A., J. D. Pruetz, M. Lewis and T. P. Young 1998a Locomotor activity differences between sympatric patas monkeys (*Erythrocebus patas*) and vervet monkeys (*Cercopithecus aethiops*): Implications for the evolution of long hindlimb length in Homo. American Journal of Physical Anthropology 105: 199-207

Isbell, L. A., J. D. Pruetz and T. P. Young 1998b Movements of vervets (*Cercopithecus aethiops*) and patas monkeys (*Erythrocebus patas*) as estimators of food resource size, density, and distribution. Behavioral Ecology and Sociobiology 42: 123-133

Jacobs, H. 2002 Like father, like son – almost. Trends in Genetics 18: 551-552

Jacobs, H. T., D. J. Elliot, V. B. Math and A. Farquharson 1988 Nucleotide sequence and gene organization of sea urchin mitochondrial DNA. Journal of Molecular Biology 202**:** 185-217

Jae-Heup, K., E. Eizirik, S. J. O'Brien and W. E. Johnson 2001 Structure and patterns of sequence variation in the mitochondrial DNA control region of the great cats. Mitochondrion 14: 279-292

Janke, A., X. Xu and U. Arnason 1997 The complete mitochondrial genome of the wallaroo (*Macropus robustus*) and the phylogenetic relationship among Monotremata, Marsupialia, and Eutheria. Proceedings of the National Academy of Sciences USA 94: 1276-1281

Jardine, E. 1833 The naturalists' library: Vol. 1. (mammalia) monkeys. W. H. Lizars and Stirling and Kenny, Edinburgh

Jeffreys, A. J., N. J. Royle, V. Wilson and Z. Wong 1988 Spontaneous mutation rates to new length alleles at tandem repetitive hypervariable loci in human DNA. Nature 332: 278-281

Jenuth J. P., A. C. Peterson, K. Fu and E. A. Shoubridge 1996 Random genetic drift in the female germline explains the rapid segregation of mammalian mitochondrial DNA. Nature Genetics 14: 146-151

Jenuth, J. P, A. C. Peterson, E. A. Shoubridge 1997 Tissue-specific selection for different mtDNA genotypes in heteroplasmic mice. Nature Genetics 16: 93-5

Jobling, M. A., V. Smara, A. Padya, N. Fretwell, B. Bernasconi, R. J. Mitchell, T. Gerelsaikhan, B, Dashnyam, A. Sajantila, P. J. Salo, Y. Nakahori, C. M. Disteche, K. Thangaraj, L. Singh, M. H. Crawford and C. Tyler-Smith 1996 Recurrent duplication and deletion polymorphisms on the long arm of the Y chromosome in normal males. Human Molecular Genetics 5: 1767-1775

Jones, A. G. and W. R. Ardren 2003 Methods of parentage analysis in natural populations. Molecular Ecology 12: 2511-2523

Jordan, B., A. Charest, J. F. Dowd, J. P. Blumenstiel, R-f. Yeh, A. Osman, D. E. Housman and J. E. Landers 2002 Genome complexity reduction for SNP genotyping analysis. Proceedings of the National Academy of Sciences USA 99: 2942-2947

Jorde L. B., A. R. Rogers, M. Bamshad, W. S. Watkins, P. Krakowiak, S. Sung, J. Kere and H. C. Harpending 1997 Microsatellite diversity and the demographic history of modern humans. Proceedings of the National Academy of Science USA 94: 3100-3103

Kaminska, B., L. Kaczmarek, S. Larocque, and A. Chaudhuri 1997 Activity-dependent regulation of cytochrome b gene expression in monkey visual cortex. Journal of Comparative Neurology 379: 271-282

Kan, Y. W. and A. M. Dozy 1978 Polymorphism of DNA sequence adjacent to human β-globin structural gene: Relationship to sickle mutation. Proceedings of the National Academy of Sciences USA 75: 5631-5635

Kandpal, R. P., G. Kandpal and S. M. Weissman 1994 Construction of libraries enriched for sequence repeats and jumping clones, and hybridization selection for region-specific markers. Proceedings of the National Academy of Sciences USA 91: 88-92

Kann, L. M., E. B. Rosenblum and D. M. Rand 1998 Aging, mating, and the evolution of germline heteroplasmy for mitochondrial DNA length variants in *Drosophila melanogaster*. Proceedings of the National Academy of Sciences USA 95: 2372–77

Kanthaswamy, S., J. J. Ely and D. G. Smith 1998. Simple sequence repeat polymorphisms for wild and captive orangutan conservation management. In: All Apes Great and Small, Vol. 2 - Asian Apes. Ed: Galdikas, B., N. Briggs, L. Sheeran, G. Shapiro and J. Goodall. Kluver NY, USA

Karagyozov, L., I. D. Kalcheva and V. M. Chapman 1993 Construction of random small-insert genomic libraries highly enriched for simple sequence repeats. Nucleic Acids Research 21: 3911-3912

Kashi, Y., D. King and M. Soller 1997 Simple sequence repeats as a source of quantitative genetic variation. Trends in Genetics 13: 74-8

Katti, M.V., P. K. Ranjekar and V. S. Gupta 2001 Differential distribution of simple sequence repeats in eukaryotic genome sequences. Molecular Biology and Evolution 18: 1161-1167

Kijas, J. M., J. C. Fowler, C. A. Garbett, M. R. Thomas 1994 Enrichment of microsatellites from the citrus genome using biotinylated oligonucleotide sequences bound to streptavidin-coated magnetic particles. Biotechniques 16: 656-662

Kimura, M 1980 A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution 16: 111-120

Kinghorn, B. and S. Kinghorn, 2005 Pedigree viewer. http://metz.une.edu.au/~bkinghor

Kinnear, M. W., J. S. Bariana, J. A. Sved and M. Frommer 1998 Polymorphic microsatellite markers for population analysis of a tephritid pest species, *Bactrocera tryoni*. Molecular Ecology 7: 1489-1495

Kilpatrick, S. R., and D. M. Rand 1995 Conditional hitchhiking of mitochondrial genetic background. Genetics 141: 1113-1124

Kingdon, J. S. 1997 The Kingdon Field Guide to African Mammals. Academic Press, London

Kittles, R. A., A. W. Bergen, M. Urbanek, M. Virkkunen, M. Linnoila, D. Goldman and J. C. long 1999 Autosomal, mitochondrial, and Y chromosome DNA variation in Finland: evidence for a male-specific bottleneck. American Journal of Physical Anthropology 108: 381-399

Kivisid, T., K. Kaldma, M. Metspalu, J. Parik, S. Papiha and R. Villems 1999 The place of the Indian mitochondrial DNA variants in the global network of maternal lineages and the peopling of the old world. p. 135-152 In: Genomic diversity: Applications in human population genetics. Ed: Papiha, S. S., R. Deka and R. Chakraborty. Kluwer Academic/Plenum Publishers, New York

Kondo, R., E.T. Mastsuura, H. Ishima, N. Takahata and S. I. Chigusa 1990 Incomplete maternal transmission of mitochondrial DNA in *Drosophila*. Genetics 126: 657-663

Krieger, M. J. B. and K. G. Ross 2002 Identification of a major gene regulating complex social behavior. Science 295: 328-332

Kruglyak, S., R. T. Durrett, M. D. Schug and C. F. Aquadro 1998 Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. Proceedings of the National Academy of Sciences USA 95: 10774-10778

Kryger, U., T. J. Robinson and P. Bloomer 2002 Isolation and characterization of six polymorphic microsatellite loci in South African hares (*Lepus saxatilis* F. Cuvier, 1823 and *Lepus capensis* Linnaeus, 1758). Molecular Ecology Notes 2: 422-424

Kumar, S. and B. Hedges 1998 A molecular timescale for vertebrate evolution. Nature 392: 917-920

Kumar, S., K. Tamura and M. Nei 2004 MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Briefings in Bioinformatics 5: 150-163

Kunst, C. B. and S. T. Warren 1994 Cryptic and polar variation of the fragile X repeat could result in predisposing normal alleles. Cell 77: 853-861

Labat, J. B. 1722 Nouveau voyage aux iles de l'Amerique. Paris. Cited by Denham, 1987

Lang, C. K. A. 2006 Primate Factsheets: Vervet (*Chlorocebus*) Taxonomy, Morphology, & Ecology. http://pin.primate.wisc.edu/factsheets/entry/vervet/taxon

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

Lansman, R. A., J. C. Avise and M. D. Huettel 1983 Critical experimental test of the possibility of paternal leakage of mitochondrial DNA. Proceedings of the National Academy of Sciences USA 80: 1969-1971

Lagerström-Fermér, M., C. Olsson, L. Forsgren and A-C Syvänen 2001 Heteroplasmy of the human mtDNA control region remains constant during life. American Journal of Human Genetics 68: 1299-1301

Lathuillière, M., N. Mèmard and B. Crouau-Roy 2001 Sequence conservation of nine Barbary macaque (*Macaca sylvanus*) microsatellite loci: Implication of specific primers for genotyping. Folia Primatology 72: 85-88

Latorre, A., C. Hernádez, D. Martinez, J. A. Castro, M. Ramó and A. Moya 1992 Population structure and mitochondrial DNA gene flow in Old World populations of *D. subobscura.* Heredity 68: 15-24

Launhardt, K., C. Borries, C. Hardt, J. T. Epplen and P. Winkler 2001 Paternity analysis of alternative male reproductive routes among the langurs (*Semnopithecus entellus*) of Ramnagar. Animal Behaviour 61: 53-64

Le, T. H., M. S. Pearson, D. Blair, N. Dai, L. H. Zhang and D. P. McManus 2002 Complete mitochondrial genomes confirm the distinctiveness of the horse-dog and sheep-dog strains of *Echinococcus granulosus*. Parasitology 124: 97-112

Lee, J. S., M. G. Hanford, J. L. Genova and R. A. Farber 1999 Relative stabilities of dinucleotide and tetranucleotide repeats in cultured mammalian cells. Human Molecular Genetics 8**:** 2567-2572

Lee, L. G., C. R. Connell and W. Bloch 1993 Allelic discrimination by nick-translation PCR with fluorogenic probes. Nucleic Acids Research 21: 3761-3766

Lell, J. T. and D. C. Wallace 2000 The peopling of Europe from the maternal and paternal perspectives. American Journal of Human Genetics 67: 1376-1381

Lertrit, P., R. M. Kapsa, M. J. Jean-Francois, D. Thyagarajan, A. S. Noer, S. Marzuki and E. Byrne 1994 Mitochondrial DNA polymorphism in disease: a possible contributor to respiratory dysfunction. Human Molecular Genetics 3: 1973-1981

Levene, H. 1949 On a matching problem arising in genetics. Annals of Mathematical Statistics 20: 91-94

Lewin, R. 1997 Human Evolution. Harvard University Press, Cambridge, MA

Lewis, R. 2002 SNPs as windows on evolution. The Scientist 16: 16

Lewontin, R. C. and L. C. Birch 1966 Hybridization as a source of variation for adaptation to new environments. Evolution 20: 315-336

Liu, Q., E. C. Thorland, J. A. Heit and S. S. Sommer 1997 Overlapping PCR for bi-directional PCR amplification of specific alleles: a rapid one-tube method for simultaneously differentiation homozygotes and heterozygotes. Genome Research. 7: 389-398

Liu, Z. Q., Y. Q. Wang and B. Su 2005 The mitochondrial genome organization of the rice frog, *Fejervarya limnocharis* (Amphibia: Anura): a new gene order in the vertebrate mtDNA. Gene 346: 145-151

Livac, K. J. 1999 Allelic discrimination using fluorogenic probes and the 5' nuclease assay. Genetic Analysis: Biomolecular Engineering 141: 143-149

Lloyd, C. J., A. P. Norton, R. A. Hufbauer, S. Bogdanowicz and S. J. Nissen 2004 Microsatellite isolation from the gall midge *Spurgia capitigena* (Diptera: Cecidomyiidae), a biological control agent of leafy spurge. Molecular Ecology Notes 4: 605–607

Lodish, H., A. Berk, S. L. Zipursky, P. Matsudaira, D. Baltimore and J. E. Darnell 1999 Molecular Cell Biology, 4th ed. W.H.Freeman & Co. York

Long, J. C. 1986 The allelic correlation structure of Gainj and Kalam speaking people. I. The estimation and interpretation of Wright's F-statistics. Genetics 112: 629-647

Lopez, M. E., N. L. Van Zeeland, D. B. Dahl, R. Weindruch and J. M. Aiken 2000 Cellular phenotypes of age-associated skeletal muscle mitochondrial abnormalities in rhesus monkeys. Mutation Research 452: 123-38

Loxdale, H. D. and G. Lushai 1998 Molecular markers in entomology. Bulletin of Entomological Research 88: 577-600

Lucotte, G., D. Gautreau, G. Galat and A. Galat-Luong 1982 Electrophoretic polymorphism of different subspecies of *Cercopithecus aethiops*. Folia Primatology 38: 183-195

Luikart, G. and P. R. England 1999 Statistical analysis of microsatellite DNA data. Trends in Ecology and Evolution 14: 253-256

Ma, C-X., M. Lin, R. C. Littell, T. Yin and R. Wu 2004 A likelihood approach for mapping growth trajectories using dominant markers in a phase-unknown full-sib family Theoretical and Applied Genetics 108: 699-705

Magoulas, A. and E. Zouros 1993 Restriction-site heteroplasmy in Anchovy (*Engraulis encrasicolus*) indicates incidental biparental inheritance of mitochondrial DNA. Molecular Biology and Evolution 10: 319-325

Mariasegaram, M., S. Pullenayegum, M. J. Ali, R. S. Shah, M. C. T. Penedo, U. Wernery and J. Sasse 2002 Isolation and characterization of eight microsatellite markers in *Camelus dromedarius* and cross-species amplification in *C. bactrianus* and *Lama pacos*. Animal Genetics 33: 385-387

Marshall, T. C., J. Slate, L. E. B. Kruuk and J. M. Pemberton 1998 Statistical confidence for likelihood-based paternity inference in natural populations. Molecular Ecology 7: 639-655

Martin, A. P. and S. R. Palumbi 1993 Body size, metabolic rate, generation time, and the molecular clock. Proceedings of the National Academy of Sciences USA 90: 4087-4091

Marvin, C. H., L. Allan, B. E. McCarry and D. W. Bryant 1993 Chemico/biological investigation of contaminated sediment from the Hamilton Harbour area of western Lake Ontario. Environmental and Molecular Mutagenesis 22: 61-70

Mc Dermid, E. M. and R. Ananthakrishnan 1972 Red cell enzymes and serum proteins of *Cercopithecus aethiops* (South African green monkeys). Folia Primatology 17: 122-131

McDonald, J. G. and M. Kreitman 1991 Adaptive protein evolution at the Adh locus in *Drosophila*. Nature 351: 652-654

McDonald, J. H., B. C. Verrelli and L. B. Geyer 1996 Lack of geographic variation in anonymous nuclear polymorphisms in the American oyster, *Crassostrea virginica*. Molecular Biology and Evolution 13: 1114-1118

McGure, M. and the members of the Behavioral Sciences Foundation, Neuropsychiatric Institute, UCLA, Las angeles, California 1974 The St. Kitts Vervet. Contributions to Primatology vol. 1. Ed: H. Kuhn, W. P. Luckett, C. R. Noback, A. H. Schultz, D. Starck and F. S. Szalay, Karger Publishers, New York, N.Y

McInnis, D. O., S. Y. T. Tam, C. Grace and D. Miyashita 1994 Population suppression and sterility rates induced by variable sex ratio, sterile insect releases of *Ceratitis capitata* (Diptera: Tepheritidae) in Hawaii. Annals of the Entomological Society of America 87: 231-240

McKechnie, S. W. 1975 Enzyme polymorphism and species discrimination in fruit flies of the genus *Dacus* (Tephritidae). Australian Journal of Biological Sciences 28: 405-411

Meats, A., P. Maheswaran, M. Frommer and J. Sved 2002. Towards a male only release system for SIT with the Queensland fruit fly *Bactrocera tryoni*, using a genetic sexing strain with a temperature-sensitive lethal mutation. Genetica 116: 97-106

Meats, A., N. Pike, X. An, K. Raphael and W. Y. S. Wang 2003 The effects of selection for early (day) and late (dusk) mating lines of hybrids of *Bactrocera tryoni* and *Bactrocera neohumeralis*. Genetica 119: 283-293

Meirelles F. V. and Smith L. C. 1997 Mitochondrial genotype segregation in a mouse heteroplasmic lineage produced by embryonic karyoplast transplantation. Genetics 145: 445-451

Meloni, R., V. Albanese, P. Ravassard, F. Treilhou and J. Mallet 1998 A tetranucleotide polymorphic microsatellite, located in the first intron of the tyrosine hydroxylase gene, acts as a transcription regulatory element in vitro. Human Molecular Genetics 7: 423-428

Merriwether, D. A., S. Huston, S. Lyengar, R. Haman, J. M. Norris, S. M. Shetterly, M. I. Kamboh, R. E. Ferrell 1997 Mitochondrial versus nuclear admixture estimates demonstrate a past history of directional mating. American Journal of Physical Anthropology 102: 153-585

Merriwether, D. A., F. A. Kaestle, B. Zemel, G. Koki, C. Mgone, M. Alpers and J. Friedlaender 1999 Mitochondrial DNA variation in the southwest Pacific. p. 153-182 In: Genomic diversity: Applications in human population genetics. Ed: Papiha, S. S., R. Deka and R. Chakraborty. Kluwer Academic/Plenum Publishers, New York

Mesa, N. R., M. C. Mondragon, I. D. Soto, M. V. Parra, C. Duque, D. Ortiz-Barrientos, L. F. Garcfa, I. D. Velez, M. L. Bravo, J. G. Munera, G. Bedoya, M-C Bortolini and A. Ruiz-Linares 2000 Autosomal, mtDNA, and Y-Chromosome diversity in Amerinds: pre- and post-Columbian patterns of gene flow in South America. American Journal of Human Genetics 67: 1277-1286

Messier, W., S-H, Li and C-B. Stewart 1996 The birth of microsatellites. Nature 381: 483

Metspaly, E., T. Kivisid, K. Kaldma, J. Parik, M. Reidla, K. Tambets and R. Villems 1999 The trans-Caucasus and the expansion of the Caucasoid-specific human mitochondrial DNA. p. 121-133 In: Genomic diversity: Applications in human population genetics. Ed: Papiha, S. S., R. Deka and R. Chakraborty. Kluwer Academic Plenum Publishers, New York

Michalakis, Y. and L. Excoffier 1996 A generic estimation of population subdivision using distances between alleles with special reference to microsatellite loci. Genetics 142: 1061-1064

Morgan, T. H. 1910 Sex-limited inheritance in *Drosophila*. Science 32: 120-122

Moritz, C., C. J. Schneider and D. Wake 1992 Evolutionary relationships within the *Ensatina eschscholtzii* complex confirm the ring species interpretation. Systematic Zoology 41: 273-291

Morrow, J., L. Scott, B. Congdon, D. Yeates, M. Frommer and J. Sved 2000 Close genetic similarity between two sympatric species of Tephritid fruit fly reproductively isolated by mating time. Evolution 54: 899-910

Moum, T. and I. Bakke 2001 Mitochondrial control region structure and single site heteroplasmy in the razorbill (*Alca torda*; Aves). Current Genetics 39: 198-203

Mountain, J. L., A. Knight, M. Jobin, C. Gignoux, A. Miller, A. A. Lin and P. A. Underhill 2002 SNPSTRs: Empirically derived, rapidly typed, autosomal haplotypes for inference of population history and mutational processes. Genome Research 12: 1766-1772

Mulligan, C. J., A. Kitchen and M. M. Miyamoto 2006 Comment on "Population Size Does Not Influence Mitochondrial Genetic Diversity in Animals" Science 314: 1390

Nachman, M. W., W. M. Brown, M. Stoneking and C. F. Aquadro 1996 Nonneutral mitochondrial DNA variation in humans and chimpanzees. Genetics 142: 953-963

Nakamoto, M., H. Takebayashi, Y. Kawaguchi, S. Narumiya, M. Taniwaki, Y. Nakamura, Y. Ishikawa, I. Akiguchi, J. Kimura and A. Kakizuka 1997 A CAG/CTG expansion in the normal population. Nature Genetics 17: 385-386

Nei, M. and S. Kumar 2000 Molecular evolution and phylogenetics. Oxford University Press, New York

Nevanlinna, H. R. 1972 The Finnish population structure: a genetic and genealogical study. Hereditas 71: 195-236

Newman, T. K., L. A. Fairbanks, D. Pollack and J. Rogers 2002 Effectiveness of human microsatellite loci for assessing paternity in a captive colony of vervets (*Chlorocebus aethiops sabaeus*). American Journal of Primatology 56: 237-243

Nielsen, J., A. A. Peixoto, A. Piccin, R. Costa, C. P. Kyriacou and D. Chalmers 1994 Big fliles, small repeats: the "Thr-Gly" region of the period gene in Diptera. Molecular Biology and Evolution 11: 839-853

Noor, M. A. F. 2003 Genes to make new species. Nature 423: 699-700

Nurnberg, P., U. Sauermann, M. Kayser, C. Lanfer, E. Manz, A. Widdig, J. Berard, F. B. Bercovitch, M. Kessler, J. Schmidtke, M. Krawczak 1998 Paternity assessment in rhesus macaques (*Macaca mulatta*): multilocus DNA fingerprinting and PCR marker typing. American Journal of Primatology 44: 1-18

Ojala, D., J. Montoya and G. Attardi 1981 tRNA punctuation model of RNA processing in human mitochondria. Nature 290: 470-474

Olson, M., L. Hood, C. Cantor and D. Botstein 1989 A common language for physical mapping of the human genome. Science 245: 1434-1435

Orita, M., Y. Suzuki, T. Sekiya and K. Hayashi 1989 Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. Genomics 5: 874-879

Orr, H. T., M. Y. Chung, S. Banfi, T. J. Kwiatkowski, A. Servadio, A. L. Beaudet, A. E. McCall, L. A. Duvick, L. P. Ranum and H. Y. Zoghbi 1993 Expansion of an unstable trinucleotide (CAG) repeat in spinocerebellar ataxia type a. Nature Genetics 4: 221-226

Osborne, R., A. Meats, M. Frommer, J. A. Sved, R. A. I. Drew and M. K. Robson 1997 Australian distribution of 17 species of fruit flies (Diptera: Tephritidae) caught in cue lure traps in February 1994. Australian Journal of Entomology 36: 45-50

Ostrander, E. A., P. M. Jong, J. Rine and G. Duyk 1992 Construction of small-insert genomic DNA libraries highly enriched for microsatellite repeat sequences. Proceedings of the National Academy of Sciences of the USA 89: 3419-3423

Paetkau, D. 1999 Microsatellites obtained using strand extension: An enrichment protocol. Biotechniques 26: 690-697.

PalsbØll, P. J., M. Bérubé and H. JØrgensen 1999 Multiple levels of single-strand slippage at cetacean tri- and tetranucleotide repeat microsatellite loci. Genetics 151: 285-296

Parsons, T. J., D. S. Muniec, K. Sullivan, N. Woodyatt, R. Alliston-Greiner, M. R. Wilson, D. L. Berry, K. A. Holland, V. W. Weedn, P. Gill and M. M. Holland 1997 A high observed substitution rate in the human mitochondrial DNA control region. Nature Genetics 15: 363-368

Pascual, M., C. F. Aquadro, V. Soto and L. Serra 2001 Microsatellite variation in colonizing and palearctic populations of *Drosophila subobscura*. Molecular Biology and Evolution 18: 731-40

Pastinen, T., J. Partanen and A. C. Syvanen 1997 Minisequencing: A specific tool for DNA analysis and diagnostics on oligonucleotide arrays. Genome Research 7: 606-614

Pastinen, T., M. Raitio, K. Lindroos, P. Tainola, L. Peltonen and A-C Syvänen 2000 A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. Genome Research 10: 1031-1042

Pemberton, J. M., J. Slate, D. R. Bancroft and J. A. Barret 1995 Nonamplifying alleles at microsatellite loci: a caution for parentage and population studies. Molecular Ecology 4: 249-252

Perelygin, A. A., C. M. Kammerer, N. C. Stowell and J. Rogers 1996 Conservation of human chromosome 18 in baboons (*Papio hamadryas*): a linkage map of eight human microsatellites. Cytogenetics and Cell Genetics 75: 207-209

Pero, M. D., M. Boniotto, D. Zuccon, P. Cervella, A. Spano, A. Amoroso and S. Crovella 2002 B-Defensin 1 gene variability among non-human primates. Immunogenetics 53: 907-913

Pesole, G., F. Mignone, C. Gissi, G. Grillo, F. Licciulli and S. Liuni 2001 Structural and functional features of eukaryotic mRNA untranslated regions. Gene 276: 73-81

Piepho, H-P. and G. Koch 2000 Codominant analysis of banding data from a dominant marker system by normal mixtures. Genetics 155: 1459-1468

Piercy, R., K. M. Sullivan, N. Benson and P. Gill 1993 The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. International Journal of Legal Medicine 106: 85-90

Pitocchelli, J., J. Piatt and M. Cronin 1995 Morphological and genetic divergence among Alaskan populations of *Brachyramphus murrelets*. The Wilson Bulletin 107: 235-250

Poirier, F. E. 1972 The St. Kitts green monkey (*Cercopithecus aethiops sabaeus*). Folia Primatology 17: 20-55

Ponamarev, M. V., M. J. Longley, D. Nguyen, T. A. Kunkel and W. C. Copeland 2002 Active site mutation in DNA polymerase g associated with progressive external ophthalmoplegia causes error-prone DNA synthesis. Journal of Biological Chemistry 277: 15225-15228

Porter, C. A., P. L. Scott, J. Czelusniak, H. Schneider, P. M. C. Schneider, I. Sampaio and M. Goodman 1997 Phylogeny and evolution of selected primates as determined by sequences of the e-globin locus and 58 flanking regions. International Journal of Primatology 18: 261-295

Porter, C. H. and F. H. Collins 1991 Species-diagnostic differences in a ribosomal DNA internal transcribed spacer from the sibling species *Anopheles freeborni* and *Anopheles*

hermsi (Diptera: Culicidae). The American Journal of Tropical Medicine and Hygiene 45: 271-279

Poulton, D. J. 1987 Trace contaminant status of Hamilton Harbour. Journal of Great Lakes Research 13: 193–201

Powell, J. R. 1997 Progress and prospects in evolutionary biology: The *Drosophila* model. Oxford University Press, Oxford.

Prevosti, A., G. Ribo, L. Serra, M. Aguade, J. Balana, M. Monclu and F. Mestres 1988 Colonization of America by *Drosophila subobscura*: experiment in natural populations that supports the adaptive role of chromosomal inversion polymorphism. Proceedings of the National Academy of Sciences USA 85: 5597-5600

Pupko, T. and D. Graur 1999 Evolution of microsatellites in the yeast *Saccharomyces cerevisiae*: role of length and number of repeated units. Journal of Molecular Evolution 48: 313-316

Queney, G., N. Ferrand, S. Weiss, F. Mougel and M. Monnerot 2001 Stationary distributions of microsatellite loci between divergent population groups of the European rabbit (*Oryctolagus cuniculus*). Molecular Biology and Evolution 18: 2169–2178

Raaum, R. L., K. N. Sterner, C. M. Noviello, C-B Stewart and T. R. Disotell 2005 Catarrhine primate divergence dates estimated from complete mitochondrial genomes: concordance with fossil and nuclear DNA evidence. Journal of Human Evolution 48: 237-257

Rand, D. M. 2001 The units of selection on mitochondrial DNA. Annual Review of Ecology and Systematics 32: 415–448

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

Raymond, M. and F. Rousset 1995b An exact test for population differentiation. Evolution 49: 1280-1283

Reynolds, J., B. S. Weir and C. C. Cockerham 1983 Estimation for the coancestry coefficient: basis for a short-term genetic distance. Genetics 105: 767-779

Reeb, C. A. and J. C. Avise 1990 A genetic discontinuity in a continuously distributed species: mitochondrial DNA in the American oyster, *Crassostrea virginica*. Genetics 124: 397-406

Richards, M., H. Corte-Real, H. Forster, V. Macaulay, H. Wilkinson-Hebbots, A. Demaine, S. Papiha, R. Hedges, H-J Bandelt, B. Sykes 1996 Paleolithic and Neolithic

lineages in the European mitochondrial gene pool. American Journal of Human Genetics 59: 185-203

Richards, M. and V. Macaulay 2001 The mitochondrial gene tree comes of age. American Journal of Human Genetics 68: 1315-1320

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

Rogers, J., S. M. Witte, C. M. Kammere, J. E. Hixson and J. W. MacCluer 1995 Linkage mapping in *Papio* baboons: conservation of a syntenic group of six markers on human chromosome 1. Genomics 28: 251-254

Rokas, A., E. Ladoukakis and E. Zouros 2003 Animal mitochondrial DNA recombination revisited. Trends in Ecology and Evolution.18: 411-417

Rosel, P. E., S. C. France, J. Y. Wang and T. D. Kocher 1999 Genetic structure of harbour porpoise *Phocoena phocoena* populations in the northwest Atlantic based on mitochondrial and nuclear markers. Molecular Ecology 8: S41-S54

Ross, K. G. and L. Keller 1998 Genetic control of social organization in an ant. Proceedings of the National Academy of Sciences USA 95:14232-14237

Ross, K. G., D. D. Shoemaker, M. J. B. Krieger, C. J. DeHeer and L. Keller 1999 Assessing genetic structure with multiple classes of molecular markers: a case study involving the introduced Fire Ant *Solenopsis invicta*. Molecular Biology and Evolution 16: 525-543

Rubinsztein, D. C., W. Amos, J. Leggo, S. Goodburn, S. Jain, S-H. Li, R. L. Margolis, C. A. Boss and M. A. Ferguson-Smith 1995a Microsatellite evolution-evidence for directionality and variation in rate between species. Nature Genetics 10: 337-343

Rubinsztein, D. C., J. Leggo and W. Amos 1995b Microsatellites evolve more rapidly in humans than in chimpanzees. Genomics 30: 610-612

Ruiz, L. A, K. Nayar, D. B. Goldstein, J. M. Hebert, M. T. Seielstad, P. A. Underhill, A. A. Lin, M. W. Feldman and L. L. Cavalli-Sforza 1996 Geographic clustering of human Y-chromosome haplotypes. Annals of Human Genetics 60: 401-408

Russo, C. A. M., N. Takezaki and M. Nei 1996 Efficiencies of different genes and different tree-building methods in recovering a known vertebrate phylogeny. Molecular Biology and Evolution 13: 525-536

Ruvolo, M. 1988 Genetic evolution in the African guenons. p: 136 In: A primate radiation: evolutionary biology of the African guenons. Ed: F. Bourlier, J. P. Gautier-Hoin and J. Kingdon. The University Press, Cambridge

Ruvolo, M., T. R. Disotell, M. W. Allard, W. M. Brown and R. L. Honeycutt 1991 Resolution of the African hominoid trichotomy by use of a mitochondrial gene sequence. Proceedings of the National Academy of Sciences USA 88: 1570-1574

Sade, D. G. and R.W. Hildrech 1965 Notes on the green monkey (*Cercopithecus aethiops sabaeus*) on St. Kitts, West Indies. Caribbean Journal of Science 5: 67-81

Sainudiin, R., R. T. Durrett, C. F. Aquadro and R. Nielsen 2004 Microsatellite mutation models: insights from a comparison of humans and chimpanzees. Genetics 103: 383-396

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

Salem, A. H., F. M. Badr, M. F. Gaballah and S. Paabo 1996 The genetics of traditional living: Y-chromosomal and mitochondrial lineages in the Sinai Peninsula. American Journal of Human Genetics 59: 741-743

Samadi, S., F. Erard, A. Estoup and P. Jarne 1998 The influence of mutation, selection and reproductive systems on microsatellite variability: a simulation approach. Genetical Research 71: 213-222

Sambrook, J., E. F. Fritsch and T. Maniatis 1989 Molecular cloning: a laboratory manual. 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Sarich, V. M. and A. C. Wilson 1967 Immunological time scale for human evolution. Science 158: 1200-1203

Satta, Y., N. Toyohara, C. Ohtaka, Y. Tatsuno, T. K. Watanabe, E. T. Matsura, S. I. Chigusa and N. Takahata 1988 Dubious maternal inheritance of mitochondrial DNA in *D. simulans* and evolution of *D. mauritaiana*. Genetical Research 52: 1-6

Schierwater, B. and A. Ender 1993 Different thermostable DNA polymerases may amplify different RAPD products. Nucleic Acids Research 21: 4647-4648

Schilthuizen, M., J. Honda and R. Stouthamer 1998 Parthenogenesis-inducing *Wolbachia* in *Trichogramma kaykai* (Hymenoptera: Trichogrammatidae) originates from a single infection. Annals of the Entomological Society of America 91: 410-414

Schlötterer, C. 2000 Evolutionary dynamics of microsatellite DNA. Chromosoma 109: 365-371

Schlötterer, C. and D. Tautz 1992 Slippage of synthesis of simple sequence of DNA. Nucleic Acid Research 20: 211-215

Schlötterer, C., C. Vogl and D. Tautz 1997 Polymorphism and locus-specific effects on polymorphism at microsatellite loci in natural *Drosophila melanogaster* populations. Genetics 146: 309-320

Schlötterer, C., R. Ritter, B. Harr and G. Brem 1998 High mutation rate of a long microsatellite allele in *Drosophila melanogaster* provides evidence for allele specific mutation rates. Molecular Biology and Evolution15: 1269-1274

Schneider, H., I. Sampaio, M. L. Harada, C. M. L Barroso, M. P. C Schneider, J. Czelusniak and M. Goodman 1996 Molecular phylogeny of the New World monkeys (Platyrrhini, Primates) based on two unlinked nuclear genes: IRBP intron 1 and epsilon-globin sequences. American Journal of Physical Anthropology 100: 153-180

Schneider, S., D. Roessli and L. Excoffier 2000 Arlequin: A software for population genetics data analysis. Ver 2.0. Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva. URL: http://anthro.unige.ch/arlequin

Schomburgk, R. H. 1848 The history of Barbados. Longmans, London. Reprinted: Frank Cass Library of West Indian Studies No. 19, London, 1971

Schug, M. D., R. F. C. Mackay and C. F. Aquadro 1997 Low mutation rates of microsatellite loci in *Drosophila melanogaster*. Nature Genetics 15: 99-102

Schug, M. D., C. M. Hutter, M. A. F. Noor and C. F. Aquadro 1998a Mutation and evolution of microsatellites in *Drosophila melanogaster*. Genetica 103: 359-367

Schug, M. D., C. M. Hutter, K. A. Wetterstrand, M. S. Gaudette, T. F. C. Mackay and C. F. Aquadro 1998b The Mutation Rates of Di-, Tri- and Tetranucleotide Repeats in *Drosophila melanogaster*. Molecular Biology and Evolution 15: 1751-1760

Schwarze, S. R., C. M. Lee, S. S. Chung, E. B. Roecker, R. Weindruch and J. M. Aiken 1995 High levels of mitochondrial DNA deletions in skeletal muscle of old rhesus monkeys. Mechanisms of Ageing and Development 83: 91-101

Schwartz, M. and J. Vissing 2002 Paternal inheritance of mitochondrial DNA. The New England Journal of Medicine 347: 576-580

Seielstad, M. 2000 Asymmetries in the maternal and paternal genetic histories of Colombian populations. American Journal of Human Genetics 67: 1062-1066

Seielstad, M. T., J. M. Hebert, A. A. Lin, P. A. Underhill, M. Ibrahim, D. Vollrath, L. L. Cavalli-Sforza 1994 Construction of human Y-chromosomal haplotypes using a new polymorphic A to G transition. Human Molecular Genetics 3: 2159-2161

Seielstad, M. T., E. Minch and L. L. Cavalli-Sforza 1998 Genetic evidence for a higher female migration rate in humans. Nature Genetics 20: 278-280

Semino, O., G. Passarino, A. Brega, M. Fellous and A. S. Santachiara-Benerecetti 1996 A view of the Neolithic demic diffusion in Europe through two Y chromosome-specific markers. American Journal of Human Genetics 59: 964-968

Shapero, M. H., K. K. Leuther, A. Nguyen, M. Scott and K. W. Jones 2001 SNP genotyping by multiplexed solid-phase amplification and fluorescent minisequencing. Genome Research 11: 1926-1934

Shen, P., F. Wang, P. A. Underhill, C. Franco, W-H. Yang, A. Roxas, R. Sung, A. A. Lin, R. W. Hyman, D. Vollrath, R. W. Davis, L. L. Cavalli-Sforza and P. J. Oefner 2000 Population genetic implications from sequence variation in four Y chromosome genes. Proceedings of the National Academy of Sciences USA 97: 7354-7359

Shen, P., H. Tang, P. Passarino, B. Do, R. Sung, K. Davis, C. Scharfe, L. L. Cavalli-Sforza, M. W. Feldman and P. J. Oefner 2001 Parallels and contrasts in female and male human evolutionary history. American Journal of Human Genetics 69 [Suppl.]: 423

Sia, E. A., R. J. Kokoska, M. Dominska, P. Greenwell and T. D. Petes 1997 Microsatellite instability in yeast dependence on repeat unit size and DNA mismatch repair genes. Molecular and Cellular Biology 17: 2851-2858

Sibley, C. G. and J. E. Ahlquist 1987 DNA hybridization evidence of hominoid phylogeny: results from an expanded data set. Journal of Molecular Evolution 26: 99-121

Sigurðardóttir, S, A. Helgason, J. R. Gulcher, K. Stefansson and P. Donnelly 2000 The mutation rate in the human mtDNA control region. American Journal of Human Genetics 66: 1599-1609

Silva, I. M. M. S., J. Honda, F. van Kan, J. Hu, L. Neto, B. Pintureau and R. Stouthamer 1999 Molecular differentiation of five *Trichogramma* species occurring in Portugal. Biological Control: Theory and Applications in Pest Management 16: 177-184

Slatkin, M. and N. H. Barton 1989 A comparison of three indirect methods for estimating average levels of gene flow. Evolution 43: 1349-1368

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

Smith, P. H. 1979 Genetic manipulation of the circadian clock's timing of sexual behavior in the Queensland fruit flies *Dacus tryoni* and *Dacus neohumeralis*. Physiological Entomology 4: 71-78

Smith, K. L., S. C. Alberts, M. K. Bayes, M. W. Bruford, J. Altmann and C. Ober 2000 Cross-Species Amplification, non-invasive genotyping, and non-mendelian inheritance of human STRPs in savannah baboons. American Journal of Primatology 51: 219-227

Smith, N. G. C and A. E-Walker 2002 Adaptive protein evolution in *Drosophila.* Nature 415: 1022-1024

Song, S. D., J. Ma and J. M. Hughes 2006 Polymorphic microsatellite DNA markers in *Bactrocera cacuminata* (Hering) (Diptera: Tephritidae). Molecular Ecology Notes 6: 47-49

Stallings, R. L., A. F. Ford, D. Nelson, D. C. orney, C. E. Hildebrand and R. K. Moyzis 1991 Evolution and distribution of (GT)n repetitive sequences in mammalian genomes. Genomics 10: 807-815

Stallings, R. L. 1995 Conservation and revolution of (CT)n/(GA)n microsatellite sequence at orthologous positions in diverse mammalian genomes. Genomics 25: 107-113

Steiper, M. E., N. M. Young and T. Y. Sukarna 2004 Genomic data support the hominoid slowdown and an Early Oligocene estimate for the hominoid–sercopithecoid divergence. Proceedings of the National Academy of Sciences USA 101: 17021-17026

Stephens, M., J. S. Sloan, P. D. Robertson, P. Scheet and D. A. Nickerson 2006 Automating sequence-based detection and genotyping of single nucleotide polymorphisms (SNPs) from diploid samples. Nature Genetics 38: 375-381

Stordeur, E. D. 1997 Nonrandom partition of mitochondria in heteroplasmic *Drosophila*. Heredity 79: 615-623

Stouthamer, R. 1997 *Wolbachia*-induced parthenogenesis. In: Influential Passenger: Inherited Microorganisms and Arthropod Reproduction, p. 102-24. Ed: S. L. O'Neill, A. A. Hoffman, J. H. Werren, Oxford University Press, New York

Stouthamer, R., J. A. J. Breeuwer and G. D. D. Hurst 1999 *Wolbachia pipientis*: microbial manipulator of arthropod reproduction. Annual Review of Microbiology 53: 71-102

Struhsaker, T. T. 1967 Ecology of vervet monkeys (*Cercopithecus aethiops*) in the Masai-Amboseli Game Reserve, Kenya. Ecology 48: 891-904

Sumitania, M., D. S. Yamamotoa, J. M. Leeb and M. Hatakeyama 2005 Isolation of white gene orthologue of the sawfly, Athalia rosae (Hymenoptera) and its functional analysis using RNA interference. Insect Biochemistry and Molecular Biology 35: 231-240

Syvänen, A-C. 2001 Accessing genetic variation: genotyping single nucleotide polymorphisms. Nature Reviews Genetics 2: 930-942

Taberlet, P., J. J. Camarra, S. Griffin, E. Uhres, O. Hanotte, L. P. Waits, C. Dubois-Paganon, T. Burke and J. Bouvet 1997 Noninvasive genetic tracking of the endangered Pyrenean brown bear population. Molecular Ecology 6: 869-876

Takenaka, O., H. Takasaki, S. Kawamoto, M. Arakawa and A. Takenaka 1993 Polymorphic microsatellite DNA amplification customized for chimpanzee paternity testing. Primates 34: 27-35

Takezaki, N., A. Rzhesky and M. Nei 1995 Phylogenetic test of the molecular clock and linearized tree. Molecular Biology and Evolution 12: 823-833

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

Takezaki, N. and T. Gojobori 1999 Correct and incorrect vertebrate phylogenies obtained by the entire mitochondrial DNA sequences Molecular Biology and Evolution 16: 590-601

Tanaka, M., A. Ando, C. Renard, P. Chardon, M, Domukai, N. Okumura, T. Awata and H. Uenishi 2005 Development of dense microsatellite markers in the entire SLA region and evaluation of their polymorphisms in porcine breeds. Immunogenetics 57: 690-696

Tappen, N. 1960 Problems of distribution and adaptation of the African monkeys. Current Anthropology 1: 91-120

Tautz, D., M. Trick and G. A. Dover 1986 Cryptic simplicity in DNA is a major source of genetic variation. Nature 322: 652-656

Taylor, E. B., M. D. Stamford and J. S. Baxter 2003 Population subdivision in westslope cutthroat trout (*Oncorhynchus clarki lewisi*) at the northern periphery of its range: evolutionary inferences and conservation implications. Molecular Ecology 12: 2609-2622

Thomson, R., J. K. Pritchard, P. Shen, P. J. Oefner and M. W. Feldman 2000 Recent common ancestry of human Y chromosomes: Evidence from DNA sequence data. Proceedings of the National Academy of Sciences USA 97: 7360-7365

Tosi, A. J., J. C. Morales and D. J. Melnick 2000 Comparison of Y chromosome and mtDNA phylogenies leads to unique inferences of Macaque evolutionary history. Molecular Phylogenetic Evolution 17: 133-144

Tóth, G., Z. Gaspári and J. Jurke 2000 Microsatellite in different eukaryotic genomes: survey and analysis. Genome Research 10: 967-981

Tully, L. A., R. J. Steighner, T. J. Parsons, M. M. Holland, V. L. Prenger and M. A. Marino 1999 A high incidence of mitochondrial DNA heteroplasmy in hypervariable region 1 in normal human tissues: implications for forensic casework. Research paper of the US National Institute of Standards and Technology (NIST). http://www.promega.com/geneticidproc/ussymp9proc/content/18.pdf

Tully, L. A., T. J. Parsons, R. J. Steighner, M. M. Holland, M. A. Marino and V. L. Prenger 2000 A sensitive denaturing gradient-gel electrophoresis assay reveals a high frequency of heteroplasmy in hypervariable region 1 of the human mtDNA control region. American Journal of Human Genetics 67: 432-443

Tychsen, P. H. 1978 The effect of photoperiod on the circadian rhythm of mating responsiveness in the fruit fly, *Dacus tryoni* Physiological Entomology 3: 65-69

Tychsen, P. H. and B. S. Fletcher 1971 Studies on the rhythm of mating in the Queensland Fruit Fly, *Dacus tryoni*. Journal of Insect Physiology 17: 2139-2156

Tyler-Smith, C. 1999 Y-Chromosomal DNA markers. In: Genomic diversity: Applications in human population genetics. Ed: Papiha, S. S., R. Deka and R. Chakraborty. Chapter 5: 65-74. Kluwer Academic/Plenum Publishers, New York

Underhill, P. A., L. Jin, R. Zemans, P. J. Oefner and L. L. Cavalli-Sforza 1996 A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. Proceedings of the National Academy of Sciences USA 93: 196-200

Underhill, P. A., L. Jin, A. A. Lin, S. Q. Mehdi, T. Jenkins, D. Vollrath, R. W. Davis, L. L. Cavalli-Sforza and P. J. Oefner 1997 Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. Genome Research 7: 996-1005

Ursing, B. M. and U. Arnason 1998 The Complete Mitochondrial DNA Sequence of the Pig (*Sus scrofa*). Journal of Molecular Evolution 47: 302-306

van der Kuyl, A. C., C. L. Kuiken, J. T. Dekker and J. Goudsmit 1995 Phylogeny of African Monkeys based upon mitochondrial 12S rRNA sequences. Journal of Molecular Evolution 40: 173-180

van der Kuyl, A. C., J. T. Dekker and J. Goudsmit 1996 St. Kitts Green Monkeys originate from West Africa: Genetic evidence from feces. American Journal of Primatology 40: 361-364

Vaňková, D., L. Barto, D. Čížová-Schröffelová, F. Nešpor and O. Jandurová 2001 Mother–offspring bonding in farmed red deer: accuracy of visual observation verified by DNA analysis. Applied Animal Behaviour Science 73: 157-165

Vazquez, J. F., T. Perez, J. Albornoz and A. Dominguez 2000 Estimation of microsatellite mutation rates in *Drosophila melanogaster*. Genetical Research 76: 323-326

Vigilant, L., R. Pennington, H. Harpending, T. D. Kocher and A. C. Wilson 1989 Mitochondrial DNA sequences in single hairs from Southern African population. Proceedings of the National Academy of Sciences USA 86: 9350-9354

Vigilant, L., M. Hofreiter, H. Siedel and C. Boesch 2001 Paternity and relatedness in wild chimpanzee communities. Proceedings of the National Academy of Sciences USA 98: 12890-12895

Vignal, A., D. Milan, M. Sancristobal and A. Eggen 2002 A review on SNP and other types of molecular markers and their use in animal genetics. Genetics, Selection, Evolution 34: 275-305

Wakeley, J. 1993 Substitution rate variation among sites in hypervariable region I of human mitochondrial DNA. Journal of Molecular Evolution 37: 613-623

Wallace D. C., J. Ye, S. N. Neckelmann, G. Singh, K. A. Webster, B. D. Greenberg 1987 Sequence analysis of cDNAs for the human and bovine ATP synthase b subunit: mitochondrial DNA genes sustain seventeen times more mutation. Current Genetics 12: 81-90

Wallace, D. C. and A. Torroni 1992 American Indian prehistory as written in the mitochondrial DNA: a review. Human Biology: an International Record of Research 64: 403-416

Wallace, D. C. 1994 Mitochondrial DNA sequence variation in human evolution and disease. Proceedings of the National Academy of Sciences USA 91: 8739-8746

Walsh, P. S., D. A. Metzger and R. Higuchi 1991 Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. Biotechniques 10: 506-512

Wang, D. G., J. B. Fan, C. J. Siao, A. Berno, P. Young, R. Sapolsky, G. Ghandour, N. Perkins, E. Winchester, J. Spencer, L. Kruglyak, L. Stein, L. Hsie, T. Topaloglou, E. Hubbell, E. Robinson, M. Mittmann, M. S. Morris, N. Shen, D. Kilburn, J. Rioux, C. Nusbaum, S. Rozen, T. J. Hudson, R. Lipshutz, M. Chee and E. S. Lander 1998 Large-scale identification mapping and genotyping of single-nucleotide polymorphisms in the human genome. Science 280: 1077-1082

Wang, Y., H. Yu, K. Raphael and A. S. Gilchrist 2003 Genetic delineation of sibling species of the pest fruit fly *Bactrocera* (Diptera: Tephritidae) using microsatellites. Bulletin of Entomological Research 93: 351-360

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

Weber, J. L. and C. Wong 1993 Mutation of human short tandem repeats. Human Molecular Genetics 2: 1123–1128

Weir, B. S. 1996 Genetic data analysis II: methods for discrete population genetic data. 2nd ed. Sinauer Associates, Sunderland, Mass.

Weir, B. S. and C. C. Cockerham 1984 Estimating F-statistics for the analysis of population structure, Evolution, 38: 1358-1370

Wheeler, D. A., C. P. Kyriacou, M. L. Greenacre, Q. Yu, J. E. Rutila, M. Rosbash and J. C. Hall 1991 Molecular transfer of species-specific behaviour from *Drosophila simulans* to *Drosophila melanogaster*. Science 251: 1082-1085

Whittaker. J. C., R. M. Harbord, N. Boxall, I. Mackay, G. Dawson and R. M. Sibly 2003 Likelihood-based estimation of microsatellite mutation rates. Genetics 164: 781-787

Wierdl, M., M. Dominska and T. D. Petes 1997 Microsatellite instability in yeast: dependence on the length of the microsatellite. Genetics 146: 769-779

Williams, R. S. 2002 Another surprise from the mitochondrial genome. The New England Journal of Medicine 347: 609-612

Wilson, A. C., S. S. Carlsonm and T. J. White 1977 Biochemical evolution. Annual Review of Biochemistry 46: 573-639

Wise, C. A., M. Sraml, D. C. Rubinsztein and S. Easteal 1997 Comparative nuclear and mitochondrial genome diversity in humans and chimpanzees. Molecular Biology and Evolution 14: 707-716

Witte, S. M. and J. Rogers 1999 Microsatellite polymorphisms in bolivian squirrel monkeys (*Saimiri boliviensis*). American Journal of Primatology 47: 75-84

Wolda, H. 1967a Reproductive isolation between two closely related species of Queensland fruit fly, *Dacus tryoni* (Frogg.) and *D. neohumeralis* Hardy (Diptera: Tephritidae). I. Variation in humeral callus pattern and the occurrence of intermediate colour forms in the wild. Australian Journal of Zoology15: 501-513

Wolda, H. 1967b Reproductive isolation between two closely related species of Queensland fruit fly, *Dacus tryoni* (Frogg.) and *D. neohumeralis* Hardy (Diptera: Tephritidae). II. Genetic variation in humeral callus pattern in each species as compared with laboratory-bred hybrids. Australian Journal of Zoology 15: 515-539

Wolfheim, J. H. 1983 Primates of the world: distribution abundance and conservation. University of Washington, Press Seattle

Wright, S. 1951 The genetical structure of populations. Annals of Eugenics 15: 323-354

Wu, W., T. R. Schmidt, M. Goodmen and L. I. Grossmen 2000 Molecular evolution of cytochrome c oxidase subunit I in primates: Is there coevolution between mitochondrial and nuclear genomes? Molecular Phylogenetics and Evolution 17: 294-304

Wuytsa, W., R. Radersmab, K. Stormb and L. Vits 2005 An optimized dHPLC protocol for molecular testing of the EXT1 and EXT2 genes in hereditary multiple osteochondromas. Clinical Genetics 68: 542-547

Xu, X. and U. Arnason 1996a A complete sequence of the mitochondrial genome of the western lowland gorilla. Molecular Biology and Evolution 13: 691-698

Xu, X. and U. Arnason 1996b The mitochondrial DNA molecule of Sumatran orangutan and a molecular proposal for two (Bornean and Sumatran) species of orangutan. Journal of Molecular Evolution 43: 431-437

Xu, X., A. Gullberg and U. Arnason 1996 The complete mitochondrial DNA (mtDNA) of the donkey and mtDNA comparisons among four closely related mammalian species-pairs. Journal of Molecular Evolution 43: 438-446

Yauk, C. L. and J. S. Quinn 1996 Multilocus DNA fingerprinting reveals high rate of heritable genetic mutation in herring gulls nesting in an industrialized urban site. Proceedings of the National Academy of Sciences USA 93: 12137-12141

Yeh, H-M, M-C Tsai, Y-N Su, R-C Shen, J-J Hwang, W-Z Sun and L-P Lai 2005 Denaturing high performance liquid chromatography screening of ryanodine receptor type 1 gene in patients with malignant hyperthermia in taiwan and identification of a novel mutation (Y522C). Anesthesia and Analgesia 101: 1401-1406

Yokobori, S., T. Ueda, G. Feldmaier-Fuchs, S. Pääbo, R. Ueshima, A. Kondow, K. Nishikawa and K. Watanabe 1999 Complete DNA Sequence of the Mitochondrial Genome of the Ascidian *Halocynthia roretzi* (Chordata, Urochordata). Genetics 153: 1851-1862

Yu, H., M. K.Frommer, M. K. Robson, A. W. Meats, D. C. A. Shearman and J. A. Sved 2001 Microsatellite analysis of the Queensland fruit fly indicates spatial structuring: implications for population control. Bulletin of Entomological Research 91: 139-147

Zane, L., L. Bargelloni and T. Patarnello 2002 Strategies for microsatellite isolation: a review. Molecular Ecology 11: 1-16

Zardoya, R. and Meyer, A. 1996 Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. Molecular Biology and Evolution 13: 933-942

Zhang, P., H. Zhou, D. Liang, Y-F Liu, Y-Q Chen and L-H Qu 2005 The complete mitochondrial genome of a tree frog, *Polypedates megacephalus* (Amphibia: Anura: Rhacophoridae), and a novel gene organization in living amphibians. Gene 346: 133-143

Zhang, R., Z. Zhu, H. Zhu, T. Nguyen, F. Yao, K. Xia, D. Liang and C. Liu 2005 SNP Cutter: a comprehensive tool for SNP PCR–RFLP assay design. Nucleic Acids Research 33: W489-W492

Zhao, J. T., M. Frommer, J. A. Sved and A. Zacharopoulou 1998 Mitotic and polytene chromosome analyses in the Queensland fruit fly, *Bactrocera tryoni* (Diptera: Tephritidae). Genome 41: 510-526

Zhao, J. T., M. Frommer, J. A. Sved and C. B. Gilles 2003 Genetic and molecular markers of the Queensland fruit fly, *Bactrocera tryoni*. Journal of Heredity 49: 416-420

Zhivotovsky, L. A. and M. W. Feldman 1995 Microsatellite variability and genetic distances. Proceedings of the National Academy of Science USA 92: 11549-11552

Zhu, Y. C. and M. H. Greenstone 1999 Polymerase chain reaction techniques for distinguishing three species and two strains of *Aphelinus* (Hymenoptera: Aphelinidae) from *Diuraphis noxia* and *Schizaphis graminum* (Homoptera: Aphididae). Annals of the Entomological Society of America 92: 71-79

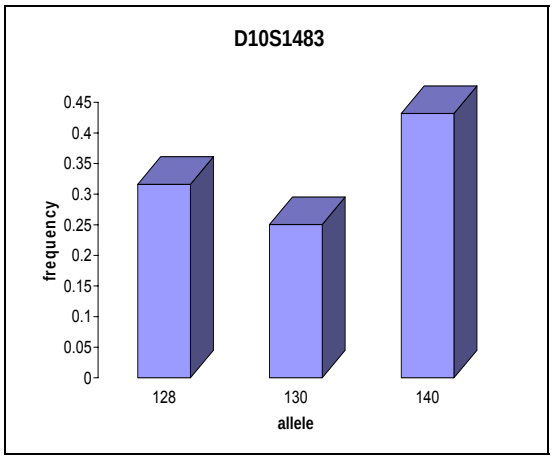
Zouros, E. and D. M. Rand 1998 Population genetics and evolution of animal mitochondrial DNA. In: Evolutionary Genetics: From Molecules to Morphology. Ed: Singh, R. S. and C. B. Krimbas Cambridge University Press Chapter 23

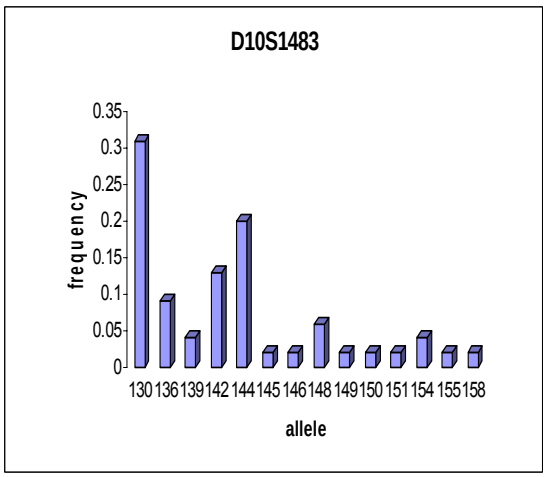
Publication

Wang, Y., Yu, H., Raphael, K. and Gilchrist, A.S. (2003) Genetic delineation of sibling species of the pest fruit fly *Bactrocera* (Diptera: Tephritidae) using microsatellites. *Bulletin of Entomological Research (2003)* ***93****, 351–360*

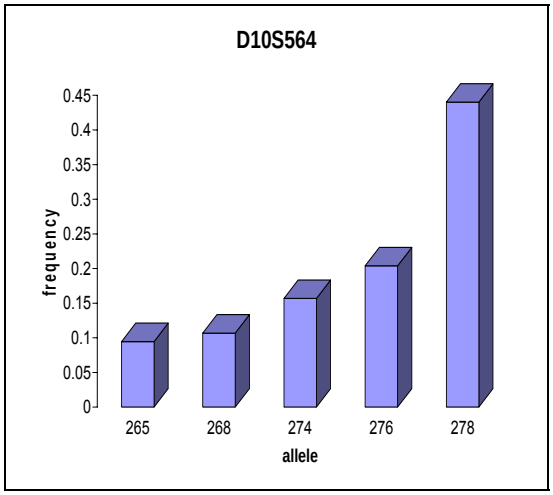
Appendix

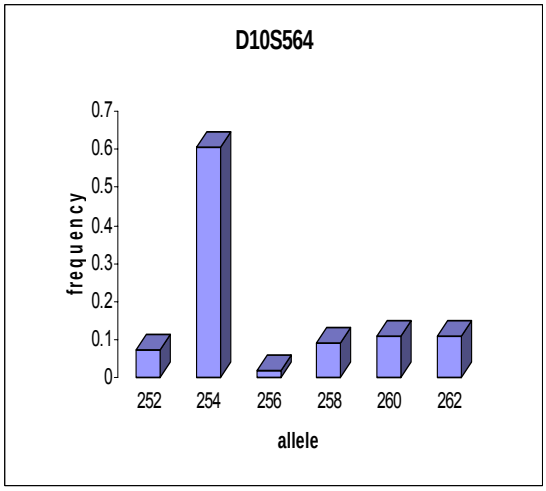
Appendix 1 Allele frequency distribution for the green monkey and human



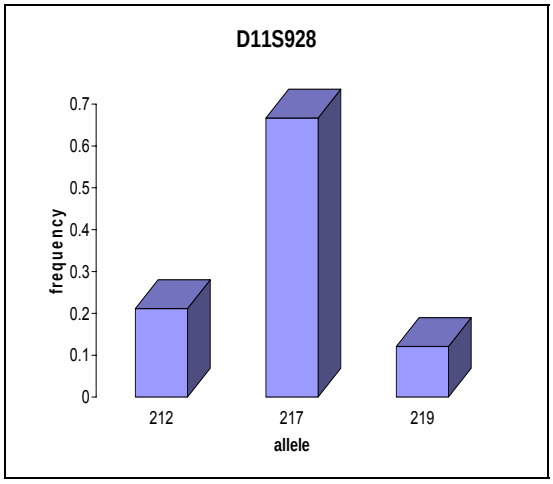


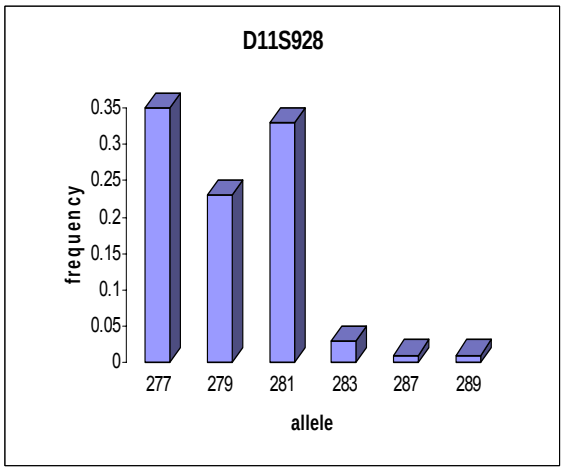
green monkey human



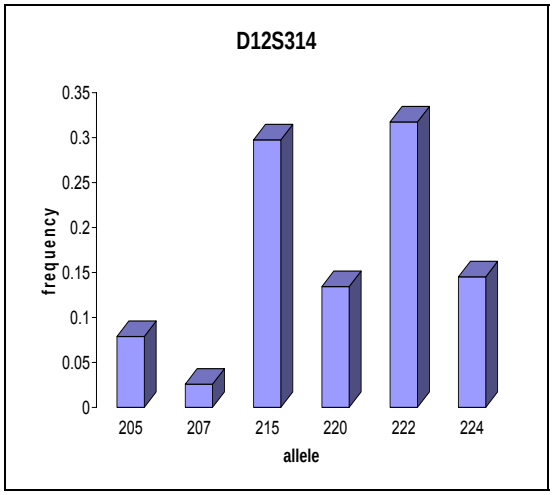


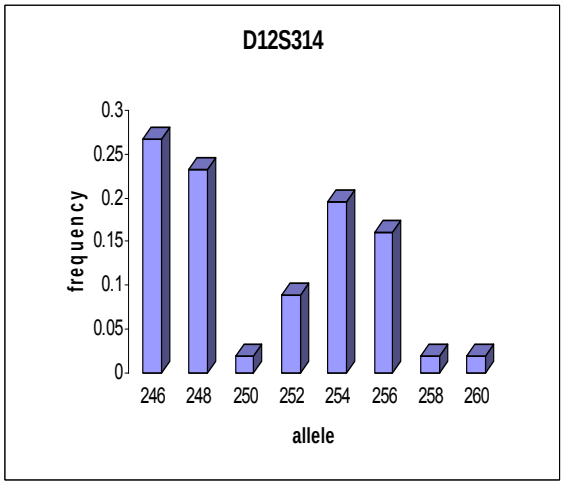
green monkey human



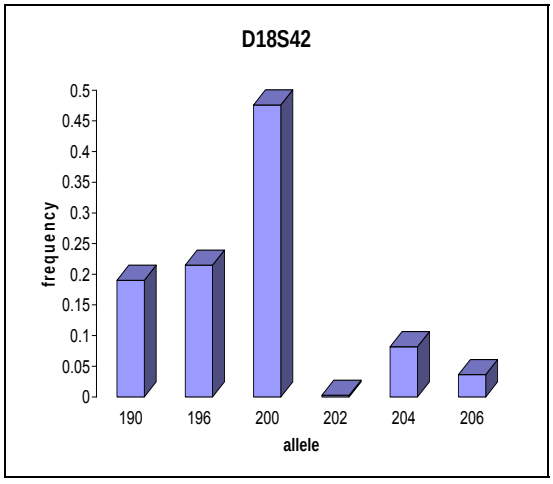


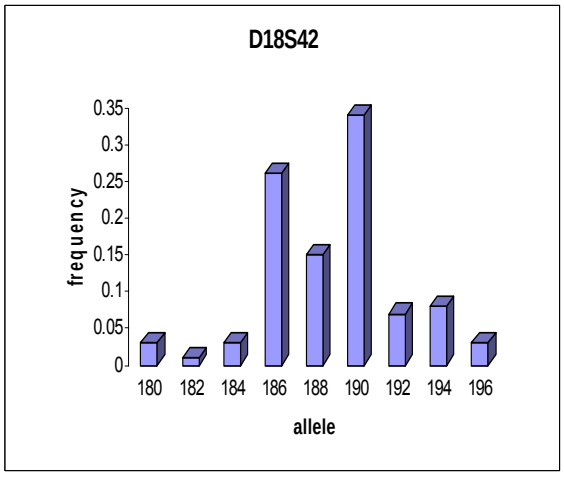
green monkey human



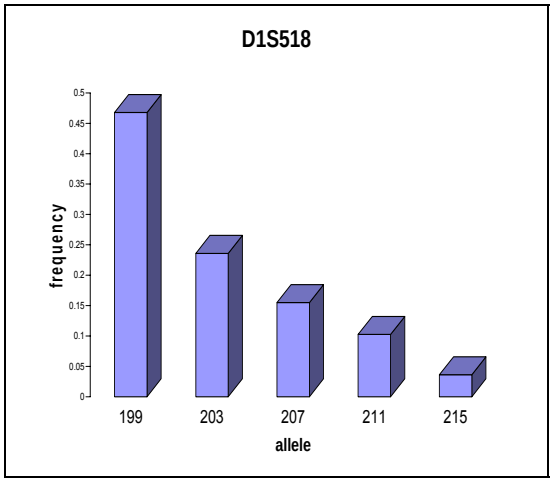


green monkey human

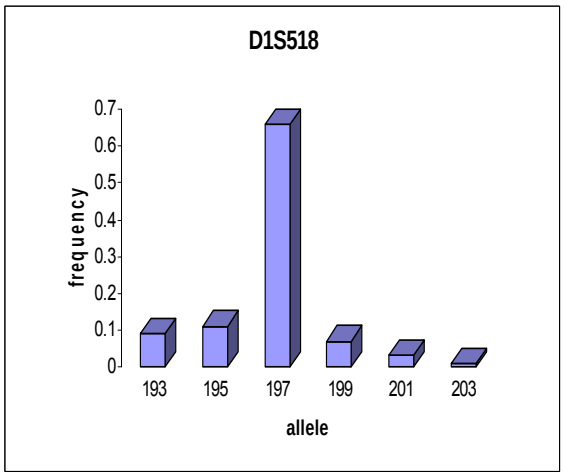




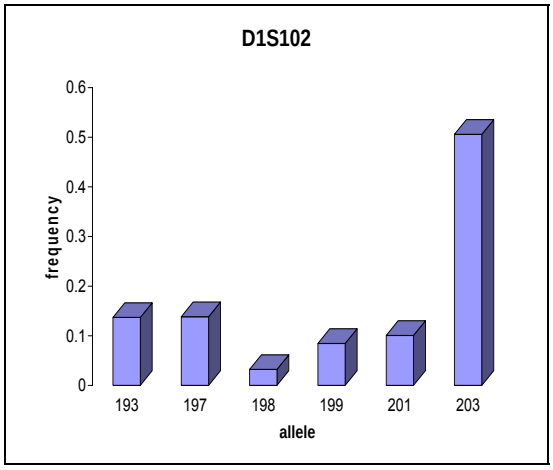
green monkey human



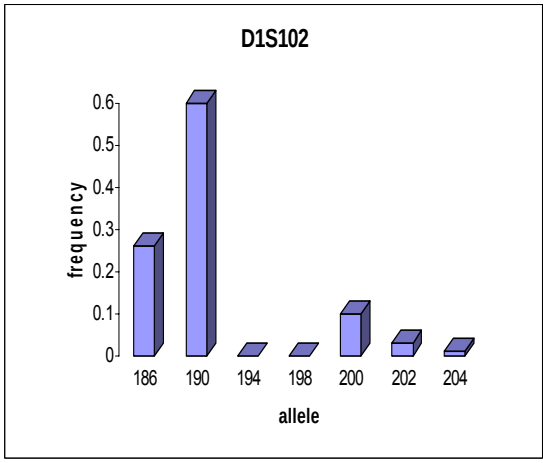
green monkey



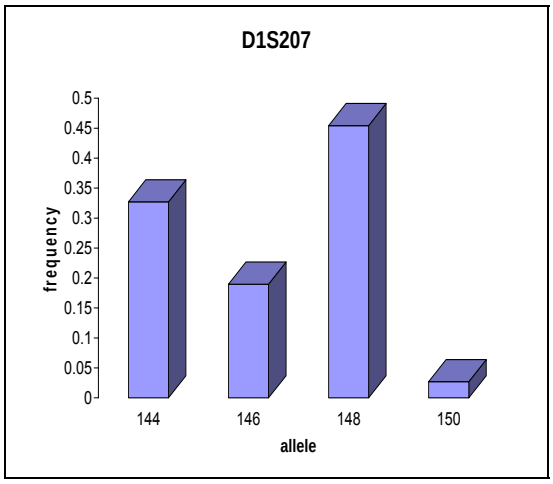
human



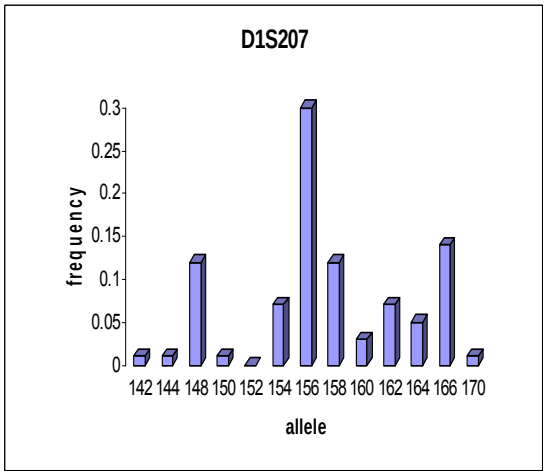
green monkey



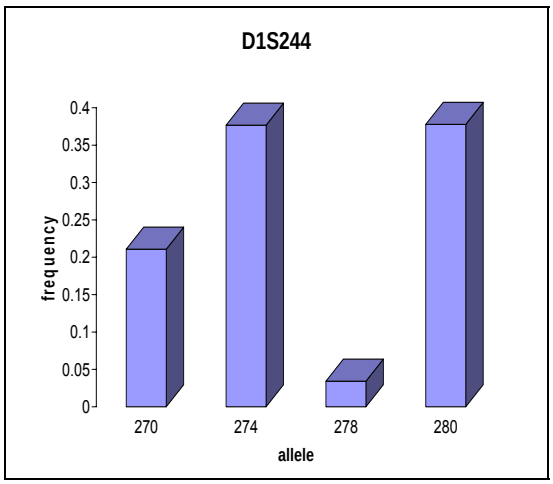
human

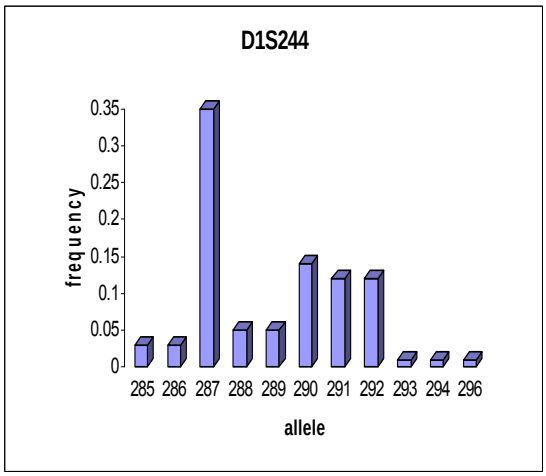


green monkey

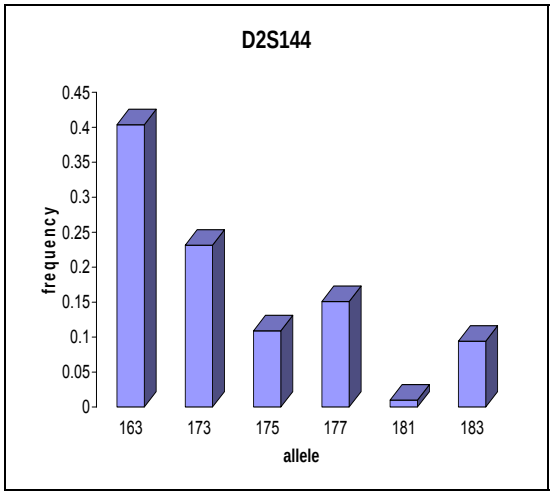


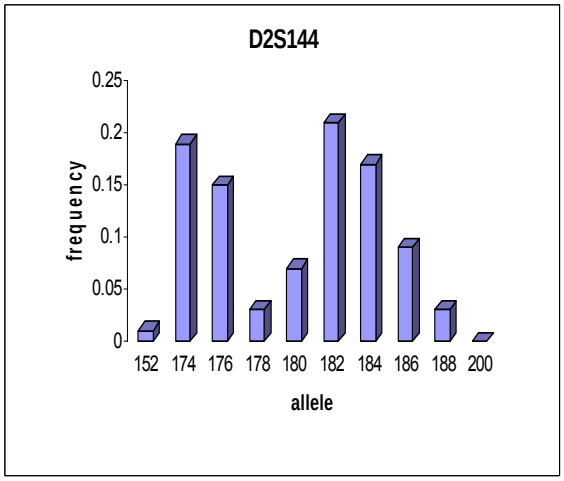
human



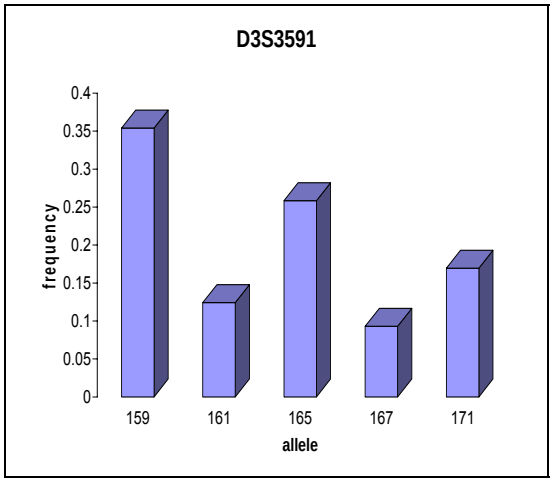


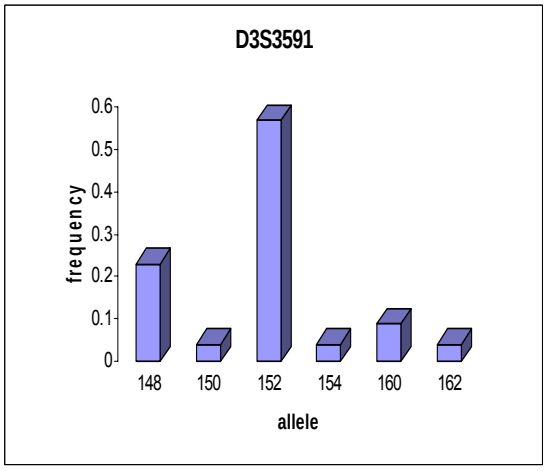
green monkey human



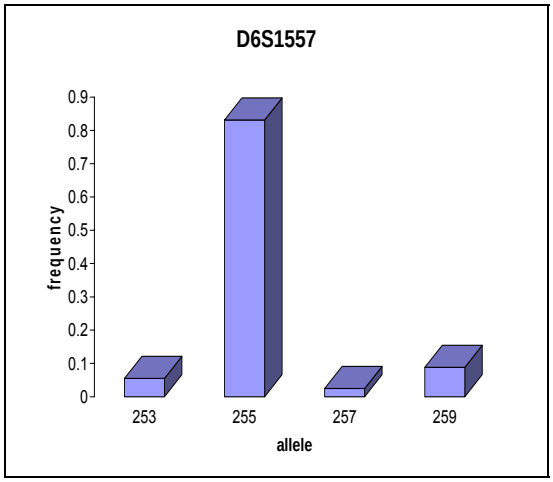


green monkey human

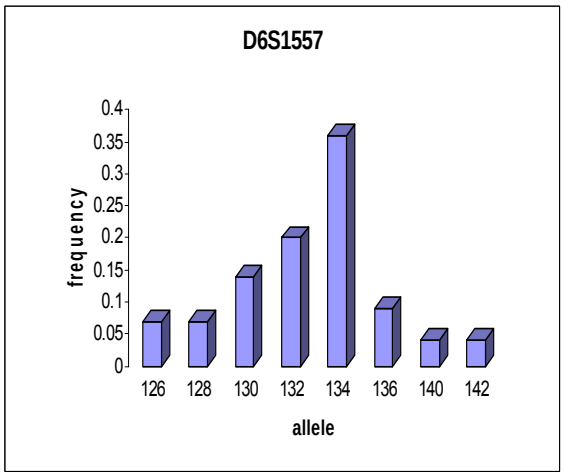




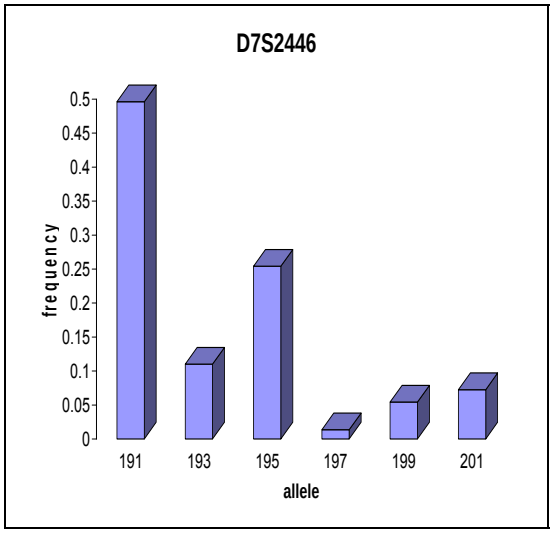
green monkey human



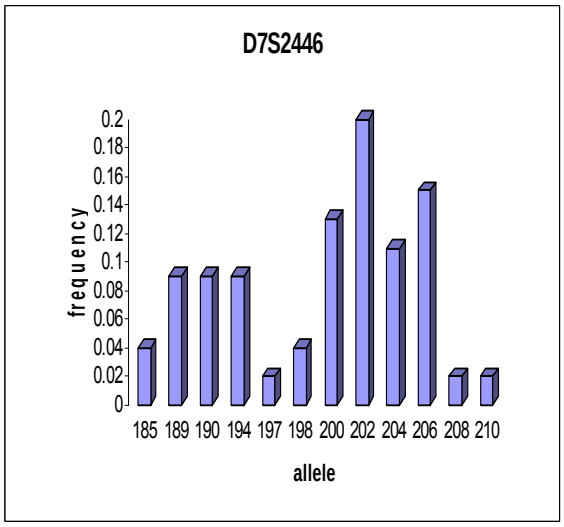
green monkey



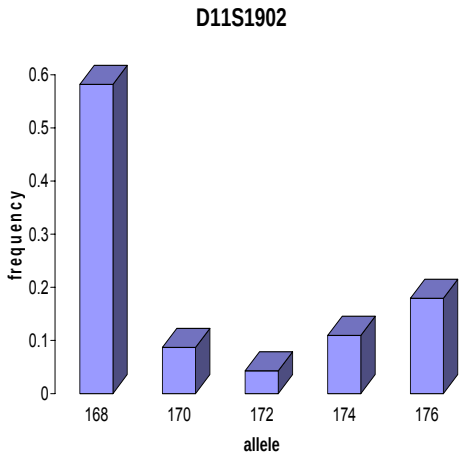
human



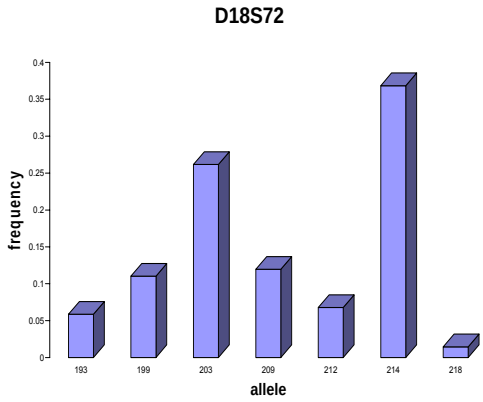
green monkey



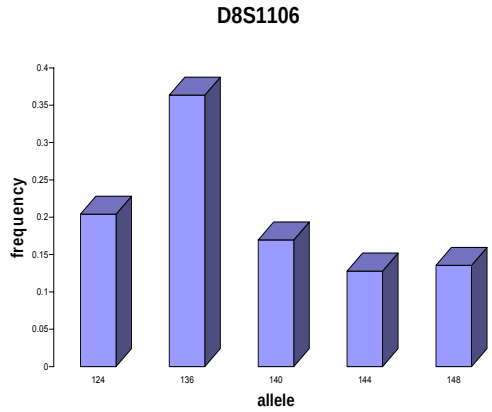
human



green monkey



green monkey



green monkey

Appendix 2 Some detailed B. tryoni microsatellite isolation methods

Genomic DNA extraction

Thirty (0.5 g) live flies were snap frozen and ground to powder in a pre-cooled mortar and pestle on liquid nitrogen or dry ice. The powder was quickly transferred to a microfuge tube and mixed with extraction buffer (100 mM Tris-HCl of pH 8, 100 mM NaCl, 50 mM EDTA and 200 mM sucrose, 0.5% sarkosyl), which was topped up to 2 ml. The tube was heated in a 60°C water bath for 15 minutes, with intermittent inversions, and cooled to 50°C. Proteinase K was added to a final concentration of 100 μg/ml, and the tube was incubated for 36-48 hours at 50°C with occasional stirring until the tube contents became red-brown and debris visibly decreased in size. Then the tube was cooled to 37°C before RNase was added to a final concentration of 100 μg/ml, then incubated at 37°C for 2 hours. An equal volume of TE-saturated phenol (equilibrated with 0.1% 8-hydroxyquinoline, 1, 0.1 and 0.01 M Tris-HCl, pH 8.0, and stored under $TE_{0.1}$ (10 mM Tris-HCl, 1.0mM EDTA, pH 8.0) in a light-proof bottle at 4°C.) was added, and the tube was mixed by gentle but complete inversion for 2 minutes, placed on ice for 1 minute, and then inverted for another 2 minutes. Phases were separated by centrifugation at 6500 rpm in a benchtop microfuge for 10 minutes. The upper aqueous layer was transferred into a fresh tube with a wide bore P1000 pipette tip, and extracted again with one volume of chloroform/isoamylalcohol (24:1). Phases were separated by centrifugation at 6500 rpm in a benchtop microfuge for 5 minute and the upper aqueous layers were transferred in a 15 ml Falcon tube and chilled on ice for 5 minutes. 1/10 volume of 3M NaOAC (pH 5.2) and 2.2 volumes of ice-cold 100% ethanol were added to precipitate high molecular weight DNA. The layer of the DNA, as a tangled mass 3-4 mm below the interface, was transferred into a fresh microfuge tube, and centrifuged for 10 minutes at 6500 rpm in a benchtop microfuge. The supernatant was removed，and the DNA pellet was washed with 1 ml ice-cold 70% ethanol and spun down to remove the supernatant. The pellet was air-dried and resuspended in 50-100 μl $TE_{0.1}$ (10 mM Tris-HCl and 0.1mM EDTA, pH 7.5) overnight at 4°C, or heated to 37°C for several hours to completely dissolve the DNA.

1 µl of the genomic DNA solution was loaded on a 0.7% agarose gel and electrophoresed against size standard to estimate DNA concentration.

DNA recovery by Gel purification

Excised gel slices with DNA fragments in size of 360 bp to 1500 bp were put into Falcon tubes after weighing. The volume of the agarose was estimated by weight (1 gm ≅ 1 ml). 1/2 volume TBE MELT™ (Bresatec Ltd., Thebarton, SA) and 4.5 volumes BRESA-SALT™ (Bresatec Ltd) was added into the tubes, incubated 5 minutes at 55°C to let the gel completely melt. The solution from the Falcon tube was divided into 2ml microfuge tubes. BRESA-BIND™ (Bresatec Ltd), the silica matrix ('glass milk'), was vortexed well before 5 µl of it was added to each microfuge tube. The tubes were incubated for 15 minutes at room temperature and mixed regularly. The complex within microfuge tubes from the same Falcon tube was pooled in one 2 ml microfuge tube and the tube was centrifuged for five seconds at 13,000 rpm in a benchtop microfuge. The supernatant was removed and the pellet was washed using ten to fifty volumes of ice-cold BRESA-WASH™ (Bresatec Ltd) solution, scraping the pellet with the 20 µl pipette tip, and pipetting several times. The tube was spun briefly and the supernatant was discarded. The pellet was resuspended in MilliQ water of approximately equal volume as the glass milk used. The tube was incubated for 5 minutes at 45-55°C, spun at 13,000 rpm in a benchtop microfuge for 1 minute to pellet the glass. The DNA supernatant was removed immediately and transferred in a 0.5 ml microfuge tube. The extract was collected in the same or a separate 0.5 ml microfuge tube and stored at 4°C. 1 µl purified size-selected DNA was electrophoresed on a 1% agarose gel against size standard to estimate DNA concentration.

Vector DNA preparation

Bacterial cells were picked from a freshly grown single colony and grown in 2 ml LB-amp media (1% tryptone, 0.5% yeast extract, 1% NaCl, pH 7.0, and 100 µg/ml ampicillin) at 37°C, in a shaking water bath for 2-4 hour at 120 rpm. This starting culture was added into 48 ml LB-amp media in a 150 ml conical flask, incubated at 37°C with

shaking at 150 rpm for at least 16 hours. The culture was centrifuged at 8000 rpm in a benchtop microfuge for 5 minutes and the supernatant removed. The pellet of cells was warmed to room temperature and resuspended in 4 ml buffer BPR (Bresatec Ltd.) containing RNase. 4 ml buffer BPL (Bresatec Ltd.) was added and incubated for 5 minutes at room temperature to perform alkaline lysis. Buffer BPN (Bresatec Ltd.) 4 ml at room temperature was added and centrifuged at 15,000 rpm in a benchtop microfuge for 10 minutes at 20°C. The supernatant was collected and applied to a pre-equilibrated column with 10 ml buffer BPQ (Bresatec Ltd.). The column was washed with 2x10 ml buffer BPW (Bresatec Ltd.) The DNA was eluted with 5 ml buffer BPE (Bresatec Ltd.), collected and precipitated using 3.5 ml of isopropanol. After centrifuging at a Beckman JA-17 rotor at 4°C, the supernatant was removed. The pellet of the DNA was washed using 2 ml 70% ethanol and re-centrifuged at 15,000 rpm in a benchtop microfuge at 4°C for 10 minutes. The supernatant was pipetted off. The DNA pellet was air-dried and redissolved in 150 µl TE0.1 buffer. The DNA solution was divided into aliquots of 0.5 ml and stored at -20°C.

Vector DNA was electrophoresed against size standard to estimate concentration.

Dephosphorylation of vector DNA

100 µl of reaction mixture containing 0.75 unit CIP (10,000 U/ml, New England Biolabs), 45 µl of each digested vector DNA, 10 µl 10x NEB buffer2 (New England Biolabs), and the appropriate volume of MilliQ water, was incubated at 37°C for 1 hour. 2 µl of 0.5 M EDTA (pH 8.0) was added to a final concentration of 5 mM and the tube was heated to 65°C for one hour to inactivate the enzyme. The resulting phosphatased linear DNA was extracted once with an equal volume of equilibrated phenol and once with an equal volume of a 1:1 mixture of phenol and chloroform/isoamylalcohol (24:1). Each time the mixture was centrifuged for 5 minutes at room temperature at 13,000 rpm in a benchtop microfuge. The upper aqueous layer was transferred to a fresh tube. The DNA was precipitated with 0.1 volume of 3 M sodium acetate (pH 7.0) and 2.5 volume of ice-cold 100% ethanol and centrifuged at 13,000 rpm in a benchtop microfuge for 10 minutes. The supernatant was carefully removed and the DNA pellet was washed with

100 µl of ice-cold 70% ethanol. The tube was centrifuged at 13,000 rpm in a benchtop microfuge for another 10 minutes and the supernatant was removed without disturbing the pellet. The DNA was air-dried and redissolved in 20 µl TE0.1 at an approximate concentration of 250 ng/µl.

The concentration of the dephosphorylated vector DNA was estimated by comparing its intensity to that of size standard on an electrophoresis gel.

Colony lifts

A dry membrane disc was carefully placed on agar with the label side down and an arrow marker at the plate orientation site. The whole membrane was spread on the agar and let sit for 2 minutes. Three holes were stabbed with a sterile needle through the membrane and the agar at different positions around the edge of the plate. The location of the holes and orientation was marked with a permanent pen at the bottom of the plate. The membrane was gently removed using forceps, put on a tray containing a filter paper pre-soaked with ~20 ml denaturing solution (0.5 N NaOH, 1.5 M NaCl) for 2 minutes, with the colony side face up. Thereafter the membrane was transferred on two trays, each containing a filter paper pre-soaked with ~20 ml neutralizing solution (1 M Tris-Cl, pH 7.6 and 1.5 M NaCl), with the colony side face up. The membrane was left in each tray for 2 minutes. The membrane disc was transferred to a large dry filter paper to remove excess buffer and air dried completely with the colony side up.

The second membrane disc was processed in the same way as described for the first membrane, except that the membrane was transferred onto a fresh (14 cm diameter) LB-AXI plate with colony side up and kept in a 37°C oven for 3-4 hours before the steps of denature, neutralize and air-dry.

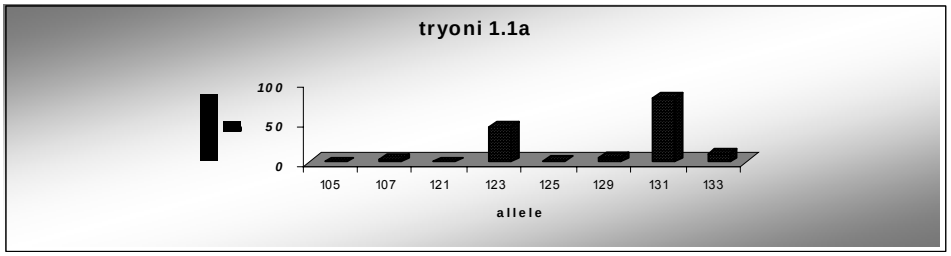
Plasmid DNA preparation

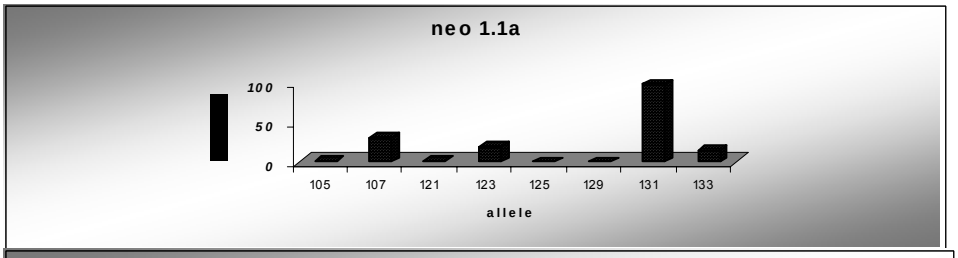
A single clone was picked up from each positive colony of the master plate, and streaked onto an LB-AXI plate using sterile tooth picks The streaked LB-AXI plates were grown

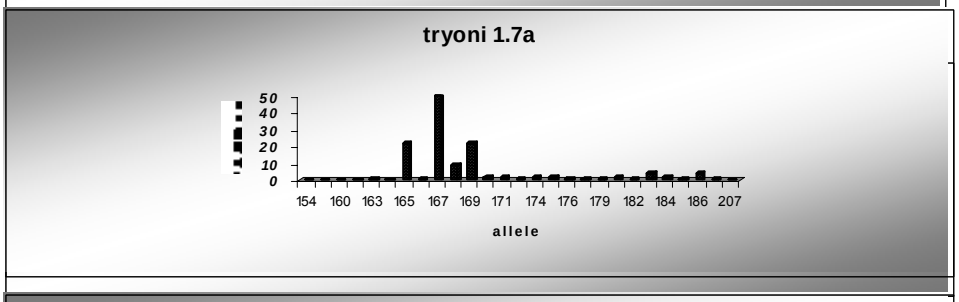
at 37°C for 16 hours. Individual white colonies were selected to inoculate test tubes containing 3 ml LB-amp media. The test tubes were incubated at 37°C with vigorous shaking at 220 rpm for 16 hours. 2 ml of overnight culture from each tube was transferred to a 2 ml microfuge tube and centrifuged at 13,000 rpm in a benchtop microfuge at room temperature for 30 seconds. The medium was removed and the pellet was resuspended in 100 µl ice-cold resuspension buffer (50mM glucose, 25 mM Tris-HCl and 10 mM EDTA, pH 8.0). The tube was left at room temperature for 2 minutes before the addition of 200 µl freshly prepared lysis solution (0.2M NaOH, 1% SDS). Neutralization solution (5 M potassium acetate and 2 M glacial acetic acid) 150 µl was added after the tube was chilled on ice for 2 minutes. The tube was vortexed, chilled on ice and centrifuged at 13,000 rpm in a benchtop microfuge for 5 minutes. 420 µl of supernatant was transferred from each tube to a fresh 1.5ml microfuge tube, without removing any of the precipitated high molecular weight material, extracted with an equal volume of a 1:1 mixture of phenol and chloroform/isoamylalcohol (24:1), vortexed for 10 seconds, centrifuged at 13,000 rpm in a benchtop microfuge for 2 minutes, and extracted again with an equal volume of chloroform/isoamylalcohol (24:1). DNA 400 µl was transferred from the upper aqueous layer to another 1.5ml fresh microfuge tube and precipitated with 2 volumes ice-cold 100% ethanol. The tube was vortexed and left at room temperature for 2 minutes before centrifuging at 13,000 rpm in a benchtop microfuge for 5 minutes. The resulting pellet was washed with ice-cold 70% ethanol and dried briefly in a vacuum desiccator (HETOVAC). The pellet was resuspended in 50 µl $TE_{0.1}$ containing 100 µg/ml RNase (Boehringer Mannheim) and incubated at 37°C for 60 minutes and stored at -20°C.

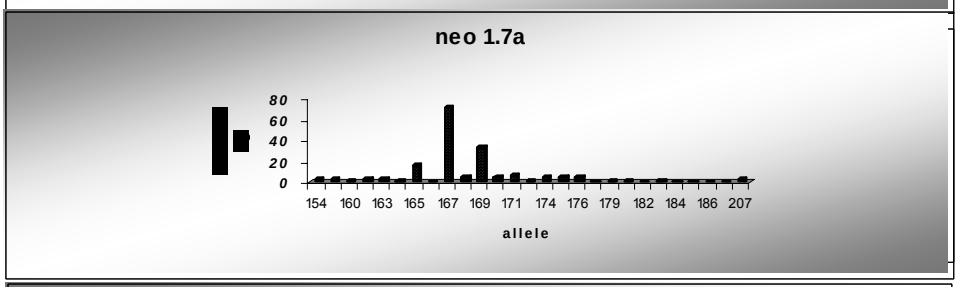
Of the remaining culture, 700 µl of each tube was taken into a 2 ml polypropylene tube with a screw cap. Autoclaved 50% glycerol 300 µl was added. The tube was mixed well by vortex, snap frozen in liquid nitrogen, and stored at -70°C.

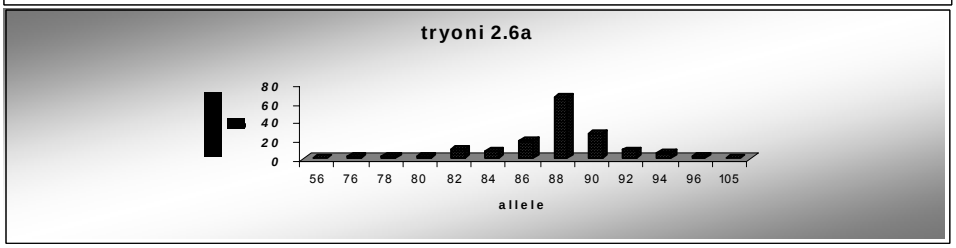
Appendix 3 Allele frequency distributions for the 22 loci in B. tryoni and B. neohumeralis

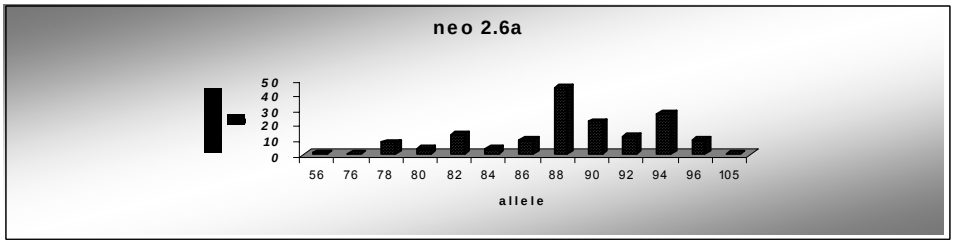


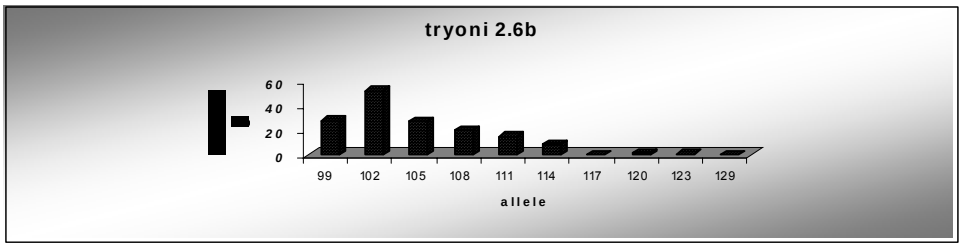


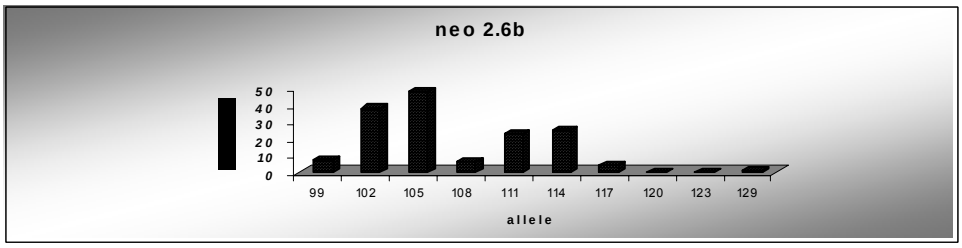


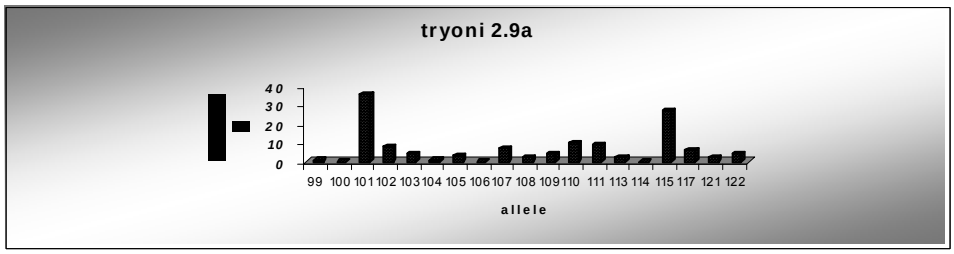


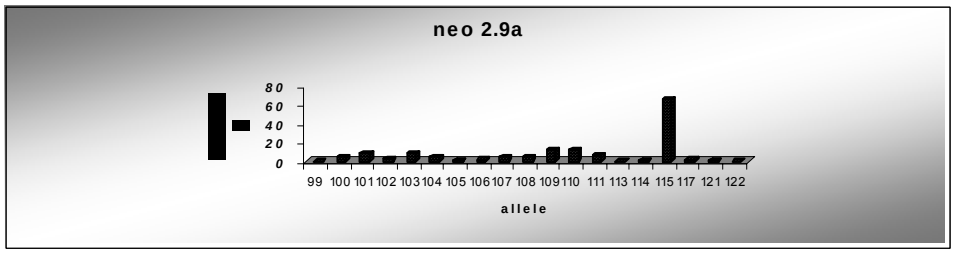


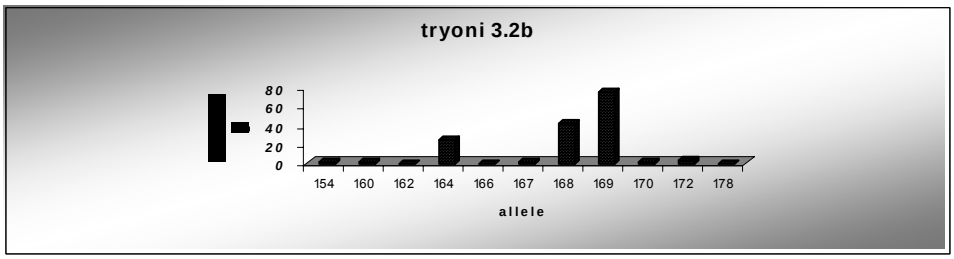


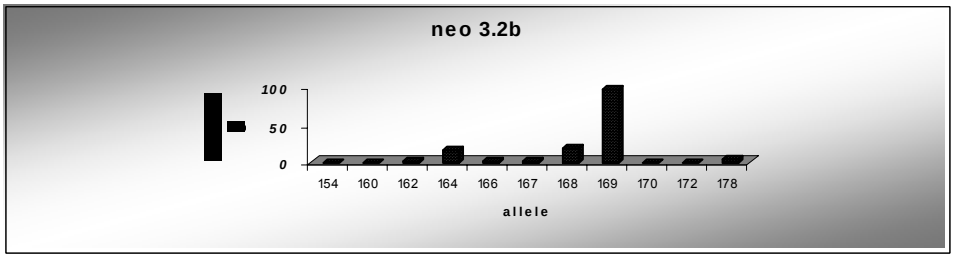

tryoni 2.6b
60
40
20
0
99 102 105 108 111 114 117 120 123 129
allele

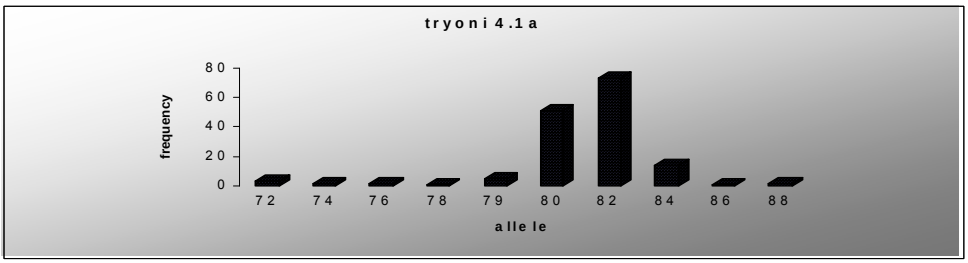

neo 2.6b
50
40
30
20
10
0
99 102 105 108 111 114 117 120 123 129
allele

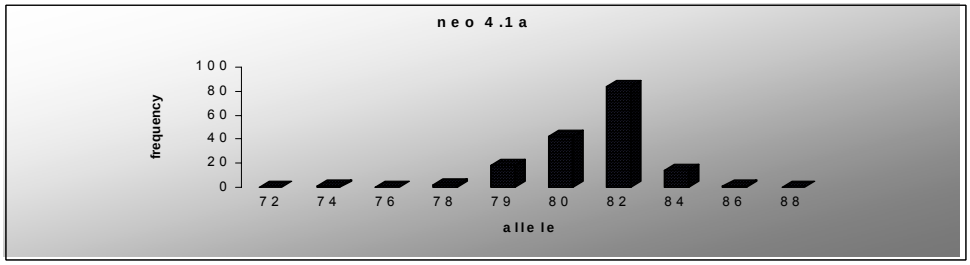

tryoni 2.9a
40
30
20
10
0
99 100 101 102 103 104 105 106 107 108 109 110 111 113 114 115 117 121 122
allele

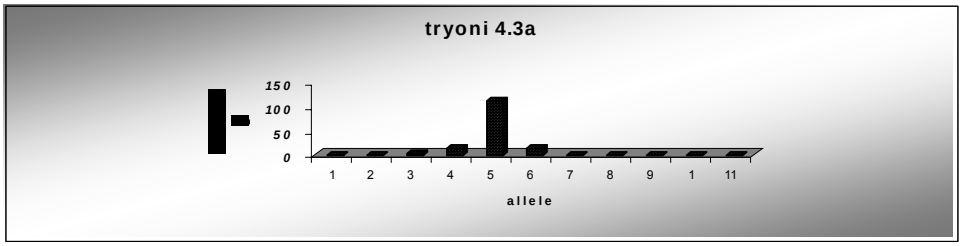

neo 2.9a
80
60
40
20
0
99 100 101 102 103 104 105 106 107 108 109 110 111 113 114 115 117 121 122
allele

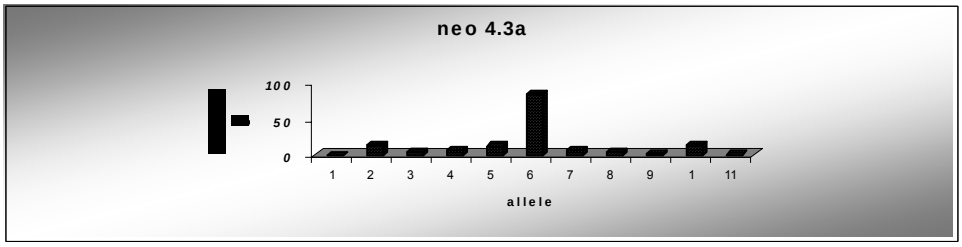

tryoni 3.2b
80
60
40
20
0
154 160 162 164 166 167 168 169 170 172 178
allele

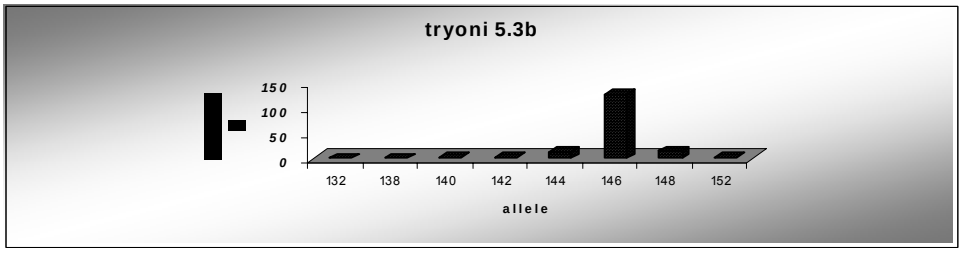

neo 3.2b
100
50
0
154 160 162 164 166 167 168 169 170 172 178
allele

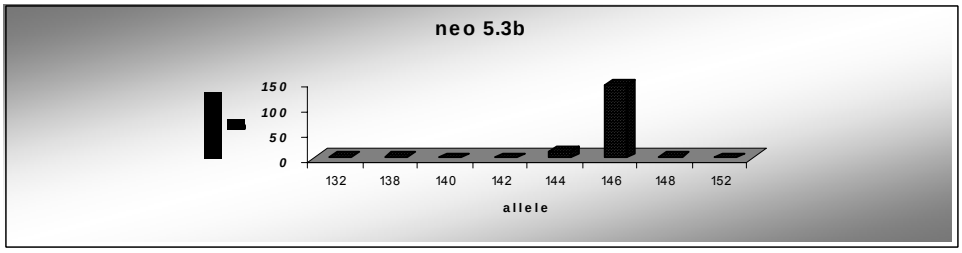

tryoni 4.1a
frequency
80
60
40
20
0
72 74 76 78 79 80 82 84 86 88
allele

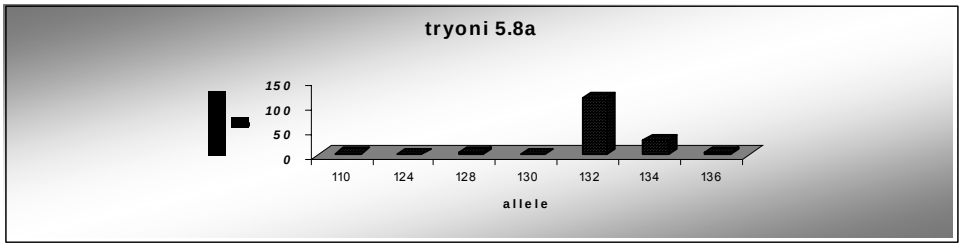

neo 4.1a
frequency
100
80
60
40
20
0
72 74 76 78 79 80 82 84 86 88
allele

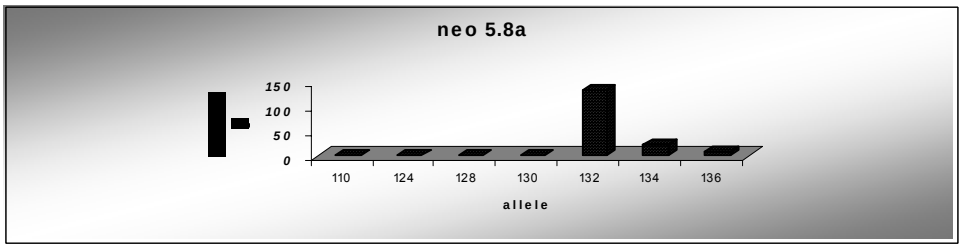

tryoni 4.3a
150
100
50
0
1 2 3 4 5 6 7 8 9 1 11
allele

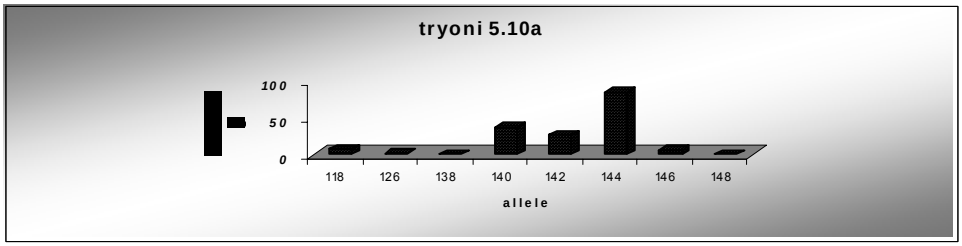

neo 4.3a
100
50
0
1 2 3 4 5 6 7 8 9 1 11
allele

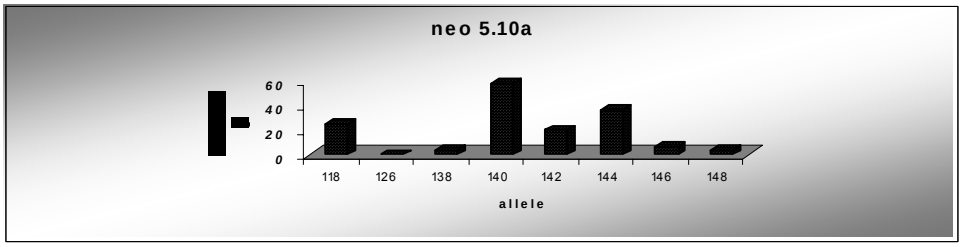

tryoni 5.3b
150
100
50
0
132 138 140 142 144 146 148 152
allele

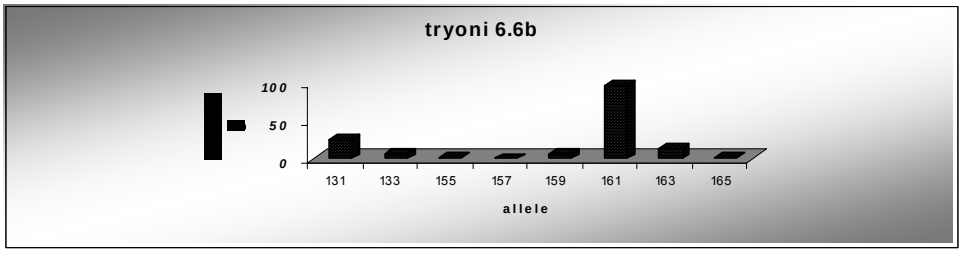

neo 5.3b
150
100
50
0
132 138 140 142 144 146 148 152
allele

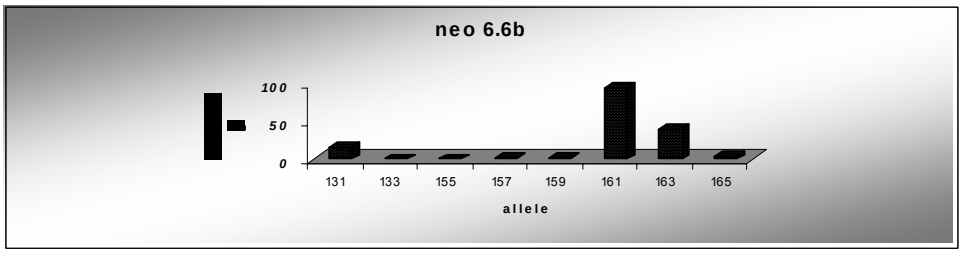

tryoni 5.8a
150
100
50
0
110
124
128
130
132
134
136
allele

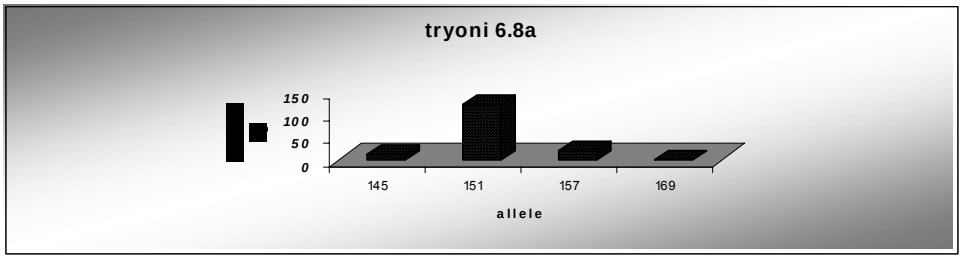

neo 5.8a
150
100
50
0
110
124
128
130
132
134
136
allele

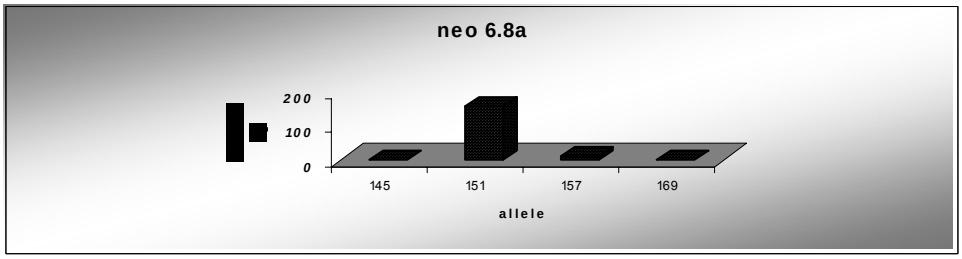

tryoni 5.10a
100
50
0
118
126
138
140
142
144
146
148
allele

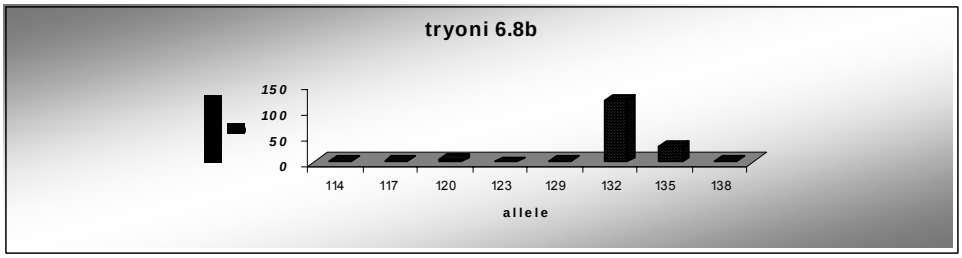

neo 5.10a
60
40
20
0
118
126
138
140
142
144
146
148
allele

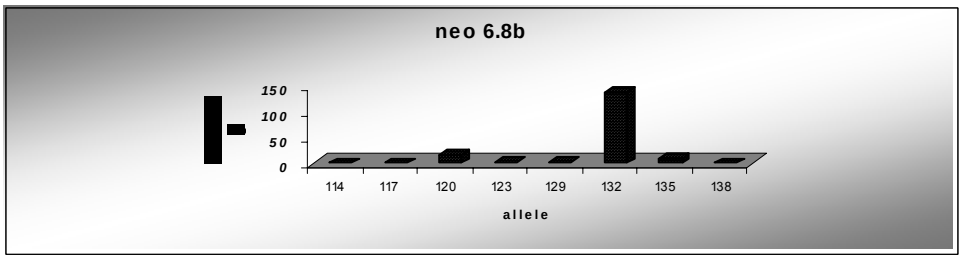

tryoni 6.6b
100
50
0
131
133
155
157
159
161
163
165
allele

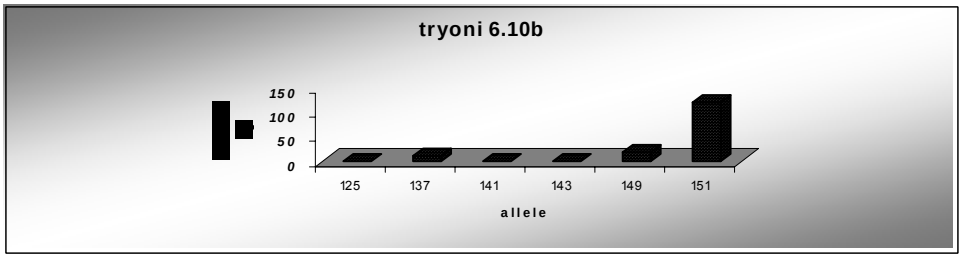

neo 6.6b
100
50
0
131
133
155
157
159
161
163
165
allele

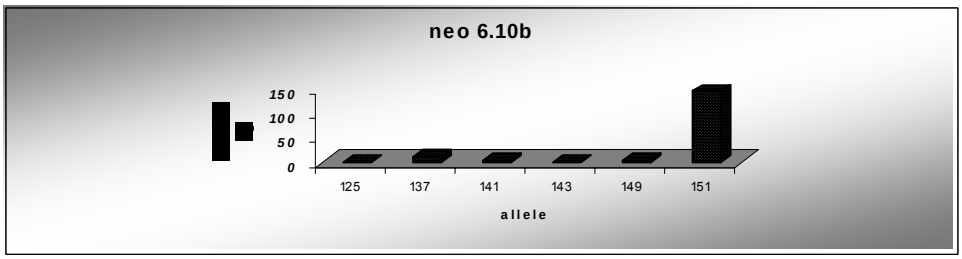

tryoni 6.8a
150
100
50
0
145
151
157
169
allele

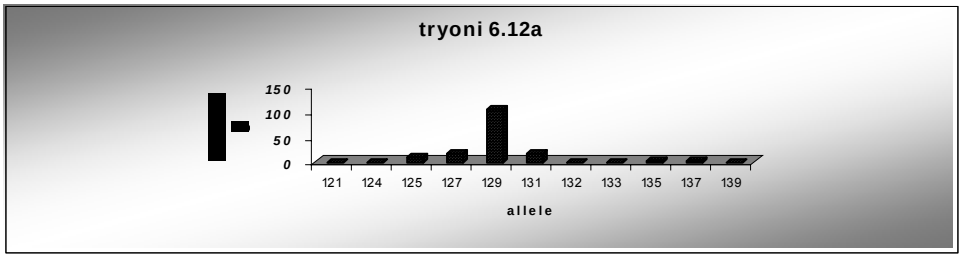

neo 6.8a
200
100
0
145
151
157
169
allele

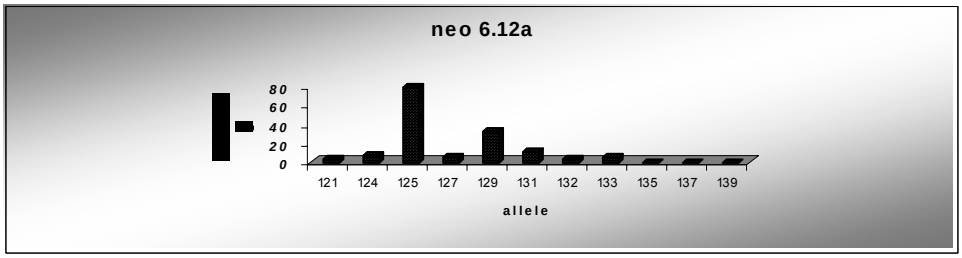

tryoni 6.8b
150
100
50
0
114
117
120
123
129
132
135
138
allele

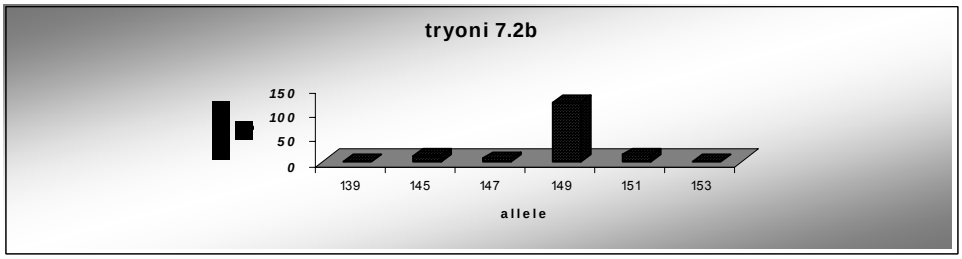

neo 6.8b
150
100
50
0
114
117
120
123
129
132
135
138
allele

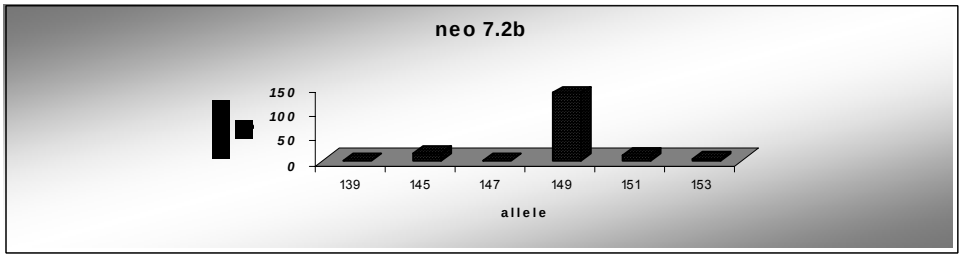

tryoni 6.10b
150
100
50
0
125
137
141
143
149
151
allele

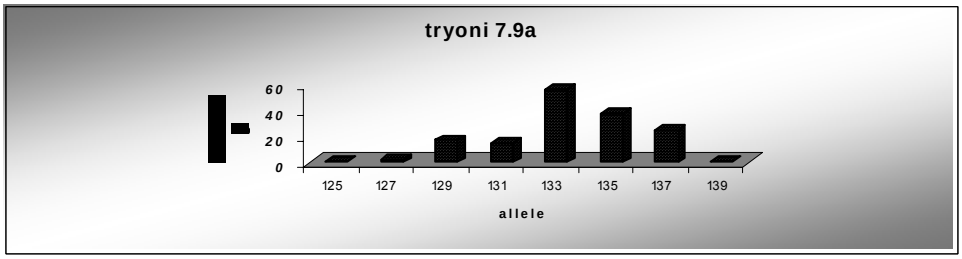

neo 6.10b
150
100
50
0
125
137
141
143
149
151
allele

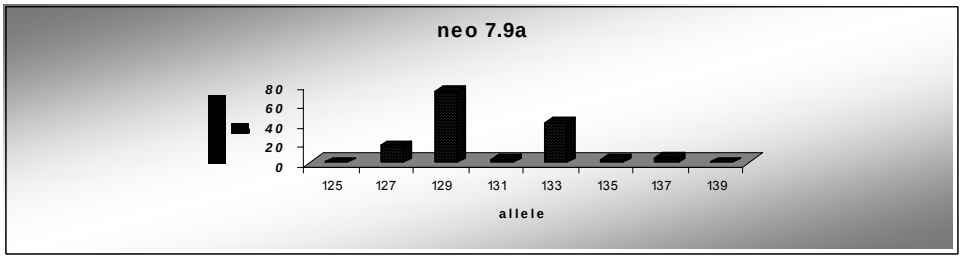

tryoni 6.12a
150
100
50
0
121 124 125 127 129 131 132 133 135 137 139
allele

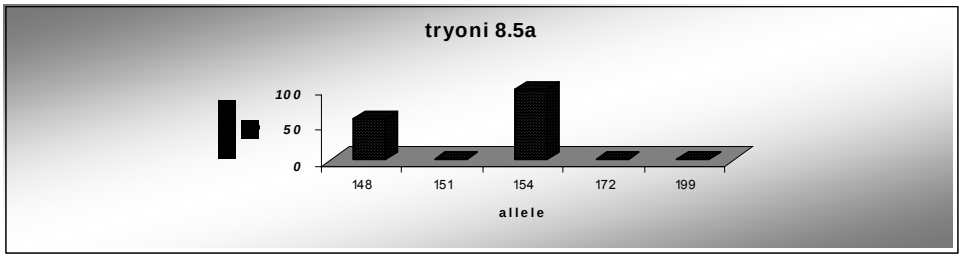

neo 6.12a
80
60
40
20
0
121 124 125 127 129 131 132 133 135 137 139
allele

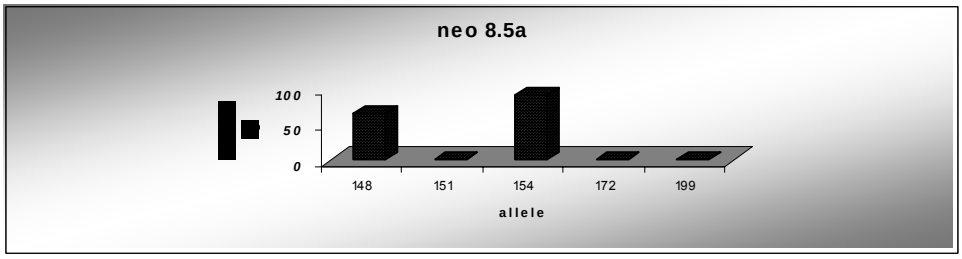

tryoni 7.2b
150
100
50
0
139 145 147 149 151 153
allele

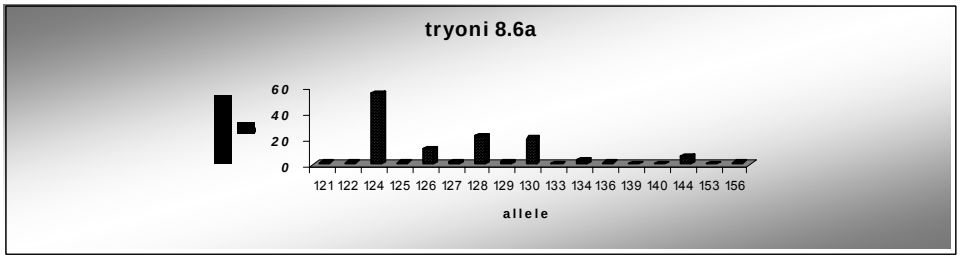

neo 7.2b
150
100
50
0
139 145 147 149 151 153
allele

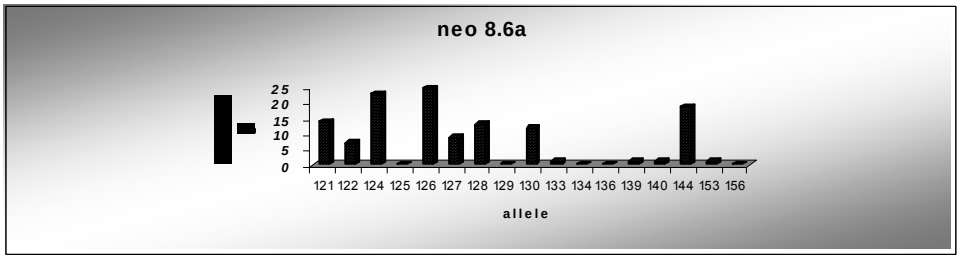

tryoni 7.9a
60
40
20
0
125 127 129 131 133 135 137 139
allele

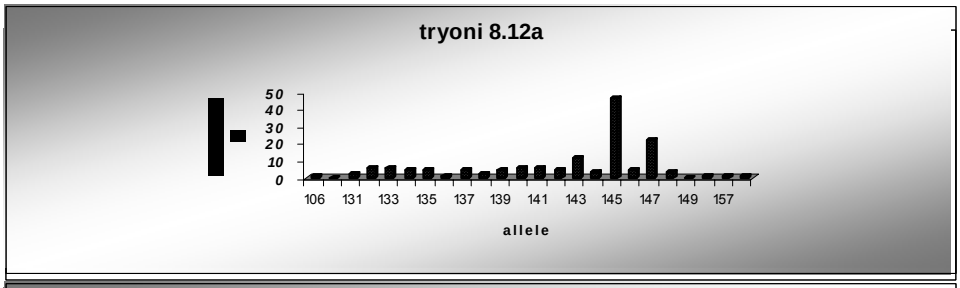

neo 7.9a
80
60
40
20
0
125 127 129 131 133 135 137 139
allele

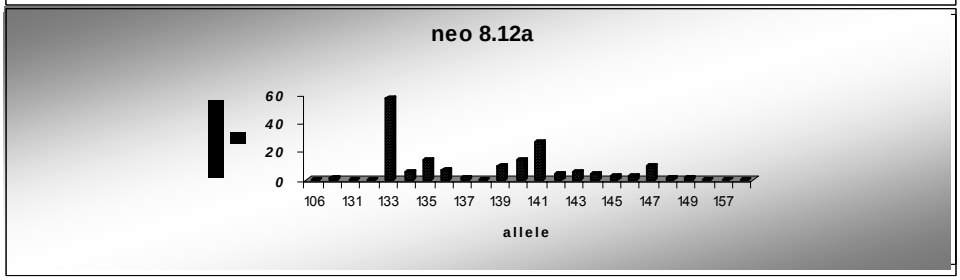

tryoni 8.5a
100
50
0
148
151
154
172
199
allele

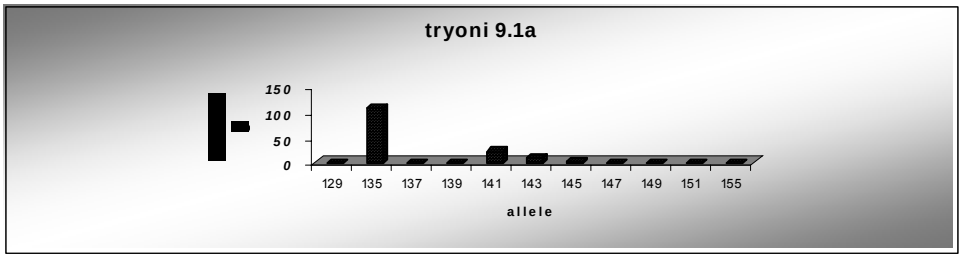

neo 8.5a
100
50
0
148
151
154
172
199
allele

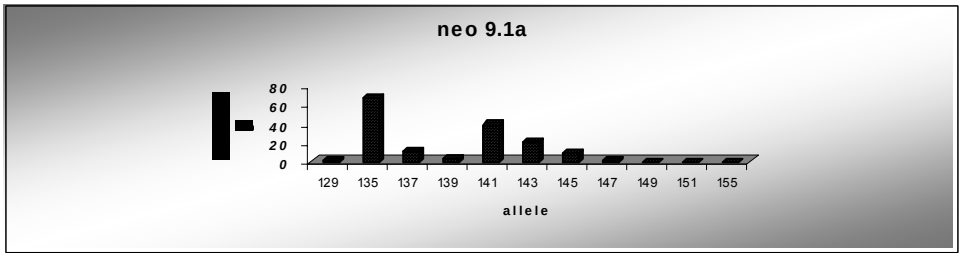

tryoni 8.6a
60
40
20
0
121 122 124 125 126 127 128 129 130 133 134 136 139 140 144 153 156
allele


neo 8.6a
25
20
15
10
5
0
121 122 124 125 126 127 128 129 130 133 134 136 139 140 144 153 156
allele


tryoni 8.12a
50
40
30
20
10
0
106 131 133 135 137 139 141 143 145 147 149 157
allele


neo 8.12a
60
40
20
0
106 131 133 135 137 139 141 143 145 147 149 157
allele


tryoni 9.1a
150
100
50
0
129 135 137 139 141 143 145 147 149 151 155
allele


neo 9.1a
80
60
40
20
0
129 135 137 139 141 143 145 147 149 151 155
allele