GENETIC CHARACTERIZATION OF HONEYBEE (*Apis mellifera*) IN COMOROS ISLANDS USING MITOCHONDRIAL AND MICROSATELLITE MARKERS

Okwaro Louis Allan (BSc.)
I56/34038/2015

A Thesis Submitted in Partial Fulfillment of the Requirements for the Award of the Degree of Master of Science (Biotechnology) in the School of Pure and Applied Sciences of Kenyatta University

June, 2019

ii

DECLARATION

This thesis is my original work and has not been presented for a degree in any other University or other award.

Signature.....**Date**.27/06/2019.....

Louis Allan Okwaro
Department of Biochemistry, Microbiology and Biotechnology,
Kenyatta University, Nairobi, Kenya

Supervisors

We confirm that the work reported in this thesis was carried out by the candidate under our supervision.

Signature.....**Date**.....

Dr Steven Runo
Department of Biochemistry, Microbiology and Biotechnology
Kenyatta University, Nairobi, Kenya

Signature.....**Date**.27/6/2019.....

Dr Elliud Muli
International Centre of Insect Physiology and Ecology
Nairobi, Kenya

Signature.....**Date**.27/06/2019.....

PD Dr Michael Lattorff
International Centre of Insect Physiology and Ecology
Nairobi, Kenya

DEDICATION

Dear Dad and Mum,

Remember the evenings you narrated folk tales? Yes! You broke them down in such a way that, young as I was, I always got the point. Those early lessons shaped my view of education as a concept of storytelling! Today, I dedicate my story to you. Broken down in a way that old as you may consider yourself, you will get the message. Thank you for having laid a strong foundation and motivating me to achieve the higher heights of education. It was not easy, but with your support, I moved on.

"*Mutio muno*" ~ Thank you!

ACKNOWLEDGEMENTS

First I thank the almighty God for the good life, health and strength which enabled me do the work with minimal hitches. It is a common say that it takes a village to raise a child. My case is not different. I would like to pass my sincere gratitude to Prof. Raina Suresh and Dr. Elliud Muli for giving me this great opportunity to carry out my research work. I am grateful to my supervisors PD Dr Michael Lattorff, Dr Elliud Muli and Dr Steven Runo for the outstanding mentorship, guidance and support throughout my research work.

My sincere appreciation to the International Center of Insect Physiology and Ecology (ICIPE) for having provided me with the research facilities without which this work would not have been possible. More precisely to the International Fund for Agricultural Department (IFAD) and ICIPE's Dissertation Research Internship Programme (DRIP) under which I was awarded the scholarship to carry out this work. Many thanks to Kenyatta University and the department of Biochemistry, Microbiology and Biotechnology where I am registered. I would also like to acknowledge the Kenyatta University Graduate School for proper guidance during my thesis preparation.

Lastly, I would like to thank all the members of the African Reference Laboratory for Bee Health for their support, positive criticisms and advice. A special mention to Kelvin Muteru, Joseph Kilonzo, Nelly Ndungu, Pamela Seda, Yoseph Hamba Tola and Gilbert Rotich who were always there to give me support when the going got tough.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABREVIATIONS

AMOVA	Analysis of Molecular Variance
ANJ	Anjouan
CaCl	Calcium chloride
CCD	Colony Collapse Disorder
COI	Cytochrome Oxidase I
COII	Cytochrome Oxidase II
DCA	Drone Congregation Areas
DNA	Deoxyribonucleic acid
EtBr	Ethidium Bromide
ICIPE	International center of insect physiology and ecology
USDA	United States Department of Agriculture
U.V	Ultra Violet
KCl	Potassium Chloride
NaCl	Sodium Chloride
NCBI	National Centre for Biotechnology Information
NGA	Ngazidja
MLG	Multilocus genotype
MOH	Moheli
PCR	Polymerase Chain Reaction
SWIO	South-West of Indian Ocean

ABSTRACT

Honeybees, *Apis mellifera*, are key contributors to the global economy, nutrition, food security and ecological biodiversity. Despite their robust benefits, declines in honeybee populations have been reported owing to a number of factors. Such include pests and diseases, climate change, poor management practices, change in land usage and low genetic diversity. High intra-colonial genetic diversity is associated with resilience and resistance to pest and diseases as well as increased colony level fitness. The use of morphological characters to characterize honeybees is limited by the fact that they are subject to manipulation by environmental selection pressures and introgression resulting from hybridization between sub-species. Hence, mitochondrial and microsatellite markers which are more powerful and neutral have been used. Honeybees native to the Comoros islands have not been characterized despite the fact that islands often possess endemic species which serve as distinct genetic reservoirs with unusual adaptations. This study aimed at generating extensive knowledge on the genetic diversity and population structure of honeybees in Comoros islands using mitochondrial and microsatellite markers. In the study, 160 wild and managed colonies were sampled across the Comoros archipelagos and characterized. Mitochondrial analyses involved PCR amplification and sequencing of the intergenic region spanning between COI- COII genes followed by phylogenetic analyses. In addition, 16 colonies distributed across the three islands were analyzed using 19 microsatellite markers. The mtDNA sequences analyses revealed the existence of three haplotypes in Comoros Islands where A1 haplotype, with a distribution of 51%, 80% and 91% in Anjouan, Moheli and Ngazidja respectively was predominant. Two other haplotypes (L1 and L2) were newly described in this study. They had a cumulative distribution of 48% in Anjouan, 20% in Moheli and 9% in Ngazidja. L1 and L2 showed high levels of variability as compared to A1 lineage. L2 arose from rare mtDNA recombination between L1 and A1 lineages. Discriminant Analysis of Principal Components, Provesti's neighbor joining tree and Bayesian clustering in STRUCTURE clustered the samples according to geography suggesting a restricted gene flow between islands. The population is strongly differentiated (G_{ST}=0.41(±0.068)) with a much stronger differentiation between Ngazidja and Moheli (G_{ST} = 0.53) than Anjouan and Moheli (G_{ST} = 0.37) or Anjouan and Ngazidja (G_{ST} =0.38)). High levels of genetic diversity evidenced by high number of alleles per locus (5.92 (±0.05)), and the high expected heterozygosity (0.66 (±0.03)). Private alleles (216 and 218) were detected at A113 locus. In conclusion, mtDNA haplotypes can coexist and undergo rare mtDNA recombination as well as hybridization at the nuclear genome. There is need to investigate mechanisms that enable mtDNA recombination in honeybees.

CHAPTER ONE

INTRODUCTION

1.1 Background information

Pollinators are major contributors to the world's economy, nutrition, food security as well as ecological biodiversity (Klein *et al.*, 2007). The global economic value of pollination services by both the wild and domesticated pollinators is estimated at USD 162 billion per annum (Gallai *et al.*, 2009). Vegetables, nuts and fruit producing crops, which are key sources of nutrients, are highly dependent on pollinators which makes them important in relation to micronutrient deficiencies in the developing world (Eilers *et al.*, 2011). Consumer surplus losses attributed to total loss of animal pollination services are estimated to range between USD 202 to 329 billion per annum (Gallai *et al.*, 2009). Apart from pollination services, the honeybee contributes to income generation through the sale of hive products such as honey, bee wax and royal jelly.

Insect pollination services are responsible for one-third of human meals. These pollination services are also vital to the ecosystem (Wallberg *et al.*, 2014). Honeybees contributes up to two thirds of the pollination services with estimates indicating that up to 35% of food consumed by human beings directly depends on insect-based pollination services (Koch, 2008). However, there has been an observed decline in the honeybees' populations over the past years (McMenamin and Genersch, 2015). In the United States (U.S) major winter losses average stands at 30% since 2006 while annual losses are were steadily approaching 50% by the year 2012 (CCD 2012). In the year 2006, United Kingdom

(U.K) experienced a 62% decrease in the honeybee population (Rac, 2006). According to a review by Pirk *et al.* (2016) on the status of honeybee health in Africa, Africa is yet to report major cases of honeybee colony losses.

The elevated honeybee losses in the U.S recorded from 2006 to 2008 was linked to a syndrome known as 'Colony Collapse Disorder' (CCD) (Johnson *et al.,* 2010; vanEngelsdorp *et al.,* 2009) (CCD Steering Committee, 2012a). Colony Collapse disorder is characterized by (i) a rapid adult worker bee losses in the affected hives resulting in either dead or weak colonies with a characteristic high brood population compared to the adults (ii) a noticeable absence of dead workers both within and around the affected colonies (iii) a delay in the invasion of the hives by pests as well as delayed cleptoparasitism by neighboring colonies and (iv) a noticeable absence of *Nosema* spp loads and *Varroa destructor* loads at levels known to cause economic damage.

Apart from CCD, several other factors have also been postulated as possible drivers of the observed colony declines. These include low genetic diversity (Tarpy, 2003), change in land usage, use of pesticides (Johnson *et al.*, 2010), parasites and pathogens (VanEngelsdorp *et al.*, 2009), climate change, poor nutrition and poor colony management strategies (Watanabe, 2008)

A report by the United States Department of Agriculture (USDA) shows that most colonies exhibiting poor health conditions as well as symptoms of CCD have a characteristic

increase in the viral and/or bacterial loads as well as pesticide residues, in comparison to their non-CCD counterparts (CCD Steering Committee, 2012b). Studies have also linked low colony fitness to a reduction in the intra-colonial genetic diversity making the colonies more susceptible to pests and diseases (Tarpy, 2003). Understanding the honeybee intracolonial diversity is important as it informs the levels of colony fitness. Genetically diverse colonies exhibit diverse sub-families of workers that result in colony behavioral diversity as well as increased colony efficiency in relation to fitness, productivity and survival (Mattila and Seeley, 2007).

The first evolutionary history of honeybee, *Apis mellifera* (Linnaeus, 1758) was defined using morphometric parameters (Ruttner *et al.,* 1978). Since then numerous approaches have gained usage in characterizing honeybees successfully classifying them into approximately 30 sub-species and four major groups (Péntek-Zakar *et al.,* 2015). These approaches have incorporated both the traditional morphometric parameters (Ruttner, Tassencourt and Louveaux, 1978) as well as the use of molecular markers such as microsatellite and mitochondrial markers (Garnery *et al.,* 1993; Solignac *et al.,* 2003a).

Various microsatellite markers have been characterized (Estoup *et al.,* 1993; Solignac *et al.,* 2003a) and since gained wide application in the determination of nuclear genetic differences that exist within various honeybee sub-species in different studies (Franck *et al.,* 2001; Desai and Currie, 2015; Kraus *et al.,* 2007; Loucif-Ayad *et al* 2015; Mattila and Seeley, 2007b; Palmer and Oldroyd, 2003; Péntek-Zakar *et al.,* 2015; Rasolofoarivao

et al., 2015; Techer *et al*., 2015a; Techer *et al*., 2015b; Techer *et al*., 2015c; Techer *et al.*, 2017a;Techer *et al*., 2017b)

Extensive studies exist on the genetic diversity of honeybees in most of the islands to the SWIO. However, limited knowledge exist on the genetic diversity and population structure of honeybees native to the Comoros islands where most colonies exist as wild populations. Attempts to study the honeybees of Comoros relied on sampling foragers on flowers at a distance of 5 km (Techer *et al* 2017a, 2017b) which might not offer a true representation of the colonies present in the islands as there is a high likelihood of oversampling individuals from the same colony. In addition, given the fact that Comoros islands is majorly populated by wild honeybee populations, it was of much interest to determine the level of genetic diversity and population structure exhibited by the honeybees population under minimal human interference. This study employed extensive sampling to determine the nuclear and mitochondrial genetic diversity as well as the population structure of honeybees across three main islands of the Comoros archipelagos (Ngazidja, Anjouan and Moheli).

1.2 Statement of the problem

Honeybees offers important pollination services vital for food production, ecological bio-diversity, world economy, nutrition, and food security (Klein *et al.*, 2007). Gradual declines in honeybee populations have been reported majorly in the northern hemisphere (McMenamin and Genersch, 2015). Consumer surplus losses associated to this decline were last estimated to range between USD 202 to 329 billion per annum (Gallai *et al*., 2009) and the figure is likely to have increased over time.

Decline in honeybee population is majorly associated with three broad factors namely pests and pathogens, environmental stressors and genetic diversity and vitality (Potts *et al.*, 2010). All these factors work as either individual entities or in synergy to impact the honeybee health (Quarles, 2008; Watanabe, 2008; VanEngelsdorp *et al.*, 2009; Johnson *et al.*, 2010). Most studies on the genetic diversity and population structure of honeybees have relied on managed honeybee populations. As such, there is limited knowledge on the genetic diversity and population structure of wild honeybee populations existing within a fragmented landscape under minimal human interference.

1.3 Justification

Genetically diverse colonies exhibit diverse worker sub-families which confers behavioral diversity to the colony, leading to an increase in colony level efficiency and fitness in terms of acquiring food, and response to pests and diseases (Mattila and Seeley, 2007; Oldroyd *et al.*, 1992; Tarpy *et al.*, 2013). The impact of low genetic diversity on most managed colonies has been studied and the results indicate that a reduced level of genetic diversity at the intra-colonial level affects the fitness of the colony making the colonies

more susceptible to pests and diseases (Tarpy *et al*., 2013) as well as less efficiency in food acquisition, which results into less productive colonies. Although the use of morphological characteristics is still considered a significant approach, it lacks the power to characterize honeybees into distinct subspecies (Franck *et al.,* 2000). Furthermore, it does not elucidate the phylogenetic relationship among them. This can be attributed to the fact that they are subject to manipulation by environmental selection pressures and introgression resulting from hybridization between sub-species (Franck *et al.,* 2000). Specific molecular markers such as the mitochondrial and neutral nuclear genome markers have given new insights into the successful classification of honeybees into subspecies. Sequencing of the COI-COII intergenic region (Garnery *et al.,* 1993) has successfully classified honeybees in accordance with the Ruttner's classification which was based on morphological characters. Furthermore, new sub-species have been identified and added to the Ruttner's list.

Although a recent study on the genetic diversity of honeybees native to the South-West Indian Ocean islands describes both the nuclear and mitochondrial diversity in Comoros islands, the study lacks power owing to low samples collected from Comoros archipelago between the year 2013 and 2015. For instance due to the small sample size, it was not possible to compute pairwise F_{ST} values. Though, the population portrayed differentiation, the Moheli samples did not significantly differentiate with the authors citing low sample size coupled with high null allele number frequencies as the possible causes (Techer et al., 2017a; 2017b). Very few samples per colony were used which limited the ability of the study to determine the diversity at the colony level.

Understanding the genetic diversity and population structure of honeybees is of significance as the knowledge generated informs conservation strategies for endemic honeybee species within their natural habitats and protect them from human interference. It also enables the identification of beneficial genotypes within a population that can be used for breeding purposes. Breeding is key to sustainable beekeeping program

1.4 Hypothesis

Honeybees of Comoros islands of Ngazidja, Anjouan and Moheli exist as a large homogeneous population with a single mitotype, low genetic diversity, non-structured and undifferentiated

1.5 Objectives

1.5.1 General objective

i. To determine the mitotypes, genetic diversity, population structure and genetic differentiation of honeybees in Comoros islands using mitochondrial and microsatellite markers

1.5.2 Specific objectives

i. To identify the mitotypes of honeybee colonizing the Comoros islands through COI-COII intergenic region sequence analysis

ii. To identify the genetic diversity, population differentiation and population structure using microsatellite DNA markers

CHAPTER TWO

LITERATURE REVIEW

2.1 Classification of honeybees

Honeybees, *Apis mellifera,* are classified under the Kingdom, Animalia: Phylum, Arthropoda: Class Insecta: Order, Hymenoptera: Family, Apidae: Genus, *Apis*: and Species, *mellifera* (Ruttner, 1988). They are also classified in the Aculeata section of hymenopterans - hymenopterans whose females possess stings. The Aculeata section include bees, wasps and ants. Bees exhibit similarities to the sphecoid group of wasps but generally more unlike the rest of Aculeata, some bees are more hairy and robust when compared to the wasps. In contrast to sphecoid wasps, honeybees do not capture other insects and spiders as their protein source but they rather depend on flowers as the sole source of proteins needed for the development of their larvae and ovaries. The sphecoid wasps and honeybees together constitute the Apoidea superfamily which is further subdivided into two distinct groups i.e. the spheciformers and the apiformers. Honeybees are grouped among the apiformers group of the Apoidea superfamily (Michener, 2000).

Apart from *Apis mellifera,* which is a cavity nesting honeybee, other honeybee species also exist. These include, *Apis koschevnikovi*; *Apis nuluensis*; *Apis nigrocincta* and *Apis cerana* classified among the cavity nesting bees*; Apis florea* and *Apis andreniformis* categorized among the dwarf honeybees, and *Apis dorsata* and *Apis laboriosa* grouped among the giant honeybees i.e. (Michener, 2000). Apart from *Apis mellifera* all the other species are endemic to Asia.

2.2 History and the origin of honeybees

The honeybee is one of the most successful species in the Kingdom Animalia due to its ease of adapting to new environmental conditions. It is widely thought that honeybees evolved in the tropics during the mid-tertiary period of the Pleistocene epoch and rapidly evolved and spread fast to various parts of the world (Ruttner, 1988)

The evolutionary origin of *Apis mellifera* was proposed based on three main scenarios. The first scenario as proposed by Ruttner *et al.* (1978) postulates that *Apis mellifera* originated from the Middle East and Northern Africa before colonizing Europe following two main routes namely (i) the western route through North Africa and (ii) through the Iberian Peninsula. This hypothesis was based on morphological analysis and suggested continuity between the African lineage (A) and the M lineage and an ancestral form of *Apis mellifera syriaca* from Israel, Lebanon and Jordan. The second hypothesis by Cornuet and Garnery (1991) and Garnery *et al*, (1992) also proposes *Apis mellifera* to have originated from Middle East but it does not suggest colonization via the Western route. This second scenario is strongly based on phylogenetic analysis in which the A lineage and the C lineage are grouped together instead of the M, contradicting the migration via the strait of Gibraltar. A third hypothesis suggests an African origin. This hypothesis capitalizes on the ability of domesticated *Apis mellifera* to form winter clusters representing a derived adaptation to temperate climates (Wilson, 1971).The hypothesis is based on the observation that *Apis mellifera* does not colonize the tropical Asia, it is more likely the proposed ancestral form had an African origin.

Biodiversity of honeybees was originally determined using morphometric analysis that was entirely based on extensive collection of samples followed by multivariate analysis of morphometrical characters (Ruttner *et al.,* 1978). This identification technique was long considered to be definitive of the honeybees. Through this classification approach, the North-Eastern Africa and the Middle East were established as the area of origin for *Apis mellifera.* From their centre of origin, the species are thought to have invaded Africa and Europe following three distinct patterns that informed their classification. These are the South and Central Africa Branch (A), North African and North European Branch (M) and the North Mediterranean branch (C). The classification was further improved by the addition of the fourth evolutionary branch known as the Oriental (O) branch, which invaded the Near and Middle Eastern sub-species (Ruttner, 1988). Most recently two more branches have been described as the Y lineage from Ethiopia and the Z lineage from Syria (Franck *et al.,* 2001; Ruttner, 1988; Alburaki *et al.*, 2013).

Despite the fact that honeybee, *Apis mellifera*, naturally inhabits Europe, Africa and the Middle East up to nine other *Apis* species are exclusively found in Asia hence , *Apis mellifera* is widely assumed to have originated from Asia. However, some hypothesis based on phylogenetic analysis using genetic markers, particularly the analysis of over 1000 single nucleotide polymorphisms (SNPs), placed the root tree of *Apis mellifera* subspecies to samples originating from Africa suggesting that *Apis mellifera* bares its origin from Africa (Whitfield *et al.,* 2006). Re-analysis of the same data by Fan *et al.* (2012), failed to support the initial findings which had suggested an out of Africa origin and moved on to strongly link its origin to Asia. The latter findings are corroborated by the

studies by Wallberg *et al.,* (2014) who suggested an out of Asia origin citing the fact that all other honeybee species are almost exclusively found in Asia.

The lineage *Apis mellifera* is thought to have split from other cavity-nesting bees and later diversified into the present subspecies eventually colonizing their present habitat. Morphometric and molecular analysis have successfully grouped the honeybees into five evolutionary lineages and further into over 29 subspecies (Ruttner, 1988; Franck *et al.,* 2001). Studies based on the mtDNA and nuclear DNA genetic diversity suggest the first split of honeybees to have occurred about 6-9 million years ago (Cornuet and Garnery, 1991; Arias and Sheppard, 1996). Contrary to this, genetic variations observed among the extant subspecies show some degree of similarity and highly suggest that they have not experienced long isolation periods. Based on genetic dating, the four major lineages (A, M, O and C) are thought to have diverged about 0.7-1.3 million years ago (Garnery *et al.,* 1992; Arias and Sheppard, 1996). More recently, Wallberg *et al.* (2014) used whole genome sequencing to give insights into the evolutionary history of honeybees. In their work, the A, C and M lineages are shown to have diverged about 300,000 years ago. Furthermore, they show that the O and C lineages separated more recently at a period of about 38,000 years from the present. However, they suggest the given dates as the minimum divergence times citing the possibility of continued gene flow among the honeybees despite longer divergence periods. Little data is available to infer the geographical ranges inhabited by the honeybees from the time they split from the other cavity nesting bees to their present habitat. Morphological and genetic relationships among the subspecies can however give a clue to the timing and location of their common origin (Fan, *et al.,* 2012).

2.3 Genetic diversity in honeybees and its importance

Hymenopterans exhibit a haplodiploid system of sex determination where males are monogamous while females are polygamous. In this system, fertilized eggs develop into females while the unfertilized eggs develop into males. Sex is determined by constitution of alleles at the sex determining locus. In the case of heterozygosity within the sex determining region, females will develop, while males develop from homozygous eggs. However, diploid eggs which are homozygous at the sex determining locus results in diploid male offspring. Such offspring are reproductively useless as they tend to produce few sperm cells hence most of them are killed at the larval stage of development among the *Apis mellifera.* Those that survive to maturity produce triploid offspring that lack reproductive potential.

The honeybee mating strategy is of increasing interest. During the mating seasons, drones fly out of the hive and congregate high in the air at distinct locations (Ruttner, 1966; Vallet and Coles, 1993). These zones of congregation are termed as Drone Congregation Areas (DCAs) and are comprised of drones that originate from different colonies (Baudry *et al.,* 1998). Generally, drones arising from apiaries that are within an interval of 5 km apart visit the same DCA. As such, each DCA has a mixed population of drones that ultimately contribute to higher diversity. Most recently, drones originating from up to 238 colonies have been detected in the same DCA (Baudry *et al*., 1998). After emergence, the young queen flies out for mating. The mating of the virgin queens usually takes place high in the air at the DCA during a mating or nuptial flight. This occurs at the age of 5-10 days from emergence of the queen. During the mating flight virgin queens perform one or

two orientation flights next to the hive, which last for about 2 minutes before finally flying to the DCA, where they copulate with several drones, usually 10-20 drones (Schlüns *et al.,* 2005), that in turn deposit their male gametes in the queen's lateral oviduct. The mating flight generally lasts for 10 -30 minutes, and upon return to the hive, 5% of the sperm stored in the oviduct is transferred to the spermatheca of the queen in a filling process which takes up to 24 hours (Woyke, 1962). The queen then uses the sperm to fertilize eggs as it lays them. The fertilized eggs becomes females and the unfertilized eggs develop into drones.

Genetically diverse colonies arising from the polyandry mating of the queen, at the DCA, exhibit diverse worker sub-families which confers to the colony behavioral diversity and this increases the colony level efficiency in terms of fitness, survival and productivity (Oldroyd *et al.,* 1992; Mattila and Seeley, 2007). Workers from colonies with low genetic diversity on the other hand portray a narrow range of behavioral threshold. Studies have indicated that a high level of genetic diversity at the intra-colony level may lower the chances of infestation with parasites and pathogens (Schmid-Hempel, 1995; Sherman *et al., 1988; Tarpy, 2003)*. It is also postulated that close kinship among colonies increases the chances of acquiring a parasitic infection among the colonies as compared to the genetically diversified colonies (Shykoff and Schmid-Hempel, 1991). Colonies with no or very low genetic diversity are advantageous in the perspective of the parasites as they are more susceptible to parasitic colonization compared to colonies with a high genetic diversity. Colonies originating from a single paternity have also been shown to have a charac-

teristic lower production of drone brood, lower honey and pollen storage, as well as a fewer numbers of queen cells (Fuchs and Schade, 1994).

A recent study has shown that genetically diverse colonies show a greater mite mortality rate as compared to genetically similar colonies, an observation that is specific to a certain season, winter, of the year (Desai and Currie, 2015). Desai and Curies (2015) also reported that for some pathogens, the pathogenic load tend to be higher in genetically similar colonies as compared to the genetically diverse colonies. Intra-colonial genetic diversity therefore benefits the colony by conferring some advantageous phenotypic characteristics that includes hygienic behavior, grooming behavior as well as other mechanisms which reduce the transmission of pathogens and parasites.

2.4 Characterization of honeybees based on mitochondrial DNA

The honeybees' mtDNA is a small circular molecule ranging between 16 500 to 17, 000 base pairs long. It is uniparentally transmitted by the queen to her offspring (drones and workers) thus expected to be non-recombining. As a result, a single honeybee is sufficient to determine the mito-haplotype of all honeybees in a colony. When using mtDNA to characterize honeybees, it is ideal to select a sample from an emerging adult or a pupa in order to eliminate the risk of picking and analyzing adult workers that may have drifted between colonies (Evans *et al.*, 2013).

The different mitochondrial lineages are inferred from the observation that mitochondrial DNA of honeybees contains a COI-COII intergenic region that exhibits an enormous

amount of length polymorphism. The region is also characterized by sequence variations that, in addition to the different fragments produced by endonuclease cleavage, are used to extensively distinguish among honeybee lineages as well as discriminate among subspecies (Garnery *et al*., 1992; Franck *et al.*, 2000; Sheppard and Smith, 2000). The COI-COII intergenic region exhibits at least seven distinct length variants that can be explained by a unique combination of three distinct sequences i.e. Po (67 bp), P (54 bp) and Q (192-197 bp). Different honeybee haplotypes contain different combinations of these sequences which are; PoQ, PoQQ, PoQQQ, PQ, PQQ, PQQQ and Q. The variability in length combined with availability of restriction site polymorphism has been utilized to develop a simple and rapid test that is used to characterize honeybee haplotypes based on mitochondrial DNA (*Dra*I test) (Garnery *et al*., 1993).

2.5 Characterization of honeybees based on microsatellite DNA

Microsatellites are short tandem repeats found in most species genomes and are used as informative markers in the field of population genetics, conservation biology and evolutionary biology. They are generally defined as short tandem repeats (STRs) or simple sequence repeats (SSRs) comprising 2-10 base pair units and present themselves as either perfect or imperfect repeats within an organism. Microsatellites exhibit high variability in terms of the number of repeats found in every microsatellite region. Current knowledge considers microsatellites as the most useful genetic markers in both population and quantitative genetics. The markers are biparentally inherited thus making them more useful tools in parental determination and genetic polymorphic analyses (Estoup *et al.*, 1995). A polymorphic microsatellite has more than one potential allele at a given locus. Given the fact that they are codominant, Mendelian inherited and neutral markers, microsatellites

are easily typed. In addition, they have a high distinctive power among closely related individuals (Abdul-Muneer, 2014) ultimately making then suitable candidates for determining the population structure.

Microsatellites have previously been used to characterize honeybee populations belonging to the African and European origin (Estoup *et al*., 1995). Estoup *et al.* (1995) scored alleles for seven (7) microsatellite loci among various subspecies and reported a high degree of genetic variation within the honeybee samples. The variation ranged between seven (7) to thirty (30) alleles per locus. The African subspecies were also shown to have a higher genetic variation as compared to those of the European origin. Also, of importance to note is that super sisters i.e. honeybees from a similar patriline, clustered together as opposed to half-sisters. This would not be the case using mtDNA given its maternal inheritance nature which would result in all super-sisters and half-sisters arising from same queen clustering together.

The microsatellite results, therefore, makes it possible to determine heterozygosity observed within a colony. Microsatellite based studies reveal moderate genetic diversity among honeybees (*Apis mellifera unicolor*) in Madagascar characterized by low allele number per locus in comparison to South Africa (Rasolofoarivao *et al*., 2015); detect no distinct genetic structure in Rodrigues island but shows the genetic diversity in the island to be closely related to continental populations (Techer *et al*., 2015). Lastly, they reveal a highly structured population among the honeybees in Seychelles (Techer *et al*., 2015).

Seychelles is an Archipelago composed of three main islands (Mahe, Praslin and La Digue) - a situation similar to Comoros islands (Ngazidja, Moheli and Anjouan). Based on microsatellite data, the honeybee population in Seychelles exhibit a significant nuclear similarity to *A.m unicolor* from Madagascar.

2.6 Genetic diversity of honeybees to the South West of Indian Ocean Islands

Current studies on the genetic diversity of the honeybees native to the South West of the Indian Ocean Islands (SWIO) give mixed results (Techer *et al*., 2014; Rasolofoarivao *et al.*, 2015; Techer *et al.,* 2015a; Techer *et al.,* 2017a;, 2017b). Techer *et al.* (2014) reported the absence of the African lineage in Rodrigues. This study also identified the existence of three haplotypes belonging to the C lineage. The C1 haplotype (*Apis mellifera liguistica*) was the most prevalent (81.3%). The C2 (*Apis mellifera carnica*) haplotype was reported in 18.0% of the study population. The study also revealed the existence of a new haplotype which was named C1-Rod.

However, the study did not detect any new genetic structure. In addition, no difference was noted in the genetic diversity in comparison to the continental populations. Further analyses of these suggested the possibility of the occurrence of a bottleneck scenario in the island. These findings are different from those in Seychelles islands (Techer *et al.,* 2015) where all the mtDNA sequences belonged to the African lineage with 96.7% of the study population belonging to the A1 (*Apis mellifera unicolor*) sub-lineage. In addition, sub-lineage Z was uniquely identified in two islands (Praslin and La Digue) that comprise part of the Seychelles archipelagos. Seychelles was also shown to have a high and well-

structured nuclear genetic diversity which signifies a restricted gene flow between the islands. A more notable observation was the high nuclear similarities between the honeybee populations in Seychelles and *Apis mellifera unicolor* that is native to Madagascar (Techer *et al.,* 2015). The study by Rasolofoarivao *et al.* (2015) showed that the honeybees of Madagascar are majorly comprised of A1 sub-species (99.5%) and 0.4% of A4 haplotype. Madagascar, also, has a unique A lineage haplotype which comprised of 0.2% of the study population. Microsatellite DNA analysis showed presence of moderate genetic diversity among the Madagascar's honeybees populations.

Most recently, large scale mitochondrial DNA sequencing has revealed the existence of a new private A1 haplotype that is shared between Madagascar and all the islands forming the SWIO archipelagos (Techer *et al.,* 2017a; 2017b). Presence of private African haplotypes in each of the archipelagos was used to suggest diversity radiation in each of the archipelagos. In addition, Comoros archipelagos were established as the possible contact point between the continental African and the insular populations (Techer *et al.,* 2017a, Techer *et al.,* 2017b). These studies showed that all the SWIO ocean islands except Rodrigues are populated by the African lineage of the honeybee sub-species. The study by Techer *et al,* (2017a) suggested complete dominance of the A lineage in Madagascar, Seychelles and Comoros islands. La Reunion and Mauritius had 95.5% and 56.1% of the A lineage respectively. The European lineages C and M were uniquely reported in the Mascarene archipelagos. A striking difference in their distribution was noted with La reunion having 4.6% whereas Mauritius had 44% (Techer *et al* 2017a).

2.7 Hardy Weinberg Equilibrium (HWE)

The Hardy Weinberg equilibrium describes a population that is not evolving (ideal population) where evolution refers to changes observed in the allele frequencies in a population or a gene pool. It is based on the Hardy Weinberg principle which assumes absence of external disturbances to a population thus inferring constant genetic variations to the population. This principle is based on seven assumptions, which include: (i) No selection (ii) Absence of mutation (iii) Absence of gene flow (iv) Presence of an infinite population (v) Random mating (vi) Each individual producing similar number of offspring and (vii) No migration occurring into and out of the population.

Mathematically a population that is in conformity to the HWE equilibrium assumes the equation

$$\boldsymbol{p2 + 2pq + q2 = 1}$$

Where $\boldsymbol{p^2}$ = probability of having the dominant homozygous alleles (**AA**), $\boldsymbol{q^2}$ = probability of having dominant recessive alleles (**aa**) while 2$\boldsymbol{pq}$ = the probability of having heterozygosity in the allele composition (**Aa**)

This equation can be further simplified to show that $\boldsymbol{p}$ represents all individuals that are homozygous dominant (AA) and half the individual who are heterozygous (Aa) while $\boldsymbol{q}$ represents individuals who are homozygous recessive and half of those who are heterozygous (aA).

In such, the equation is rewritten as

$$\boldsymbol{p = AA + \frac{1}{2} Aa}$$

While

$$\boldsymbol{q = aa + \frac{1}{2}aA}$$

Therefore

$$\boldsymbol{p + q = 1}$$

A population is assumed to be in a HWE if the frequency of **A** (*p*) and **a** (*q*) are both maintained at 50 % over generations if that was the case represented in the parental genotype. Deviations from this equilibrium is an indicator that the population is evolving.

2.8 Advantage of using G_{ST} to determine population differentiation

G_{ST} is an improvement to the Wright's F_{ST} statistics that was originally developed with the aim of analyzing biallelic data. G_{ST} has a higher capacity to analyze multiallelic data than F_{ST}. Based on this measure of diversity $G_{ST} = 0$ indicates that the population is not differentiated and the more the G_{ST} approaches 1 the more differentiated the population. A population with a G_{ST} value of 1 indicates the fixation of that given allele in the population.

G_{ST} is mathematically represented as (Nei, 1987);

$G_{ST} = (H_T - H_S)/H_S$

In the equation, H_T represents the total genetic diversity, while H_S represents the diversity within the population (this is the equivalent of expected population's heterozygosity in the case of diploids). From the equation above it is clear that G_{ST} value arises from calcu-

lation of genetic diversity (expected heterozygosity) within and among the study populations. Hedrick's G_{ST} is an improvement to the initial Nei's G_{ST} (Nei, 1972) with corrections for the latter's weaknesses (Hedrick, 2005).

2.9 The Principle of Discriminant Analysis of Principle Components

Discriminant Analysis of Principal Components (DAPC) is a multivariate STRUCTURE-like statistical approach of inferring the population structure among genetically related individuals (Jombart, Devillard and Balloux, 2010). The method works based on an assumption that all markers are in linkage disequilibrium and that a panmictic population exists (Pritchard *et al.,* 2000) i.e. all individuals in the populations are randomly mating.

The approach divides the sample variance into within-group and between-groups components as a way of maximizing discrimination between groups. The method first transforms data using principal component analysis (PCA) and then employs discriminant analysis (DA) to identify subsequent clusters.

CHAPTER THREE

MATERIALS AND METHODS

3.1 Study site

Comoros, an archipelago made up of volcanic islands, is located off the South-East coast of Africa to the eastern side of Mozambique and North-West of Madagascar (12°10′S 44°15′E). It is made up of three main islands namely Grande Comore or Ngazidja (1,025 km^2), Anjouan or Nzwani (424 km^2) and Moheli or Mwali (211 km^2). The sampling sites on the three main islands are represented below (Figure 3.1)

3.2 Sampling

One hundred and sixty (160) honeybee colonies were sampled across the different geographic locations in the Comoros archipelagos in June 2017. These included forty-eight (48) colonies from Ngazidja, fifty two (52) colonies from Anjouan and sixty colonies (60) from Moheli. The sampling was carried out on both managed and feral colonies. Where available, apiaries were sampled and for an apiary to be considered for sampling, it had to be five kilometers away from the previously sampled one since honeybees are known to have a flight range of 5 km and sampling within that range increases the chances of sampling close relatives. Three (3) colonies were randomly sampled from each apiary, and in cases where apiaries had three or less colonies, all colonies were sampled. Feral colonies were also sampled at an approximate distance of 5 km apart and where the feral colonies were within a close proximity they were all treated as an apiary.

One emerging adult per colony was sampled for mitochondrial DNA analysis. The choice of an emerging adult per colony reduced chances of sampling a cleptomaniac from neighboring colonies as would be the case if workers at the entrance were to be considered. Since mtDNA is maternally inherited, a single individual per colony was sufficient to infer the maternal haplotype.

For microsatellite DNA analyses 288 workers were sampled from 12 colonies across the three islands. During sampling 24 workers per colony were randomly sampled from four colonies in each island. The colonies were located in four geographically distinct zones on every island. These included Iconi (*n=24*), Bweni (*n=24*), Ipvwani (*n=24*) and Tsinimoichongo (*n =24*) in Ngazidja (*N=96*); Jimilime (*n=24*), Kambalahari (*n=24*), Nyombeni (*n=24*), and Bougweni (*n=24*) in Anjouan (*N=96*) and Siryziroudrani (*n=24*), Badralaju (*n=24*), Domoni (*n=24*) and Dahoni (*n=24*) in Moheli (*N=96*). (Appendix I). All honeybee samples were immediately stored in 96 % ethanol and placed in cool boxes in the field and transported to the African Reference Laboratory for Bee Health at the International Center of Insect Physiology and Ecology (ICIPE), Nairobi, Kenya where they were stored at -20 °C prior to molecular analyses.

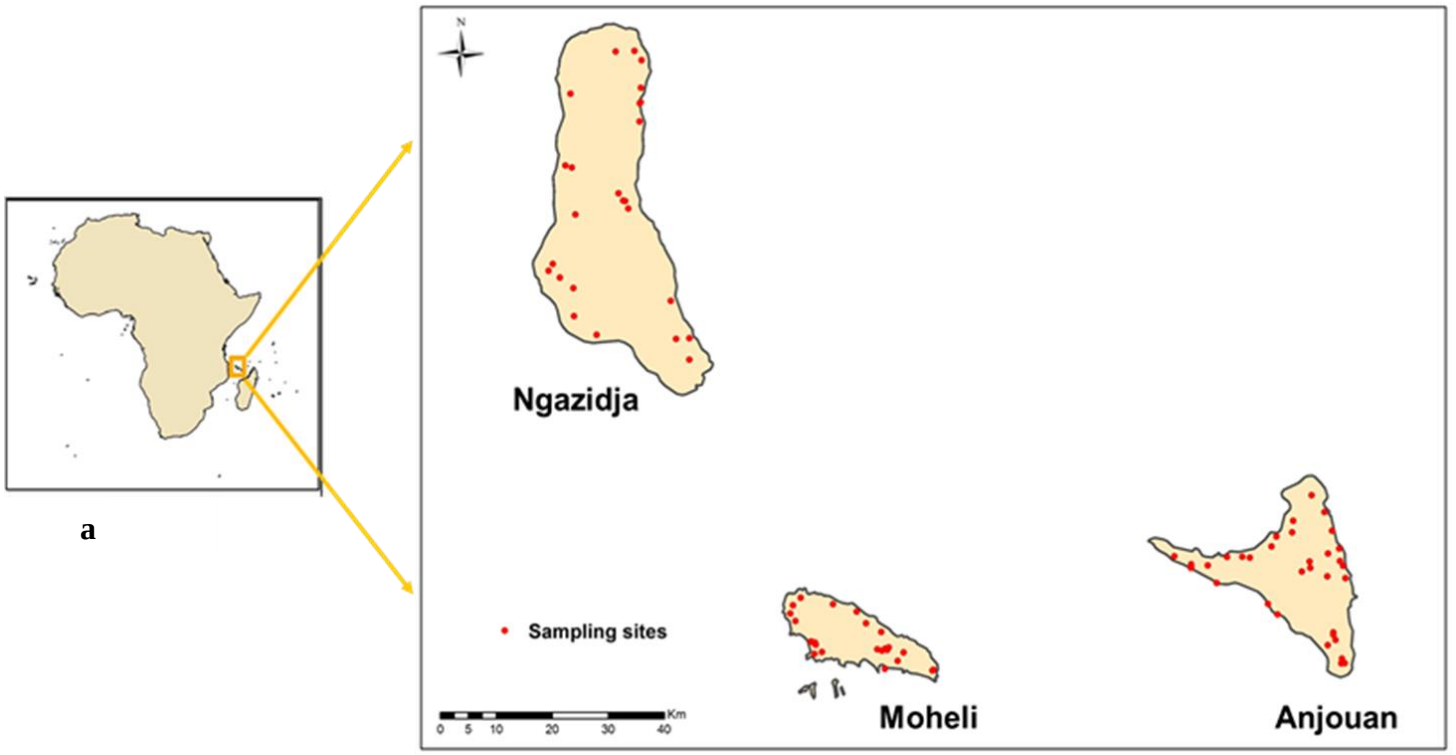

Figure 3.1: Sampling sites on the three main islands of Comoros islands. Letter **a** represents location of Comoros on the map of Africa while **b** is an enlargement of the three islands indicating the exact sampling locations. (Source: Emily, icipe GIS Unit)

3.3 DNA Extraction from honeybees

The honeybee samples were placed in a laminar flow chamber which was turned on for 30 minutes to dry off the ethanol before DNA extraction. For mtDNA analysis DNA was extracted from the honeybee thoraces while for microsatellite DNA analyses, DNA was extracted from 24 individual honeybees per colony using Qiagen DNA extraction kit (Qiagen Inc. Germany) following the manufacturer's instructions with slight modifications. Briefly, the thoracic muscle tissue of individual honeybee or the entire bee for mtDNA and microsatellite DNA analyses, respectively, were placed in a 2µl microcentrifuge tube containing 150µl of pre-chilled Phosphate Buffered Saline (PBS) (pH=7.4) (Kakumanu *et al.,* 2016) and agitated using a TissueLyser II (Qiagen Inc. Germany). DNA was extracted from 100µl of the homogenate following all the manufacturer's instructions. The integrity of the extracted DNA was checked through gel electrophoresis prior to PCR amplification. The concentration of the extracted DNA was checked using a nanodrop and standardized to 50ng/µl.

3.4 Mitotypes of honeybee colonizing the Comoros islands

3.4.1 PCR amplification of COI-COII intergenic region

Total DNA was subjected to PCR amplification using the E2 (forward) (5'-GGCAGAATAAGTGCATTG-3') and H2 (reverse) (5'-CAATATCATTGATGACC-3') primers described in Garnery *et al*. (1993). The primers targeted the mitochondrial COI-COII intergenic region that has been extensively used in determining the subspecies and evolutionary lineages among honeybees (Garnery *et al.,* 1993; Magnus *et al.,* 2011; Rasolofoarivao *et al.,* 2015; Techer *et al*., 2015a, 2015b).

A 15µl PCR reaction containing 1 x Phusion High Fidelity Taq polymerase with HF buffer, 0.4 µM of each of the primers, 0.5µl DNA at 50ng/ µl was set up **(**Table 3.1). DNA from a previously amplified COI-COII intergenic region was used as a positive control nuclease free water was served as template control in the negative control reaction. The reaction profile was set up with an initial denaturation step of 98°C for 30 seconds followed by 40 cycles of 98°C for 10 seconds, 48°C for 45 seconds 62°C for 30 seconds and a final extension step at 72°C for 10 minutes in a ProFlex PCR machine (Applied Biosystems).

Table 3.1: Components of the 15µl reaction mix used for mitochondrial DNA amplifications

Component	**Initial Concentration**	**Volume per reaction (µl)**	**Final Concentration**
Phusion High Fidelity Taq polymerase with HF buffer	2 x	7.5	1 x
Forward Primer (H2)	10 picomoles	0.6	0.4 picomoles
Reverse Primer (E2)	10 picomoles	0.6	0.4 picomoles
DMSO	100%	0.45	3%
Nuclease free water	-	5.35	-
Template	50ng/µl	0.5	1.67 ng/µl

3.4.2 Agarose gel electrophoresis and sample purification

Two percent (2%) agarose gel was prepared by adding 2 grams of agarose powder into 100ml of Tris Acetate Ethylenediaminetetraacetic acid **(**TAE) buffer in a glass conical flask. The mixture was heated in a microwave for 3 minutes to dissolve the agarose powder. The solution was allowed to cool to about 60°C before 2µl of Ethidium Bromide

(EtBr) DNA staining dye was added. It was then swirled and poured on a preset gel casting tray where it was allowed to cool for 20 minutes before being transferred into a gel electrophoresis tank pre-filled with TAE buffer and the combs removed.

The first lane of the gel was loaded with 5µl of 100 bp DNA ladder premixed with 1 µl of a 6 x loading dye (New England Biolabs). One microliter of the 6 x loading dye was also mixed with 5µl of the sample on a parafilm before the mixture was loaded into the wells. The gel was run for one hour at 80 volts and later visualized under a UVP GelDoc-It^{TS2} Imager (Avantor). The 600bp and 380bp positive samples were selected for purification and subsequent sequencing.

Purification was done using the ExoSapIT purification kit (New Englands Biolabs) following the manufacturer's instructions. Briefly, a 13µl microliter reaction was prepared by adding 3 µl of exonuclease I enzyme and Alkaline phosphatase premixed in the ratio of 2:1, respectively into 10µl of the PCR amplicon. The samples were then incubated at 37°C for 15 minutes followed by 80°C for 30 minutes to inactivate unused enzymes.

3.4.3 Sequencing of the COI-COII intergenic region

Ten microliters (10µl) of the purified samples was packaged in a 1.5µl microcentrifuge tube and sent to Macrogen Inc. (Amsterdam, Netherlands) where bidirectional sequencing was done using the Sanger chain termination technology.

3.4.4 COI-COII intergenic region sequence analyses

The resultant sequences were edited and consensus sequences for the forward and reverse strands generated using BioEdit version 7.2.5 (Hall, 1999). COI-COII sequences belonging to different species of honeybees *Apis mellifera* (AY587542.1), *Apis dorsata* (AY588415.1), *Apis Koschevnikovi* (AY587546.1), *Apis laboriosa* (AY587548.1), *Apis cerana* (AY587544.1*), Apis nigrocincta* (AY587545.1), *Apis nuluensis* (AY587543.1), *Apis florea* (AY588416.1) and *Apis andreniformis* (AY588417.1) were mined from the GenBank database in NCBI and included to the consensus sequences for phylogenetic analyses.

In addition, different *Apis mellifera* subspecies i.e. A1_MAD3 (KT828418), A1_MYT2 (KT828439), A1_SEY1 (KT828433), Z2_SEY1 (KT828478), A1_MAD15 (KT828430), A1_GCO1 (KT828435), A1_GCO2 (KT828436), A1_ANJ1 (KT828437), A1_MAD12 (KT828427), Z7_EGY1 (KT828447), A4_SEN1 (KT828455), A4_SEN2 (KT828456), A4_STP1 (KT828457), A4_a (KT828448), A66_STP1 (KT828475), C2_b (KT828496), C1_a (KT828493), M7'_a (KT828492), C1_Rod1 (KT828494), A4_ZAF5 (KT828469), M4_a (KT828487), A6_TCD1 (KT828474), A1_MAD15 (KT828430), A1_MAD12 (KT828427), A11_PRT1 (KT828479), M7_a (KT828490) were mined from the GeneBank database and included among the consensus sequences for multiple sequence comparison

These sequences were subsequently aligned using Muscle (Edgar, 2004) on MEGA 7.0.26 platform (Kumar *et al.,* 2012) and a phylogenetic tree constructed with *A. florea* as the outgroup. Prior to constructing the phylogenetic tree, the best nucleotide substitu-

tion model was predicted based on maximum likelihood statistics and a neighbor-joining tree. Gaps/missing data in the sequence alignment were subjected to partial deletion set at a threshold of 95% site coverage. Through this approach the model with the lowest Bayesian Information Criterion (BIC), Tamura 3-parameter, was considered to best describe the nucleotides substitution model (Kumar *et al.*, 2012). The sequences from dwarf honeybees (*Apis florea* and *Apis andreniformis*), which are the most ancient honeybee species (Ruttner, 1988) were used as outgroups. The sequences were grouped according to the various clusters obtained from the phylogenetic tree, and the genetic distances within and between groups calculated in MEGA.

3.5 Genetic diversity, population differentiation and population structure

3.5.1 Microsatellite DNA amplification

Nineteen previously published honeybees' polymorphic microsatellite loci were scored. These included A113, A24, AC306, AP81, A88, UN351,A56, SEX1, AP273, UN16603, UN4987, UN370, A28, AP289, A124, A35, A8, AP33, and A43 (Estoup, 1994; Estoup *et al.*, 1995; Pierre Franck, Garnery *et al.*, 1998; Shaibi *et al.*, 2008; Solignac *et al.*, 2003a). Forward primers were tagged using fluorescent dyes (FAM, HEX and TET) that absorb and emit light at different wavelengths (Appendix I).

The amplification reaction set up was distributed into six multiplex PCR reactions (Table 3.2). Selection of the plexes was based on the fluorescent dyes and differences in the product size. A fifteen microliter (15µl) multiplex PCR reactions was ran using MyTaq™ HS DNA Polymerase (BIOLINE). The final reaction volume contained 1 x My Taq Polymerase reaction buffer, 0.067 units per microliter of MyTaq Hot start Polymerase,

0.2µm of each of the primers, 1µl of DNA at a concentration of 50ng/µl. The reaction was set up with an initial denaturation step of 95°C for 1 minute, followed by 40 cycles of 95°C for 30s, 52°C for 15 seconds, 72°C for 30 seconds and a final elongation step of 72°C for 10 minutes.

Table 3. 2: List of microsatellite markers used for multiplex PCR reaction showing how they were combined in the different plexes based on fragment size and the fluorochrome.

Marker	Product range (bp)	Fluorochrome	Multiplex
A113	202-224	FAM	1
AP81	124-136	TET	1
UN351	147-166	FAM	1
A56	270-300	TET	2
AP289	174-288	HEX	2
A88	136-149	HEX	2
AC306	165-181	FAM	3
A28	128-134	FAM	3
A24	93-116	TET	3
A8	165-181	TET	4
AP33	225-247	HEX	4
A35	94-123	HEX	4
SEX1	142-187	HEX	4
UN467.16603	261-282	FAM	5
UN462.4987	168-178	FAM	5
AP273	106-110	FAM	5
A43	124-154	TET	6
B124	216-232	HEX	6
UN467.370	174-192	HEX	6

3.5.2 Fragment analysis and allele scoring

The resultant PCR products from different plexes were co-loaded into three different mixes in 96-well plates. Mixes were first selected based on labeling dye, where products labelled with different dyes were distributed across the plates (Table 3.3). In the event that two products labelled using similar dye had to be co-loaded on one plate, then they had to be of different sizes so as to avoid overlapping of peaks during the genotyping process. The mixes were packaged in a 96 well plate and sent to the Institute of Biology, Martin Luther University Halle-Wittenberg in Germany where capillary sequencing was done using MegaBace automatic sequencer.

Table 3.3: The primer mix present in each reaction as co-loaded for fragment analysis in the MegaBace automatic sequencer.

MIX 1	MIX 2	MIX 3
A113	AC306	UN467-16603
UN351	A28	UN462-4987
A56	A8	AP273
AP81	A24	A43
AP289	AP33	B124
A88	A35	UN467-370
	SEX1	

MegaBace Fragment Profiler Version 1.2 was used to score the allele sizes at the various loci. Briefly, a peak filter algorithm was generated using expected products ranges shown in Table 3.2. Three filters were generated in accordance with the primer information. Using the algorithm, alleles at various loci were scored. Additionally, alleles were verified by eye and scoring of alleles sizes was manually double-checked and finally exported to an excel sheet. Micro-Checker 2.2.3 (Van Oosterhout *et al.,* 2014) was used to check for genotyping errors including null alleles, stuttering and large allele drop outs. Queen alleles for each locus per colony were manually scored. Scoring was based on an assump-

tion that every member of a colony arises from a single queen hence queen alleles are expected to be represented in each and every worker. Homozygosity and heterozygosity among the workers were inferred based on the methods previously described (Neumann *et al.*, 1999). A homozygous queen was scored when an allele was present across all sampled workers whilst heterozygous queens was considered in the event that different workers in a colony bared either one of the two alleles present in the heterozygous queen.

After scoring the queen allele, the remaining allele was scored as the drone alleles. For homozygous workers, one of the homozygotes was automatically scored as the drone allele while for heterozygous workers the allele different from either of the queen's allele was scored as the drone allele. Determining drone alleles in heterozygous workers portraying both the heterozygous queen alleles involved a multiloci genotyping in order to score drone allele. When multilocus genotyping failed to separate the alleles, the remaining alleles were divided into half and scored as different drones.

3.5.3 Determination of queen mating frequency

Multilocus genotyping approach was used to determine the number of drones in each sample population. Drones were considered super siblings when they portrayed identical alleles at all loci scored while those exhibiting variations at different loci were considered half-siblings. Following this criteria the number of drones that mate with each queen in the different colonies was determined. Due to finite sample size, the observed drones were used to calculate estimate actual number of drones that mate with the queen by fitting in a Poisson distributions while using drones data to calculate the queens mating frequency super siblings were only represented once in order to avoid overestimation of the

mating frequency. Variations in the queen mating frequency between islands was determined using Kruskal-Wallis Test (MacDonald, 2009).

3.5.4 Prediction of queens using drones

The observed drones per island were also used to infer their maternal lineage and sibship assignment in Colony version 2.0.6.4 (Jones and Wang, 2010). For super siblings, only one was used in to avoid overrepresentation of a particular drone genotype. Following this process the queen genotypes in each subpopulation (Anjouan, Moheli and Ngazidja) were scored. Manually scored queen genotypes were included among the predicted queen genotypes parenting more than two drones and used for downstream analyses.

3.5.5 Genetic diversity and linkage disequilibrium

Using *poppr* package in R, observed individuals (N), observed multilocus genotypes (MLG), expected multilocus genotypes (eMLG) given the smallest sample size ($N \geq 10$) based on rarefaction, Shannon-waiver index of MLG diversity (Shannon, 2001), Stoddart and Taylor's Index of MLG diversity (Stoddart and Taylor, 1988), Simpson's index (Simpson, 1949), Evenness, E5 (Grünwald *et al.*, 2003), Nei's unbiased gene diversity (Nei, 1978), the index of association, I_A (Brown *et al.*, 1980; Smith *et al.*, 1993) and the standardized index of association were determined. Using *poppr* and *magrittr* packages (Kamvar *et al.*, 2014), alleles per locus, allele mean numbers (*Na*) and unbiased expected heterozygosity (H_{exp}) were estimated based on Nei's method (Nei, 1978).

To test for deviation from the HWE, *pegas* package was used to compute chi-square (χ^2) statistics and the p-values for the entire dataset (Comoros) as well as for every subpopulation (Anjouan, Moheli and Ngazidja). Using *lattice* package a heatmap showing

deviations from the HWE ($p<0.05$) for every locus in each sub-population was computed. Linkage disequilibrium was determined using *poppr* and *magrittr* packages. The test assumed that all alleles observed at the different loci were linked hence they do not freely recombine during sexual reproduction. The index of association (I_A) (Brown *et al.*, 1980) was computed assuming 999 permutations test the hypothesis. Pairwise linkage disequilibrium was further calculated to determine alleles that were in strong linkage.

3.5.6 Population differentiation and population structure:

3.5.6.1 Population differentiation

Population differentiation was calculated using *mmod* package where Hedrick's standardized measure of genetic differentiation (G_{ST}) (Hedrick, 2005) was calculated. The G_{ST} per locus as well as the global G_{ST} value per subpopulation were determined. Also calculated was a pairwise G_{ST} between the subpopulations (Anjouan, Moheli and Ngazidja).

3.5.6.2 Analysis of Molecular Variance (AMOVA)

Using *poppr* package in R, AMOVA test (Excoffier, Smouse and Quattro, 1992) was performed to detect the level of population differentiation within samples, within and between populations. Test for significance was carried out using the *ade4* package where both the Monte-Carlo tests (Excoffier, Smouse and Quattro, 1992) were performed with 999 permutations.

3.5.6.3 Population structure

Queen and genotypes were used determine the population structure. Using *poppr* package a neighbor joining tree based on Provesti's distance with1000 permutations was calculated and visualized using *ape* package. Besides, the population structure was determined using DAPC in R (Jombart, Devillard and Balloux, 2010). Additionally, the population

structure was determined using Bayesian clustering approach implemented in STRUCTURE 2.3.3 (Pritchard, Stephens and Donnelly, 2000)

CHAPTER FOUR

RESULTS

4.1 Mitotypes of honeybee colonizing the Comoros islands

4.1.1 Amplification of the COI-COII Intergenic Region

One hundred and fourteen (114) samples from Ngazidja (*n*=40), Anjouan (*n*=35) and Moheli (*n*=39) were successfully amplified and sequenced. Two band populations were observed across the three islands (4.1.1). One was 380 bp (*n*=28) while the other was 600 bp (*n*=86). The 380 bp mitotypes were distributed such that 16 were in Anjouan, 7 in Moheli and 5 in Ngazidja while the 600 bp mitotypes were distributed in the frequency of 19 in Anjouan, 32 in Moheli and 35 in Ngazidja. A total of 45 samples, though PCR amplified, were not successfully sequenced. These included 17 from Anjouan, 21 from Moheli, and 8 from Ngazidja.

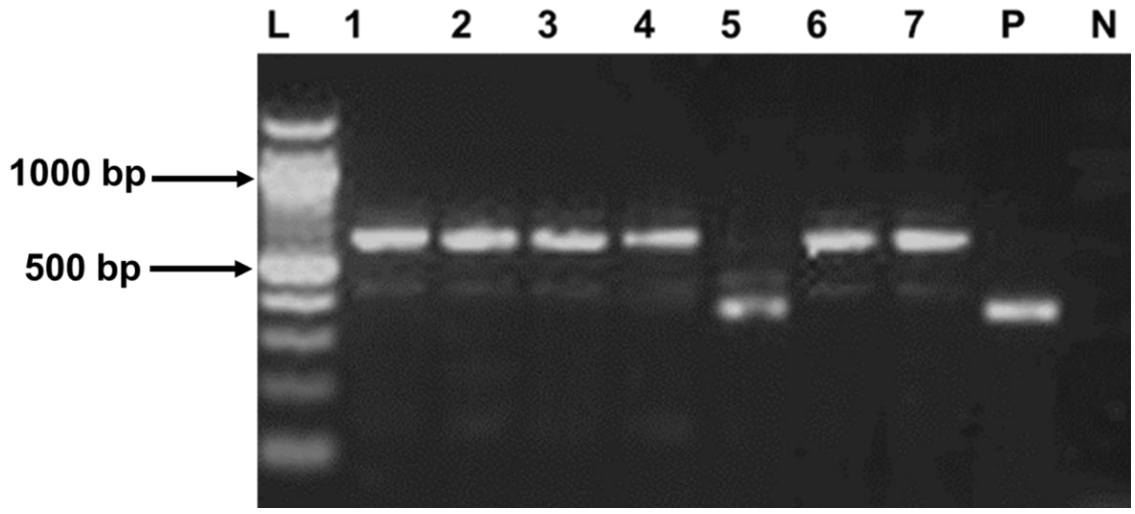

Figure 4.1: Gel image showing variations in mitochondrial DNA COI-COII intergenic region of honeybee samples from Comoros islands. **L** represents 100 bp molecular marker, **1-7** represents the samples from different sites**, P** is the positive control and **N** is the negative control.

4.1.2 COI-COII intergenic region sequence analyses

Two sequence populations 600bp and 380bp, respectively were obtained. Sequence alignment revealed the presence of three major differences among these sequences. Based on these differences the sequences were grouped into L1, L2 and A1 haplotypes (Figures 3). L1 and L2 were newly described while A1 corresponded to *A. m. unicolor* haplotypes common to the SWIO. L1 had a distinctive TAGGT indel starting from position 3697-3702 and a second major AAATCATAATATTG indel at position 3791-3805. L2 was characterized by a TCAAAT indel starting from position 3654-3659. It also had a distinctive TTCA insert from position 3669 - 3672.Both L2 and L1 shared common inserts and an indel among themselves. These included a CTT insertion at positions 3861-3863; an AA insertion at position 3886-3887 and an AATTA indel at position 3613-3617. A1 portrayed similarity to the reference sequence. Apart from the described insertions and indels, the sequences expressed distinct single nucleotide polymorphisms (SNPs) among them. The L2 uniquely contained T3666A, T3701A, T3744C and A3759G SNPs. The L1 haplotype exclusively had SNPs T3665C, T3692G, T3738C and G3814A. Both the L2 and L1 haplotypes uniquely had SNPs that distinguished them which were absent in the A1 haplotype. These included C366T, T3643C, G3651C, T669C, T3734A, A3737T, T3776C, T3821A, A3839T, T3841C, G3850C, A3857T, T3897C, T3904A, T3907C, A3918T and A3922T (Appendix III).

The three haplotypes shared regions of high conservation among themselves. Briefly, both L1 and A1, unlike L2, portrayed high conservation of the TCAAAT ancestral sequence at position 3791-3805. L2 and A1 on the other hand, unlike L1, showed high con-

servation of the ancestral TAG located at position 3697-3699. Furthermore, they had a conservation of AAATCATAATATTG at position 3791-3805. The L2 haplotype appeared to have arisen from a rare mtDNA recombination between L1 and the *A. mellifera* haplotype (4.2).

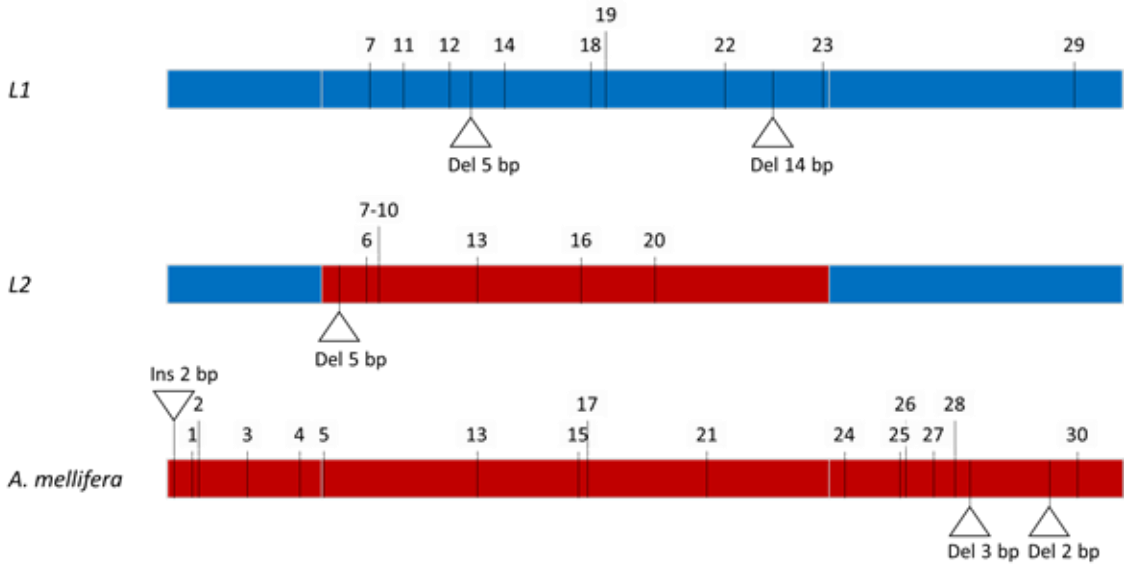

Figure 4.2: Recombinant mtDNA haplotype (L2) as indicated by the shared unique indels and single nucleotide polymorphisms between *A. mellifera* haplotype and L1. The different numerical numbers indicates either unique single nucleotide polymorphisms (SNPs) where the number are unique to each segment or shared SNPs where the numbers are similar between two or more segments. Ins means insertion, while Del means deletion.

A maximum likelihood phylogenetic tree rooted at *A. florea* strongly (bootstrap = 97%) separated the giant honeybees *(A. dorsata, A. laboriosa and A. dorsata bingami)* from the rest. The next separation (bootstrap = 75%) distinguished L1, L2, *A. koschevnikovi* and *A. mellifera* from other cavity nesting bees namely *A. nigrocincta, A. cerana and A. nuluensis*. *A. mellifera* group was partially separated (bootstrap = 23%) from *A. koschevnikovi,* L1 and L2. *A. koschevnikovi* was partially (bootstrap=31%) separated from L1 and L2. Notably, there was 100% separation between L1 and L2. (Figure 4.3). L1 and L2 clades exceptionally contained sequences from Comoros islands. The L1 and L2 clades corre-

sponded to the L1 and L2 haplotypes respectively whilst the *A. mellifera* group had sequences from Comoros and other *A. mellifera* sequences belonging to the A, C and O lineages.

The sequences form other honeybee species were included in the analysis because preliminary blast search of sequences belonging to L1 and L2 haplotypes revealed close similarity to *A. koschevnikovi*. It was thus important to understand how they cluster among all other honeybee species.

Pairwise genetic distances between groups (Table 4.1) showed that L1 and L2 were closely related but distant from the other groups (*Apis mellifera*, cavity nesting bees, giant honeybees and dwarf honeybees). The genetic distance within groups (Table 4.2) revealed high genetic variability within L2 and L1 compared to Apis mellifera represented by samples from A, O and C lineage.

Pairwise genetic distance between groups showed that the distance between L2 and L1 is closer (0.077) than the distance between either L1 (0.107) or L2 (0.116) and *Apis mellifera*. *Apis mellifera* was also shown to be genetically close to the other cavity nesting bees (0.073) as compared to the distance between either of the L2 (0.139) and L1 (0.123) honeybees (Table 2.1). Computing the genetic distance within groups indicated high variability within both L1 (0.0070) and L2 (0.0193) as compared to the variability exhibited in the *Apis mellifera* group (0.0058) (Table 4.2). The variation within the group is how-

ever lower as compared to cavity nesting bees (0.0809), giant honeybees (0.0534) and dwarf honeybees (0.0912) (Table 4.2).

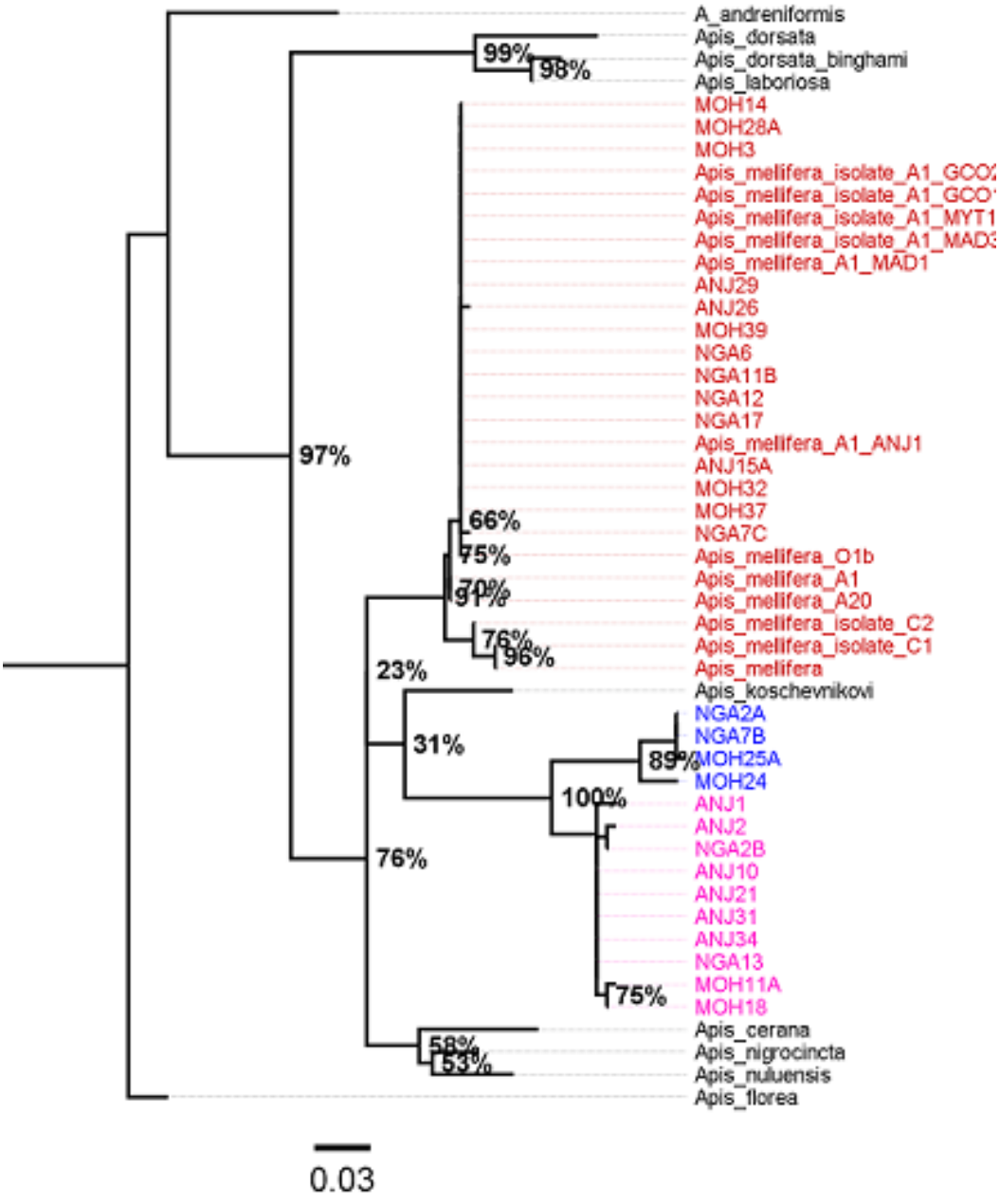

Figure 1.3: Maximum likelihood phylogenetic tree showing clustering of samples from Comoros L1 (Pink), L2 (Blue) and *A. mellifera* (Red) sub-species among other honeybee species. The evolutionary history was inferred by using the Maximum Likelihood method and Tamura 3-parameter model. The tree with the highest log likelihood (-970.63) is

shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.5414)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 9.19% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 49 nucleotide sequences. All positions with less than 95% site coverage were eliminated, i.e., fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position (partial deletion option). There were a total of 217 positions in the final dataset. Evolutionary analyses were conducted in MEGA X.

Table 4.1: Pairwise genetic distances between groups showing a close relationship between L1 and L2 as opposed to the other groups.

		1	**2**	**3**	**4**	**5**	**6**
1.	L2						
2.	L1	0.077					
3.	*Apis mellifera*	0.116	0.107				
4.	Cavity nesting bees	0.139	0.123	0.073			
5.	Giant honeybees	0.180	0.184	0.140	0.180		
6.	Dwarf honeybees	0.208	0.188	0.140	0.150	0.150	

Table 4.2: Genetic distances within groups indicating high genetic variability within L1 and L2 compared to *Apis mellifera* represented by samples from A, C and O lineages.

		Genetic distance
1.	L1	0.0070
2.	L2	0.0193
3.	*Apis mellifera*	0.0058
4.	Cavity nesting bees	0.0809
5.	Giant honeybees	0.0534
6.	Dwarf honeybees	0.0912

The various haplotypes were distributed in the three islands with L1 and the *A. mellifera* haplotype occurring in Anjouan in an almost equal frequency (4.4) The haplotype (s) were distributed such that Ngazidja had 88% (95% CI = 0.80-0.97) of the *A. mellifera* A1 haplotype; 5% (95% CI = 0.0-0.15) of L1 haplotype and of 7% (95% CI = 0.0-0.17) L2 haplotype. Moheli was characterised by 77% (95% CI = 0.68-0.93) of the normal A1 haplotype, 13% (95% CI = 0.03-0.29) of the L1 haplotype and 10% (95% CI = 0.0-0.26) of the L2 haplotype. Anjouan had 52% (95% CI = 0.36-0.70) of the normal A. mellifera haplotype, 48% (95% CI = 0.33 – 0. 67) of the L1 haplotype and 0% (95% CI = 0.00 - 0.10) of L2 haplotype.

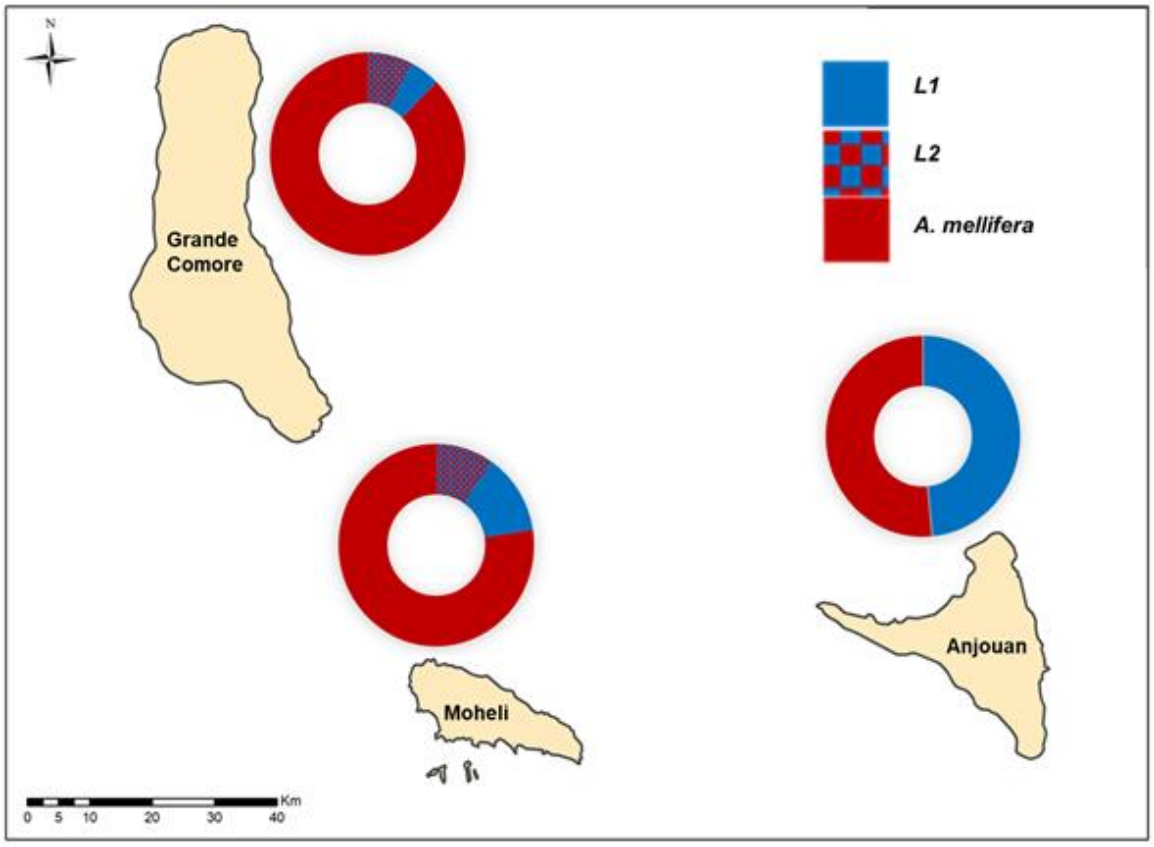

Figure 4.4: Distribution of the three mtDNA haplotypes across the Comoro islands. (Map Credit: Emily Kimathi, icipe Geo-informatics unit)

4.2 Genetic diversity, population differentiation and population structure

4.2.1 Microsatellite DNA amplification

Several microsatellite loci were successfully amplified in a multiplex PCR reaction (Figure 4.5).

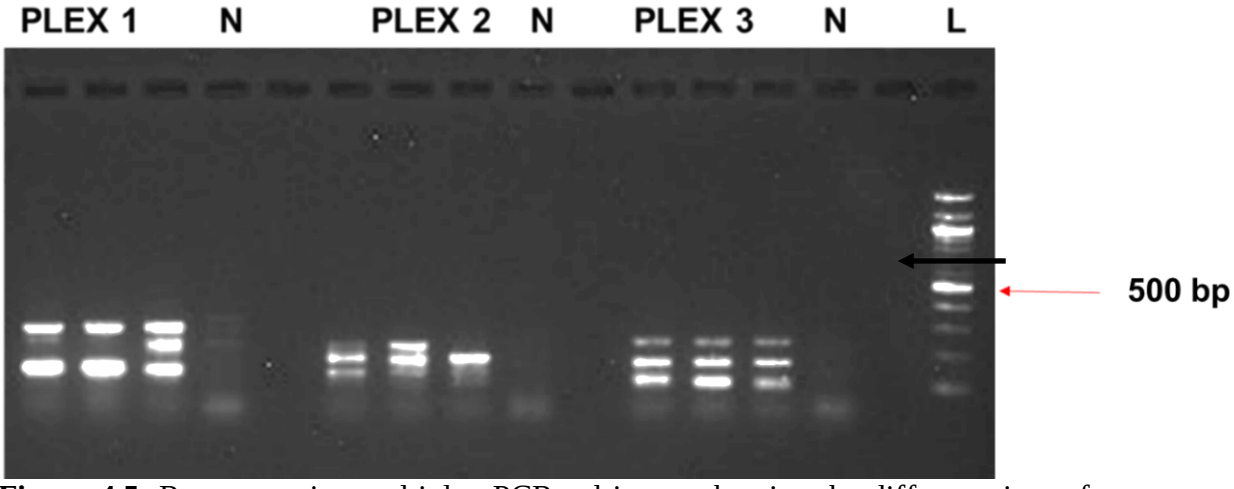

Figure 4.5: Representative multiplex PCR gel image showing the different sizes of amplified fragments using microsatellite markers. **N** represents the negative control while **L** represents a 100 base pairs molecular marker (New England Biolabs)

4.2.2 Fragment analysis and allele scoring

All the sixteen amplified colonies, four from each island, were successfully genotyped. Upon fragment analysis, twelve (12) out of the nineteen (19) loci were successfully scored and alleles generated. From the alleles obtained, four queen genotypes per island were manually scored. A minimum of nineteen (19) and a maximum of twenty four (24) drone genotypes in the various sampled colony were recorded (Table 4.3).

Table 4.3: Table showing the observed and estimated drone genotypes from the different colonies sampled in individual islands.

Colony	Island	Observed drones	Estimated Drones
ANJ16	Anjouan	22	34
ANJ32	Anjouan	22	34
ANJ5	Anjouan	19	27
ANJ8	Anjouan	22	34
MOH12	Moheli	21	31
MOH18	Moheli	22	34
MOH5	Moheli	21	31
MOH23	Moheli	24	38
NGA15	Ngazidja	20	30
NGA23	Ngazidja	21	31
NGA4	Ngazidja	23	36
NGA7B	Ngazidja	19	27

4.2.3 Determination of queen mating frequency

The observed queen mating frequency ranged between 19 and 24 drones (Table 4.3). Due to finite samples size, estimated queen mating frequency was calculated by fitting in a Poisson distribution. The estimated queen mating frequency ranged between 27 and 38 drones. Kruskal-wallis test found no significant difference in the mating frequencies among the three islands ($p=0.395$).

4.2.4 Prediction of queens using drones

Following the input of drones data into Colony software and manual prediction of queens from workers dataset, 145 queens were retrieved with 59 arising from Anjouan, 38 from

Ngazidja and 48 form Moheli. From these data set, 61 queens parented more than two drones, i.e.17 from Anjouan, 24 from Moheli and 20 from Ngazidja (Appendix II).

4.2.4.1 Genetic diversity and linkage disequilibrium

Considering Comoros as a single population, the least number of alleles was recorded at the AP23 locus (2) whilst AP33 locus had the highest number of alleles (12) (Table 9). The mean number of alleles detected was 5.92 (±0.05). Simpson's diversity index per locus ranged between 0.29 (AP23) and 0.84 (A306) with a mean of 0.65 (±0.05) (Table 9). Expected heterozygosity (H_{exp}) per locus ranged between 0.29 (AP23) and 0.85 (AC306).

Comparable genetic diversity was detected between sites. In Moheli the mean number of alleles detected was 3.92 (±0.04) whilst Anjouan and Ngazidja had a mean number of alleles of 3.75 (±0.05) and 3.67 (±0.05), respectively. Simpson's diversity index showed that Ngazidja (0.65±0.08) had the highest mean index of diversity followed by Moheli (0.59±0.04) with Anjouan 0.55±0.08) having the least. Using Nei's 1978 diversity Moheli had the highest expected heterozygosity (0.61±0.03), followed by Anjouan (0.57±0.07) and Ngazidja (0.50±0.10).

Private alleles were recorded at A113 locus where allele 216 was private to Anjouan whilst allele 218 was private to Ngazidja (Figure 4.6). Though not private, allele 108 on AP23 locus was notably absence in Ngazidja (Figure 4.7). Most of the sampled loci showed significant deviations from HWE (Figure 8)

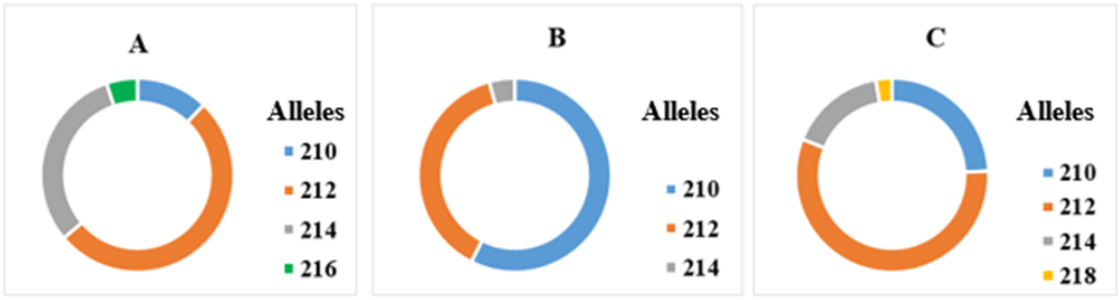

Figure 4.6: Allele distribution at locus A113 in Anjouan (**A**), Moheli (**B**) and Ngazidja (**C**) showing the presence of private alleles 216 and 218 in Anjouan and Ngazidja respectively

Figure 4.7: Allele distribution at locus AP23 in Anjouan (A), Moheli (B) and Ngazidja (C) showing the absence of allele 108 in Ngazidja.

Table 4.4: Table indicating various indices of diversity within the entire Comoros population and within individual islands.

Locus	Entire Comoros Population					Ngazidja				Anjouan				Moheli			
	N_a	Simpson's diversity Index	H_{exp}	Evenness	G_{ST}	N_a	Simpson's diversity Index	H_{exp}	Evenness	N_a	Simpson's diversity Index	H_{exp}	Evenness	N_a	Simpson's diversity Index	H_{exp}	Evenness
A113	3	0.63	0.64	0.93	0.14	3	0.62	0.64	0.91	3	0.61	0.63	0.89	3	0.55	0.57	0.88
A56	4	0.64	0.64	0.90	0.68	4	0.67	0.69	0.91	2	0.36	0.37	0.77	2	0.30	0.31	0.71
A88	6	0.69	0.70	0.80	0.28	3	0.5	0.51	0.74	6	0.78	0.81	0.86	5	0.67	0.69	0.81
AP81	8	0.78	0.79	0.76	0.19	7	0.71	0.73	0.70	6	0.76	0.79	0.80	5	0.77	0.80	0.93
A28	5	0.66	0.67	0.76	0.27	0	1	0.00	0.00	3	0.54	0.56	0.83	5	0.67	0.70	0.78
A8	3	0.32	0.32	0.60	0.36	0	1	0	0.00	1	0.00	0.00	0.00	3	0.49	0.50	0.76
AC306	10	0.84	0.85	0.88	0.76	6	0.67	0.69	0.74	5	0.67	0.70	0.75	4	0.71	0.73	0.91
AP33	12	0.77	0.78	0.61	0.76	5	0.69	0.71	0.82	5	0.61	0.64	0.67	4	0.36	0.38	0.56
A43	5	0.68	0.69	0.81	0.53	4	0.51	0.53	0.72	2	0.45	0.47	0.91	4	0.67	0.69	0.90
AP23	2	0.29	0.29	0.69	0.31	1	0	0	0	2	0.26	0.27	0.66	2	0.48	0.50	0.97
B124	8	0.80	0.81	0.84	0.40	6	0.75	0.77	0.82	5	0.72	0.75	0.86	5	0.74	0.76	0.85
UN4987	5	0.74	0.74	0.84	0.19	5	0.70	0.72	0.81	5	0.78	0.81	0.94	5	0.62	0.64	0.68
Mean± SEM	5.92 ±0.05	0.65 ±0.05	0.66 ±0.03	0.79 ±0.06	0.41±0.068	3.67±0.05	0.65±0.08	0.50±0.10	0.60±0.53	3.75 ±0.05	0.55±0.08	0.57±0.07	0.75± 0.34	3.92±0.04	0.59±0.04	0.61±0.03	0.81± 0.03

N_a, number of alleles; **H*exp***, expected unbiased heterozygosity; **G_{ST}**, Hendrick,s standardised G_{ST}. **SEM**, standard error of Mean

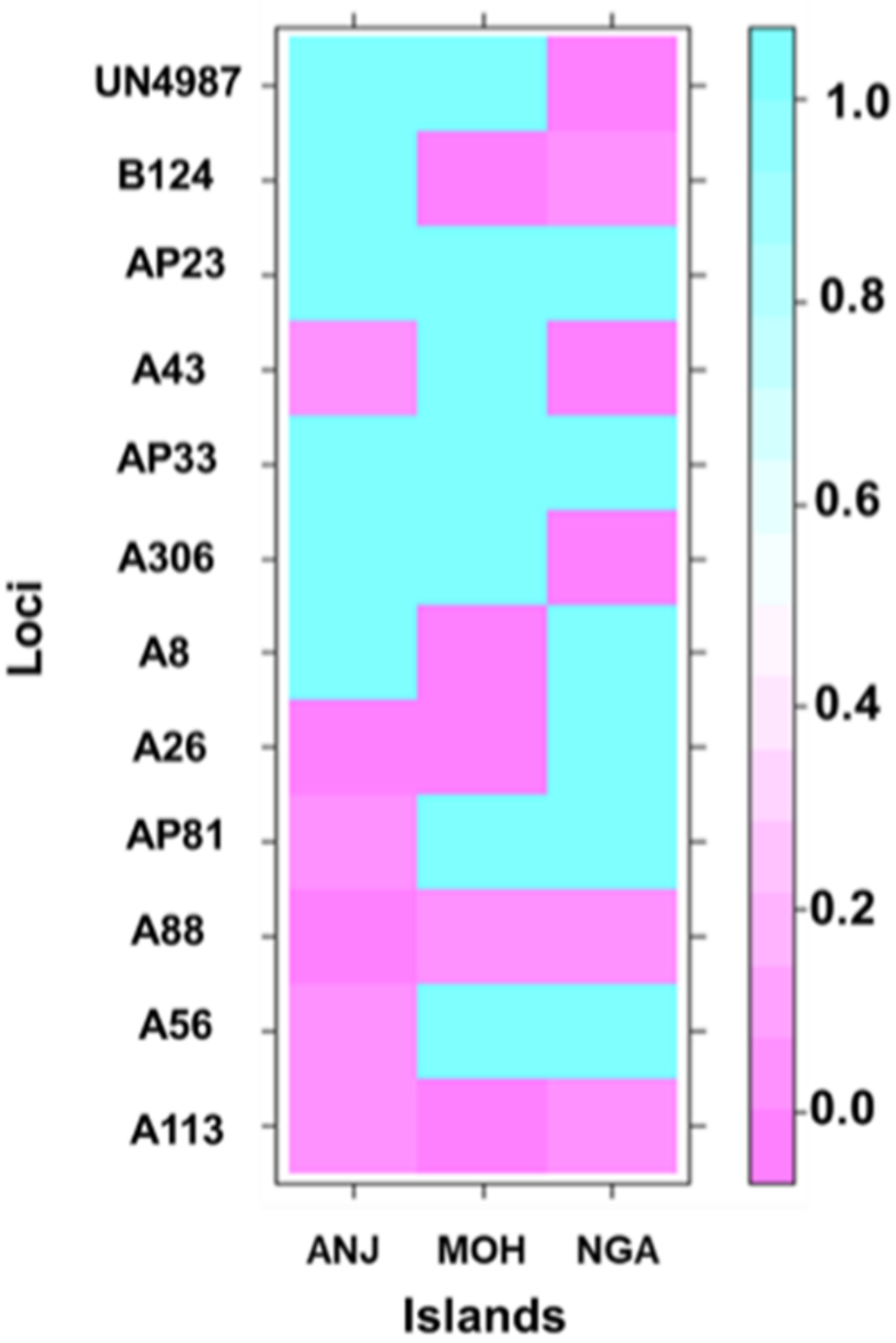

Figure 4.8: Heatmap showing deviations from the HWE. Loci in pink are those suspected of not being in HWE while all the loci in blue conform to the HWE.

Most of the loci were in a linkage disequilibrium. However, markers AP23 and A56 were strongly linked in Anjouan honeybee populations (Figure 4.9). In Ngazidja, markers AP33 and B124; A113 and A88 and AC306 and UN4987 were strongly linkage (Figure 4.10). In Moheli, strong linkage was not detected in any of the loci (Figure 4.11).

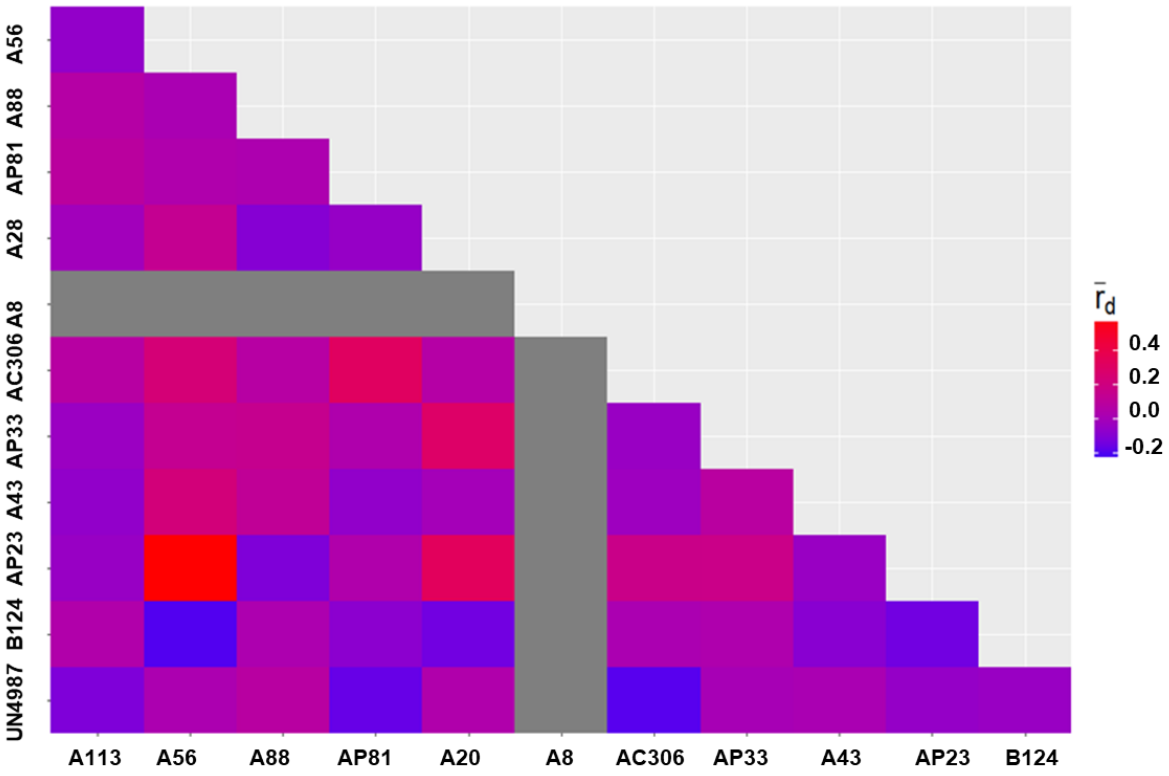

Figure 4.9: Heatmap showing pairwise linkage disequilibrium pattern in honeybees collected in Anjouan. Blue indicates no linkage, red indicates strong linkage between markers whilst grey indicates missing data.

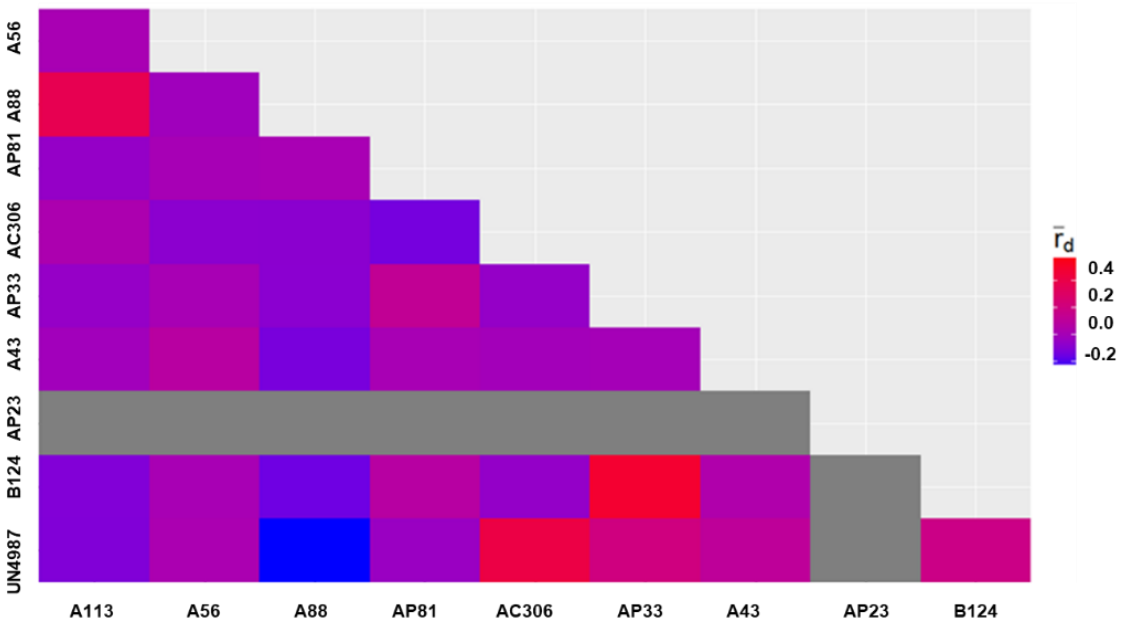

Figure 4.10: Heatmap showing Pairwise linkage disequilibrium pattern in honeybees collected in Ngazidja. Blue indicates no linkage, red indicates strong linkage between markers whilst grey is missing data.

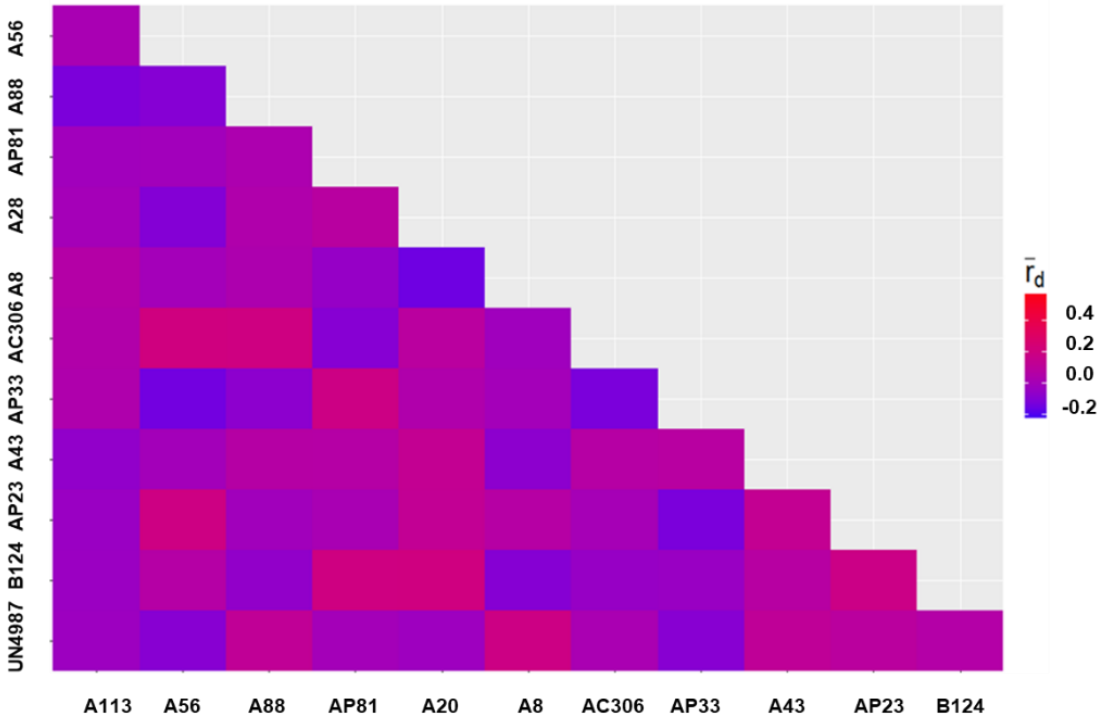

Figure 4.11 Heatmap showing Pairwise linkage disequilibrium pattern in honeybees collected in Moheli. Blue indicates no linkage while red indicates strong linkage between markers

4.2.4.2 Population differentiation and population structure

4.2.4.2.1 Population differentiation and Analysis of Molecular Variance

Using G_{ST} to infer entire population differentiation revealed a strong differentiation (G_{ST}=0.41(±0.068)) (Table 4.4) for the Comoros population. A113 (G_{ST}=0.14) was the least differentiated whilst locus AP33 was strongly differentiated (G_{ST}=0.76) (Table 4.4). Pairwise G_{ST} comparisons between islands showed strong differentiation between all the islands (G_{ST}>0.2) with a much stronger differentiation occurring between Moheli and Ngazidja (0.52) (Table 4.5). Pairwise genetic distance based on Hedrick's G_{ST} showed that Ngazidja and Anjouan were much closer than Ngazidja and Moheli while was almost equidistant from both Ngazidja and Anjouan (Table 4.5). AMOVA test revealed significant variations within and between samples as well as between populations (p = 0.001) (Appendix VI).

Table 4.5: Pairwise Nei's G_{ST} Statistics revealing stronger differentiation between Ngazidja and Moheli as compared to the differentiation between Moheli and Anjouan or Anjouan and Ngazidja.

	ANJ	MOH
MOH	0.368	
NGA	0.380	0.525

Several loci in the different populations showed significant deviations from the HWE (Figure 4.8). (Appendix V).

4.2.4.2.2 Population structure

A neighbor joining tree constructed revealed a population structured according to geography (Figure 4.13). The first level of separation strongly (bootstrap = 100%) separated samples from Anjouan and Ngazidja from Moheli. Samples from Anjouan were partially separated from the Ngazidja. In the DAPC analysis (Figure 4.12), PC 1 separated An-

jouan and Moheli from the Ngazidja, whilst PC 2 separated Anjouan and Moheli populations. The most significant STRUCTURE results (K = 3) also clustered the samples in accordance to the islands of origin (Figure 4.14). Notably, Ngazidja population with 51.9% bootstrap support divided into two sub-populations in the neighbor joining tree. This separation was also seen in the PC 2 of DAPC analysis (Figure 4.12) and STRUCTURE analysis at K=4 (Figure 4.14).

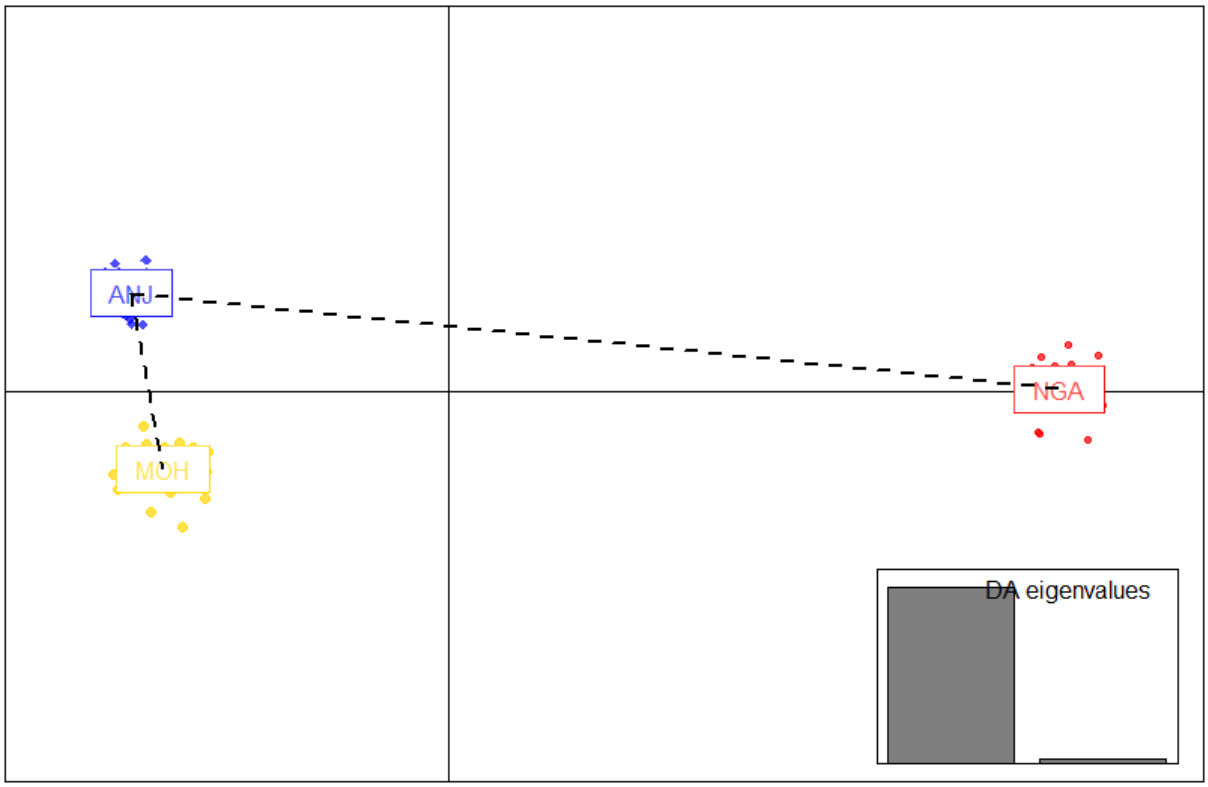

Figure 4.12: Discriminant Analysis of Principal Component (DAPC) image showing how samples from the three islands in Comoros interact in space. ANJ=Anjouan; NGA= Ngazidja; MOH=Moheli.

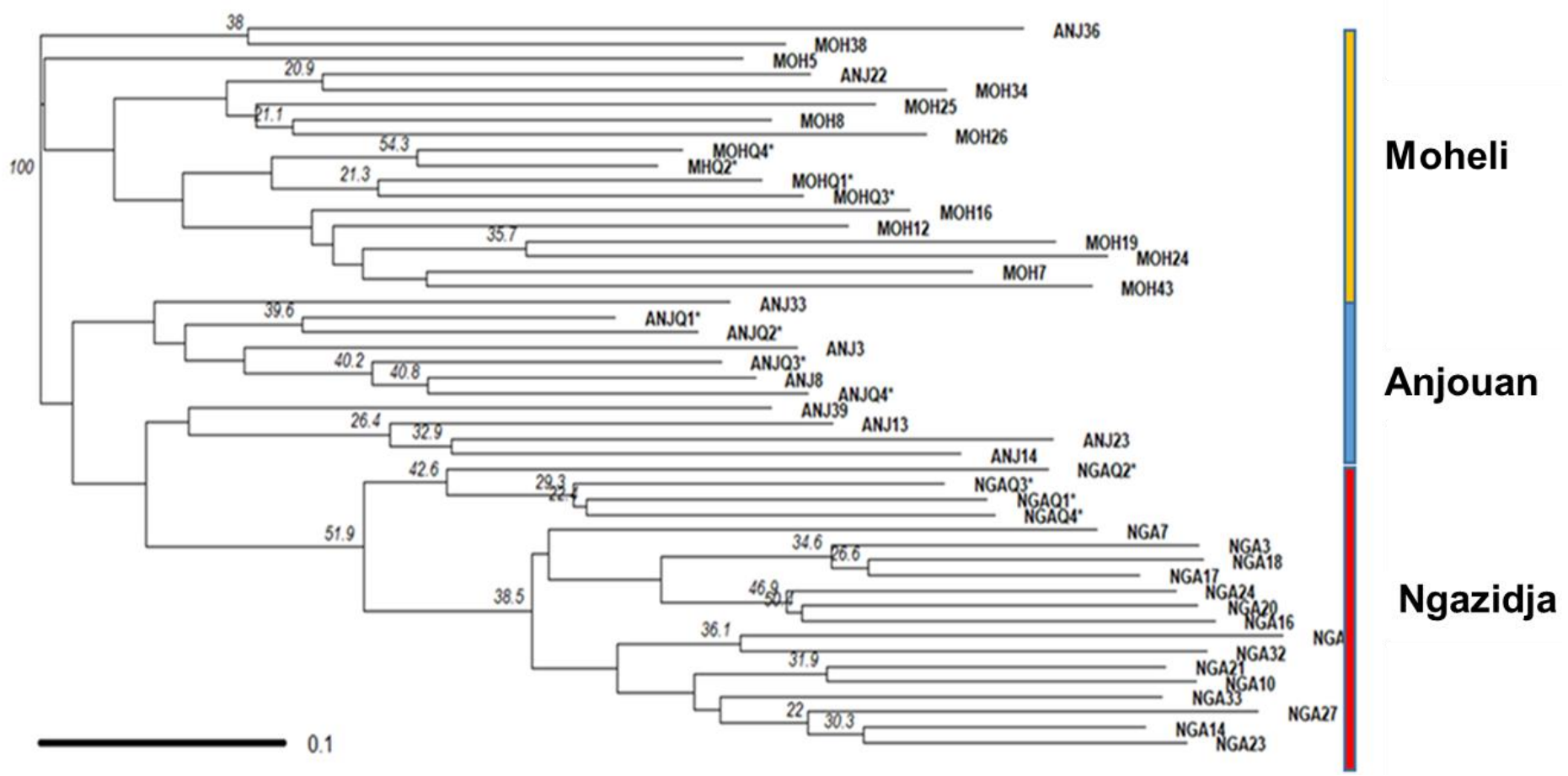

Figure 4.13: A neighbor-joining tree calculated based on 12 microsatellite loci and 1000 bootstraps using Provesti's distance showing a structured population in Comoros islands.

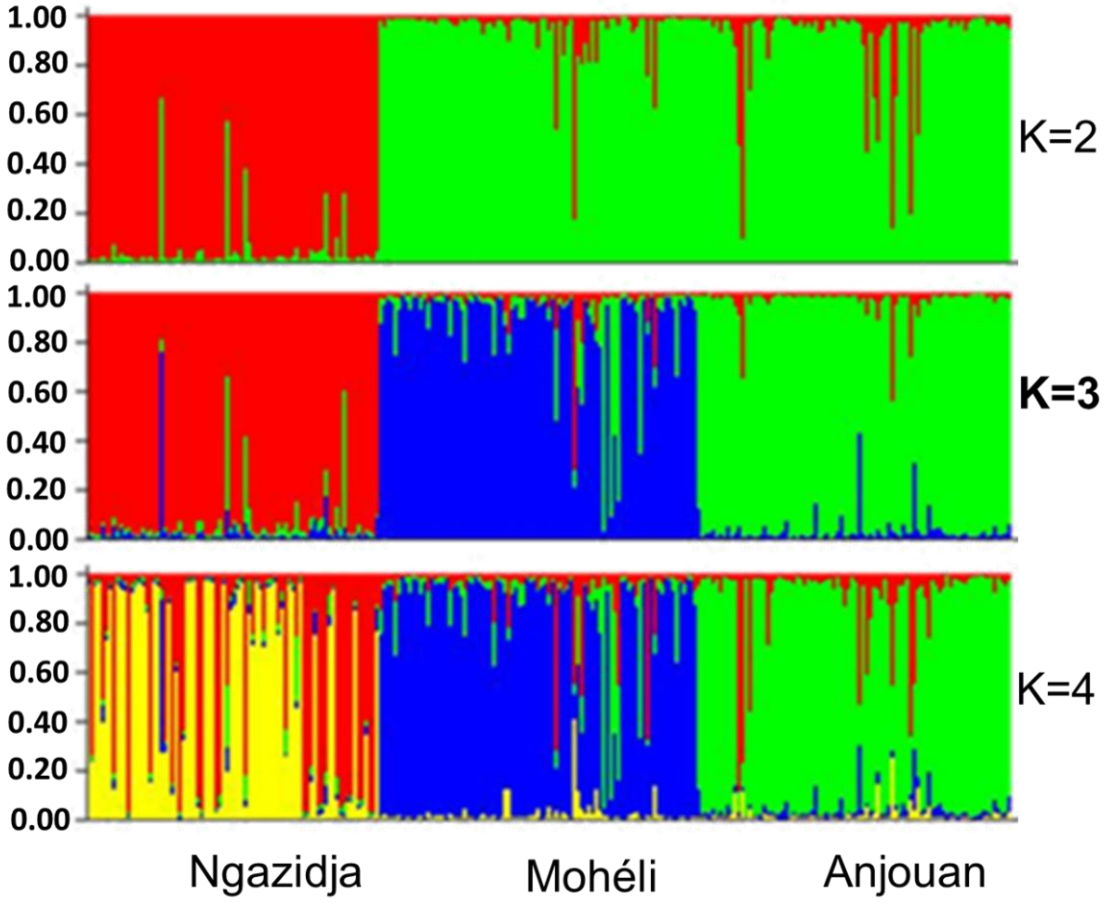

Figure 4.14: Population structure of honeybees in Comoros islands. In the image, the best population structure is achieved at K=3 which indicates structuring according to the island of origin namely, Ngazidja (red), Moheli (blue) and Anjouan (green). K=2 clusters both Moheli and together whilst K=4 shows partial separation in Ngazidja (red and yellow) but does not achieve significant separation in Anjouan.

CHAPTER FIVE

DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS

5.1 Discussion

5.1.1 Mitochondrial DNA haplotypes in Comoros Islands

Mitochondrial DNA sequence analysis showed that the Comoros Islands are colonized by the A lineage of honeybees that are 100 % similar to *A. mellifera unicolar* native to Madagascar (Rasolofoarivao *et al.*, 2015). This results support Techer *et al.* (2017a, 2017b) findings which reported the exclusive presence of A1 lineage in most of the islands to the SWIO except in the Rodrigues islands where the European C lineage was predominant (Techer *et al.*, 2017a). Contrary to Techer *et al.* (2017a, 2017b), these study revealed existence of a honeybee subspecies (L haplotype(s)) that closely clustered alongside the *A. koschevnikovi*, a honeybee subspecies native to Southeast Asia. Close relationship between L1 and L2 with *A. koschevnikovi* needs an explanation.

Human movement from Southeast Asia to Madagascar and Comoros is well documented with Archaeobotanical evidence showing the introduction of Asian crops to the islands during the period (Boivin *et al.*, 2016). It is likely that Southeast Asia honeybees were also introduced in the Comoros islands during that time and have since colonized the island to the present time. The A1 haplotype is distinguishable from both L1 and L2 based on the length of the COI-COII amplified fragment (Figure 4.1). Whereas the A1 fragment is 600 bp the other two are 380 bp in size.

Recombination, the principle mechanism underlying nuclear reorganization is a rare phenomenon in the mtDNA genomes. This study shows that L2 arose from a rare mtDNA recombination between A1 and L1 (Figure 4.2). Cases of recombining mtDNA genomes are well documented in plants, fungi (Barroso and Labarère, 1997) and animals (Ladoukakis and Zouros, 2001; Ciborowski *et al.*, 2007; Good *et al.*, 2008). Heteroplasmy, occurrence of more than one mtDNA or plastid DNA within a cell, has been suggested by previous studies to play a central role in ensuring recombination in the mtDNA genome (Baptista-Ferreira, Economou and Casselton, 1983; Barroso and Labarère, 1997; Saville, Kohli and Anderson, 1998). Though the sperm tail, which harbors the mitochondria, in many animals is excluded from penetrating the egg, there exists evidence of a breach to this barrier in some animals, mostly insects. This result in the incorporation of the male mitochondria into the egg (Hildereth and Lucchesi, 1963). In addition, cases of more than one sperm penetrating the egg among insects are well documented, a scenario known as polyspermy.

Honeybees are known to be characterized by both mechanisms i.e. complete penetration and polyspermy ultimately leading to a more paternal contribution of mtDNA (Meusel and Moritz, 1993) which could then lead to heteroplasmy. Furthermore, the possible leakage of considerable amounts of paternal mtDNA into the egg among the honeybees has been shown (Meusel and Moritz, 1993). Direct evidence of mitochondrial DNA recombination in fungi, plants and protists (Gray, 1989) exists. Besides, fusion of the mtDNA has been shown in *Drosophila* (Yaffe, 1999), Crickets (Rand and Harrison, 1986), scallops(Snyder *et al.*, 1987) as well as sturgeons (Buroker *et al.*, 1990) with non-

homologous recombination being put forward as the probable recombination mechanism involved.

The study also showed, for the first time, the coexistence of two honeybee haplotypes without preferential replacement of one haplotype as reported in different studies (Hall and Muralidharan, 1989; Wragg *et al.*, 2018). The African honeybee *A. mellifera scutellata* was introduced in South America in the early 1950s and has since spread to colonize most countries in the South America slowly replacing the European *A. mellifera mellifera* haplotype. The same scenario was recently observed in La Reunion where European honeybee species of the C lineage were introduced 30 years ago. In the study, the A lineage haplotype was shown to have replaced the C lineage (Wragg *et al.*, 2018).

The coexistence of two haplotypes in Comoros islands suggest that they both exist as part of mitonuclear coadaptation complexes (Hill, 2015) necessary for survival in their habitat. Mitonuclear coadaptation refers to cases where specific nuclear genes are adapted to survive alongside specific mtDNA genes and this necessitates the need for cotransmission of genes involved in the complex formation (Rand, Haney and Fry, 2004; Rogell *et al.*, 2014; Hill, 2015). Nuclear gene conveying key survival adaptations could be linked to the different haplotypes thus the need for their coexistence. Maintenance of two haplotypes in a near equilibrium as observed in Anjouan (4.) could arise from a negative frequency-depended selection (NFDS) mechanism (Kazancıoğlu and Arnqvist, 2014). In this type of selection mechanism, the mitotype with the lowest population frequency increases whilst the mitotype with the highest population frequency decreases until a bal-

ance is achieved (Hill, 2015). An experimental study involving seed beetles (Kazancıoğlu and Arnqvist, 2014) as well as few observation in drosophila (MacRae and Anderson, 1988; Oliver *et al.*, 2005; Andrianov *et al.*, 2008) associates NFDS to the coexistence of two mitochondrial haplotypes in a near stable equilibrium.

5.1.2 Genetic diversity, population differentiation and population structure

The Comoros honeybees exhibit high genetic diversity evidenced by high average expected heterozygosity (0.66 (±0.03)), alleles per locus (5.92 (±0.05)) and Simpson's diversity index (0.65 (±0.05)). The average allele per locus for individual populations ranging between 3.67 (±0.05) and 3.92 (±0.04) is much higher than in Madagascar, where an average of 2.47 and 3.18 alleles per locus within populations were reported (Rasolofoarivao *et al.*, 2015). However, the average alleles per locus for the entire population (5.92 (±0.05)) is below the average in the study in Madagascar (7.76) (Rasolofoarivao *et al.*, 2015).

High diversity within the Comoros honeybee populations could be the result of high queen mating frequency (27 to 38 drones). Increased polyandry mating results in increased offspring heterozygosity (Taylor, Price and Wedell, 2014). The high mating frequency increases the number of reproductively successful males that give rise to increased effective population size (Trontti *et al.*, 2007). On average, honeybees mate with up to 16 drones (Schlüns et al., 2005). The high queen mating frequency recorded in Comoros could arise from the need to reduce the chances of inbreeding and its associated cost. Generally, Islands are inhabited by small populations and given the division of labor nature of Hymenopterans, the census population does not translate to the effective popu-

lation. In order to maintain the small population, it is important to keep the inbreeding rate as low as possible and a higher queen mating frequency would be one check mechanism (Nomura, 2018). The haplodiploid sex determination mechanism in social insects renders inbreeding costly as homozygosity at the sex determining locus results in sterile diploid males which can neither contribute to the workforce nor reproduction (Trontti *et al.*, 2007).

Higher average alleles per locus in Moheli could explain the high diversity (H_{exp} = 0.61 (±0.03)) in the population as compared to Anjouan (H_{exp} = 0.57 (±0.07)) and Ngazidja (H_{exp} = 0.50 (±0.10)) both of which have low average alleles per locus. One possible explanation for the high variability is the small human population in Moheli (50,854) and slow urbanization (Andjib, 2015) which exert minimal fragmentation on the honeybee natural habitats. Anthropogenic disturbance resulting in fragmentation of natural habitats is associated with reduction in species abundance and diversity (Hung, Ascher and Holway, 2017). During sampling, wild honeybee colonies were found nesting on house roofs and walls, old vehicle and around human settlement which suggest the lack of appropriate nesting sites following human interference.

Broadly, high variability at the microsatellite loci corroborates previous studies which showed high polymorphism levels at different microsatellite loci among the African honeybees (Estoup *et al.*, 1995; Franck *et al.*, 1998). The high polymorphism can be explained based on the suggestion by Frank *et al*. (1998), who linked high levels of polymorphism among the African honeybees to their pronounced swarming tendency and

migratory behavior. In addition, it could be as a result of climate change (Frank *et al.*, 2001) or the presence of a large effective population size that gives room for the maintenance of more alleles in a population (Estoup *et al.*, 1995). Large effective population could result from the high queen mating frequency. The high genetic diversity within populations, suggests that the population is panmictic characterised by random mating within breeding populations. Increased polymorphism among the microsatellite loci are a result of random mating (Loucif-Ayad *et al.*, 2015). High nuclear diversity coupled to restricted gene flow within population was reported among honeybees of Seychelles archipelagos whose geographical organization is almost similar to the Comoros Archipelagos (Techer *et al.*, 2015). Findings in honeybees from Comoros thus corroborates Techer *et al.* (2015) suggestions that high variations might arise due to existence of a permanent geographic barrier-in this case the ocean water.

The presence of private alleles 216 and 218 at the A113 locus in Anjouan and Moheli is of interest to note (Figure 4.6). Though present, their limited frequency cannot be used to give a conclusive argument regarding their occurrence. The complete absence of allele 108 and dominance of allele 104 at the AP23 locus in Ngazidja needs an explanation. Genetic drift which is usually associated with random fluctuations in the number of alleles at specific loci is the most likely explanation. In genetic drift, alleles either decrease or increase over time leading to complete loss and fixation of a specific alleles (Sarma *et al.*, 2019). It is also expected that Ngazidja with the lowest genetic diversity is likely to experience a genetic drift among the infrequently occurring alleles as compared on the other islands. The absence of allele 104 in Ngazidja is an indicator of restricted gene flow be-

tween either Anjouan or Moheli and Ngazidja. Significant deviation from the HWE within individual populations suggests that the population is undergoing evolution and that there is inbreeding within the islands.

The geographic structuring, strong pairwise genetic differentiation as well as the significant variance between islands based on AMOVA suggests a restricted gene flow between islands. Isolation by distance of over 50 km is the most possible explanation since honeybees cannot fly over a large water mass separated by such long distances (Techer *et al.*, 2016). Poorly developed beekeeping practices in the islands where most colonies exist as wild colonies deter human aided movement of bees between islands.

The study results suggest that hybridization gene flow occurs in the nuclear genome within islands since no sub-structuring is observed within islands. Sub-structuring of the Ngazidja is non-significant based on STRUCTURE analysis was shown to be non- significant and cannot be used to describe any population structure in Ngazidja. Although, the same sub-structuring was detected in PC 2 of DAPC analysis and neighbor joining tree, the two analyses were based on inferred queen genotypes which have been shown to have negative impact when used to infer population structure (Lepais *et al.*, 2010). Besides, the STRUCTURE result was based on true drone genotypes thus more powerful. These findings are similar to what is reported in Seychelles archipelagos where the population is structured according to the islands but no structure is noted within individual islands (Techer *et al*., 2015).

5.2 Conclusions

This work has shown that fragmentation to a certain scale can result in population structure as the nuclear genome of the honeybees in Comoros is structured according to the islands. Through the work, it is clear that honeybees have the ability to undergo recombination and resulting in unique mitotypes. Through this work it is shown, for the first time, the possibility of existence of a new haplotype of honeybees in the Comoros islands which could have separated from the other cavity nesting bees before *A. mellifera*.

5.3 Recommendations

Since the current study has shown the existence of unique mitotypes Comoros islands, there is need to ensure that the honeybee are adequately conserved in order to retain the endemic species.

5.4 Future studies

Future studies need to investigate the presence of honeybee pathogens in the Comoros islands and compare he rate of infection, and pathogen load among the described haplotypes. Future studies should also investigate the mechanism that govern mtDNA recombination in honeybees. Lastly, there is need to carry out a phenotypic study on the Comoros honeybees and correlate it to the genotypic data presented in this study.

REFERENCES

Abdul-Muneer, P. M. (2014) 'Application of Microsatellite Markers in Conservation Genetics and Fisheries Management: Recent Advances in Population Structure Analysis and Conservation Strategies', *Genetics Research International*. Hindawi Publishing Corporation, 2014.: 1–11. doi: 10.1155/2014/691759.

Alburaki, M., Bertrand, B., Legout, H., Moulin, S., Alburaki, A., Sheppard, W. S. and Garnery, L. (2013) 'A fifth major genetic group among honeybees revealed in Syria', *BMC Genetics*, 14. doi: 10.1186/1471-2156-14-117.

Andjib, S. A. (2015) *Vice President in charge of the Ministry of Territory Development, of Infrastructures, Urbanization and Housing*. Available at: http://habitat3.org/wp-content/uploads/National-Reports-Comoros-English.pdf.

Andrianov, B. V., Sorokina, S. Y., Mugue, N. S., Reznik, N. L. and Mitrofanov, V. G. (2008) 'Dynamics of mitochondrial polymorphism in a natural population of *Drosophila littoralis*', *Russian Journal of Genetics*, 44(2), : 159–164. doi: 10.1007/s11177-008-2006-2.

Arias, M. C. and Sheppard, W. S. (1996) 'Molecular Phylogenetics of Honey Bee Subspecies (*Apis mellifera* L.) Inferred from Mitochondrial DNA Sequence', *Molecular Phylogenetics and Evolution*, 5(3), : 557–566. doi: 10.1006/mpev.1996.0050.

Baptista-Ferreira, J. L. C., Economou, A. and Casselton, L. A. (1983) 'Mitochondrial genetics of *Coprinus*: recombination of mitochondrial genomes', *Current genetics*, 7(5), : 405–407. doi: 10.1038/414745a.

Barroso, G. and Labarère, J. (1997) 'Genetic evidence for nonrandom sorting of mitochondria in the basidiomycete *Agrocybe aegerita*', *Applied and Environmental Microbiology*, 63(12), : 4686–4691. doi: 10.1534/genetics.112.140483.

Baudry, E., Solignac, M., Garnery, L., Gries, M., Cornuet, J. and Koeniger, N. (1998) 'Relatedness among honeybees (*Apis mellifera*) of a drone congregation', *Proceedings of the Royal Society B: Biological Sciences*, 265(1409), : 2009–2014. doi: 10.1098/rspb.1998.0533.

Boivin, N. L., Lucas, L., Fuller, D. Q., Radimilahy, C., Walshaw, S., Pawlowicz, M., Shipton, C., Helm, R., *et al.* (2016) 'Ancient crops provide first archaeological signature of the westward Austronesian expansion', *Proceedings of the National Academy of Sciences*, 113(24),: 6635–6640. doi: 10.1073/pnas.1522714113.

Brown, A. H., Feldman, M. W. and Nevo, E. (1980) 'Multilocus Structure of Natural Populations of *Hordeum spontaneum*.', *Genetics*, 96(2), : 523–536.

Buroker, N. E., Brown, J. R., Gilbert, T. A., O'Hara, P. J., Beckenbach, A. T., Thomas, W. K. and Smith, M. J. (1990) 'Length heteroplasmy of sturgeon mitochondrial DNA: An illegitimate elongation model', *Genetics*, 124(1), : 157–163.

Ciborowski, K. L., Consuegra, S., García de Leániz, C., Beaumont, M. A., Wang, J. and Jordan, W. C. (2007) 'Rare and fleeting: an example of interspecific recombination in animal mitochondrial DNA', *Biology Letters*, 3(5), : 554–557. doi: 10.1098/rsbl.2007.0290.

Cornuet and Garnery (1991) 'Mitochondrial DNA variability in honeybees and its phylogeographic implications', *Apidologie*, 22(6), : 627–642. doi: 10.1051/apido:19910606.

Cornuet, J. and Garnery, L. (1991) 'Putative Origin and Function of the Intergenic Region Between COI and COII of Apis melliferu L. Mitochondrial DNA'.

Desai, S. D. and Currie, R. W. (2015) 'Genetic diversity within honey bee colonies affects pathogen load and relative virus levels in honey bees, Apis mellifera L', *Behavioral Ecology and Sociobiology*, 69(9), pp. 1527–1541. doi: 10.1007/s00265-015-1965-2.

Edgar, R. C. (2004) 'MUSCLE: Multiple sequence alignment with high accuracy and high throughput', *Nucleic Acids Research*, 32(5), : 1792–1797. doi: 10.1093/nar/gkh340.

Eilers, E. J., Kremen, C., Greenleaf, S. S., Garber, A. K. and Klein, A. M. (2011) 'Contribution of pollinator-mediated crops to nutrients in the human food supply', *PLoS ONE*, 6(6). doi: 10.1371/journal.pone.0021363.

Estoup, A., Presa, P., Krieg, F., Vaiman, D. and Guyomard, R. (1993) '(CT)n and (GT)n microsatellites: A new class of genetic markers for *Salmo trutta* L.(brown trout)', *Heredity*, 71(5), : 488–496. doi: 10.1038/hdy.1993.167.

Estoup, A., Garnery, L., Solignac, M. and Cornuet, J. M. (1995) 'Microsatellite variation in honey bee (*Apis mellifera* L.) populations: Hierarchical genetic structure and test of the infinite allele and stepwise mutation models', *Genetics*, 140(2), : 679–695. doi: 10.1089/hum.2006.17.1165.

Estoup, A., Solignac, M. and Cornuet, J.-M. (1994) 'Precise Assessment of the Number of Patrilines and of Genetic Relatedness in Honeybee Colonies', *Proceedings of the Royal Society B: Biological Sciences*, 258(1351), : 1–7. doi: 10.1098/rspb.1994.0133.

Evans, J. D., Schwarz, R. S., Chen, Y. P., Budge, G., Cornman, R. S., De la Rua, P., de Miranda, J. R., Foret, S., *et al.* (2013) 'Standard methods for molecular research in *Apis mellifera*', *Journal of Apicultural Research*, 52(4), : 1–54. doi: 10.3896/IBRA.1.52.4.11.

Excoffier, L., Smouse, P. E. and Quattro, J. M. (1992) 'Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data.', *Genetics*, 131(2), : 479–91. doi: 10.3354/meps198283.

Franck, P., Garnery, L., Solignac, M. and Cornuet, J.-M. (1998) 'The origin of West European subspecies of honeybees (*Apis mellifera*): New insights from

microsattellite and mitochondrial data', *Evolution*, 52(4), : 119–1134.

Franck, P., Garnery, L., Solignac, M. and Cornuet, J. (2000) 'Molecular confirmation of a fourth lineage in honeybees from the Near East', *Apidologie*, 31, : 167–180. doi: 10.1051/apido:2000114.

Franck, P., Garnery, L., Loiseau, A., Oldroyd, B. P., Hepburn, H. R., Solignac, M. and Cornuet, J. M. (2001) 'Genetic diversity of the honeybee in Africa: Microsatellite and mitochondrial data', *Heredity*, 86(4), : 420–430. doi: 10.1046/j.1365-2540.2001.00842.x.

Fuchs, S. and SCHADE, V. (1994) 'Lower Performance in Honeybee Colonies of Uniform Paternity', *Apidologie*, *Springer Verlag*, 25(2), : 155–168. doi: 10.1051/apido:19940204.

Gallai, N., Salles, J. M., Settele, J. and Vaissière, B. E. (2009) 'Economic valuation of the vulnerability of world agriculture confronted with pollinator decline', *Ecological Economics*. Elsevier B.V., 68(3), pp. 810–821. doi: 10.1016/j.ecolecon.2008.06.014.

Garnery, L., Solignac, M., Celebrano, G. and Cornuet, J. (1993) 'A simple test using restricted PCR-amplifled mitochondrial DNA to study the genetic structure of *Apis mellifera* L .', (18).

Garnery, L., Cornuet, J.-M. and Solignac, M. (1992) 'Evolutionary history of the honey bee Apis mellifera inferred from mitochondrial DNA analysis', *Molecular Ecology*, 1(3), : 145–154. doi: 10.1111/j.1365-294X.1992.tb00170.x.

Good, J. M., Hird, S., Reid, N., Demboski, J. R., Steppan, S. J., Martin-Nims, T. R. and Sullivan, J. (2008) 'Ancient hybridization and mitochondrial capture between two species of chipmunks', *Molecular Ecology*, 17(5), : 1313–1327. doi: 10.1111/j.1365-294X.2007.03640.x.

Gray, M. W. (1989) 'Origin and Evolution of Mitochondrial DNA', *Annual Reviews Cell Biology*, : 25–50.

Grünwald, N. J., Goodwin, S. B., Milgroom, M. G. and Fry, W. E. (2003) 'Analysis of Genotypic Diversity Data for Populations of Microorganisms', *Phytopathology*, 93(6), : 738–746. doi: 10.1094/PHYTO.2003.93.6.738.

Hall, H. G. and Muralidharan, K. (1989) 'Evidence from mitochondrial DNA that African honey bees spread as continuous maternal lineages', *Nature*, 339(6221), : 211–213. doi: 10.1038/339211a0.

Han, F., Wallberg, A. and Webster, M. T. (2012) 'From where did the western honeybee (*Apis mellifera*) originate?', *Ecology and Evolution*, 2(8), pp. 1949–1957. doi: 10.1002/ece3.312.

Hedrick, P. W. (2005) 'A standardized genetic differentiation measure.', *Evolution; international journal of organic evolution*, 59(8), : 1633–1638. doi: Doi 10.1554/05-076.1.

Hildereth, P. E. and Lucchesi, J. C. (1963) 'Fertilization in Drosophila I. Evidence for

the Regular Occurence of Monospermy', *Developmental Biology*, 6, : 262–278.

Hill, G. E. (2015) 'Mitonuclear Ecology', *Molecular Biology and Evolution*, 32(8), :1917–1927. doi: 10.1093/molbev/msv104.

Hung, K.-L. J., Ascher, J. S. and Holway, D. A. (2017) 'Urbanization-induced habitat fragmentation erodes multiple components of temporal diversity in a Southern California native bee assemblage', *PLOS ONE*. Edited by N. E. Raine, 12(8), p. e0184136. doi: 10.1371/journal.pone.0184136.

Johnson, R. M., Ellis, M. D., Mullin, C. A. and Frazier, M. (2010) 'Pesticides and honey bee toxicity – USA', *Apidologie*, 41(3), :. 312–331. doi: 10.1051/apido/2010018.

Jombart, T., Devillard, S. and Balloux, F. (2010) 'Discriminant analysis of principal components: a new method for the analysis of genetically structured populations', *BMC Genetics*, 11(1), p. 94. doi: 10.1186/1471-2156-11-94.

Jones, O. R. and Wang, J. (2010) 'COLONY: A program for parentage and sibship inference from multilocus genotype data', *Molecular Ecology Resources*, 10(3), : 551–555. doi: 10.1111/j.1755-0998.2009.02787.x.

Kakumanu, M. L., Reeves, A. M., Anderson, T. D., Rodrigues, R. R. and Williams, M. A. (2016) 'Honey bee gut microbiome is altered by in-hive pesticide exposures', *Frontiers in Microbiology*, 7(AUG), pp. 1–11. doi: 10.3389/fmicb.2016.01255.

Kamvar, Z. N., Tabima, J. F. and Grünwald, N. J. (2014) '*Poppr* : an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction', *PeerJ*, 2, p. e281. doi: 10.7717/peerj.281.

Kazancıoğlu, E. and Arnqvist, G. (2014) 'The maintenance of mitochondrial genetic variation by negative frequency-dependent selection', *Ecology Letters*. Edited by D. Ebert, 17(1), : 22–27. doi: 10.1111/ele.12195.

Klein, A.-M., Vaissiere, B. E., Cane, J. H., Steffan-Dewenter, I., Cunningham, S. A., Kremen, C. and Tscharntke, T. (2007) 'Importance of pollinators in changing landscapes for world crops', *Proceedings of the Royal Society B: Biological Sciences*, 274(1608), : 303–313. doi: 10.1098/rspb.2006.3721.

Delaplane, K. S. andamp; Mayer, D. F. (2000). Crop Pollination by Bees. - New York, Oxon (CABI Publishing). ISBN 0-85199-448-2

Kraus, F. B., Franck, P. and Vandame, R. (2007) 'Asymmetric introgression of African genes in honeybee populations (*Apis mellifera* L.) in Central Mexico', *Heredity*, 99(2),: 233–240. doi: 10.1038/sj.hdy.6800988.

Kumar, S., Stecher, G., Peterson, D. and Tamura, K. (2012) 'MEGA-CC: Computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis', *Bioinformatics*, 28(20), : 2685–2686. doi: 10.1093/bioinformatics/bts507.

Ladoukakis, E. D. and Zouros, E. (2001) 'Direct Evidence for Homologous

Recombination in Mussel (*Mytilus galloprovincialis*) Mitochondrial DNA', *Molecular Biology and Evolution*, (May), : 1168–1175.

Lepais, O., Darvill, B., O'Connor, S., Osborne, J. L., Sanderson, R. A., Cussans, J., Goffe, L. and Goulson, D. (2010) 'Estimation of bumblebee queen dispersal distances using sibship reconstruction method', *Molecular Ecology*, 19(4), : 819–831. doi: 10.1111/j.1365-294X.2009.04500.x.

Loucif-Ayad, W., Achou, M., Legout, H., Alburaki, M. and Garnery, L. (2015) 'Genetic assessment of Algerian honeybee populations by microsatellite markers', *Apidologie*, 46(3), : 392–402. doi: 10.1007/s13592-014-0331-0.

MacDonald, J. H. (2009) 'Kruskal-Wallis Test', *Biological Handbook of Statistics*, (1), : 165–172. doi: 10.1002/9780470479216.corpsy0491.

MacRae, A. F. and Anderson, W. W. (1988) 'Evidence for non-neutrality of mitochondrial DNA haplotypes in *Drosophila pseudoobscura*.', *Genetics*, 120(2), : 485–494.

Magnus, R. ., Tripodi, A. . and Szalanski, A. . (2011) 'Mitochondrial DNA diversity of honey bees, *Apis mellifera* L. (Hymenoptera: Apidae) from queen breeders in the United States', *Journal of Apicultural Science*, 55(1), : 37–47.

Mattila, H. R. and Seeley, T. D. (2007) 'Genetic diversity in honey bee colonies enhances productivity and fitness', *Science*, 317(5836), pp. 362–364. doi: 10.1126/science.1143046.

McMenamin, A. J. and Genersch, E. (2015) 'Honey bee colony losses and associated viruses', *Current Opinion in Insect Science*. Elsevier Inc, 8, : 121–129. doi: 10.1016/j.cois.2015.01.015.

Meusel, M. S. and Moritz, R. F. A. (1993) 'Transfer of paternal mitochondrial DNA during fertilization of honeybee (*Apis mellifera* L.) eggs', *Current Genetics*, 24(6), : 539–543. doi: 10.1007/BF00351719.

Michener, C. D. (2000) *The Bees of the World*. 2715 North Charles Street Baltimore, Maryland 21218-4363: The John Hopkins University Press. Available at: www.press.jhu.edu.

Nei, M. (1972) 'Genetic Distance between Populations', *The American Naturalist*, 106(949), : 283–292. doi: 10.1086/282771.

Nei, M. (1978) 'Estimation of average heterozygosity and genetic distance from a small number of individuals', *Genetics*, 89(3), : 583–590. doi: 10.3390/ijms15010277.

Nei, M. (1987) *Molecular Evolutionary Genetics*, *Columbia University Press*.

Neumann, P., Moritz, R. F. A., Mautz, D. and Ökologie, M. (1999) 'Using DNA microsatellites for maternity testing in honeybees (*Apis mellifera* L .) To cite this version : HAL Id : hal-00891644 Using DNA microsatellites for maternity testing in honeybees (*Apis mellifera* L .) a'.

Nomura, T. (2018) 'Maximum avoidance of inbreeding in haplodiploid populations', *Mathematical Biosciences*. Elsevier, 306 (7), : 49–55. doi: 10.1016/j.mbs.2018.10.006.

Oldroyd, B. P., Rinderer, T. E. and Buco, S. M. (1992) 'Intra-colonial foraging specialism by honey bees (*Apis mellifera*) (Hymenoptera: Apidae)', *Behavioral Ecology and Sociobiology*, 30, : 291–295. doi: 10.1007/BF00170594.

Oliver, P., Balanya, J., Ramon, M., Picornell, A., Serra, L., Moya, A. and Castro, J. A. (2005) 'Population dynamics of the 2 major mitochondrial DNA haplotypes in experimental populations of *Drosophila subobscura*', *Genome*, 48(6), : 1010–1018. doi: 10.1139/g05-077.

Palmer, K. A. and Oldroyd, B. P. (2003) 'Evidence for intra-colonial genetic variance in resistance to American foulbrood of honey bees (*Apis mellifera*): Further support for the parasite/pathogen hypothesis for the evolution of polyandry', *Naturwissenschaften*, 90(6), : 265–268. doi: 10.1007/s00114-003-0418-3.

Péntek-Zakar, E., Oleksa, A., Borowik, T. and Kusza, S. (2015) 'Population structure of honey bees in the Carpathian Basin (Hungary) confirms introgression from surrounding subspecies', *Ecology and Evolution*, 5(23), : 5456–5467. doi: 10.1002/ece3.1781.

Pirk, C. W. W., Strauss, U., Yusuf, A. A., Démares, F. and Human, H. (2016) 'Honeybee health in Africa—a review', *Apidologie*, 47(3), : 276–300. doi: 10.1007/s13592-015-0406-6.

Potts, S. G., Biesmeijer, J. C., Kremen, C., Neumann, P., Schweiger, O. and Kunin, W. E. (2010) 'Global pollinator declines: trends, impacts and drivers', *Trends in Ecology and Evolution*. Elsevier Ltd, 25(6), : 345–353. doi: 10.1016/j.tree.2010.01.007.

Pritchard, J. K., Stephens, M. and Donnelly, P. (2000) 'Inference of population structure using multilocus genotype data', *Genetics*, 155, : 945–959. doi: 10.1111/j.1471-8286.2007.01758.x.

Quarles, W. (2008) 'Pesticides and honey bee colony collapse disorder.', *Ipm practitioner*, 30(9/10), : 1–10.

Rand, D. M., Haney, R. A. and Fry, A. J. (2004) 'Cytonuclear coevolution: The genomics of cooperation', *Trends in Ecology and Evolution*, 19(12), : 645–653. doi: 10.1016/j.tree.2004.10.003.

Rand, D. M. and Harrison, R. G. (1986) 'Mitochondrial DNA transmission genetics in crickets.', *Genetics*, 114(3), : 955–970.

Rasolofoarivao, H., Clémencet, J., Techer, M. A., Ravaomanarivo, L. H. R., Reynaud, B. and Delatte, H. (2015) 'Genetic diversity of the endemic honeybee: *Apis mellifera unicolor* (Hymenoptera: Apidae) in Madagascar', *Apidologie*, 46(6), : 735–747. doi: 10.1007/s13592-015-0362-1.

Rogell, B., Dean, R., Lemos, B. and Dowling, D. K. (2014) 'Mito-nuclear interactions

as drivers of gene movement on and off the X-chromosome', *BMC Genomics*, 15(1), : 1–9. doi: 10.1186/1471-2164-15-330.

Ruttner, F. (1966) 'The Life and Flight Activity of Drones', *Bee World*, 47(3), : 93–100. doi: 10.1080/0005772X.1966.11097111.

Ruttner, Friedrich (1988) *Biogeography and Taxonomy of Honeybees*. 1st edn. Berlin: Springer-Verlag Berlin Heidelberg GmbH.

Ruttner, F., Tassencourt, L. and Louveaux, J. (1978) 'Biometrical-statistical analysis of the geographical variability of Apis mellifera L.', *Apidolodie*, 9(4), : 363–381. doi: 10.1051/apido:19780408.

Sarma, H., Pradhan, S., Mattaparthi, V. S. K. and Kaushik, S. (2019) 'Phylogenetic Analysis: Early Evolution of Life', in *Encyclopedia of Bioinformatics and Computational Biology*. Elsevier, : 938–952. doi: 10.1016/B978-0-12-809633-8.20171-4.

Saville, B. J., Kohli, Y. and Anderson, J. B. (1998) 'mtDNA recombination in a natural population', *Proceedings of the National Academy of Sciences*, 95(February), : 1331–1335.

Schlüns, H., Moritz, R. F. A., Neumann, P., Kryger, P. and Koeniger, G. (2005) 'Multiple nuptial flights, sperm transfer and the evolution of extreme polyandry in honeybee queens', *Animal Behaviour*, 70(1), : 125–131. doi: 10.1016/j.anbehav.2004.11.005.

Schmid-Hempel, P. (1995) 'Parasites and social insects', 26(3), : 255–271.

Shaibi, T., Lattorff, H. M. G. and Moritz, R. F. A. (2008) 'A microsatellite DNA toolkit for studying population structure in *Apis mellifera*', *Molecular Ecology Resources*, 8(5), : 1034–1036. doi: 10.1111/j.1755-0998.2008.02146.x.

Sheppard, W. S. and Smith, D. R. (2000) 'Identification of African-Derived Bees in the Americas: A Survey of Methods', *Annals of the Entomological Society of America*, 93(2), : 159–176. doi: 10.1603/0013-8746(2000)093[0159:IOADBI]2.0.CO;2.

Sherman, P. W., Seeley, T. D. and Reeve, H. K. (1988) 'Parasites , Pathogens , and Polyandry in Social Hymenoptera Author (s): Paul W . Sherman , Thomas D . Seeley and Hudson K . Reeve Published by : The University of Chicago Press for The American Society of Naturalists Stable URL : http://www.jstor.org/st', 131(4), : 602–610.

Simpson, E. H. (1949) 'Measurement of Diversity', *Nature*, : 688–688. doi: 10.1038/163688a0.

Smith, J. M., Smith, N. H., O'Rourke, M. and Spratt, B. G. (1993) 'How clonal are bacteria?', *Proceedings of the National Academy of Sciences*, 90(10), : 4384–4388. doi: 10.1073/pnas.90.10.4384.

Snyder, M., Fraser, A. R., LaRoche, J., Gartner-Kepkay, K. E. and Zouros, E. (1987) 'Atypical mitochondrial DNA from the deep-sea scallop Placopecten

magellanicus', *Proceedings of the National Academy of Sciences*, 84(21), : 7595–7599. doi: 10.1073/pnas.84.21.7595.

Solignac, M., Vautrin, D., Loiseau, A., Mougel, F., Baudry, E., Estoup, A., Garnery, L., Haberl, M. and Cornuet, J. M. (2003) 'Five hundred and fifty microsatellite markers for the study of the honeybee (*Apis mellifera* L.) genome', *Molecular Ecology Notes*, 3(2), : 307–311. doi: 10.1046/j.1471-8286.2003.00436.x.

Stoddart, J. A. and Taylor, J. F. (1988) 'Genotypic diversity: Estimation and prediction in samples', *Genetics*, 118(4), : 705–711.

Tarpy, D. R. (2003) 'Genetic diversity within honeybee colonies prevents severe infections and promotes colony growth', *Proceedings of the Royal Society B: Biological Sciences*, 270(1510), : 99–103. doi: 10.1098/rspb.2002.2199.

Tarpy, D. R., Vanengelsdorp, D. and Pettis, J. S. (2013) 'Genetic diversity affects colony survivorship in commercial honey bee colonies', *Naturwissenschaften*, 100(8), : 723–728. doi: 10.1007/s00114-013-1065-y.

Taylor, M. L., Price, T. A. R. and Wedell, N. (2014) 'Polyandry in nature: A global analysis', *Trends in Ecology and Evolution*. Elsevier Ltd, 29(7), : 376–383. doi: 10.1016/j.tree.2014.04.005.

Techer, M. A., Clémencet, J., Turpin, P., Volbert, N., Reynaud, B. and Delatte, H. (2015) 'Genetic characterization of the honeybee (*Apis mellifera*) population of Rodrigues Island, based on microsatellite and mitochondrial DNA', *Apidologie*, 46(4), : 445–454. doi: 10.1007/s13592-014-0335-9.

Techer, M. A., Clémencet, J., Simiand, C., Portlouis, G., Reynaud, B. and Delatte, H. (2016) 'Genetic diversity of the honeybee (*Apis mellifera* L.) populations in the Seychelles archipelago', *Insect Conservation and Diversity*. Edited by S. R. Leather and D. Roubik, 9(1), : 13–26. doi: 10.1111/icad.12138.

Techer, M. A., Clémencet, J., Simiand, C., Turpin, P., Garnery, L., Reynaud, B. and Delatte, H. (2017a) 'Genetic diversity and differentiation among insular honey bee populations in the southwest Indian Ocean likely reflect old geographical isolation and modern introductions', *PLoS ONE*, 12(12), : 1–26. doi: 10.1371/journal.pone.0189234.

Techer, M. A., Clémencet, J., Simiand, C., Preeaduth, S., Azali, H. A., Reynaud, B. and Hélène, D. (2017b) 'Large-scale mitochondrial DNA analysis of native honey bee *Apis mellifera* populations reveals a new African subgroup private to the South West Indian Ocean islands', *BMC Genetics*. BMC Genetics, 18(1), : 1–21. doi: 10.1186/s12863-017-0520-8.

Trontti, K., Thurin, N., Sundstrom, L. and Aron, S. (2007) 'Mating for convenience or genetic diversity? Mating patterns in the polygynous ant *Plagiolepis pygmaea*', *Behavioral Ecology*, 18(2), : 298–303. doi: 10.1093/beheco/arl083.

Vallet, A. M. and Coles, J. A. (1993) 'The perception of small objects by the drone honeybee', *Journal of Comparative Physiology A: Sensory, Neural and Behavioral*

Physiology, 172(2), : 183–188. doi: 10.1007/BF00189395.

VanEngelsdorp, D., Evans, J. D., Saegerman, C., Mullin, C., Haubruge, E., Nguyen, B. K., Frazier, M., Frazier, J., *et al.* (2009) 'Colony collapse disorder: A descriptive study', *PLoS ONE*, 4(8). doi: 10.1371/journal.pone.0006481.

Wallberg, A., Han, F., Wellhagen, G., Dahle, B., Kawata, M., Haddad, N., Luz, Z., Simões, P., *et al.* (2014) 'A worldwide survey of genome sequence variation provides insight into the evolutionary history of the honeybee *Apis mellifera*', *Nature Publishing Group*. Nature Publishing Group, 46(10), : 1081–1088. doi: 10.1038/ng.3077.

Watanabe, M. E. (2008) 'Colony Collapse Disorder: Many Suspects, No Smoking Gun', *BioScience*, 58(5),: 384. doi: 10.1641/B580503.

Whitfield, C. W., Behura, S. K., Berlocher, S. H., Clark, A. G., Johnston, J. S., Sheppard, W. S., Smith, D. R. and Suarez, A. V (2006) 'Thrice Out of Africa : Ancient and Recent Expansions of the Honey Bee, *Apis mellifera*', *Science*, 314(October), : 642–645. doi: 10.1126/science.1132772.

Wilson, E. O. (1971) 'The Insect Societies', in *The Insect Societies*, p. 548.

Woyke, J. (1962) 'Natural and Artificial Insemination of Queen Honeybees', *Bee World*, 43(1), : 21–25. doi: 10.1080/0005772X.1962.11096922.

Wragg, D., Techer, M. A., Canale-Tabet, K., Basso, B., Bidanel, J.-P., Labarthe, E., Bouchez, O., Le Conte, Y., Clémencet, J., Delatte, H. and Vignal, A. (2018) 'Autosomal and Mitochondrial Adaptation Following Admixture: A Case Study on the Honeybees of Reunion Island', *Genome Biology and Evolution*, 10(1), : 220–238. doi: 10.1093/gbe/evx247.

Yaffe, M. P. (1999) 'The machinery of mitochondrial inheritance and behavior', *Science*, : 1493–1497. doi: 10.1126/science.283.5407.1493.

APPENDICES

APPENDIX I: Sampled colonies in the three main islands of Comoros. Mitochondrial DNA analysis was performed on all the sampled colonies whilst the colonies in with * (*N=12*) were used for microsatellite DNA analysis.

Colony	Site Name	Latitude	Longitude	Type of colony	Altitude (m)
Ngazidja 1	-	S11° 44.176’	E43° 14.884’	Wild	63
Ngazidja 2	-	S11° 46.442’	E43° 16.868’	Apiary	307
Ngazidja 3	Bweni	S11° 45.456’	E43° 15.567’	Wild	59
Ngazidja 4*	Iconi	S11° 44.823’	E43°14.464’	Wild	14
Ngazidja 5	Iconi	S11°49.063’	E43°16.925’	Wild	41
Ngazidja 6	Singani	S11°50.817’	E43° 19.103’	Wild	140
Ngazidja 7*	Tsinimoichongo	S11° 51.137’	E43° 28.027’	Apiary	111
Ngazidja 8	Panda	S11° 52.586’	E43° 21.979’	Wild	112
Ngazidja 9	Simbusa	S11° 53.138’	E43° 28.027’	Wild	538
Ngazidja 10	Nyuma Milima	S11°51.201’	E43° 26.758’	Wild	669
Ngazidja 11	Tsini Mapanga	S11° 47.636’	E43° 26.233’	Apiary	63
Ngazidja 12	Usipvo	S11°34.980’	E43° 16.087’	Wild	31
Ngazidja 13	Usipvo	S11° 35.157’	E43° 16.712’	Wild	155
Ngazidja 14	Kwambani	S11° 37.576’	E43° 21.212’	Apiary	503
Ngazidja 15*	Bweni	S11° 38.252’	E43° 21.656’	Wild	482
Ngazidja 16	Sada	S11°38.301’	E43° 21.855’	Wild	406
Ngazidja 17	-	S11° 39.009’	E43° 22.150’	Wild	370
Ngazidja 18	Dzahani	S11° 39.554	E43° 17.044’	Traditional hive	369
Ngazidja 19	Ifundihe	S11° 30.851’	E43° 23.231’	Wild	255
Ngazidja 20	Babadyani	S11° 29.038’	E43° 23.340’	Wild	253
Ngazidja 21	Babadyani	S11° 29.160’	E43° 23.246’	Wild	268
Ngazidja 22	Mwadja	S11°27.704’	E43° 23.362’	Wild	332
Ngazidja 23*	Ipvwani	S11° 24.150’	E43° 23.959’	Wild	19
Ngazidja 24	Trepezini	S11°25.123’	E43° 23.423’	Wild	189
Ngazidja 25	Uziyo	S11°24.310’	E43° 20.927’	Wild	325
Ngazidja 26	Fomboni	S11°24.242’	E43° 22.752’	Wild	226

Colony	Site Name	Latitude	Longitude	Type of colo-	Altitude (m)

				ny	
Moheli 1	Mshegi	S12° 17.795’	E43° 45.052’	Wild	93
Moheli 2	Mahurani	S12° 18.615’	E43° 46.525’	Wild	36
Moheli 3	Sephadar Pomboni	S12° 16.723’	E43° 44.144’	Wild	15
Moheli 4	Batse	S12°16.021’	E43° 41.875’	Wild	17
Moheli 5*	Domoni	S12°15.428’	E43° 38.755’	Wild	37
Moheli 6	Djitroni	S12°16.130’	E43° 38.018’	Wild	67
Moheli 7	Pakouni	S12°16.884’	E43° 37.759’	Wild	138
Moheli 8	Barakani	S12° 17.588’	E43° 38.244’	Wild	122
Moheli 9	Wallah	S12° 19.508’	E43° 39.741’	Wild	18
Moheli 10	Wallah	S12°19.561’	E43° 39.878’	Wild	24
Moheli 11	Bandani	S12°19.606’	E43° 40.145’	Wild	15
Moheli 12*	Dahoni	S12° 19.682’	E43° 40.084’	Wild	22
Moheli 13	Wallah	S12° 19.805’	E43° 40.204’	Wild	28
Moheli 14	Wallah	S12° 20.672’	E43° 40.071’	Wild	24
Moheli 15	Ndronoroni	S12° 20.491’	E43° 40.809’	Wild	116
Moheli 16	Bandani	S12° 21.319’	E43° 48.089’	Wild	221
Moheli 17	Mwahani	S12° 22.061’	E43° 46.851’	Wild	22
Moheli 18*	Badralaju	S12° 22.218’	E43° 51.449’	Wild	39
Moheli 19	Shisiwani	S12° 22.419’	E43° 52.119’	Wild	44
Moheli 20	Badralaju	S12° 22.159’	E43° 51.549’	Wild	83
Moheli 21	Mlabada	S12° 20.524’	E43° 48.668’	Wild	279
Moheli 22	Wanani	S12° 20.294’	E43° 47.082’	Wild	275
Moheli 23*	Siryziroudrani	S12° 20.351’	E43° 46.574’	Wild	304
Moheli 24	Singana	-	-	Wild	-
Moheli 25	Crde Mibani	S12° 20.241’	E43° 46.122’	Wild	277
Moheli 26	Dangoni	S12° 20.164’	E43° 46.890’	Wild	299
Moheli 27	Uhoni	S12° 20.066’	E43° 47.230’	Wild	275
Moheli 28	Barakani	-	-	Wild	-
Moheli 29	Oungoni	-	-	Wild	-
Moheli 30	Miremani	-	-	Wild	-
Moheli 31	Moja oume	-	-	Wild	-
Moheli 32	Mlabanda	-	-	Wild	-

Moheli 33	Ndremeyani	-	-	Wild	-
Moheli 34	Salemani	-	-	Wild	-
Moheli 35	Mirngoni	-	-	Wild	-
Moheli 36	Nimachioi	-	-	Wild	-
Moheli 37	Mrabao	-	-	Wild	-
Moheli 27	Bandasalmini	-	-	Wild	-
Moheli 38	Mirngoni	-	-	Wild	-
Moheli 40	Banahari	-	-	Wild	-
Moheli 41	Mtrouni	-	-	Wild	-
Moheli 42	Dargoube	-	-	Wild	-
Moheli 43	Nkagani	-	-	Wild	-

Colony	**Site Name**	**Latitude**	**Longitude**	**Type of colony**	**Altitude (m)**
Anjouan 1	Patsy	S12° 09.307’	E44° 26.080’	Wild	263
Anjouan 2	Cuvette/Ntsameni	S12° 12.033’	E44° 27.765’	Wild	439
Anjouan 3	Tsembenou/Cuvette	S12° 12.620’	E44° 27.830’	Wild	468
Anjouan 4	Dindri	S12° 12.963’	E44° 27.024’	Wild	566
Anjouan 5*	Kambalahari	S12°13. 417’	E44° 29.475’	Wild	476
Anjouan 6	Jeje	S12° 13.587’	E44° 31.218’	Wild	64
Anjouan 7	Bambao Matsanga	S12° 12.436’	E44° 30.989’	Wild	14
Anjouan 8*	Jimilime	S12° 05.828’	E44° 27.960’	Wild	284
Anjouan 9	Hajaho	S12° 07.394’	E44° 29.196’	Wild	22
Anjouan 10	Mahale	S12° 09.142’	E44° 29.929’	Wild	49
Anjouan 11	Ongoni ya Marahani	S12° 10.819’	E44° 30.632’	Wild	38
Anjouan 12	Bambao Matsanga (Hospital)	S12° 11.980’	E44° 30.683’	Wild	47
Anjouan 13	Mromvovo	S12°11.275’	E44° 29.542’	Wild	197
Anjouan 14	Moudiriyani/Nyaboimio	S12° 21.526’	E44° 31.152’	Wild	230
Anjouan 15	Bweju	S12° 21.107’	E44° 30.888’	Wild	327
Anjouan 16*	Nyombeni	S12° 21.561’	E44° 30.807’	Wild	257

Anjouan 17	Tshoroni	-	-	Wild	-
Anjouan 18	Dagi	S12° 19.855’	E44° 29.516’	Wild	542
Anjouan 19	Baoramafouga	S12° 22.419’	E43° 52.119’	Wild	44
Anjouan 20	Liwara	S12° 19.360’	E44° 30.259’	Wild	609
Anjouan 21	Zipwepweri	S12° 18.866’	E44° 30.057’	Wild	674
Anjouan 22	Crde mremani	S12° 18.650	E44° 30.043’	Wild	-
Anjouan 23	Lavaniju	-	-	Wild	-
Anjouan 24	Kohani	-	-	Wild	-
Anjouan 25	Pomoni	S12° 16.968’	E44° 24.685’	Wild	-
Anjouan 26	Shitsacouni	S12° 15.988’	E44° 23.770’	Wild	-
Anjouan 27	Maranare	S12° 14.031’	E44° 18.824’	Wild	-
Anjouan 28	Badramji	S12° 12.619’	E44° 16.341’	Wild	-
Anjouan 29	Simia mvonyi	S12° 12.543’	E44° 16.378’	Wild	-
Anjouan 30	Nyobeni	S12° 11.523’	E44° 14.741’	Wild	129
Anjouan 31	Sima	S12° 12.275	E44° 16.367	Wild	-
Anjouan 32*	Bougweni	S12° 12.375	E44° 17.964	Wild	-
Anjouan 33	Hombo	S12° 10.627	E44° 24.108	Wild	-
Anjouan 34	Foubouni	S12° 11.597	E44° 19.806	Wild	-
Anjouan 35	Mvoure	S12° 11.573	E44° 21.294	Wild	-
Anjouan 36	Njimandra	S12°11.660	E44° 22.004	Wild	-
Anjouan 37	Micontsy	S12° 09.678	E44° 24.569	Wild	-
Anjouan 38	Nyatraga	-	-	Wild	13
Anjouan 39	-	S12° 08.216	E44° 26.186	Wild	53

APPENDIX II: Primers used for microsatellite DNA analyses indicating both the forward and reverse primer sequences, expected product size ranges, dye name and the source.

Primer Name	Upper sequence (5'-3')	Lower sequence (5'-3')	(Tm in °C)	Product size Range (bp)	Fluorochrome	Source
A29	CAACTTCAACTGAAATCCG	AAACAGTACATTT-GTGACCC	52	128-175	FAM	(Solignac *et al.*, 2003)
Ap33	TTTCTTTTTGTG-GACAGCG	AAATATGGCGAAAC-GTGTG	52	225-247	HEX	(Solignac *et al.*, 2003)
AP289	AGCTAGGTCTTTCTAA-GAGTGTTG	TTCGACCGCAA-TAACATTC	52	174-288	HEX	(Solignac *et al.*, 2003)
A56	CCCCAGATGTCGCCATTC	TCATTTCTTCGCGAAACCG	55	282	TET	(Solignac *et al.*, 2003)
A35	GTACACGGTTGCAC-GGTTG	CTTCGATGGTCGTT-GTACCC	52	94-123	HEX	(Solignac *et al.*, 2003)
B124	GCAACAGGTCGGGTTA-GAG	CAGGATAGGG-TAGGTAAGCAG	52	216-232	HEX	(Solignac *et al.*, 2003)
A008	CGAAGGTAAGGTAAATGGAAC	GGCGGTTAAAGTTCTGG	52	165-181	TET	(Solignac *et al.*, 2003)

A113	CTCGAATCGTGGCGTCC	CCTGTATTTTGCAAC-CTCGC	55	202-234	FAM	(Shaibi, Lattorff and Moritz, 2008)
AP273	GATCTT-GTGTTAAACAGCCG	GATCTCTGGCAGAC-GAAGAG	52	106-110	FAM	(Solignac *et al*., 2003)
A088	CGAATTAACCGATTT-GTCG	GATCG-CAATTATTGAAGGAG	52	136-149	HEX	(Solignac *et al*., 2003)
AC306	GAA-TATGCCGCTGCCACC	TTTCGTT-GCATCCGAGCG	55	165-185	FAM	(Solignac *et al*., 2003)
A28	GAAGAGCGTTGGTT-GCAGG	GCCGTTCATGGTTAC-CACG	52	128-134	FAM	(Solignac *et al*., 2003)
AP081	GGATCGTCGAGGCGTT-GA	GAAAGTATTCCGCCGAG CA	55	124-136	TET	(Solignac *et al*., 2003)
A43	CACCGAAACAA-GATGCAAG	CCGCTCATTAAGA-TATCCG	52	124-154	TET	(Solignac *et al*., 2003)
Sex 1	AGTG-CAAAATCCAAATCATC	ATTCGATCACCCAAA-GAA	52	142-187	HEX	(Shaibi, Lattorff and Moritz, 2008)
UN467-	TTCCACAATAGA-	AATTTGGAGAACACAG-	52	261-288	FAM	(Shaibi, Lattorff and

16603	TAAAACAATACG	CATC				Moritz, 2008)
UN467-370	TGTAATAC-GTGATGTCAAAACA	GAAAAACTGTGCTCGTGAA	52	174-192	HEX	(Shaibi, Lattorff and Moritz, 2008)
UN462.4987	AAAATGACAAAAAC-GGAGAA	AATCGTTGCCAAGA-GAATC	48	168-178	FAM	(Shaibi, Lattorff and Moritz, 2008)
A24	CACAAGTTCCAACAATGC	CACATTGAGGATGAGCG	55	93-116	TET	(Solignac *et al*., 2003)
UN351	AGCATACTTCTTCAC-CGAACCAC	TCCGTTTATGCTTCATTTTCGA	52	147-166	FAM	(Shaibi, Lattorff and Moritz, 2008)

APPENDIX III: ClustalO multiple sequence alignment representing all the sequenced COI-COII intergenic spanning the region between positions 3441 to 3880 of the reference sequence (NC_001566.1).

```
NC_001566.1_(RefSeq)      AATTCAACCAATCATCACTTGAATGATTAAATTTTTTACCACCTCTAGATCATTCACATT  3300
MOH30                     ------------------------------------------------------------  0
NGA25_(reversed)          ------------------------------------------------------------  0
MOH36B                    ------------------------------------------------------------  0
ANJ18B                    ------------------------------------------------------------  0
ANJ8B                     ------------------------------------------------------------  0
NGA34                     ------------------------------------------------------------  0
MOH3                      ------------------------------------------------------------  0
ANJ15A                    ------------------------------------------------------------  0
MOH32                     ------------------------------------------------------------  0
ANJ26A                    ------------------------------------------------------------  0
ANJ22                     ------------------------------------------------------------  0
ANJ25                     ------------------------------------------------------------  0
ANJ28                     ------------------------------------------------------------  0
ANJ23                     ------------------------------------------------------------  0
ANJ19B                    ------------------------------------------------------------  0
NGA17                     ------------------------------------------------------------  0
NGA23                     ------------------------------------------------------------  0
MOH25B                    ------------------------------------------------------------  0
NGA22                     ------------------------------------------------------------  0
NGA26A_(reversed)         ------------------------------------------------------------  0
ANJ33                     ------------------------------------------------------------  0
MOH24                     ------------------------------------------------------------  0
ANJ35A_2                  ------------------------------------------------------------  0
ANJ35B                    ------------------------------------------------------------  0
ANJ34                     ------------------------------------------------------------  0
ANJ31                     ------------------------------------------------------------  0
ANJ30                     ------------------------------------------------------------  0
Anj21                     ------------------------------------------------------------  0
MOH28A                    ------------------------------------------------------------  0
NGA11B                    ------------------------------------------------------------  0
ANJ18A                    ------------------------------------------------------------  0
MOH10                     ------------------------------------------------------------  0
NGA26B_(reversed)         ------------------------------------------------------------  0
NGA16                     ------------------------------------------------------------  0
MOH12C                    ------------------------------------------------------------  0

NGA9                      ----------ATGCATTGAACTTAAGATTCAAATATAAAGTATTTTTAAACTTTTAT  47
NGA6                      ----------TGCATTGAACTTAAGATTCAAATATAAAGTATTTTTAAACTTTTAT  46
NGA14A                    ----------TGCATTGAACTTAAGATTCAAATATAAAGTATTTTTAAACTTTTAT  46
NGA7A                     ----------TGCATTGAACTTAAGATTCAAATATAAAGTATTTTTAAACTTTTAT  46
NGA11A                    ----------TGCATTGAACTTAAGATTCAAATATAAAGTATTTTTAAACTTTTAT  46
NGA15                     ----------TGCATTGAACTTAAGATTCAAATATAAAGTATTTTTAAACTTTTAT  46
NGA14B                    ----------TGCATTGAACTTAAGATTCAAATATAAAGTATTTTTAAACTTTTAT  46
NGA20                     ----------TGCATTGAACTTAAGATTCAAATATAAAGTATTTTTAAACTTTTAT  46
NGA10                     ----------TGCATTGAACTTAAGATTCAAATATAAAGTATTTTTAAACTTTTAT  46
NGA5B                     ----------TGCATTGAACTTAAGATTCAAATATAAAGTATTTTTAAACTTTTAT  46
ANJ35A                    ----------CTT---TTATTAAATTAATATTATAAATATTATTTATTAAAAT----  40
ANJ26B                    ------------------------------------------------------------  0
NGA7C                     ------AGAATAAGTGCATTGAACTTAAGATTCAAATATAAAGTATTTTTAAACTTTTAT  54
NGA31                     ----------TTAAGATTCAAATATAAAGTATTTTTAAACTTTTAT  36
MOH29                     ----------TTTGATTAAAAAATAATGTTTTATTTATCTTTTAT  35
NGA1                      ----------TAAGATTAAAAAATAAAGTATTAATATACTTTTAT  35
MOH15B                    ------------------------------------------------------------  0
MOH40                     ----------------------------------------------------TTTTA  5
MOH42B                    ----------------------------------------------------TTTTA  5
MOH39                     ----------------------------------------------------TTTTA  5
ANJ39                     ------------------------------------------------------------  0
ANJ17                     ------------------------------------------------------------  0
ANJ37                     ------------------------------------------------------------  0
ANJ10                     ----------GC--------------------------------------------  2
MOH42A                    ------------------------------------------------------------  0
MOH14                     ------------------------------------------------------------  0
MOH27                     ------------------------------------------------------------  0
MOH7                      ------------------------------------------------------------  0
NGA8B                     ----------GAACTTAAGATTCAAATATAAAGTATTTTTAAACTTTTAT  40
ANJ36                     ------------------------------------------------------------  0
MOH23C                    ------------------------------------------------------------  0
M0H8                      ------------------------------------------------------------  0
MOH17B                    ------------------------------------------------------------  0
```

```
MOH16B                  ----------                                                  0
MOH11A                  ----------                                                  0
NGA7B                   ----------                                                  0
NGA2A                   ----------                                                  0
NGA2B                   ----------                                                  0
ANJ20B                  ----------                                                  0
MOH25A_(reversed)       ----------                                                  0
NGA28A                  --GCACTGAACT----------TATGATT-CAAAAATAAGGTTTTTATAAATTTTTAT  45
MOH9                    --GCATTGAACT----------TAAGATT-CAAATATAAAGTATTTTTAAACTTTTAT  45
NGA21                   --GCATTGAACT----------TAAGATT-CAAATATAAAGTATTTTTAAACTTTTAT  45
NGA11C                  --GCATTGAACT----------TAAGATT-CAAATATAAAGTATTTTTAAACTTTTAT  45
NGA3                    --GCATTGAACT----------TAAGATT-CAAATATAAAGTATTTTTAAACTTTTAT  45
ANJ29                   ----------                                                  0
KT828418.1_(A1_MAD3)    --------AACT----------TAAGATTCAAATATAAAGT-ATTTTTAAACTTTTAT  39
KT828427.1_(A1_MAD12)   --------AACT----------TAAGATTCAAATATAAAGT-ATTTTTAAACTTTTAT      39
NGA14C                  ----------AATTCAAATATAAAGTATTTATAAAACTTTTAT                  33
NGA12                   ----------TCAAATATAAAGTATTTTTAAACTTTTAT                  29
NGA5A                   ----------TGCATTGAACTTAAGATTCAAATATAAAGTATTTTTAAACTTTTAT  46
NGA8A                   ----------ATGCATTGAACTTAAGATTCAAATATAAAGTATTTTTAAACTTTTAT  47
ANJ12                   ----------GTGCATTGAACTTAAGATTCAAATATAAAGTATTTTTAAACTTTTAT  47
MOH11B                  ----------GTGCATTGAACTTAAGATTCAAATATAAAGTATTTTTAAACTTTTAT  47
NGA4                    ----------AGTGCATTGAACTTAAGATTCAAATATAAAGTATTTTTAAACTTTTAT  48
MOH34                   ----------AGTGCATTGAACTTAAGATTCAAATATAAAGTATTTTTAAACTTTTAT  48
NGA15B                  ----------AATGCATTGAAATTAAGATTCAAATATAAAGTATTTTTAAACTTTTAT  48
MOH19                   ----------GTGCATTGAACTTAAGATTCAAATATAAAGTATTTTTAAACTTTTAT  47
MOH23B                  ----------                                                  0
MOH36A                  ----------ATTCAAATATAAAGTATTTTTAAACTTTTAT                    31
ANJ2                    ----------                                                  0
ANJ27                   ----------CATTGAATTTAAGCTTCAA                        19
MOH37                   ----------TTTCTTTTAT                                 10
MOH26                   ----------TTTTTAATTTTCTTTTAAT                        19
AJ20C                   ----------                                                  0
MOH18                   ----------                                                  0
ANJ06                   ----------                                                  0
ANJ5                    ----------                                                  0
ANJ7B                   ----------                                                  0
ANJ1                    ----------                                                  0
MOH4                    ----------                                                  0
NGA19                   -----CCATGTCAGTGCATTGAACTTAAGATTCAAATATAAAGTATTTTTAAACTTTTAT  55
MOH15A                  -----CAGAATAAGTGCATTGAACTTAAGATTCAAATATAAAGTATTTTTAAACTTTTAT  55
NGA13                   ----------                                                  0
NC_001566.1_(RefSeq)    TAGAAATTCCATTATTAATTAAAAATTTAAATTTAAAATCAATTTTAATTAAATTTTAAT  3360
MOH30                   ----------AAAAAAGGGTTTT-----TTTAATTTTTCT                 25
NGA25_(reversed)        ----------CTTAAGATTCAAATATAAAGTATTT-----TTAAACTTTTAT      37
MOH36B                  ----------TTTTAACTTTTA                                   12
ANJ18B                  ----------ACTTTTA                                    7
ANJ8B                   ----------CTTTTAT                                    7
NGA34                   ----------CTTTTAT                                    7
MOH3                    ----------CTTTTAT                                    7
ANJ15A                  ----------CTTTTAT                                    7
MOH32                   ----------CTTTTAT                                    7
ANJ26A                  ----------CTTTTAT                                    7
ANJ22                   ----------CTTTTAT                                    7
ANJ25                   ----------CTTTTAT                                    7
ANJ28                   ----------CTTTTAT                                    7
ANJ23                   ----------                                                  0
ANJ19B                  ----------CTTTTA                                     6
NGA17                   ----------CTTTTA                                     6
NGA23                   ----------CTTTTA                                     6
MOH25B                  ----------CTTTTA                                     6
NGA22                   ----------CTTTTA                                     6
NGA26A_(reversed)       ----------TTTTAT                                     6
ANJ33                   ----------TTTTAT                                     6
MOH24                   ----------                                                  0
ANJ35A_2                ----------                                                  0
ANJ35B                  ----------                                                  0
ANJ34                   ----------                                                  0
ANJ31                   ----------                                                  0
ANJ30                   ----------                                                  0
Anj21                   ----------                                                  0
MOH28A                  ----------ATTGAACTTAAGATTCAAATATA-----AAGTATTTTTAAACTTTTAT  43
NGA11B                  ----------ATTGAACTTAAGATTCAAATATA-----AAGTATTTTTAAACTTTTAT  43
ANJ18A                  ----------ATTGAACTTAAGATTCAAATATA-----AAGTATTTTTAAACTTTTAT  43
MOH10                   ----------ATTGAACTTAAGATTCAAATATA-----AAGTATTTTTAAACTTTTAT  43
NGA26B_(reversed)       ----------CATTGAACTTAAGATTCAAATATA-----AAGTATTTTTAAACTTTTAT  44
NGA16                   ----------CATCGAACTTAAGATTCAAATATA-----AAGTATTTTTAAACTTTTAT  44
MOH12C                  ----------                                                  0

NGA9                    TAAAATTAATAAATTAATA-TAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA  105
```

```
NGA6                TAAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA      105
NGA14A               TAAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA      105
NGA7A                TAAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA      105
NGA11A               TAAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA      105
NGA15               TAAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA      105
NGA14B               TAAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA      105
NGA20               TAAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA      105
NGA10               TAAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA      105
NGA5B               TAAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA      105
ANJ35A              ---------------------------------------------------- 40
ANJ26B              ----------------------------------AATATA-ACGAAATATATTTATTAAAA 26
NGA7C               TAAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA      113
NGA31               TAAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA      95
MOH29               TAAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA      94
NGA1                TAAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA      94
MOH15B               ------------------TAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA 40
MOH40               TTAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA      64
MOH42B               TTAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA      64
MOH39               TTAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA      64
ANJ39               ---------------------------------------------------- 0
ANJ17               ---------------------------------------------------- 0
ANJ37               ---------------------------------------------------- 0
ANJ10               ---------------------------------------------------- 2
MOH42A               -TAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA      58
MOH14                ----------TAAATTTATATAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA      49
MOH27                -----TTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA      54
MOH7                -----TTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA      54
NGA8B                TAAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA      99
ANJ36               ------------------------------ACAAAATATA-ACAAAATATATTTATTAAAA 30
MOH23C               ---------------------------------------TGGCAGAA---------TAAGT 13
M0H8                ------------------------------------TAACAAAATATATTTATTAAAA 22
MOH17B               ------------------------------------TGGCAGAATAGTGCATTGAATTTAA 25
MOH16B               ------------------------------------AAATATATTTATTATAATTTAAT--- 23
MOH11A               ---------------------------------------------------------- 0
NGA7B               ---------------------------------------------------- 0
NGA2A               ---------------------------------------------------- 0
NGA2B               ---------------------------------------------------- 0
ANJ20B              ------------------------------------------CAAATATATTTATTAAAA 18
MOH25A_(reversed)          ---------------------------------------------------- 0
NGA28A               TAAAATTAATAAATTAAT-ATAAAATAAAAACAAAATATAACAAAATATATTTATTAAAA      104
MOH9                TAAAATTAATAAATTAAT-ATAAAATAAAAACAAAATATAACAAAATATATTTATTAAAA      104
NGA21               TAAAATTAATAAATTAAT-ATAAAATAAAAACAAAATATAACAAAATATATTTATTAAAA      104
NGA11C               TAAAATTAATAAATTAAT-ATAAAATAAAAACAAAATATAACAAAATATATTTATTAAAA      104
NGA3                TAAAATTAATAAATTAAT-ATAAAATAAAAACAAAATATAACAAAATATATTTATTAAAA      104
ANJ29               ------------------------------CAAAATATAACAAAATATATTTATTAAAA 29
KT828418.1_(A1_MAD3)      TAAAATTAATAAATTAAT-ATAAAATAAAAACAAAATATAACAAAATATATTTATTAAAA      98
KT828427.1_(A1_MAD12)     TAAAATTAATAAATTAAT-ATAAAATAAAAACAAAATATAACAAAATATATTTATTAAAA      98
NGA14C               TAAAATTAATAAATTAAT-ATAAAATAAAAACAAAATATAACAAAATATATTTATTAAAA      92
NGA12               TAAAATTAATAAATTAAT-ATAAAATAAAAACAAAATATAACAAAATATATTTATTAAAA      88
NGA5A               TAAAATTAATAAATTAATATAAAATAAGAAACAAAATATAACAAAATATATTTATTAAAA      106
NGA8A               TAAAATTAATAAATTAATAT-AAAATAAAAACAAAATATAACAAAATATATTTATTAAAA      106
ANJ12               TAAAATTAATAAATTAATAT-AAAATAAAAACAAAATATAACAAAATATATTTATTAAAA      106
MOH11B               TAAAATTAATAAATTAATAT-AAAATAAAAACAAAATATAACAAAATATATTTATTAAAA      106
NGA4                TAAAATTAATAAATTAATATAAAATAAAAACAAAATATAACAAAATATATTTATTAAAAT      108
MOH34               TAAAATTAATAAATTAATATAAAATAAAAACAAAATATAACAAAATATATTTATTAAAAT      108
NGA15B               TAAAATTAATAAATTAATATAAAATAAAAACAAAATATAACAAAATATATTTATTAAAAT      108
MOH19               TAAAATTAATAAATTAATATAAAATAAAAACAAAATATAACAAAATATATTTATTAAAAT      107
MOH23B               ---------------------------------------------------- 0
MOH36A               TAAAT-TA-ATAAATTAATATAAAATAAAAACAAAATATAACAAAATATATTTATTAAAA      89
ANJ2                ---------------------------------------------------- 0
ANJ27               ATATA-AA-GTTGATAAACTTTTATTAAAATTAATATTATAA---------- 59
MOH37               TAAAT-TA-ATAAATTAATATAAAATAAAAACAAAATATAACAAAATATATTTATTAAAA      68
MOH26               TAAAT-TA-ATAAATTAATATAAAATAAAAACAAAATATAACAAAATATATTTATTAAAA      77
AJ20C               ---------------------------------------------------- 0
MOH18               ---------------------------------------------------- 0
ANJ06               ---------------------------------------------------- 0
ANJ5                ---------------------------------------------------- 0
ANJ7B               ---------------------------------------------------- 0
ANJ1                ---------------------------------------------------- 0
MOH4                ------------ATGAGTATGAAATAGAAACAAAATATA-ACAAAATATATTTATTAAAA 47
NGA19               TAAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA      114
MOH15A               TAAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA      114
NGA13               ---------------------------------------------------- 0
NC_001566.1_(RefSeq)      ATGGCAGAATAAGTGCATTGAAC--------TTAAGATT-CAAATATAAA-GTATTTTTA      3410
MOH30               TTAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA      84
NGA25_(reversed)         TAAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA      96
MOH36B               TTAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA      71
ANJ18B              TTAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA      66
ANJ8B               TAAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA      66
NGA34               TAAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA      66
```

```
MOH3                TAAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA   66
ANJ15A              TAAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA   66
MOH32               TAAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA   66
ANJ26A              TAAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA   66
ANJ22               TAAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA   66
ANJ25               TAAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA   66
ANJ28               TAAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA   66
ANJ23               ------------------------------------------------------------  0
ANJ19B              TTAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA   65
NGA17               TTAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA   65
NGA23               TTAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA   65
MOH25B               TTAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA   65
NGA22               TTAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA   65
NGA26A_(reversed)       TAAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA   65
ANJ33               TAAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA   65
MOH24               ------------------------------------------------------------  0
ANJ35A_2             ------------------------------------------------------------  0
ANJ35B              ------------------------------------------------------------  0
ANJ34               ------------------------------------------------------------  0
ANJ31               ------------------------------------------------------------  0
ANJ30               ------------------------------------------------------------  0
Anj21              ------------------------------------------------------------  0
MOH28A               TAAAATTAATAAATTAATATAGAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA   102
NGA11B               TAAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA   102
ANJ18A               TAAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA   102
MOH10               TAAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA   102
NGA26B_(reversed)       TAAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA   103
NGA16               TAAAATTAATAAATTAATATAAAATAAAAACAAAATATA-ACAAAATATATTTATTAAAA   103
MOH12C               ----------AAGAACAAAATATA-ACAAAATATATTTATTAAAA                34


NGA9                TTTAATTTATTAAAATTCCC----CACTTAATTCATATTA-ATTTAAAATAAATTAATAA   160
NGA6                TTTAATTTATTAAAATTCCC----CACTTAATTCATATTA-ATTTAAAATAAATTAATAA   160
NGA14A               TTTAATTTATTAAAATTCCC----CACTTAATTCATATTA-ATTTAAAATAAATTAATAA   160
NGA7A               TTTAATTTATTAAAATTCCC----CACTTAATTCATATTA-ATTTAAAATAAATTAATAA   160
NGA11A               TTTAATTTATTAAAATTCCC----CACTTAATTCATATTA-ATTTAAAATAAATTAATAA   160
NGA15               TTTAATTTATTAAAATTCCC----CACTTAATTCATATTA-ATTTAAAATAAATTAATAA   160
NGA14B               TTTAATTTATTAAAATTCCC----CACTTAATTCATATTA-ATTTAAAATAAATTAATAA   160
NGA20               TTTAATTTATTAAAATTCCC----CACTTAATTCATATTA-ATTTAAAATAAATTAATAA   160
NGA10               TTTAATTTATTAAAATTCCC----CACTTAATTCATATTA-ATTTAAAATAAATTAATAA   160
NGA5B               TTTAATTTATTAAAATTCCC----CACTTAATTCATATTA-ATTTAAAATAAATTAATAA   160
ANJ35A              ------------------------------------------------------------  40
ANJ26B              TTTAATTTATTAAAATTCCC----CACTTAATTCATATTA-ATTTAAAATAAATTAATAA   81
NGA7C               TTTAATTTATTAAAATTCCC----CACTTAATTCATATTA-ATTTAAAATAAATTAATAA   168
NGA31               TTTAATTTATTAAAATTCCC----CACTTAATTCATATTA-ATTTAAAATAAATTAATAA   150
MOH29               TTTAATTTATTAAAATTCCC----CACTTAATTCATATTA-ATTTAAAATAAATTAATAA   149
NGA1                TTTAATTTATTAAAATTCCC----CACTTAATTCATATTA-ATTTAAAATAAATTAATAA   149
MOH15B               TTTAATTTATTAAAATTCCC----CACTTAATTCATATTA-ATTTAAAATAAATTAATAA   95
MOH40               TTTAATTTATTAAAATTCCC----CACTTAATTCATATTA-ATTTAAAATAAATTAATAA   119
MOH42B               TTTAATTTATTAAAATTCCC----CACTTAATTCATATTA-ATTTAAAATAAATTAATAA   119
MOH39               TTTAATTTATTAAAATTCCC----CACTTAATTCATATTA-ATTTAAAATAAATTAATAA   119
ANJ39               ------------------------------------------------------------  0
ANJ17               ------------------------------------------------------------  0
ANJ37               ------------------------------------------------------------  0
ANJ10               ------------------------------------------------------------  2
MOH42A               TTTAATTTATTAAAATTCCC----CACTTAATTCATATTA-ATTTAAAATAAATTAATAA   113
MOH14               TTTAATTTATTAAAATTCCC----CACGTAATTCATATTA-ATTTAAAATAAATTAATAA   104
MOH27               TTTAATTTATTAAAATTCCC----CACTTAATTCATATTA-ATTTAAAATAAATTAATAA   109
MOH7                TTTAATTTATTAAAATTCCC----CACTTAATTCATATTA-ATTTAAAATAAATTAATAA   109
NGA8B               TTTAATTTATTAAAATTCCC----CACTTAATTCATATTA-ATTTAAAATAAATTAATAA   154
ANJ36               TTTAATTTATTAAAATTCCC----CACTTAATTCATATTA-ATTTAAAATAAATTAATAA   85
MOH23C               GCATAAAATTTAAGCATAAAATATAAAGTTAATAAAATTAAAT-TAAAATTATATAATAG   72
M0H8                TTTAATTTATTAAAATTCCCCACTGAAT----TCATATTAATT-TAAAATAAATTAATAA   77
MOH17B               GCTTCAAATATAAAGTTGAT----AAACTTTTAT--------T-AAAATTAATATTATAA  72
MOH16B               ------CTATTAAAATTCCC----CACTTAATTCATATTAATT-TAAAATAAATTAATAA72
MOH11A               ------------------------------------------------------------  0
NGA7B               ------------------------------------------------------------  0
NGA2A               ------------------------------------------------------------  0
NGA2B               ------------------------------------------------------------  0
ANJ20B              TTTAATTTATTAAAATTCCC----CACGTAATTCATATTAATT-TAAAATAAATTAATAA   73
MOH25A_(reversed)       ------------------------------------------------------------  0
NGA28A               TTTAATTTATTAAAATTCCC----CACTTAATTCATATTAATT-TAAAATAAATTAATAA   159
MOH9                TTTAATTTATTAAAATTCCC----CACTTAATTCATATGAATT-TAAAATAAATTAATAA   159
NGA21               TTTAATTTATTAAAATTCCC----CACTTAATTCATATTAATT-TAAAATAAATTAATAA   159
NGA11C               TTTAATTTATTAAAATTCCC----CACTTAATTCATATTAATT-TAAAATAAATTAATAA   159
NGA3                TTTAATTTATTAAAATTCCC----CACTTAATTCATATTAATT-TAAAATAAATTAATAA   159
ANJ29               TTTAATTTATTAAAATTCCC----CACTTAATTCATATTAATT-TAAAATAAATTAATAA   84
KT828418.1_(A1_MAD3)     TTTAATTTATTAAAATTCCC----CACTTAATTCATATTAATT-TAAAATAAATTAATAA   153
KT828427.1_(A1_MAD12)     TTTAATTTATTAAAATTCCC----CACTTAATTCATATTAATT-TAAAATAAATTAATAA   153
NGA14C               TTTAATTTATTAAAATTCCC----CACTTAATTCATATTAATT-TAAAATAAATTAATAA   147
```

```
NGA12                TTTAATTTATTAAAATTCCC----CACTTAATTCATATTAATTTA-AAATAAATTAATAA  143
NGA5A                TTTAATTTATTAAAATTCCC----CACTTAATTCATATTAATTTA-AAATAAATTAATAA  161
NGA8A                TTTAATTTATTAAAATTCCC----CACTTAATTCATATTAATTTA-AAATAAATTAATAA  161
ANJ12               TTTAATTTATTAAAATTCCC----CACTTAATTCATATTAATTTA-AAATAAATGAATAA  161
MOH11B               TTTAATTTATTAAAATTCCC----CACTTAATTCATATTAATTTA-AAATAAATTAATAA  161
NGA4                T-TAATTTATTAAAATTCCC----CACTTAATTCATATTAATTTA-AAATAAATTAATAA  162
MOH34               T-TAATTTATTAAAATTCCC----CACTTAATTCATATTAATTTA-AAATAAATTAATAA  162
NGA15B              T-TAATTTATTAAAATTCCC----CACTTAATTCATATTAATTTA-AAATAAATTAATAA  162
MOH19               TTTAATTTATTAAAATCCCC----CACTTAATTCATATTAATTTA-AAATAAATTAATAA  162
MOH23B               ------------------------------------------------------------ 0
MOH36A               TTTAATTTATTAAAATTCCC----CACTTAATTCATATTAATTTA-AAATAAATTAATAA  144
ANJ2                ------------------------------------------------------------ 0
ANJ27               ------------------------------------------------------------ 59
MOH37                TTTAATTTATTAAAATTCCC----CACTTAATTCATATTAATTTA-AAATAAATTAATAA  123
MOH26                TTTAATTTATTAAAATTCCC----CACTTAATTCATATTAATTTA-AAATAAATTAATAA  132
AJ20C               ------------------------------------------------------------ 0
MOH18                ------------------------------------------------------------ 0
ANJ06               ------------------------------------------------------------ 0
ANJ5                ------------------------------------------------------------ 0
ANJ7B               ------------------------------------------------------------ 0
ANJ1                ------------------------------------------------------------ 0
MOH4                 TTTAATTTATTAAAATTCCC----CACTGA--ATTGGATTAATTTAAAATAAATTAATAA  101
NGA19                TTTAATTTATTAAAATTCCC----CACTTAATTCATATTAATTTTAAAATAAATTAATAA  170
MOH15A               TTTAATTTATTAAAATTCCC----CACTTAATTCATATTAATTTA-AAATAAATTAATAA  169
NGA13               ------------------------------------------------------------ 0
NC_001566.1_(RefSeq)      AACTTTTATTAAAATTTCCC----CACTTAATTCATATTAATTTAAAAATAAATTAATAA  3466
MOH30                TTTAATTTATTAAAATTCCC----CACTTAATTCATATT-AATTTAAAATAAATTAATAA  139
NGA25_(reversed)          TTTAATTTATTAAAATTCCC----CACTTAATTCATATT-AATTTAAAATAAATTAATAA  151
MOH36B               TTTAATTTATTAAAATTCCC----CACTTAATTCATATT-AATTTAAAATAAATTAATAA  126
ANJ18B               TTTAATTTATTAAAATTCCC----CACTTAATTCATATT-AATTTAAAATAAATTAATAA  121
ANJ8B               TTTAATTTATTAAAATTCCC----CACTTAATTCATATT-AATTTAAAATAAATTAATAA  121
NGA34               TTTAATTTATTAAAATTCCC----CACTTAATTCATATT-AATTTAAAATAAATTAATAA  121
MOH3                TTTAATTTATTAAAATTCCC----CACTTAATTCATATT-AATTTAAAATAAATTAATAA  121
ANJ15A               TTTAATTTATTAAAATTCCC----CACTTAATTCATATT-AATTTAAAATAAATTAATAA  121
MOH32                TTTAATTTATTAAAATTCCC----CACTTAATTCATATT-AATTTAAAATAAATTAATAA  121
ANJ26A               TTTAATTTATTAAAATTCCC----CACTTAATTCATATT-AATTTAAAATAAATTAATAA  121
ANJ22               TTTAATTTATTAAAATTCCC----CACTTAATTCATATT-AATTTAAAATAAATTAATAA  121
ANJ25               TTTAATTTATTAAAATTCCC----CACTTAATTCATATT-AATTTAAAATAAATTAATAA  121
ANJ28               TTTAATTTATTAAAATTCCC----CACTTAATTCATATT-AATTTAAAATAAATTAATAA  121
ANJ23               ----------CATTCCC----CTTTGAATTCATATT-AATTTAAAATAAATTAATAA  42
ANJ19B               TTTAATTTATTAAAATTCCC----CACTTAATTCATATT-AATTTAAAATAAATTAATAA  120
NGA17                TTTAATTTATTAAAATTCCC----CACTTAATTCATATT-AATTTAAAATAAATTAATAA  120
NGA23                TTTAATTTATTAAAATTCCC----CACTTAATTCATATT-AATTTAAAATAAATTAATAA  120
MOH25B               TTTAATTTATTAAAATTCCC----CACTTAATTCATATT-AATTTAAAATAAATTAATAA  120
NGA22                TTTAATTTATTAAAATTCCC----CACTTAATTCATATT-AATTTAAAATAAATTAATAA  120
NGA26A_(reversed)         TTTAATTTATTAAAATTCCC----CACTTAATTCATATT-AATTTAAAATAAATTAATAA  120
ANJ33               TTTAATTTATTAAAATTCCC----CACTTAATTCATATT-AATTTAAAATAAATTAATAA  120
MOH24                ------------------------------------------------------------ 0
ANJ35A_2             ------------------------------------------------------------ 0
ANJ35B               ------------------------------------------------------------ 0
ANJ34               ------------------------------------------------------------ 0
ANJ31               ------------------------------------------------------------ 0
ANJ30               ------------------------------------------------------------ 0
Anj21               ------------------------------------------------------------ 0
MOH28A               TTTAATTTATTAAAATTCCC----CACTTAACTCATATT-AATTTAAAATAAATTAATAA  157
NGA11B               TTTAATTTATTAAAATTCCC----CACTTAATTAATATT-AATTTAAAATAAATTAATAA  157
ANJ18A               TTTAATTTATTAAAATTCCC----CACTTAATTCATATT-AATTTAAAATAAATTAATAA  157
MOH10                TTTAATTTATTAAAATTCCC----CACTGAATTCATATT-AATTTAAAATAAATTAATAA  157
NGA26B_(reversed)         TTTAATTTATTAAAATTCCC----CACTTAATTCATATT-AATTTAAAATAAATTAATAA  158
NGA16                TTTAATTTATTAAAATTCCC----CACTTAATTCATATT-AATTTAAAATAAATTAATAA  158
MOH12C               TTTAATTTATTAAAATTCCC----CACTTAATTCATATT-AATTTAAAATAAATTAATAA  89


NGA9                CAATTTTTAATAAAATAAATAATTAATTTTATTTTTATATTGAATTTTTAATTCAATCT-  219
NGA6                CAATTTTTAATAAAATAAATAATTAATTTTATTTTTATATTGAATTTTTAATTCAATCT-  219
NGA14A               CAATTTTTAATAAAATAAATAATTAATTTTATTTTTATATTGAATTTTTAATTCAATCT-  219
NGA7A                CAATTTTTAATAAAATAAATAATTAATTTTATTTTTATATTGAATTTTTAATTCAATCT-  219
NGA11A               CAATTTTTAATAAAATAAATAATTAATTTTATTTTTATATTGAATTTTTAATTCAATCT-  219
NGA15               CAATTTTTAATAAAATAAATAATTAATTTTATTTTTATATTGAATTTTTAATTCAATCT-  219
NGA14B               CAATTTTTAATAAAATAAATAATTAATTTTATTTTTATATTGAATTTTTAATTCAATCT-  219
NGA20               CAATTTTTAATAAAATAAATAATTAATTTTATTTTTATATTGAATTTTTAATTCAATCT-  219
NGA10               CAATTTTTAATAAAATAAATAATTAATTTTATTTTTATATTGAATTTTTAATTCAATCT-  219
NGA5B               CAATTTTTAATAAAATAAATAATTAATTTTATTTTTATATTGAATTTTTAATTCAATCT-  219
ANJ35A               ------------------------------------------------------------ 40
ANJ26B               CAATTTTTAATAAAATAAATAATTAATTTTATTTTTATATTGAATTTTTAATTCAATCT-  140
NGA7C                CAATTTTTAATAAAATAAATAATTAATTTTATTTTTATATTGAATTTTTAATTCAATCT-  227
NGA31               CAATTTTTAATAAAATAAATAATTAATTTTATTTTTATATTGAATTTTTAATTCAATCT-  209
MOH29               CAATTTTTAATAAAATAAATAATTAATTTTATTTTTATATTGAATTTTTAATTCAATCT-  208
NGA1                CAATTTTTAATAAAATAAATAATTAATTTTATTTTTATATTGAATTTTTAATTCAATCT-  208
MOH15B               CAATTTTTAATAAAATAAATAATTAATTTTATTTTTATATTGAATTTTTAATTCAATCT-  154
```

```
MOH40                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  178
MOH42B                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  178
MOH39                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  178
ANJ39                  ------------------------------------------------------------  0
ANJ17                  ------------------------------------------------------------  0
ANJ37                  ------------------------------------------------------------  0
ANJ10                  ------------------------------------------------------------  2
MOH42A                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  172
MOH14                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  163
MOH27                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  168
MOH7                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  168
NGA8B                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  213
ANJ36                 CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  144
MOH23C                  TATATTTATTAAAATTTCTACATGATTTATATTTACATTTCAACAATCAAATTCATACTA  132
M0H8                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  136
MOH17B                  A-------------------------------------------------------  73
MOH16B                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  131
MOH11A                  ------------------------------------------------------------  0
NGA7B                  ------------------------------------------------------------  0
NGA2A                  ------------------------------------------------------------  0
NGA2B                  ------------------------------------------------------------  0
ANJ20B                 CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  132
MOH25A_(reversed)      ----------------------------------------------GGGCAGAAT-  9
NGA28A                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  218
MOH9                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  218
NGA21                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  218
NGA11C                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  218
NGA3                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  218
ANJ29                 CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  143
KT828418.1_(A1_MAD3)      CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  212
KT828427.1_(A1_MAD12)      CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  212
NGA14C                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  206
NGA12                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  202
NGA5A                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  220
NGA8A                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  220
ANJ12                 CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  220
MOH11B                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  220
NGA4                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  221
MOH34                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  221
NGA15B                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  221
MOH19                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  221
MOH23B                  ------------------------------------------------------------  0
MOH36A                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  203
ANJ2                   ------------------------------------------------------------  0
ANJ27                  ------------------------------------------------------------  59
MOH37                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  182
MOH26                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  191
AJ20C                  ------------------------------------------------------------  0
MOH18                  ------------------------------------------------------------  0
ANJ06                  ------------------------------------------------------------  0
ANJ5                   ------------------------------------------------------------  0
ANJ7B                  ------------------------------------------------------------  0
ANJ1                   ------------------------------------------------------------  0
MOH4                  CGATTTTTAATAAAAGAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  160
NGA19                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  229
MOH15A                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  228
NGA13                  ------------------------------------------------------------  0
NC_001566.1_(RefSeq)      CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTAAATTCAATCT-  3525
MOH30                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  198
NGA25_(reversed)          CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  210
MOH36B                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  185
ANJ18B                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  180
ANJ8B                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  180
NGA34                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  180
MOH3                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  180
ANJ15A                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  180
MOH32                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  180
ANJ26A                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  180
ANJ22                 CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  180
ANJ25                 CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  180
ANJ28                 CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  180
ANJ23                 CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  101
ANJ19B                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  179
NGA17                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  179
NGA23                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  179
MOH25B                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  179
NGA22                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  179
NGA26A_(reversed)          CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  179
ANJ33                 CAATTTTTAATAAAATAAATAATTAATTTTATTTTTTATATTGAATTTTTTAATTCAATCT-  179
MOH24                  ------------------------------------------------------------  0
```

```
ANJ35A_2                ------------------------------------------------------------ 0
ANJ35B                  ------------------------------------------------------------ 0
ANJ34                   ------------------------------------------------------------ 0
ANJ31                   ------------------------------------------------------------ 0
ANJ30                   ------------------------------------------------------------ 0
Anj21                   ------------------------------------------------------------ 0
MOH28A                  CAATTTCTAATAAAATAAATAATTAATTTTATTTTTATATTGAATTTTTAATTCAATCT- 216
NGA11B                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTATATTGAATTTTTAATTCAATCT- 216
ANJ18A                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTATATTGAATTTTTAATTCAATCT- 216
MOH10                   CAATTTTTAATAAAATAAATAATTAATTTTATTTTTATATTGAATTTTTAATTCAATCT- 216
NGA26B_(reversed)       CAATTTTTAATAAAATAAATAATTAATTTTATTTTTATATTGAATTTTTAATTCAATCT- 217
NGA16                   CAATTTTTAATAAAATAAATAATTAATTTTATTTTTATATTGAATTTTTAATTCAATCT- 217
MOH12C                  CAATTTTTAATAAAATAAATAATTAATTTTATTTTTATATTGAATTTTTAATTCAATCT- 148


NGA9                    ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAA-ATA 275
NGA6                    ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAA-ATA 275
NGA14A                  ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAA-ATA 275
NGA7A                   ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAA-ATA 275
NGA11A                  ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAA-ATA 275
NGA15                   ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAA-ATA 275
NGA14B                  ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAA-ATA 275
NGA20                   ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAA-ATA 275
NGA10                   ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAA-ATA 275
NGA5B                   ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAA-ATA 275
ANJ35A                  ------------------------------------------------------------ 40
ANJ26B                  ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 197
NGA7C                   ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 284
NGA31                   ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 266
MOH29                   ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 265
NGA1                    ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 265
MOH15B                  ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 211
MOH40                   ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 235
MOH42B                  ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 235
MOH39                   ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 235
ANJ39                   ----------------ATGCATTGAATTTAAGCTTCAAATATAAAGTTGATAAACTTTT  43
ANJ17                   ----------------GTGCATTGAATTTAAGCTTCAAATATAAAGTTGATAAACTTTT  43
ANJ37                   ----------------ATGCATTGAATTTAAGCTTCAAATATAAAGTTGATAAACTTTT  43
ANJ10                   -------------------ATTGAATTTAAGCTTCAAATATAAAGTGGATAAATATTT  41
MOH42A                  ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAAT---------AAT 220
MOH14                   ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 220
MOH27                   ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 225
MOH7                    ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 225
NGA8B                   ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 270
ANJ36                   ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 201
MOH23C                  TGCTGATAATTTAATTGCATTCCAGAATAATAATAATT---ATTAATATAATTTCAACAT 189
M0H8                    ---TAAAGATTTAATCTTTTTATT-AAAATTAATAAATTAATATAAAATAAAACAAAATA 192
MOH17B                  ------------------------------------------------------------ 73
MOH16B                  ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAAT-AAAACAAAATA 187
MOH11A                  ------TTGGCAGAATAAGTGCATTGAATTTAAGCTTCAAATATAAAGTTGATAAAATTT 54
NGA7B                   ----AGTGCATTGAATTTAAGCTTCAAATATAAAGTTGATAAACTTTTATTAAAATTAAT 56
NGA2A                   ----GGTGCATTGAATTTAAGCTTCAAACATTAAGTTGATAAACTTTTATTAAAATTAAT 56
NGA2B                   --------------------CCT-GGAAATAACATTCAAATATAAAGTTGATAAACTTT  38
ANJ20B                  ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAAT-AAAACAAAATA 188
MOH25A_(reversed)       ---AAGTGCATTGAATTTAAGCTTCAAATATAAAGTTGATAAACTTTTATTAAAATTAAT 66
NGA28A                  ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAAT-AAAACAAAATA 274
MOH9                    ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAAT-AAAACAAAATA 274
NGA21                   ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAAT-AAAACAAAATA 274
NGA11C                  ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAAT-AAAACAAAATA 274
NGA3                    ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAAT-AAAACAAAATA 274
ANJ29                   ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAAT-AAAACAAAATA 199
KT828418.1_(A1_MAD3)    ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAAT-AAAACAAAATA 268
KT828427.1_(A1_MAD12)   ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAAT-AAAACAAAATA 268
NGA14C                  ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAAT-AAAACAAAATA 262
NGA12                   ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 259
NGA5A                   ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 277
NGA8A                   ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 277
ANJ12                   ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 277
MOH11B                  ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 277
NGA4                    ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 278
MOH34                   ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 278
NGA15B                  ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 278
MOH19                   ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 278
MOH23B                  -------------ATAGTGCATTGAATTTAAGCTTCAAATATAAAGTTGATAAACTTTT  46
MOH36A                  ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 260
ANJ2                    ------------------------------------------------------------ 0
ANJ27                   ------------------------------------------------------------ 59
MOH37                   ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 239
MOH26                   ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAAT---------AAT 239
AJ20C                   ----------AGAATAGTGCATTGAATTTAAGCTTCAAATATAAAGTTGATAAACTTTT  49
```

```
MOH18                ----------GAATAAGTGCATTAAATTTAAGCTTCAAATATAAAGTTGATAAACTTTT 49
ANJ06                ---------------------------------------------------------- 0
ANJ5                 ---------------------------------------------------------- 0
ANJ7B                ---------------------------------------------------------- 0
ANJ1                 ---------------------------------------------------------- 0
MOH4                 ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 217
NGA19                ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 286
MOH15A               ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 285
NGA13                ----------TTGAATTTAAGCTTCAAATATAAAGTTGATAAACTTTT 38
NC_001566.1_(RefSeq) ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 3582
MOH30                ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 255
NGA25_(reversed)     ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 267
MOH36B               ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 242
ANJ18B               ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 237
ANJ8B                ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 237
NGA34                ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 237
MOH3                 ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 237
ANJ15A               ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 237
MOH32                ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 237
ANJ26A               ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 237
ANJ22                ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 237
ANJ25                ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 237
ANJ28                ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 237
ANJ23                ---TAAAGATTTAATCTTTTTATTAAAATTAATAAGTTAATATAAAATAAAACAAAATAT 158
ANJ19B               ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 236
NGA17                ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 236
NGA23                ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 236
MOH25B               ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 236
NGA22                ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 236
NGA26A_(reversed)    ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 236
ANJ33                ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 236
MOH24                ----------GCAGAATAAGTG----------CATTGAATTTAACTTTT 29
ANJ35A_2             ----------TGCATTGAATTTAAGCTTCAAATATAAAGTTGATAAACTTTT 42
ANJ35B               ----------TGCATTGAATTTAAGCTTCAAATATAAAGTTGATAAACTTTT 42
ANJ34                ----------TGCATTGAATTTAAGCTTCAAATATAAAGTTGATAAACTTTT 42
ANJ31                ----------TGCATTGAATTTAAGCTTCAAATATAAAGTTGATAAACTTTT 42
ANJ30                ----------TGCATTGAATTTAAGCTTCAAATATAAAGTTGATAAACTTTT 42
Anj21                ----------TGCATTGAATTTAAGCTTCAAATATAAAGTTGATAAACTTTT 42
MOH28A               ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 273
NGA11B               ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 273
ANJ18A               ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 273
MOH10                ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 273
NGA26B_(reversed)    ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 274
NGA16                ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 274
MOH12C               ---TAAAGATTTAATCTTTTTATTAAAATTAATAAATTAATATAAAATAAAACAAAATAT 205


NGA9                 TAACAGAATATATTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 335
NGA6                 TAACAGAATATATTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 335
NGA14A               TAACAGAATATATTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 335
NGA7A                TAACAGAATATATTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 335
NGA11A               TAACAGAATATATTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 335
NGA15                TAACAGAATATATTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 335
NGA14B               TAACAGAATATATTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 335
NGA20                TAACAGAATATATTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 335
NGA10                TAACAGAATATATTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 335
NGA5B                TAACAGAATATATTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 335
ANJ35A               ----------TTCTACATGATTTATATTTACAT 63
ANJ26B               AA-CAGAATATATTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 256
NGA7C                AA-CAGAATATATTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 343
NGA31                AA-CAGAATATATTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 325
MOH29                AA-CAGAATATATTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 324
NGA1                 AA-CAGAATATATTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 324
MOH15B               AA-CAGAATATATTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 270
MOH40                AA-CAGAATATATTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 294
MOH42B               AA-CAGAATATATTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 294
MOH39                AA-CAGAATATATTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 294
ANJ39                AT-TAAAATTAATATTATAAATATTATTTATTAAAATTTCTACATGATTTATATTTACAT 102
ANJ17                AT-TAAAATTAATATTATAAATATTATTTATTAAAATTTCTACATGATTTATATTTACAT 102
ANJ37                AT-TAAAATTAATATTATAAATATTATTTATTAAAATTTCTACATGATTTATATTTACAT 102
ANJ10                AT-TAAAATTAATATTATAAATATTATTTATTAAAATTTCTACATGATTTATATTTACAT 100
MOH42A               AA-CAGAATATATTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 279
MOH14                AA-CAGAATATATTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 279
MOH27                AA-CAGAATATATTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 284
MOH7                 AA-CAGAATATATTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 284
NGA8B                AA-CAGAATATATTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 329
ANJ36                AA-CAGAATATATTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 260
MOH23C               TAACAGTTCAAATTATTTTAAATTTATTTATAAATAAATTTTCAAACTTATTATTAATAA 249
M0H8                 TAACAGAATATATTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 252
MOH17B               ----------TATTATTTATTAAAATTTCTACATGATTTATATTTACAT 112
```

```
MOH16B                TAACAGAATATATTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 247
MOH11A                TATTAAAATTAATATTATAAATATTATTTATTAAAATTTCTACATGATTTATATTTACAT 114
NGA7B                 TAATTAACTTAATATTATAAATATTATTTATTAAAATTTCTACATGATTTATATTTACAT 116
NGA2A                 TAATTAACTTAATATTATAAATATTATTTATTAAAATTTCTACATGATTTATATTTACAT 116
NGA2B                 TATTAAAATTAATATGATAAATATTATTTATTAAAATTTCTACATGATTTATATTTACAT 98
ANJ20B                TAACAGAATATATTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 248
MOH25A_(reversed)     TAATTAACTTAATATTATAAATATTATTTATTAAAATTTCTACATGATTTATATTTACAT 126
NGA28A                TAACAGAATATATTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 334
MOH9                  TAACAGAATATATTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 334
NGA21                 TAACAGAATATATTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 334
NGA11C                TAACAGAATATATTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 334
NGA3                  TAACAGAATATATTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 334
ANJ29                 TAACAGAATATATTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 259
KT828418.1_(A1_MAD3)  TAACAGAATATATTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 328
KT828427.1_(A1_MAD12) TAACAGAATATATTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 328
NGA14C                TAACAGAATATATTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 322
NGA12                 AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 318
NGA5A                 AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 336
NGA8A                 AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 336
ANJ12                 AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 336
MOH11B                AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 336
NGA4                  AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 337
MOH34                 AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 337
NGA15B                AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 337
MOH19                 AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 337
MOH23B                ATTAAAATTAA-TATTATAAATATTATTTATTAAAATTTCTACATGATTTATATTTACAT 105
MOH36A                AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 319
ANJ2                  --------------------------TTTAGTAAAATTTCTACATGATTTATATTTACAT34
ANJ27                 --------------------ATATTATTTATTAAAATTTCTACATGATTTATATTTACAT 99
MOH37                 AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 298
MOH26                 AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 298
AJ20C                 ATTA-AAATTAATATTATAAATATTATTTATTAAAATTTCTACATGATTTATATTTACAT 108
MOH18                 ATTA-AAATTAATATTATAAATAGCATTTATTAAAATTTCTACATGATTTATATTTACAT 108
ANJ06                 ----------------------TAATTTATTAAAATTTCCACATGATTTATATTTATAT 37
ANJ5                  ---------------------TTATTTGTTAAAATTTCTACATGATTTATGTTTACAT 37
ANJ7B                 ----------------------TTATTTATTAAAATTTCTACATGATTTATATTTACAT 37
ANJ1                  ----------------------TTATTTATTAAAATTTATACATGATTTATATTTACAT 37
MOH4                  AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 276
NGA19                 AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 345
MOH15A                AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 344
NGA13                 ATT-AAAATTAATATTATAAATATTATTTATTAAAATTTCTACATGATTTATATTTACAT 97
NC_001566.1_(RefSeq)  AACAGAATATATTTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 3642
MOH30                 AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 314
NGA25_(reversed)      AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 326
MOH36B                AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 301
ANJ18B                AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 296
ANJ8B                 AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 296
NGA34                 AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 296
MOH3                  AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 296
ANJ15A                AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 296
MOH32                 AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 296
ANJ26A                AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 296
ANJ22                 AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 296
ANJ25                 AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 296
ANJ28                 AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 296
ANJ23                 AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 217
ANJ19B                AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 295
NGA17                 AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 295
NGA23                 AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 295
MOH25B                AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 295
NGA22                 AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 295
NGA26A_(reversed)     AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 295
ANJ33                 AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 295
MOH24                 ATTAAAATTAA-TATTATAAATATTATTTATTAAAATTTATACATGATTTATATTTACAT 88
ANJ35A_2              ATTAAAATTAA-TATTATAAATATTATTTATTAAAATTTCTACATGATTTATATTTACAT 101
ANJ35B                ATTAAAATTAA-TATTATAAATATTATTTATTAAAATTTCTACATGATTTATATTTACAT 101
ANJ34                 ATTAAAATTAA-TATTATAAATATTATTTATTAAAATTTCTACATGATTTATATTTACAT 101
ANJ31                 ATTAAAATTAA-TATTATAAATATTATTTATTAAAATTTCTACATGATTTATATTTACAT 101
ANJ30                 ATTAAAATTAA-TATTATAAATATTATTTATTAAAATTTCTACATGATTTATATTTACAT 101
Anj21                 ATTAAAATTAA-TATTATAAATATTATTTATTAAAATTTCTACATGATTTATATTTACAT 101
MOH28A                AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 332
NGA11B                AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 332
ANJ18A                AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 332
MOH10                 AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 332
NGA26B_(reversed)     AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 333
NGA16                 AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 333
MOH12C                AACAGAATATA-TTATTAAAATTTAATTTATTAAAATTTCCACATGATTTATATTTATAT 264
                                  *  **  ** * * * * *

NGA9                  TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA 395
```

```
NGA6                    TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  395
NGA14A                   TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  395
NGA7A                    TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  395
NGA11A                   TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  395
NGA15                   TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  395
NGA14B                   TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  395
NGA20                   TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  395
NGA10                   TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  395
NGA5B                    TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  395
ANJ35A                   TTCAACAATCAAATTCATACTATGCTGATAACTTAATTTCAT-----TCCAGAATAATAA  118
ANJ26B                   TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  316
NGA7C                    TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  403
NGA31                    TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  385
MOH29                    TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  384
NGA1                    TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  384
MOH15B                   TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  330
MOH40                    TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  354
MOH42B                   TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  354
MOH39                    TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  354
ANJ39                   TTCAACAATCAAATTCATACTATGCTGATAACTTAATTTCATTCCAG-----AATAATAA  157
ANJ17                   TTCAACAATCAAATTCATACTATGCTGATAACTTAATTTCATTCCAG-----AATAATAA  157
ANJ37                   TTCAACAATCAAATTCATACTATGCTGATAACTTAATTTCATTCCAG-----AATAATAA  157
ANJ10                   TTCAACAATCAAATTCATACTATGCTGATAACTTAATTTCATTCCA-----GAATAATAA  155
MOH42A                   TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  339
MOH14                    TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  339
MOH27                    TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  344
MOH7                    TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  344
NGA8B                    TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  389
ANJ36                   TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  320
MOH23C                   ----------------------AAAATTATTTAAACAGAAATTCCTATTATTATTCTTT  286
M0H8                    TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  312
MOH17B                   TTCAA------CAATCATATAATTTCAATAACTTAATTTCATTCGATAATATAGTAATAA  166
MOH16B                   TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  307
MOH11A                   TTCGACAATCAAATTCATACTATGCTGATAACTTAATTGCATTC-----CAGAATAATAA  169
NGA7B                    TTCAACAATCATATAATTTCAATAATTTAA--TTTCATTCCATA-----ATATAGAATAA  169
NGA2A                    TTCAACAATCATATAATTTCAATAATTTAA--TTTCATTCCATA-----ATATAGAATAA  169
NGA2B                    TTCAACAATCAAATTCATACGATGCTGATAACTTAATTTCATTCC-----AGAATAATAA  153
ANJ20B                   TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  308
MOH25A_(reversed)        TTCAACAATCATAT------AATTTCAATAATTTAATTTCATTCCATA-ATATAGAATAA  179
NGA28A                   TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  394
MOH9                    TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  394
NGA21                    TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  394
NGA11C                   TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  394
NGA3                    TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  394
ANJ29                   TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  319
KT828418.1_(A1_MAD3)     TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  388
KT828427.1_(A1_MAD12)    TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  388
NGA14C                   TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  382
NGA12                    TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  378
NGA5A                    TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  396
NGA8A                    TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  396
ANJ12                   TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  396
MOH11B                   TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  396
NGA4                    TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  397
MOH34                    TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  397
NGA15B                   TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  397
MOH19                    TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  397
MOH23B                   TTCGACAATCAAATTCATACTATGCTGATAACTTAATTTCATTCCAG-----AATAATAA  160
MOH36A                   TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  379
ANJ2                    TTCAACAATCAGATTCATACGGTGCTGATAACT-----TAATTTCATTCCAGAATAATAA  89
ANJ27                   TTCAACAATCAAATTCATACTATGCTGATAACTTAATTTCATTCCAG-----AATAATAA  154
MOH37                    TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  358
MOH26                    TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  358
AJ20C                   TTCAACAATCAAATTCATACTATGCTGATAACTTAATTTCATTCCAGA-----ATAATAA  163
MOH18                    TTCAACAATCAAATTCATACTATGCTGATAACTTAATTGCATTCCAGA-----ATAATAA  163
ANJ06                   TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  97
ANJ5                    TTGAACAATCAAATTCATACTATGCTGATAACTTAATTTCATTCCAGAA-----TAATAG  92
ANJ7B                    TTCAACAATCAAATTCATACTATGCTGATAACTTAATTTCATTCCAGAA-----TAATAA  92
ANJ1                    TTCAACAATCAAATTCATATTATGCTGATAACTTAATTTCATTCCAGAA-----TAATAA  92
MOH4                     TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  336
NGA19                    TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  405
MOH15A                   TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  404
NGA13                    TTCAACAATCAAATTCATACTATGCTGAT-----AACTTAATTTCATTCCAGAATAATAA  152
NC_001566.1_(RefSeq)     TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  3702
MOH30                    TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  374
NGA25_(reversed)         TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  386
MOH36B                   TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  361
ANJ18B                   TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  356
ANJ8B                    TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  356
NGA34                    TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  356
```

```
MOH3                    TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  356
ANJ15A                  TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  356
MOH32                   TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  356
ANJ26A                  TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  356
ANJ22                   TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  356
ANJ25                   TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  356
ANJ28                   TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  356
ANJ23                   TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  277
ANJ19B                  TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  355
NGA17                   TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  355
NGA23                   TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  355
MOH25B                  TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  355
NGA22                   TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  355
NGA26A_(reversed)       TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  355
ANJ33                   TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  355
MOH24                   TTCAACAATCATATAATTTCAAT-AACTT-----AATTTCATTCCATAATATAGTAATAA  142
ANJ35A_2                TTCAACAATCAAATTCATACTATGCTGAT-----AACTTAATTTCATTCCAGAATAATAA  156
ANJ35B                  TTCAACAATCAAATTCATACTATGCTGAT-----AACTTAATTTCATTCCAGAATAATAA  156
ANJ34                   TTCAACAATCAAATTCATACTATGCTGAT-----AACTTAATTTCATTCCAGAATAATAA  156
ANJ31                   TTCAACAATCAAATTCATACTATGCTGAT-----AACTTAATTTCATTCCAGAATAATAA  156
ANJ30                   TTCAACAATCAAATTCATACTATGCTGAT-----AACTTAATTTCATTCCAGAATAATAA  156
Anj21                   TTCAACAATCAAATTCATACTATGCTGAT-----AACTTAATTTCATTCCAGAATAATAA  156
MOH28A                  TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  392
NGA11B                  TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  392
ANJ18A                  TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  392
MOH10                   TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  392
NGA26B_(reversed)       TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  393
NGA16                   TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  393
MOH12C                  TTCAAGAATCAAATTCATATTATGCTGATAATTTAATTTCATTTCATAATATAGTTATAA  324
                                                                                *

NGA9                    TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA  455
NGA6                    TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA  455
NGA14A                  TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA  455
NGA7A                   TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA  455
NGA11A                  TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA  455
NGA15                   TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA  455
NGA14B                  TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA  455
NGA20                   TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA  455
NGA10                   TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA  455
NGA5B                   TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA  455
ANJ35A                  TAATTATTATTATAATTTCAACATTAACAGTTCATATTATTTTAGATTTATTTATAAATA  178
ANJ26B                  TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA  376
NGA7C                   TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA  463
NGA31                   TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAAATTTATTTATAAATA  445
MOH29                   TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA  444
NGA1                    TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA  444
MOH15B                  TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA  390
MOH40                   TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA  414
MOH42B                  TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA  414
MOH39                   TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA  414
ANJ39                   TAATTATTATTATAATTTCAACATTAACAGTTCATATTATTTTAGATTTATTTATAAATA  217
ANJ17                   TAATTATTATTATAATTTCAACATTAACAGTTCATATTATTTTAGATTTATTTATAAATA  217
ANJ37                   TAATTATTATTATAATTTCAACATTAACAGTTCATATTATTTTAGATTTATTTATAAATA  217
ANJ10                   TAATTATTATTATAATTTCAACATTAACAGTTCATATTATTTTAGATTTATTTATAAATA  215
MOH42A                  TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA  399
MOH14                   TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA  399
MOH27                   TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA  404
MOH7                    TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA  404
NGA8B                   TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA  449
ANJ36                   TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA  380
MOH23C                  CAATTATTTCTTTTCCTTCACTT---------TTAAAAATTTTATATTTAATTGAAATGA  337
M0H8                    TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA  372
MOH17B                  TAATTATTATTATAATTTCAACATTAACAGTTTATATCATTTTAGATTTATTTGTAAATA  226
MOH16B                  TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA  367
MOH11A                  TAATTATTATTATAATTTCAACATTAACAGTTCATATTATTTTAGATTTATTTATAAATA  229
NGA7B                   TAATCATTATTATAATTTCAACATTAACAATTTATATCATTTTAGATTTATTTGTAAATA  229
NGA2A                   TAATCATTATTATAATTTCAACATTAACAATTTATATCATTTTAGATTTATTTGTAAATA  229
NGA2B                   TAATTATTATTATAATTTCAACATTAACAGTTCATATTATTTTAGATTTATTTATAAATA  213
ANJ20B                  TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA  368
MOH25A_(reversed)       TAATCATTATTATAATTTCAACATTAACAATTTATATCATTTTAGATTTATTTGTAAATA  239
NGA28A                  TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA  454
MOH9                    TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA  454
NGA21                   TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA  454
NGA11C                  TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA  454
NGA3                    TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA  454
ANJ29                   TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA  379
KT828418.1_(A1_MAD3)    TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA448
KT828427.1_(A1_MAD12)   TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA  448
NGA14C                  TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA  442
```

```
NGA12                TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          438
NGA5A                TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          456
NGA8A                TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          456
ANJ12               TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          456
MOH11B               TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          456
NGA4                TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          457
MOH34                TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          457
NGA15B               TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          457
MOH19                TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          457
MOH23B               TAATTATTATTATAATTTCAACATTAACAGTTCATATTATTTTAGATTTATTTATAAATA          220
MOH36A               TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          439
ANJ2                TAATTATTATTATAATTTCAACATTAACAGTTCATATTATTTTAGATTTATTTATAAATA          149
ANJ27               TAATTATTATTATAATTTCAACATTAACAGTTCATATTATTTTAGATTTATTTATAAATA          214
MOH37                TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          418
MOH26                TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          418
AJ20C               TAATTATTATTATAATTTCAACATTAACAGTTCATATTATTTTAGATTTATTTATAAATA          223
MOH18                TAATTATTATTATAATTTCAACATTAACAGTTCATATTATTTTAGATTTATTTATAAATA          223
ANJ06               TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          157
ANJ5                TAATTATTATTATAATTTCAACATTAACAGTTCATATTATTTTAGATGTATTTATAAATA          152
ANJ7B                TAATTATTATTATAATTTCAACATTAACAGTTCATATTATTTTAGATTTATTTATAAATA          152
ANJ1                TAATTATTATTATAATTTCAACATTAACAGTTCATATTATTTTAGATTTATTTATAAATA          152
MOH4                 TAATCATTATTATAATTTCAACATTAACAGTATATATTATTTTAGATTTATTTATAAATA          396
NGA19                TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          465
MOH15A               TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          464
NGA13                TAATTATTATTATAATTTCAACATTAACAGTTCATATTATTTTAGATTTATTTATAAATA          212
NC_001566.1_(RefSeq)      TAATTATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAACA 3762
MOH30                TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          434
NGA25_(reversed)          TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          446
MOH36B               TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          421
ANJ18B               TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          416
ANJ8B               TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          416
NGA34                TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          416
MOH3                 TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          416
ANJ15A               TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          416
MOH32                TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          416
ANJ26A               TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          416
ANJ22               TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          416
ANJ25               TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          416
ANJ28               TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          416
ANJ23               TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          337
ANJ19B               TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          415
NGA17                TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          415
NGA23                TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          415
MOH25B               TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          415
NGA22                TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          415
NGA26A_(reversed)         TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA 415
ANJ33               TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          415
MOH24                TAATTATTATTATAATTTCAACATTAACAGTTTATATCATTTTAGATTTATTTGTAAATA          202
ANJ35A_2             TAATTATTATTATAATTTCAACATTAACAGTTCATATTATTTTAGATTTATTTATAAATA          216
ANJ35B               TAATTATTATTATAATTTCAACATTAACAGTTCATATTATTTTAGATTTATTTATAAATA          216
ANJ34               TAATTATTATTATAATTTCAACATTAACAGTTCATATTATTTTAGATTTATTTATAAATA          216
ANJ31               TAATTATTATTATAATTTCAACATTAACAGTTCATATTATTTTAGATTTATTTATAAATA          216
ANJ30               TAATTATTATTATAATTTCAACATTAACAGTTCATATTATTTTAGATTTATTTATAAATA          216
Anj21               TAATTATTATTATAATTTCAACATTAACAGTTCATATTATTTTAGATTTATTTATAAATA          216
MOH28A               TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          452
NGA11B               TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          452
ANJ18A               TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          452
MOH10                TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          452
NGA26B_(reversed)         TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA 453
NGA16                TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          453
MOH12C               TAATCATTATTATAATTTCAACATTAACTGTATATATTATTTTAGATTTATTTATAAATA          384
                    *** *** * *  ****          *  ****** ** ** **  **  *

NGA9                 AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          513
NGA6                AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          513
NGA14A               AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          513
NGA7A                AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          513
NGA11A               AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          513
NGA15               AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          513
NGA14B               AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          513
NGA20                AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          513
NGA10                AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          513
NGA5B                AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          513
ANJ35A               AATTTTCAAACTTATTATT-ATTAAAAAATTA----------TTTAAACAGTAAT       222
ANJ26B               AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          434
NGA7C                AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTTAACAGTTAT          521
NGA31                AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          503
MOH29                AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          502
NGA1                AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          502
MOH15B               AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          448
```

```
MOH40                AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          472
MOH42B               AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          472
MOH39                AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          472
ANJ39               AATTTTCAAACTTATTATT-ATTAAAAAATTAT----------TTAAACAGTAAT        261
ANJ17               AATTTTCAAACTTATTATT-ATTAAAAAATTAT----------TTAAACAGTAAT        261
ANJ37               AATTTTCAAACTTATTATT-ATTAAAAAATTAT----------TTAAACAGTAAT        261
ANJ10               AATTTTCAAACTTATTATT-ATTAAAAAATTA----------TTTAAACAGTAAT        259
MOH42A               AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          457
MOH14                AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          457
MOH27                AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          462
MOH7                 AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          462
NGA8B                AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          507
ANJ36               AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          438
MOH23C               A-ATTGCAAATCCATTCTTTTCTATCCATTGGTCATCAATGAAATTGATGG---------          387
M0H8                AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          430
MOH17B               AATTTTCAAACTTATTATT-ATTAAAAAATCATGA-TATTGAAATTATTTGAACAGTAAT          284
MOH16B               AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          425
MOH11A               AATTTTCAAACTTATTATT-ATTA----------AAAAATTATTTAAACAGTAAT      273
NGA7B                AATTTTCAAACTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTAAT          287
NGA2A                AATTTTCAAACTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTAAT          287
NGA2B                AATTTTCAAACTTATTATT-ATTAAAAA----------ATTATTTAAACAGTAAT      257
ANJ20B               AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          426
MOH25A_(reversed)        AATTTTCAAACTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTAAT          297
NGA28A               AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          512
MOH9                AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          512
NGA21                AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          512
NGA11C               AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          512
NGA3                AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          512
ANJ29               AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          437
KT828418.1_(A1_MAD3)     AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT 506
KT828427.1_(A1_MAD12)    AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT 506
NGA14C               AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          500
NGA12                AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          496
NGA5A                AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          514
NGA8A                AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          514
ANJ12               AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          514
MOH11B               AATTTTCAAATTTATTTTT-ATTAAAAAATCTTAA-TATTGAAATTATTTGAACAGTTAT          514
NGA4                AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          515
MOH34                AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          515
NGA15B               AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          515
MOH19                AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          515
MOH23B               AATTTTCAAACTTATTATT-ATTAAAAAAT----------TATTTAAACAGTAAT      264
MOH36A               AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          497
ANJ2                AATTTTCAAACTTATTATT-ATTAAAAAAT----------TATTTAAACAGTAAT      193
ANJ27               AATTTTCAAACTTATTATT-ATTAAAAAATTA----------TTTAAACAGTAAT      258
MOH37                AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          476
MOH26                AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          476
AJ20C               AATTTTCAAACTTATTATT-ATTAAAAAATT----------ATTTAAACAGTAAT      267
MOH18                AATTTTCAAACTTATTATT-ATTAAAAAATT----------ATTTAAACAGTAAT      267
ANJ06               AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          215
ANJ5                AATTTTCAAACTTATTATT-ATTAAAAAATTA----------TTTAAACAGTAAT      196
ANJ7B               AATTTTCAAACTTATTATT-ATTAAAAAATTA----------TTTAAACAGTAAT      196
ANJ1                AATTTTCAAACTTATTATT-ATTAAAAAATTA----------TTTAAACAGTAAT      196
MOH4                 AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          454
NGA19                AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          523
MOH15A               AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          522
NGA13               AATTTTCAAACTTATTATT-ATTAAAAAATT----------ATTTAAACAGTAAT      256
NC_001566.1_(RefSeq)     AATTCTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAATTAT 3820
MOH30                AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          492
NGA25_(reversed)         AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          504
MOH36B               AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          479
ANJ18B               AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          474
ANJ8B               AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          474
NGA34                AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          474
MOH3                 AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          474
ANJ15A               AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          474
MOH32                AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          474
ANJ26A               AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          474
ANJ22               AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          474
ANJ25               AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          474
ANJ28               AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          474
ANJ23               AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          395
ANJ19B               AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          473
NGA17                AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          473
NGA23                AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          473
MOH25B               AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          473
NGA22                AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          473
NGA26A_(reversed)        AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          473
ANJ33               AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT          473
MOH24                AATTTTCAAACTTATTATT-ATTAAAAAATCATGA-TATTGAAATTATTTGAACAGTAAT          260
```

```
ANJ35A_2              AATTTTCAAACTTATTATT-ATTAAAAAATT----------ATTTAAACAGTAAT     260
ANJ35B               AATTTTCAAACTTATTATT-ATTAAAAAATT----------ATTTAAACAGTAAT     260
ANJ34               AATTTTCAAACTTATTATT-ATTAAAAAATT----------ATTTAAACAGTAAT     260
ANJ31               AATTTTCAAACTTATTATT-ATTAAAAAATT----------ATTTAAACAGTAAT     260
ANJ30               AATTTTCAAACTTATTATT-ATTAAAAAATT----------ATTTAAACAGTAAT     260
Anj21              AATTTTCAAACTTATTATT-ATTAAAAAATT----------ATTTAAACAGTAAT     260
MOH28A               AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT     510
NGA11B               AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT     510
ANJ18A               AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT     510
MOH10                AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT     510
NGA26B_(reversed)        AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT     511
NGA16                AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT     511
MOH12C               AATTTTCAAATTTATTTTT-ATTAAAAAATCATAA-TATTGAAATTATTTGAACAGTTAT     442
                  * *  ****   *** **   **                *

NGA9                 TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA     573
NGA6                 TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA     573
NGA14A               TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA     573
NGA7A                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA     573
NGA11A               TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA     573
NGA15                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA     573
NGA14B               TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA     573
NGA20                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA     573
NGA10                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA     573
NGA5B                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA     573
ANJ35A               TCCAATTATTATTCTTTCAATTATTTCTTTTCCTTCACTTTTAAAAATTTTATATTTAAT     282
ANJ26B               TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAA---AATTTTATATTTAAT     491
NGA7C                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAA---AATTTTATATTTAAT     578
NGA31                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAA---AATTTTATATTTAAT     560
MOH29                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAA---AATTTTATATTTAAT     559
NGA1                 TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAA---AATTTTATATTTAAT     559
MOH15B               TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAA---ATTTTATATTTAAT     505
MOH40                TCCCATTATTATTCTATTAATTATTTGTTTTCCATCATTA---AAAATTTTATATTTAAT     529
MOH42B               TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTA---AAAATTTTATATTTAAT     529
MOH39                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTA---AAAATTTTATATTTAAT     529
ANJ39               TCCAATTATTATTCTTTCAATTATTTCTTTTCCTTCACTTTTAAAAATTTTATATTTAAT     321
ANJ17               TCCAATTATTATTCTTTCAATTATTTCTTTTCCTTCACTTTTAAAAATTTTATATTTAAT     321
ANJ37               TCCAATTATTATTCTTTCAATTATTTCTTTTCCTTCACTTTTAAAAATTTTATATTTAAT     321
ANJ10               TCCAATTATTATTCTTTCAATTATTTCTTTTCCTTCACTTTTAAAAATTTTATATTTAAT     319
MOH42A               TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAA---AAATTTTATATTTAAT     514
MOH14                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAA---AAATTTTATATTTAAT     514
MOH27                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAA---AAATTTTATATTTAAT     519
MOH7                 TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAA---AAATTTTATATTTAAT     519
NGA8B                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAA---AAATTTTATATTTAAT     564
ANJ36               TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAA---AAATTTTATATTTAAT     495
MOH23C               ------------------------------------------------------------     387
M0H8                 TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA     490
MOH17B               TCCAATTATTATTCTTTTAATTATTTCTTTTCCTTTACTTTTAAAAATTTTATATTTAAT     344
MOH16B               TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTT---TATATTT--A     480
MOH11A               TCCAATTATTATTCTTTCAATTATTTCTTTTCCTTCACTTTTAAAAATTTTATATTTAAT     333
NGA7B                TCCAATTATTATTCTTTCAATTATTTCTTTTCCTTCACTTTTAAAAATTTTATATTTAAT     347
NGA2A                TCCAATTATTATTCTTTCAATTATTTCTTTTCCTTCACTTTTAAAAATTTTATATTTAAT     347
NGA2B                TCCAATTATTATTCTTTCAATTATTTCTTTTCCTTCACTTTTAAAAATTTTATATTTAAT     317
ANJ20B               TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA     486
MOH25A_(reversed)        TCCAATTATTATTCTTTCAATTATTTCTTTTCCTTCACTTTTAAAAATTTTATATTTAAT     357
NGA28A               TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA     572
MOH9                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA     572
NGA21                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA     572
NGA11C               TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA     572
NGA3                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA     572
ANJ29               TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA     497
KT828418.1_(A1_MAD3)      TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATT----------  545
KT828427.1_(A1_MAD12)     TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATT----------  545
NGA14C               TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA     560
NGA12                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATT---TTATATTTAAT     553
NGA5A                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATT---TTATATTTAAT     571
NGA8A                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATT---TTATATTTAAT     571
ANJ12               TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATT---TTATATTTAAT     571
MOH11B               TCCAATTATTATTTG-----------------------------------------     529
NGA4                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATT---TTATATTTAAT     572
MOH34                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATT---TTATATTTAAT     572
NGA15B               TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATT---TTATATTTAAT     572
MOH19                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATT---TTATATTTAAT     572
MOH23B               TCCAATTATTATTCTTTCAATTATTTCTTTTCCTTCACTTTTAAAAATTTTATATTTAAT     324
MOH36A               TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAAT---TTTATATTTAAT     554
ANJ2                TCCAATTATTATTCTTTCAATTATTTCTTTTCCTTCACTTTTAAAAATTTTATATTTAAT     253
ANJ27               TCCAATTATTATTCTTTCAATTATTTCTTTTCCTTCACTTTTAAAAATTTTATATTTAAT     318
MOH37                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAA---AATTTTATATTTAAT     533
MOH26                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAA---AATTTTATATTTAAT     533
AJ20C               TCCAATTATTATTCTTTCAATTATTTCTTTTCCTTCACTTTTAAAAATTTTATATTTAAT     327
```

```
MOH18                 TCCAATTATTATTCTTTCAATTATTTCTTTTCCTTCACTTTTAAAAATTTTATATTTAAT 327
ANJ06                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAA---AAATTTTATATTTAAT 272
ANJ5                TCCAATTATTATTCTTTCAATTATTTCTTTTCCTTCACTTTTAAAAATTTTATATTTAAT 256
ANJ7B                TCCAATTATTATTCTTTCAATTATTTCTTTTCCTTCACTTTTAAAAATTTTATATTTAAT 256
ANJ1                TCCAATTATTATTCTTTCAATTATTTCTTTTCCTTCACTTTTAAAAATTTTATATTTAAT 256
MOH4                 TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA 514
NGA19                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA 583
MOH15A                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA 582
NGA13                TCCAATTATTATTCTTTCAATTATTTCTTTTCCTTCACTTTTAAAAATTTTATATTTAAT 316
NC_001566.1_(RefSeq)      TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA 3880
MOH30                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA 552
NGA25_(reversed)          TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA 564
MOH36B                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA 539
ANJ18B                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA 534
ANJ8B                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA 534
NGA34                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTTA 534
MOH3                 TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA 534
ANJ15A                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA 534
MOH32                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA 534
ANJ26A                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCAATAAAAATTTTATATTTATTTGA 534
ANJ22                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCACTAAAAATTTTATATTTAATTGA 534
ANJ25                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCACTAAAAATTTTATATTTAATTGA 534
ANJ28                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCACTAAAAAATTTATATTTAATTGA 534
ANJ23                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA 455
ANJ19B                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCAATAAAAAATTTATATTTAATTGA 533
NGA17                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA 533
NGA23                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA 533
MOH25B                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA 533
NGA22                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA 533
NGA26A_(reversed)          TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA 533
ANJ33                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA 533
MOH24                TCCAATTATTATTCTTTTAATTATTTCTTTTCCTTTACTTTTAAAAATATATTATA---- 316
ANJ35A_2                TCCAATTATTATTCTTTCAATTATTTCTTTTCCTTCACTTTTAAAAATTTTATATTTAAT 320
ANJ35B                TCCAATTATTATTCTTTCAATTATTTCTTTTCCTTCACTTTTAAAAATTTTATATTTAAT 320
ANJ34                TCCAATTATTATTCTTTCAATTATTTCTTTTCCTTCACTTTTAAAAATTTTATATTTAAT 320
ANJ31                TCCAATTATTATTCTTTCAATTATTTCTTTTCCTTCACTTTTAAAAATTTTATATTTAAT 320
ANJ30                TCCAATTATTATTCTTTCAATTATTTCTTTTCCTTCACTTTTAAAAATTTTATATTTAAT 320
Anj21                TCCAATTATTATTCTTTCAATTATTTCTTTTCCTTCACTTTTAAAAATTTTATATTTAAT 320
MOH28A                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA 570
NGA11B                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA 570
ANJ18A                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA 570
MOH10                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA 570
NGA26B_(reversed)          TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA 571
NGA16                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA 571
MOH12C                TCCAATTATTATTCTATTAATTATTTGTTTTCCATCATTAAAAATTTTATATTTAATTGA 502


NGA9                TGA-----AATTGTAAATCCTTTTTTTTTCAATTAAATCAATTGGTCAC---------- 616
NGA6                ATGA----AATTGTAAATCCTTTTTTTTCAATTAAATCAATTGGTCATC---------- 618
NGA14A                TG-A----AATTGTAAATCCTTTTTTTTCAATTAAATCAATTGGTCATCAA--------- 619
NGA7A                TG-A----AATTGTAAATCCTTTTTTTTCAATTAAATCAATTGGTCTC---------- 616
NGA11A                TG-A----AATTGTAAATCCTTTTTTTTCAATTAAATCAATTGGTCTC---------- 616
NGA15                TG-A----AATTGTAAATCCTTTTTTTTCAATTAAATCAATTGGTCAC---------- 616
NGA14B                TG-A----AATTGTAAATCCTTTTTTTTCAATTAAATCAATTGGTC---------- 614
NGA20                TG-A----AATTGTAAATCCTTTTTTTTCAATTAAATCAATTGGTC---------- 614
NGA10                TG-A----AATTGTAAATCCTTTTTTTTCAATTAAATCAATTGGTCACAATGAATGGGGA 628
NGA5B                TG-A----AATTGTAAATCCTTTTTTTTCAATTAAATCAATTGGTCACAA---------- 618
ANJ35A                TGAAATGAAATTGCAAATCCATTCTTTT-----CTATCTATTGGTCAT---------- 325
ANJ26B                TGA--TGAAATTGTAAATCCTTTTTTTTCAATTAAATCAATTGGTCT---------- 536
NGA7C                TTA--TGAAATTGTAAATCCCTTTTTTTCAATTAAATCAATTGGTCATC---------- 625
NGA31                TTA--TAAATTTT-AAATCCTTTTTTTTCAATTAAATCAATTGGTCATCAATGAATTT-- 615
MOH29                TGA--TGAAATTGTAAATCCTTTTTTTTCAATTAAATCAATTGGTCATCG---------- 607
NGA1                TGA--TGAAATTGTAAATCCCTTTTTTTC---------- 586
MOH15B                TGA--TGAAATTGTAAATCCTTTTTTTTCAATTAAATCAATTGGTCATCAATG------- 556
MOH40                TGA--TGAAAA---------- 538
MOH42B                TGA--TGAAATTGTAAATCCTTTTTTTTCAATTAAATCAATTGGTCATAA---------- 577
MOH39                TG---------- 531
ANJ39                TGAAATGAAATTGCAAATCCATTCTTTTCT-----ATCTATTGGTCTCAAATGATATTG- 375
ANJ17                TGAAATGAAATTGCAAATCCATTCTTTTCT-----ATCTATTGGTCATCAATGATATTG- 375
ANJ37                TGAAATGAAATTGCAAATCCATTCTTTTCT-----ATCTATTGGTCATA---------- 365
ANJ10                TGAAATGAAATTGCAAATCCATTCTTTTC-----TATCTATTGGTCATCAATGATATTGA 374
MOH42A                TGA--TGAAATTGTAAATCCTTTTTTTTCAATTAAATCAATTGGTCATC---------- 561
MOH14                TGA--TGAAATTGTAAATCCTTTTTTTTCAATTAAATCAATTGGTCATCA---------- 562
MOH27                TGA--TGAAATTGTAAATCCATTTTTTTCAATTAAATCAATTGGTCATC---------- 566
MOH7                TGA--T---------- 523
NGA8B                TGA--TGAAATTGTAAATCCTTTTTTTTCAATTAAATCAATTGGTCTCAGTGAATGGGGG 622
ANJ36                TGA--TGAAAATGTAAATCCTTTTTTTTCAAT---------- 525
MOH23C                ---------- 387
M0H8                TGAAATTGTAAATCCTTTTT-----TTTCAATTAAATCAATTGGTCATCAA---------- 536
MOH17B                TGAAATGAAATTGCAAATCTATTCTTTTCTATTAAATGCATTGGTCATCAATGATATTGA 404
```

```
MOH16B                ATTGATGAAATTGTAAATCCTTTTTTTTTCAATTAAATCAATTGGTCATCAATGA------  534
MOH11A                TGAAATGAAATTGCAAATCC-----ATTCTTTTCTATCTATTGGTCACAAATGATATT--  386
NGA7B                 TGAAATGAAATTACAAATCT-----ATTCTTTTCTATCTATTGGTCAA------------  390
NGA2A                 TGAAATGAAATTACAAATCT-----ATTCTTTTCTATCTATTGGTCA-------------  389
NGA2B                 TGAAATGAAATTGCAAATCC-----ATTCTTTTCTATCTATTGGTCAT------------  360
ANJ20B                TGAAATTGTAAATCCTTTTT-----TTTCAATTAAATCAATTGGTCACA-----------  530
MOH25A_(reversed)     TGAAATGAAATTACAAATCT-----ATTCTTTTCTATCTATTGGTCATCAATGATATT--  410
NGA28A                TGAAATTGTAAATCCTTTTT-----TTTCAATTAAATCAATTGGTCATCAA---------  618
MOH9                  TGAAATTGTAAATCCTTTTT-----TTTCAATTAAATCAATTGGTC--------------  613
NGA21                 TGAAATTGTAAATCCTTTTT-----TTTCAATTAAATCAATTGGT--------------  612
NGA11C                TGAAATTGTAAATCCTTTTT-----TTTCAATTAAATCAATTGGTC--------------  613
NGA3                  TGAAATTGTAAATCCTTTTT-----TTTCAATTAAATCAATTGGTCAC------------  615
ANJ29                 TGAAATTGTAAATCCTTTTT-----TTTCAATTAAATCAATTGGTCATC-----------  541
KT828418.1_(A1_MAD3)  ---------- 545
KT828427.1_(A1_MAD12) ---------- 545
NGA14C                TGAAATTGTAAATCCTTTTT-----TTTCAATTAAATCAATT-----------------  597
NGA12                 TGA--TGAAATTGTAAATCCTTTTTTTTCAATTAAATCAATTGGTCACA-----------  600
NGA5A                 TGA--TGAAATTGTAAATCCTTTTTTTTCAATTAAATCGATTGGTCATC-----------  618
NGA8A                 TGA--TGAAATTGTAAATCCTTTTTTTTCAATTAAATCAATTGGTCTCAGTGAAT-----  624
ANJ12                 TGA--TGAAATTGTAAATCCTTTTTTTTCAATTAAATCAATTGGTCAC------------  617
MOH11B                ---------- 529
NGA4                  TGA--TGAAATTGTAAATCCTTTTTTTTCAATTAAATCAATTGGTCACAGT---------  621
MOH34                 TGA--TGAAATTGTAAATCCTTTTTTTTCAATTAAATCAATTGGTCACAA----------  620
NGA15B                TGA--TGAAATTGTAAATCCTTTTTTTTCAATTAAATCAATTGGTCATC-----------  619
MOH19                 TGA--TGAAATTGTAAATCCTTTTTTTTTTCAATTAAAAATCAAAA--------------  616
MOH23B                TGAAATGAAATTGCAAATCCATTCTTTTCTATCTA-----TTGGTCATCAATGATATTGA  379
MOH36A                TGA--TGAAATTGTAAATCCTTTTTTTTCAATTAAATCAATTGGTCATCA----------  602
ANJ2                  TGAAATGAAATTGCAAATCCATTCTTTTCTTCTATTGGT---------------------  292
ANJ27                 TGAAATGAAATTGCAAATCCATTCTTTT-----CTATCTATTGGTCAT------------  361
MOH37                 TGA--TGAAATTGTAAATCCTTTTTTTTCAATTAAATCAATTGGTCATCAATG-------  584
MOH26                 TGA--TGAAATTGTAAATCCTTTTTTTTCAATTAAATCAATTGGTCAT------------  579
AJ20C                 TGAAATGAAATTGCAAATCCATTCTTTTCTATC-----TATTGGTCAT------------  370
MOH18                 TGAAATGAAATTGCAAATCCATTCTTTTCTATC-----TATTGGTCATCAATGATATTGA  382
ANJ06                 TG--ATGAAATTGTAAATCCTTTTTTTTCAATT-----AAATCAATTGGT----------  315
ANJ5                  TGAAATGAAATTGCAAATCCATTCTTTTCTATC-----TA-----TTGGT----------  296
ANJ7B                 TGAAATGAAATTGCAAATCCATTCTTTTCTATC-----TA-----TTGGT----------  296
ANJ1                  TGAAATGAAATTGCAAATCCATTCTTTTCTATC-----TA-----TTGGT----------  296
MOH4                  TGAAATTGTAAATCCTTTTTTTTCAATTAAATC-----AAT------------------  550
NGA19                 TGAAATTGTAAATCCTTTTTTTTCAATTAAATC-----AATTGGTC-------------  624
MOH15A                TGAAATTGTAAATCCTTTTTTTTCAATTAAATC-----AATTGGTCA------------  624
NGA13                 TGAAATGAAATTGCAAATCCATTCTTTTCTATC-----TATTGGTCATC-----------  360
NC_001566.1_(RefSeq)  TGAAATTGTAAATCCTTTTTTTTCAATTAAATC-----AATTGGTCATCAATGATATTGA  3935
MOH30                 TGAAAATGTAAATCCATTTTTTTCAATTAAA-----------------------------  583
NGA25_(reversed)      TGAAATTGTAAATCCTTTTTTTTCA---------- 589
MOH36B                TGAAATTGTATATCCTTTTTTTTCAATTAAATC-----AATTGGTCATCAG---------  585
ANJ18B                TGAAATTGTAAATCCTTTTTTTTCAATTAAATC-----AATTGGTCATCAA---------  580
ANJ8B                 TGAAATTGTAAATCCTTTTTTTTCTTTTAAATC-----AATTGGTCA-------------  576
NGA34                 TGAAATTTTAAATCCTTTTTTTTCAATTAAATC-----AATTGGTCATCAATGATATT--  587
MOH3                  TGAAATTGTAAATCCTTTTTTTTCAATTAAATC-----AATTGGTCAT------------  577
ANJ15A                TGAAATTGTAAATCCTTTTTTTTCAATTAAATC-----AATTGGTCATCAAT--------  581
MOH32                 TGAAATTGTAAATCCTTTTTTTTCAATTAAATC-----AATTGGTCATCAATGA------  583
ANJ26A                TGAAATTGTAAAATCCCTTT---------- 554
ANJ22                 TGAAATTGTAGATCCATTTTTTTCAATTAAATC-----AATTGGGTCTCC----------  579
ANJ25                 TGAAATTGTAGATCCATTTTTTTCT---------- 559
ANJ28                 TGAAA---------- 539
ANJ23                 TGAAATTGTAAATCCTTTTTTTTCAATTAAATC-----AATTGGTC--------------  496
ANJ19B                TGAAAATGAAGATCAATTTTTTTCAATTAAATA-----AAT------------------  569
NGA17                 TGAAATTGTAAATCCTTTTTTTTCAATTAAATC-----AATTGGTCATAAT---------  579
NGA23                 TGAAATTGTAAATCCTTTTTTTTCAATTAAATC-----AATTGGTCATCAATGAA-----  583
MOH25B                TGAAATTGTAAATCCTTTTTTTTCAATTAAATC-----AATTGGTCATC-----------  577
NGA22                 TGAAATTGTAAATCCTTTTTTTTCAATTAAATC-----AATTGGTCATC-----------  577
NGA26A_(reversed)     TGAAATTGTAAATCCTTTTTTTTCAATTAAATC-----AATT------------------  570
ANJ33                 TGAAATTGTAAATCCATTTTTTTCAATTAAATC-----AAT------------------  569
MOH24                 ---------- 316
ANJ35A_2              TGAAATGAAATTGCAAATCCATTCTTTTCTATC-----TATTGGTCATAATGATATTGA-  374
ANJ35B                TGAAATGAAATTGCAAATCCATTCTTTTCTATC-----TATTGGTCAC------------  363
ANJ34                 TGAAATGAAATTGCAAATCCATTCTTTTCTATC-----TATTGGTCATCAATGAT-----  370
ANJ31                 TGAAATGAAATTGCAAATCCATTCTTTTCTATC-----TATTGGT---------------  360
ANJ30                 TGAAATGAAATTGCAAATCCATTCTTTTCTATC-----TATTGGTC-------------  361
Anj21                 TGAAATGAAATTGCAAATCCATTCTTTTCTATC-----TATTGGTCAT------------  363
MOH28A                TGAAATTGTAAATCCTTTTTTTTCAATTAAATC-----AATTGG----------------  609
NGA11B                TGAAATTGTAAATCCTTTTTTTTCAATTAAATC-----AATTGGTCATAATGAATTT---  622
ANJ18A                TGAAATTGTAAATCCTTTTTTTTCAATTAAATC-----AATTGGTC-------------  611
MOH10                 TGAAATTGTAAATCCTTTTTTTTCAATTAAATC-----AATTGGTCACA----------  614
NGA26B_(reversed)     TGAAATTGTAAATCCTTTTTTTTCAATTAAATC-----AATTGGTC-------------  612
NGA16                 TGAGATTGTAAATCCTTTTTTTTCAATTAAATC-----GGTTGGTCATCCGTGAA-----  621
MOH12C                TGAAATTGTAAATCCTTTTTTTTCAATTAAATC-----AATTGGTCATCAATG-------  550
```

APPENDIX IV: GeneAlex file for predicted queen genotypes data. The data was used for genetic diversity, population differentiation and population structure analysis. Individuals with * were manually predicted from the genotyped workers whilst the rest were obtained as a result of sibship analysis in Colony version 2.0.6.4. 1 represent missing data

12	73	3																						
Ind	pop	A113		A56		A88		AP81		A28		A8		AC306		AP33		A43		AP23		B124		UN4987	
ANJ13	ANJ	212	212	283	283	150	152	134	136	139	139	169	169	177	177	240	240	142	142	104	104	239	239	175	175
ANJ14	ANJ	212	212	283	283	150	150	138	138	131	139	169	169	173	177	248	248	152	152	104	104	239	239	171	171
ANJ2	ANJ	210	210	283	291	140	140	134	134	131	131	169	169	171	177	242	242	152	152	104	104	233	233	175	175
ANJ21	ANJ	212	212	283	291	140	152	138	138	139	139	169	169	171	171	242	242	152	152	104	104	233	239	171	171
ANJ22	ANJ	212	212	291	291	152	152	134	134	139	139	169	169	171	177	240	240	142	142	104	108	233	235	157	157
ANJ23	ANJ	212	212	283	291	160	160	134	134	131	139	169	169	173	177	238	238	142	142	104	104	239	239	157	157
ANJ24	ANJ	212	212	283	291	154	154	136	136	139	139	169	169	171	173	240	242	152	152	104	104	239	239	157	157
ANJ25	ANJ	214	214	283	283	160	160	136	136	133	139	169	169	171	177	236	236	152	152	104	104	237	237	169	169
ANJ3	ANJ	210	210	283	283	149	149	134	134	139	139	169	169	171	177	240	242	152	152	104	104	235	235	171	171
ANJ33	ANJ	214	214	283	283	154	154	136	136	139	139	169	169	177	181	240	242	152	152	104	104	233	233	157	157
ANJ36	ANJ	214	214	283	283	140	140	138	138	133	133	169	169	171	173	242	242	152	152	108	108	233	239	173	175
ANJ37	ANJ	214	214	283	283	140	140	136	136	131	131	169	169	173	173	240	242	152	152	104	108	233	239	169	169
ANJ39	ANJ	214	214	291	291	150	150	128	128	139	139	169	169	173	173	240	240	142	142	104	104	233	233	169	171
ANJ43	ANJ	216	216	283	283	150	150	136	136	139	139	169	169	171	177	240	240	142	142	104	104	219	219	169	171
ANJ51	ANJ	212	214	283	283	150	150	136	138	133	133	169	169	177	181	240	240	154	154	104	104	235	235	171	171
ANJ6	ANJ	210	212	283	283	152	152	136	136	133	139	169	169	181	181	242	248	152	152	104	104	237	237	169	169
ANJ8	ANJ	212	212	283	283	140	152	134	134	131	131	169	169	177	177	236	240	152	152	104	104	233	233	169	169
ANJQ1*	ANJ	212	214	283	283	140	152	130	136	139	139	169	169	171	177	240	240	142	152	104	108	231	235	169	171
ANJQ2*	ANJ	212	214	283	283	140	140	132	136	131	139	169	169	177	181	240	240	152	152	104	104	231	235	171	175
ANJQ3*	ANJ	210	210	283	283	140	152	134	136	131	131	169	169	171	177	240	248	152	152	104	104	233	235	171	173
ANJQ4*	ANJ	212	214	283	283	140	152	130	132	131	131	169	169	175	177	240	248	152	152	104	104	233	237	169	173
MOH1	MOH	210	210	285	285	152	152	138	138	133	133	169	169	173	173	240	240	152	152	104	104	233	233	175	175
MOH12	MOH	210	210	285	285	140	160	134	136	131	131	169	169	177	177	240	240	148	152	104	104	223	223	157	169
MOH14	MOH	210	210	285	285	160	160	126	126	133	139	169	169	181	181	240	240	144	144	104	108	223	233	175	175

MOH15	MOH	210	210	285	285	152	154	132	136	125	125	169	171	181	181	240	240	148	152	104	108	221	223	169	173
MOH16	MOH	210	210	285	285	152	152	132	136	133	133	169	169	181	181	231	240	152	152	104	108	223	223	169	171
MOH17	MOH	210	210	285	285	150	152	126	134	133	133	169	169	181	181	240	240	152	152	108	108	233	233	173	173
MOH18	MOH	210	210	285	285	140	153	134	134	135	135	169	169	183	183	246	246	144	144	104	104	233	233	171	171
MOH19	MOH	210	210	285	285	150	150	126	136	133	133	171	171	171	171	240	240	144	144	104	104	223	237	169	175
MOH23	MOH	212	212	283	283	140	140	128	128	133	133	169	169	171	181	240	240	144	144	104	104	233	233	169	173
MOH24	MOH	212	212	285	285	140	154	126	136	133	133	171	171	181	181	240	240	148	148	104	104	223	223	157	157
MOH25	MOH	212	212	285	285	150	150	138	138	139	139	169	169	171	181	240	240	144	148	108	108	225	233	169	171
MOH26	MOH	212	212	285	285	140	140	132	134	139	139	169	171	171	181	246	246	152	152	104	108	233	233	169	169
MOH29	MOH	212	212	285	285	140	150	138	138	133	133	165	171	181	181	240	240	144	144	104	104	237	237	173	173
MOH30	MOH	212	212	285	285	140	140	134	136	139	139	169	169	177	177	240	246	144	152	104	104	237	237	173	173
MOH31	MOH	212	212	285	285	152	154	134	136	133	133	169	169	171	171	240	240	152	152	104	104	245	245	171	173
MOH34	MOH	212	212	283	285	152	152	126	126	139	139	173	173	177	181	240	240	144	152	108	108	233	233	157	157
MOH38	MOH	214	214	283	285	140	140	138	138	133	133	169	169	173	177	240	246	144	144	104	104	235	235	169	169
MOH4	MOH	210	212	285	285	140	140	134	136	133	133	166	166	181	181	240	240	144	144	104	108	237	237	169	173
MOH41	MOH	212	212	285	285	154	154	140	140	133	133	169	169	181	181	242	242	148	148	104	108	235	235	169	171
MOH42	MOH	212	212	291	291	152	152	128	132	135	135	169	169	173	173	240	240	148	148	104	108	225	225	169	173
MOH43	MOH	210	212	285	285	150	150	134	134	123	123	169	169	171	171	240	240	148	148	104	104	225	225	173	175
MOH5	MOH	210	210	283	283	140	140	136	136	123	123	169	171	173	173	240	240	152	152	108	108	233	233	169	169
MOH7	MOH	210	210	285	285	150	150	126	126	135	135	169	169	171	171	240	240	148	148	104	104	223	223	157	169
MOH8	MOH	210	210	285	285	152	152	132	134	139	139	169	169	171	171	240	240	142	144	104	108	233	233	169	169
MOHQ1*	MOH	212	212	285	285	140	140	126	136	139	139	169	171	173	177	240	240	144	152	104	108	223	223	169	173
MHQ2*	MOH	210	212	283	285	140	152	126	136	133	133	169	169	171	177	240	240	144	152	104	108	235	237	169	169
MOHQ3*	MOH	212	212	283	285	140	140	136	138	139	139	169	171	171	177	240	246	144	144	104	104	223	237	169	173
MOHQ4*	MOH	210	210	285	285	140	140	136	138	139	139	169	169	171	177	240	240	144	152	104	108	235	237	157	169
NGA10	NGA	212	214	283	285	150	150	128	128	1	1	1	1	168	168	222	235	142	142	104	104	223	230	169	169
NGA11	NGA	212	214	285	285	150	152	128	128	1	1	1	1	169	169	222	222	142	142	104	104	225	225	169	169
NGA12	NGA	212	212	285	285	150	150	128	128	1	1	1	1	170	170	237	237	152	152	104	104	233	237	169	169
NGA13	NGA	212	212	283	283	150	150	128	128	1	1	1	1	173	173	222	222	126	126	104	104	223	223	169	171
NGA14	NGA	212	212	285	291	150	150	132	136	1	1	1	1	170	170	222	222	142	142	104	104	223	223	169	171
NGA16	NGA	212	212	283	283	150	150	130	134	1	1	1	1	168	168	237	237	152	152	104	104	225	225	157	157
NGA17	NGA	212	212	291	291	150	150	134	136	1	1	1	1	170	170	235	235	142	142	104	104	235	235	169	171

NGA18	NGA	212	212	285	285	150	150	134	134	1	1	1	1	170	170	235	235	148	152	104	104	235	235	171	171
NGA2	NGA	212	212	285	285	150	150	130	130	1	1	1	1	170	174	222	235	126	126	104	104	223	235	169	171
NGA20	NGA	212	212	283	283	150	150	134	136	1	1	1	1	172	172	235	237	152	152	104	104	237	237	173	173
NGA21	NGA	212	214	283	291	150	152	136	136	1	1	1	1	168	168	235	235	142	142	104	104	235	235	169	169
NGA23	NGA	212	214	283	283	150	152	136	136	1	1	1	1	170	170	222	222	142	142	104	104	223	223	171	175
NGA24	NGA	212	212	283	283	150	150	134	142	1	1	1	1	170	170	233	237	152	152	104	104	233	233	171	171
NGA26	NGA	218	218	285	291	140	140	128	128	1	1	1	1	170	170	237	237	126	152	104	104	223	223	171	171
NGA27	NGA	214	214	285	285	150	150	128	134	1	1	1	1	170	170	222	222	142	142	104	104	225	225	169	171
NGA3	NGA	212	212	283	283	150	150	128	136	1	1	1	1	170	182	235	235	126	126	104	104	235	235	169	171
NGA32	NGA	214	214	283	285	150	150	136	136	1	1	1	1	173	173	233	235	142	142	104	104	223	233	175	175
NGA33	NGA	214	214	285	291	150	150	134	136	1	1	1	1	168	168	222	222	152	152	104	104	223	223	169	169
NGA5	NGA	210	210	291	291	140	150	130	134	1	1	1	1	173	173	237	237	142	142	104	104	223	223	175	175
NGA7	NGA	210	210	283	283	140	140	136	136	1	1	1	1	168	168	235	235	142	142	104	104	235	235	169	171
NGAQ1*	NGA	210	212	285	291	140	152	132	136	1	1	1	1	168	170	222	224	142	142	104	104	223	225	157	169
NGAQ2*	NGA	210	212	291	287	140	140	134	142	1	1	1	1	168	168	222	233	142	142	104	104	233	237	169	169
NGAQ3*	NGA	210	210	283	291	140	152	136	138	1	1	1	1	170	174	222	235	142	152	104	104	223	235	1	1
NGAQ4*	NGA	210	214	285	291	140	140	136	136	1	1	1	1	170	170	235	235	142	142	104	104	223	235	169	171

APPENDIX V: Table showing the deviations from HWE equilibrium. The table was used to generate the heatmap

	Anjouan			**Moheli**			**Ngazidja**		
	χ^2	df	Pr(χ^2)	χ^2	df	Pr(χ^2)	χ^2	df	Pr(χ^2)
A113	40.386	6	3.823750e-07	44.9382716	3	9.536458e-10	35.99902	6	2.757833e-06
A56	3.070	1	7.974608e-02	32.7769376	3	3.589166e-07	12.88838	6	4.484319e-02
A88	78.448845	15	1.340488e-10	46.0302217	15	5.258484e-05	13.39649	3	3.853113e-03
AP81	54.622631	15	2.066271e-06	67.6741650	21	8.251074e-07	33.97822	21	3.643426e-02
A28	16.102978	3	1.080168e-03	102.9570370	15	3.552714e-15	0.00000	0	1.000000e+00
A8	0.000000	0	1.000000e+00	66.1378512	10	2.454145e-10	0.00000	0	1.000000e+00
AC306	10.534224	10	3.949371e-01	56.8925000	10	1.396351e-08	91.22911	21	9.896095e-11
AP33	34.213994	10	1.698971e-04	16.5711662	6	1.099546e-02	23.36000	10	9.492987e-03
A43	32.293333	3	4.539011e-07	15.8556984	6	1.454981e-02	30.77310	6	2.800700e-05
AP23	2.135895	1	1.438865e-01	0.4288307	1	5.125633e-01	0.00000	0	1.000000e+00
B124	51.709126	15	6.309339e-06	87.0313933	21	5.221859e-10	43.36761	15	1.380352e-04
UN4987	34.157791	10	1.553367e-02	21.9168842	10	1.553367e-02	46.82195	10	1.017350e-06

APPENDIX VI: AMOVA results showing significant variations between populations, between samples within populations and within samples with populations

	Df	Sigma	Covariance (%)	Phi	p value
Between populations	2	2.27	24.97	0.64	0.001
Between samples within population	70	3.56	39.12	0.52	0.001
Within samples within populations	73	3.27	35.91	0.25	0.001
Total variations	145	9.11	100.00		