THE POPULATION GENETIC STRUCTURE OF THE MALARIA MOSQUITO *ANOPHELES MELAS* THROUGHOUT ITS WEST-AFRICAN RANGE

A Thesis

by

KEVIN CANNING DEITZ

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

December 2011

Major Subject: Entomology

THE POPULATION GENETIC STRUCTURE OF THE MALARIA MOSQUITO *ANOPHELES MELAS* THROUGHOUT ITS WEST-AFRICAN RANGE

A Thesis

by

KEVIN CANNING DEITZ

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Chair of Committee,	Michel A. Slotman
Committee Members,	J. Spencer Johnston
	Konstantin V. Krutovsky
	Raul F. Medina
Head of Department,	David W. Ragsdale

December 2011

Major Subject: Entomology

ABSTRACT

The Population Genetic Structure of the Malaria Mosquito *Anopheles melas* Throughout Its West-African Range. (December 2011)
Kevin Canning Deitz, B.S., State University of New York
Chair of Advisory Committee: Dr. Michel A. Slotman

Anopheles melas is a brackish water mosquito found along the coast of West-Africa where it can be the dominant malaria vector locally. In order to facilitate genetic studies of this species and to examine the usefulness of microsatellite markers when used in a sibling species, 45 microsatellite loci originally developed for *Anopheles gambiae* were sequenced in *An. melas*. These loci were evaluated on their suitability as polymorphic markers based on repeat structure, length, and polymorphism in wild *An. melas* populations. Of the 45 loci, 18 were not considered promising markers in *An. melas*. A total of 48 out of 90 *An. gambiae* primers contained at least one mismatch with the *An. melas* annealing site. *An. melas*-specific primers were designed for 27 loci, and their variability was examined in two wild populations from Equatorial Guinea. Based on a low level of polymorphism, Hardy-Weinberg disequilibrium, or poor amplification, a further 12 loci were excluded. The remaining fifteen loci were screened in four additional wild populations from a wider geographic region including Equatorial Guinea, Cameroon, The Gambia, and Guinea Bissau. These loci showed an average heterozygosity ranging from 0.18 to 0.79, with 2.5 to 15 average alleles per locus, yielding 13 highly polymorphic markers and two loci with more limited variability in a wide geographic region. To examine the effects of cross species amplification, five of the original *An. gambiae* markers were also amplified in the *An. melas* populations. Null alleles were found for one of these *An. gambiae* markers. We discuss the pitfalls of using microsatellite loci even in a very closely related species, and conclude that in addition to the well-known problem of null alleles associated with this practice, many loci may

prove to be of very limited use as polymorphic markers even when used in a sibling species.

Fifteen *An. melas*-specific markers were subsequently amplified and analyzed in 11 wild *An. melas* populations from throughout the range of this species, including Bioko Island, Equatorial Guinea. We analyzed pair-wise population differentiation between all populations, and found that all but two comparisons were significant (p-val.<0.05), and populations clustered into three distinct groups representing Bioko Island, Central Africa, and West Africa populations. A Bayesian clustering analysis found little, if any, evidence for migration from mainland to Bioko Island populations, although there was evidence of migration from Bioko Island to the West population cluster, and from the Central to the West population cluster. Simulations of historical gene followed these same patterns and further support our predictions of unidirectional gene flow. Comparison of 1161 nucleotides amplified and sequenced from the ND4 and ND5 regions of the mtDNA showed that differentiation between *An. melas* population clusters is on par with levels of differentiation between member species of the *An. gambiae* complex, with low support for internal nodes in a maximum likelihood tree, which suggests that observed *An. melas* clusters are not monophyletic. From this we hypothesize that Bioko Island *An. melas* populations are derived from Tiko, Cameroon, and that these populations became isolated from one another when sea levels rose after the last glaciation period (≥10,000-11,000 years ago), cutting off Bioko Island populations from the mainland and significantly reducing migration. Our conclusions have implications for vector control within the region, as Bioko Island is the subject of an intensive malaria control campaign, and the lack of migration from mainland West Africa to Bioko Island make it unlikely that eradicated populations of this malaria vector will be repopulated by mainland immigrants.

DEDICATION

For Anna and my family

ACKNOWLEDGEMENTS

I would like to thank my advisor and committee chair Dr. Michel A. Slotman for his support and guidance throughout the course of this research project, and my committee members Dr. J. Spencer Johnston, Dr. Konstantin V. Krutovksy, and Dr. Raul F. Medina for their insight and direction. Thanks also to my fellow graduate students and lab members who were always willing to answer questions and offer a helping hand.

I am also grateful to Dr Gregory Lanzaro for providing *An. melas* samples from Guinea Bissau and to Drs Frederic Simard, C. Antonio-Nkondjio and Parfait H. Aweono-Ambene for logistical support during collections in Ipono, Cameroon. I also offer my thanks to Dr. Chris Schwabe, Dr. Gloria Nseng, Dr. Luis Segura, Ed Aldrich and Jaime Kuklinski for logistical support during field collections in Equatorial Guinea. Additional logistical support was provided by Medical Care Development International Inc. and Marathon Oil Corporation. This study was supported by an operational research grant awarded by the Bioko Island Malaria Control Program to Dr. Michel A. Slotman and Texas Agrilife Research funds.

TABLE OF CONTENTS

Page

CHAPTER I
INTRODUCTION AND LITERATURE REVIEW

Anopheles (An.) melas is a member of the *An. gambiae* species complex comprised of seven morphologically indistinguishable species thought to be a single species until crosses revealed the presence of male hybrid sterility (Davidson 1962, White 1974, Hunt *et al.* 1998). All members of this complex are competent vectors of *P. falciparum,* the human malaria parasite. Malaria is estimated to have caused 781,000 deaths worldwide in 2009, over 90% (709,000) of which occurred in Africa (World Health Organization). Members of the *An. gambiae* species complex range in host specificity from almost entirely anthropophillic (*An. gambiae*) to almost entirely zoophilic (*An. quadriannulatus*), and vary widely in their contribution to malaria transmission and importance as malaria vectors (Garret-Jones *et al.* 1980, Hunt *et al.* 1998, White *et al.* 1980).

The evolution of the complex is characterized by recent origin (as recent as 5,000 years ago, see Coluzzi *et al.* 2002), introgression, adaptive polymorphic inversions and incipient speciation (della Torre *et al.* 1997, Besansky *et al.* 1994, Besansky *et al.* 2003, Slotman *et al.* 2005). Nonrandom distribution of polymorphic chromosomal inversions provides evidence for further genetic division within *An. gambiae, An. arabiensis,* and *An. melas* (Coluzzi *et al.* 2002). While these chromosomal forms are well characterized in the former two species, little information is known about the distribution of *An. melas* chromosomal forms in the wild. On the basis of these inversions several chromosomal forms have been described in *An. gambiae,* termed Bamako, Bissau, Forest, Mopti, and Savanna (Bryan *et al.* 1982, Coluzzi *et al.* 1979, 1985). These were originally hypothesize to represent reproductively isolated populations, but are now widely recognized as ecologically adaptive forms within *An. gamb*iae. The recent recognition of two ribosomal DNA (rDNA) molecular forms, M and S, revealed that *An. gambiae* is an

---

This thesis follows the style of *Molecular Ecology.*

example of an incipient speciation in action (Gentile *et al.* 2001, Mukabayire *et al.* 2001). M and S molecular forms were originally characterized due to their association with the Mopti and Savanna chromosomal forms. However, further studies demonstrated that both the M and S rDNA genotypes were found in individuals with the Savanna and Forest chromosomal forms (della Torre *et al.* 2001, 2002). Strong assortative mating between molecular forms (also referred to as races) does occur, although reproductive isolation is not complete as M/S heterozygotes are found in very low frequencies in sympatric populations (della Torre *et al.* 2001, Tripet *et al.* 2001, Wondji *et al.* 2002). Yet another layer of genetic complexity was revealed in *An. gambiae* when analyses demonstrated further subdivision within the M and S molecular forms (Slotman *et al.* 2007a). These authors demonstrated that *An. gambiae* individuals that shared the S ribosomal molecular form, but differed in their chromosomal form (Savanna or Forest), had little genetic differentiation between them (F_{ST}=0.0053). In contrast, significant genetic differentiation (F_{ST}=0.0406) was found between M molecular form individuals with different chromosomal forms (Forest and Mopti).

The majority of evolutionary and population genetic studies of members of the *An. gambiae* complex have concentrated on the two most prominent malaria vectors, *An. gambiae* and *An. arabiensis*. This has resulted in a dearth of information about the genetic diversity and population structure of other species in the *An. gambiae* complex. Knowledge of the population genetic structure of vector populations can better inform current control methods; isolated populations that show low levels of migration may be optimal candidates for increased vector control measures. Additionally, knowledge of the movement of insecticide resistant alleles within and between mosquito populations and species can help inform effective malaria control. By gaining a thorough understanding of gene flow and migration between vector populations, we can identify those which are may serve as reservoirs for genetic diversity, and those that may act as source populations from which migrants have emigrated. Lehmann *et al.* (2003) utilized 11 microsatellite loci to analyze the population genetic structure of *An. gambiae* across its sub-Sahara Africa range, and for the most part found low genetic differentiation

between populations across this vast geographic scale. The differentiation that was found was attributed to assortative mating among M and S molecular forms (Favia *et al.* 1997, Gentile *et al.* 2001, Mukabayire *et al.* 2001) and a geographical barrier to gene flow, viz. the Rift Valley.

Microsatellite markers have been widely used in studies of population genetics, molecular ecology, evolutionary genetics, and genetic mapping. Although recent advances in genomics have made available other methods of populations-scale genotyping (Thomas & Klaper 2004, Stinchcombe & Hoekstra 2008, Ekblom & Galindo 2011), microsatellite markers remain a cost-effective and useful tool. Although microsatellite markers are developed for specific species, it is common for such markers to be used across species boundaries in closely related taxa. Such use of these markers is often complicated by the occurrence of null alleles in the data set (Callen *et al.* 1993, Paetkau & Strobeck 1995, Pompanon *et al.* 2005), which occur when specific alleles fail to amplify due to base pair mis-matches in the annealing site of one or both primers.

Data sets containing null alleles will be deficient in heterozygotes and such errors can result in incorrect parentage assignment or exclusion (Dakin & Avise 2004), and biases in population genetic data. Null alleles have been shown to underestimate intrapopulation variance, and overestimate interpopulation genetic differentiation (F_{ST}) and genetic distance (Chapuis & Estoup 2007). These problems can be ameliorated through careful selection and screening of microsatellite loci, the identification and quantification of genotyping errors (Guichoux *et al.* 2011, Pompanon *et al.* 2005), screening for the presence of null alleles in the dataset by comparing allele frequencies and heterozygote proportions, and the subsequent correction of allele frequencies (van Oosterhout *et al.* 2004). Despite the availability of methods to detect and correct for null alleles (Brookfield 1996, Chakraborty *et al.* 1992), Dakin and Avise (2004) found that out of 233 reviewed articles, 90% of the studies that identified null alleles merely reported their presence and failed to take corrective action. This is in part due to the inability to use null-corrected allele frequencies in a variety of analyses that are based upon individual multi-locus genotypic data; for example, the popular Bayesian-

assignment tests implemented in the programs STRUCTURE 2.1 (Falush *et al.* 2003, Pritchard *et al.* 2000) and BAPS 5.4 (Corander *et al.* 2006, Corander *et al.* 2008), and the inference of parentage and kinship (Queller *et al.* 1993).

Some authors have made an effort to develop microsatellite markers for use in multiple species. For example, Dawson *et al.* (2010) identified microsatellites conserved across passerine birds by characterizing Expressed Sequence Tag (EST) loci with sequence homology between the zebra finch and chicken genomes. The most common strategy, however, is to simply screen a number of non-specific markers in a target species, and to discard those loci that do not consistently amplify, are monomorphic, or deviate from HWE (Chambers *et al.* 2004, Kim *et al.* 2004, Miles *et al.* 2009, Wilson *et al.* 2004).

If microsatellite loci are used across species boundaries, the presence of null alleles is increasingly likely as genetic distances increase (Dakin & Avise 2004, Jarne & Lagoda 1996). Barbará *et al.* (2007) found that the cross-species utility of microsatellite loci, including the level of polymorphism, was inversely correlated with taxonomic distance in a large number of animals, fungi, and plants. Similarly, Hendrix *et al.* (2010) showed microsatellite cross-species amplification success was negatively correlated ($r = -0.84$) with increasing mtDNA sequence divergence between species within the family Salamandridae. These authors also reported the amplification of a large number of loci over considerable genetic distances. Of 20 tested loci, 12 or 13 amplified in species with mtDNA divergence of 14% - 16%. However, these authors only tested a few individuals per species to see if the loci amplified, and did not test for the presence of null alleles in population samples.

Despite the well-known null allele problem associated with using microsatellite markers across species boundaries (Panova *et al.* 2008), this practice is common in studies of a wide variety of organisms, including studies of the malaria vectors within the *An. gambiae* complex. Of the seven morphologically indistinguishable species within the *An. gambiae* complex (Coetzee 2004, Coluzzi *et al.* 1979, Favia *et al.* 1997, Hunt *et al.* 1998), two are major malaria vectors and are intensively studied. Insight into the

population genetics of these same species can inform the potential introduction of transgenes into populations (e.g., Tabachnick 2003, Tripet *et al.* 2005), help to monitor and predict the spread of insecticide resistance alleles (e.g. Chandre *et al.* 1999, Reimer *et al.* 2005, Weill *et al.* 2000), and gauge the efficacy of malaria control campaigns (e.g. Pinto *et al.* 2002). Studies to date were facilitated by the development of 131 microsatellite markers for *An. gambiae* by Zheng *et al.* (1996). These loci have since been used in the closely related malaria vector *An. arabiensis* for genetic mapping (Slotman *et al.* 2004), population structure (Donnelly *et al.* 1999, Kamau *et al.* 1999, Onyabe *et al.* 2001, Kent *et al.* 2007), and effective population size estimation studies (Simard *et al.* 2000, Wondji *et al.* 2005). There are however, potential problems associated with this approach. Kent *et al.* (2007) evaluated 20 of the *An. gambiae* loci for use in *An. arabiensis* and found that only 12 amplified well. Furthermore, two of the 12 loci that amplified failed to produce PCR products in a previous study (Donnelly *et al.* 1999), demonstrating that cross-amplification success, and by extension, null alleles, can be population specific.

An. melas is an understudied member species of the *An. gambiae* complex (Besansky *et al.* 2003), and is an important local vector of malaria (Sharp *et al.* 2007). Nothing is known about the population genetic structure of *An. melas*, and little is known about the patterns of evolutionary divergence between *An. melas* and other species in the *An. gambiae* complex (Besansky *et al.* 1994). Based upon a documented history of introgression between *An. gambiae* and *An. arabiensis* (Besansky *et al.* 2003, Slotman *et al.* 2005), as well as evidence from several immune genes (Parmakelis *et al.* 2008, Parmakelis *et al.* 2010), the genetic distance between *An. gambiae* and *An. arabiensis* appears to be less than between *An. gambiae* and *An. melas*. Knowledge of the population genetic structure of *An. melas,* and estimates of migration rates between Bioko Island, Equatorial Guinea and mainland Africa, will better inform current control measures in this region. Knowledge of *An. melas* migration and population structure will allow control programs to predict if migrants will re-populate extirpated populations, or add genetic variability and potentially beneficial alleles to depressed populations.

An. melas is a brackish-water breeding mosquito whose range is confined to the west coast of Africa where it breeds primarily in mangrove swamps and tidal marshes (Sinka *et al.* 2010). Because of this, population sizes fluctuate with seasonal tides and rainfall. These environmental dynamics regulate the distribution and migration of *An. melas*, confining continuous populations to permanent saline water bodies, and supposedly limiting the dispersal capabilities of individual mosquitoes (Bryan *et al.* 1987). Due to its geographical, environmental, and biological confines, *An. melas* has a disjunct, patchy distribution. Information about the rate of *P. falciparum* transmission by *An. melas* is limited in comparison to other members of the *An. gambiae* complex. A study performed in the Republic of Benin found that when the two species are in sympatry, the infection rate of *P. falciparum* in *An. melas* is significantly less than *An. gambiae* (Akogbeto & Ramano 1999). Bryan *et al.* (1987) found that local populations of *An. melas* in The Gambia can comprise up to 100% of the sampled *An. gambiae* complex mosquitoes, and had a mean *P. falciparum* sporozoite infection rate of 1.5%.

Although not considered a primary vector of malaria because of its limited distribution, studies in Ghana (Tuno *et al.* 2010) and The Gambia (Bryan 1983, Bryan *et al.* 1987, Bøgh *et al.* 2007) have shown that *An. melas* can be an abundant vector in localized areas adjacent to suitable breeding habitat. Furthermore, on Bioko Island, Equatorial Guinea, *An. melas* is an important, and in a several locations, dominant malaria vector (Sharp *et al.* 2007, Slotman *et al.* unpublished data). Although the population genetic structure of *An. gambiae* has been examined in Equatorial Guinea (Moreno *et al.* 2007), no study of the genetic diversity and population structure within *An. melas* in E.G or elsewhere is available, although the recognition of nonrandom assortment of polymorphic chromosomal inversions between allopatric populations has been recognized (Coluzzi *et al.* 2002).

Bioko Island, Equatorial Guinea, is this country's population, political, and economic center. Bioko Island and mainland Equatorial Guinea are currently the focus of two malaria control programs implemented by Medical Care Development International (MCDI) in partnership with the Government of Equatorial Guinea and

funded by Marathon Oil Corp. (Houston, TX). This consortium coordinates the Bioko Island Malaria Project (BIMCP) and Equatorial Guinea Malaria Control Initiative (EGMCI), which distribute insecticide treated bed nets (ITN) and perform indoor residual spraying (IRS) of pyrethroid insecticides to reduce the rate of malaria transmission and protect against *Plasmodium falciparum* infective mosquito bites. In 2008, these efforts resulted in the protection of 79% of the population through the use of IRS. Additionally, IRS and/or ITNs protected 95% of children under the age of five. In the first four years after its implementation in 2003, the BIMCP reduced mortality in children under the age of five from an average of 197 deaths per 1,000 births during the four years before the control program was implemented, to 61deaths per 1,000 births during the first four years of malaria control (Kleinschmidt *et al.* 2009).

Bioko Island is a candidate for a future malaria eradication campaign and of particular interest to the BIMCP is the level of migration between mainland and Bioko Island vector populations. While Moreno *et al.* (2007) provided an analysis of the population structure of *An. gambiae* in Equatorial Guinea, no information is available on the other main vector on the island, *An. melas*. To provide insight into the migration of *An. melas* to and from the Bioko Island and to contribute to our understanding of the evolution of the *An. gambiae* complex, we investigated the population structure of *An. melas* on Bioko Island and African mainland population throughout its range from The Gambia to Angola.

As part of the operational research component of the Bioko Island Malaria Control Program (BIMCP), we have adapted *An. gambiae* microsatellite loci (Zheng *et al.* 1996) for use in *An. melas* by re-sequencing and evaluating 45 loci, and designing *An. melas*-specific primers for 24 of these loci. We identified polymorphic markers by amplifying each in two *An. melas* populations in Equatorial Guinea: Bomé (BOM), which is located on the mainland, and Cacahual (CAC), which is located on Bioko Island. We subsequently examined 15 polymorphic loci in populations from a wider geographic range: Luba, Equatorial Guinea (LUB), Ipono, Cameroon (IPO), Ballingho, The Gambia (GAM), and Ponta Anabaca, Guinea Bissau (GUI) (Figure 1). To examine

the extent of the null allele problem when using microsatellite loci across species-boundaries in closely related taxa, we also amplified five of these loci in LUB, IPO, GAM, and GUI using the original *An. gambiae* primers. We compare the two data sets and discuss the implications for cross-amplification of microsatellite loci across species boundaries.

We analyzed genetic variation at 15 microsatellite loci and a portion of two mitochondrial DNA genes (ND4 and ND5) in 11 *An. melas* populations. Our study addresses the following questions: Is the patchy distribution of *An. melas* reflected in its population genetic structure? Is there gene flow between Bioko Island and mainland populations? What is the level of divergence between *An. melas* and other members of the *An. gambiae* complex?

CHAPTER II
MATERIALS AND METHODS

Primer Development and Cross-Species Amplification Analysis

Mosquito Collections

Adult female *Anopheles melas* were collected from four locations along the West African Coast. Female mosquitoes from Equatorial Guinea were collected using CDC light traps and human landing catches from Cacahual and Bomé, in October 2008, and Luba in April 2009. Female mosquitoes were collected from Ipono, Cameroon, using human landing catches in December 2005. Resting female mosquitoes were collected using aspirators inside residences in Ballingho, The Gambia in February 2010, and Ponta Abanaca, Guinea Bissau in December 2009.

Molecular Methods

Mosquito DNA was extracted from mosquito abdomens or whole mosquitoes using Qiagen Biosprint 96 DNA extraction (Qiagen Inc., Valencia, CA). Species diagnostics were performed following Scott *et al*. (1993). A total of 48 *An. gambiae* microsatellite markers from Zheng *et al.* (1996) were selected based on chromosomal location, distance from each other, and location relative to chromosomal inversions on right arms of the second ($2Rm^1$, 2Rn, and $2Rn^1$) and third (3Re and 3Rc) chromosomes (Bryan *et al.* 1987). Markers inside polymorphic inversions known to be present in *An. melas* were excluded from this study, as these can create an illusion of population subdivision. Sequences containing the published microsatellite loci were downloaded from the *An. gambiae* genome (Holt *et al.* 2002) using Vectorbase (Lawson *et al.* 2009). Primer3 0.4.0 (Rozen and Skaletsky 2000) was used to design primers located outside the original *An. gambiae* primer annealing sites. PCR was performed on three *An. melas* individuals from Ipono, Cameroon. Reactions contained 10-20 ng DNA template, with 1X PCR buffer (10 mM Tris-HCl pH 8.5, 50 mM KCl), 2.5 mM $MgCl_2$, 200 µM of each dNTP, 2.0 µM of each forward (F) and reverse (R) primer, 0.03 U of Promega GoTaq DNA Polymerase (Promega Co., Madison, WI), and ddH_2O to produce the final 20 µL

reaction volume. PCRs were performed with an initial denaturing time of 2 min at 94˚C followed by five cycles of 30 s at 94˚C, 30 s at 50˚C, 35 s at 72˚C, 30 cycles of 30 s at 94˚C, 30 s at 52˚C, 35 s at 72˚C, followed by a 15 min extension step at 72˚C.

PCR products were ligated in a pGEM-T vector (Promega Co., Madison, WI) and transformed into *E. coli* competent cells. Colonies were grown in LB media overnight and extracted using a Qiagen QIAprep Spin Miniprep Kit (Qiagen Inc., Valencia, CA). Plasmid DNA was used as template in 10μL sequencing reactions, which were performed in forward and reverse directions using BigDye Terminator 3.1 Cycle Sequencing Kit (Life Technologies Corporation, Carlsbad, CA), with a final primer concentration of 250 nM. Sequences will be submitted to Genbank. *An. melas*-specific primers were designed from the above sequences using Primer3 0.4.0 (Rozen & Skaletsky 2000). As part of an initial screen, twenty-four *An. melas* specific loci were amplified using the new primers in 35 individuals from Cacahual and 69 from Bomé, Equatorial Guinea, under the conditions above using a fluorescently labeled forward primer. PCR products were run on a 96-cappillary Applied Biosystems 3730*xl* DNA Genetic Analyzer (Life Technologies Corporation, Carlsbad, CA). A second screen was conducted on four populations from a wider geographic region for 15 loci.

To demonstrate the importance of developing *An. melas* specific microsatellite primers, we amplified five of these 15 microsatellite loci in four *An. melas* populations using the original *An. gambiae* specific primers published by Zheng *et al.* (1996). These five markers were chosen because they vary in the number of mismatches (1-3 bp) between the *An. gambiae* primer and the *An. melas* sequence (Table 1). By testing loci that have a variable number of mis-matches between the *An. gambiae* primer and *An. melas* sequence, we were able to examine how different numbers of mis-matches between primer and annealing site may result in the presence of null alleles. PCRs were performed and analyzed as outlined above.

Data Analyses

Sequence data was aligned using Sequencher 4.9 (GeneCodes, Ann Arbor, USA), and genotypes were assigned using GeneMarker 1.85 (SoftGenetics, LLC, State

College, PA). Observed heterozygosity (H_O), expected heterozygosity (H_E), tests for deviation from Hardy Weinberg Equilibrium (HWE), and population pair-wise F_{ST} values were calculated using Arlequin 3.5.1.2 (Excoffier & Lischer 2010). Micro-Checker 2.2.3 was used to examine the data for the presence of null alleles (van Oosterhout *et al.* 2004), and allele frequencies were corrected for the presence of null alleles using the method of van Oosterhout *et al.* (2004). Null corrected allele frequencies were then used to re-calculate population pairwise-F_{ST} values. MEGA 5 (Tamura *et al.* 2011) was used to calculate the Kimura two-parameter genetic distance (Kimura 1980) between previously published mitochondrial DNA ND4-ND5 (1445 bp) sequences for two *An. gambiae, An. arabiensis,* and *An. melas* individuals (GenBank accession nos. U10123, U10124, U10125, U10127, U10128, U10129) (Besansky *et al.* 1994) (Table 2). By examining the level of mtDNA genetic divergence between these species, we can observe the relationship between cross-microsatellite locus amplification success and genetic distance, and compare these findings to those in other taxa (i.e. Hendrix *et al.* 2010).

Population and Evolutionary Genetics

Mosquito Collections

Adult female *An. melas* mosquitoes from Equatorial Guinea were collected using CDC light traps and human landing catches from Cacahual and Bomé, in October 2008, and Riaba and Luba in April 2009. Female mosquitoes were collected from Ipono, Cameroon, using human landing catches in December 2005. Resting female mosquitoes were collected using aspirators inside residences in Ballingho, The Gambia in February 2010, and Ponta Abanaca, Guinea Bissau in December 2009. Mosquitoes were collected in Mateba, Angola in 2002, and Port Gentil, Gabon in 1999 by aspirating indoor resting females, and from Tiko, Cameroon in October 2010 using human landing catches. Larval mosquitoes were collected in July 2010 in Ada Foah, Ghana from a roadside lagoon. Adult female *An. gambiae* mosquitoes were collected from Ukomba, Bioko Island, Equatorial Guinea in March 2007 using human landing catches and light traps, and from Mongola, Equatorial Guinea in 2009 using human landing catches.

Additional sequences published by Besanky *et al.* (1994) were utilized for comparative mtDNA evolutionary analysis between *An. melas* and other members of the *An. gambiae* complex. Sequences for the ND4 and ND5 genes of two *An. melas*, three *An. gambiae,* two *An. arabiensis,* two *An. merus,* and two *An. quadriannulatus* strains were downloaded from GenBank (Accession numbers U10123-U10133) (Table 2).

Molecular Methods

Microsatellite DNA

Mosquito DNA was extracted from mosquito abdomens or whole mosquitoes using Qiagen Biosprint 96 DNA extraction (Qiagen Inc., Valencia, CA). Species diagnostics were performed following Scott *et al*. (1993). A polymerase chain reaction (PCR) was used to amplify 15 *An. melas* specific polymorphic microsatellite loci (Table 3) in 6-96 individuals from each of the 11 populations included in this study (Table 4). Each microsatellite locus was amplified using a fluorescently labeled forward primer. PCR reactions contained 10-20 ng DNA template, with 1X PCR buffer (10 mM Tris-HCl pH 8.5, 50 mM KCl), 2.5 mM $MgCl_2$, 200 μM of each dNTP, 2.0 μM of each forward (F) and reverse (R) primer, 0.03 U of Promega GoTaq DNA Polymerase (Promega Co., Madison, WI), and ddH_2O to the final 20 μL reaction volume. PCRs were performed with an initial denaturing time of 2 min at 94˚C followed by five cycles of 30 s at 94˚C, 30 s at 50˚C, 35 s at 72˚C, 30 cycles of 30 s at 94˚C, 30 s at 52˚C, 35 s at 72˚C, followed by a 15 min extension step at 72˚C.

Confirmation of the amplification of PCR products was performed on a 1% agarose gel. Varying volumes (0.5 μL - 1.0 μL) of PCR products were combined with 0.5 μL GeneScan 500 ROX size standard (Life Technologies Corporation, Carlsbad, CA). This mixture was dried at 37˚C, and run on a 96-cappillary Applied Biosystems 3730*xl* DNA Genetic Analyzer (Life Technologies Corporation, Carlsbad, CA). Fragment sizes were analyzed using GeneMarker ver. 1.85 (SoftGenetics LLC., State College, PA).

Mitochondrial DNA

Primers were designed to amplify a 1,425 bp region spanning part of the ND4 and ND5 genes of the mitochondrial genome based on the *An. gambiae* mitochondrial DNA (mtDNA) sequence (Beard *et al.* 1993). PCR amplification was performed at a volume of 20 µL at the following reagent concentrations: 1X PCR buffer (10 mM Tris-HCl pH 8.5, 50mM KCl), 2.5 mM $MgCl_2$, 200 µM of each dNTP, 2.0 µM of each primer, and 0.03 U of Promega GoTaq DNA Polymerase (Promega Co., Madison, WI). PCRs were run with an initial 10 min. 94˚C denaturation, followed by 35 cycles of 1 min. at 94˚C, 2 min. at 53˚C, 3 min. at 72˚C, then a 15 min. extension at 72˚C, and a hold step at 4˚C. PCR products were purified using the PEG purification method (Lis 1980), and amplification was confirmed on a 2% agarose gel. Amplified, PEG purified PCR products were used as template in 10µL sequencing reactions, which were performed in forward and reverse directions using BigDye Terminator 3.1 Cycle Sequencing Kit (Life Technologies Corporation, Carlsbad, CA), with a final primer concentration of 250 nM. Forward and reverse sequence data was aligned using Sequencher 4.9 (GeneCodes, Ann Arbor, USA), and had approximately 600 bp of overlap between the forward and reverse sequences. Consensus sequences were trimmed to 1161 bp.

Analytical Methods

Microsatellite DNA Data

Micro-Checker 2.2.3 was used to examine the microsatellite data set for the presence of null alleles (van Oosterhout *et al.* 2004). If null alleles were detected in a population, allele frequencies were corrected using the method of van Oosterhout *et al.* (2004). Observed heterozygosity (H_O), expected heterozygosity (H_E), and other population genetic and diversity parameters were calculated in Arlequin ver. 3.5.1.2 (Excoffier and Lischer 2010) using both null-corrected and non-null corrected allele frequencies. Tests for deviation from Hardy Weinberg Equilibrium (HWE) were run using 1 million steps in the Markov chain, and linkage disequilibrium (LD) was calculated using 10,000 permutations. Population pair-wise F_{ST} values were calculated

using 10,000 permutations. Corrected and uncorrected genotype frequencies were used to calculate population pairwise G_{ST} values (Nei 1973) using 10,000 permutations in the program Gentix ver. 4.05 (Belkhir *et al.* 1996-2004). G_{ST} values were then standardized by the maximum value according to Hedrick (2005) to calculate G_{ST}'. Pairwise G_{ST}' values were used to calculate an un-rooted neighbor-joining tree in the program QuickTree (Howe *et al.* 2002). The tree was visualized in the program FigTree ver. 1.3.1 (Rambaut 2009).

A Bayesian assignment test was implemented in the program STRUCTURE ver. 2.3.3 (Pritchard *et al.* 2000) to estimate the most likely number of populations in our dataset, and to identify potential hybrids and migrants between populations. Only the uncorrected data set can be used to perform this analysis, as it is based upon individual multi-locus genotypes. The most likely number of populations (K) was inferred using the method of Evanno *et al.* (2005) (ΔK), as implemented in the program Structure Harvester ver. 0.6.8 (Earl 2011). The individuals of each predicted cluster (K) were then grouped and an Analysis of Molecular Variance (AMOVA) (Excoffier *et al.* 1992) was performed in Arlequin ver. 3.5.1.2 (Excoffier and Lischer 2010).

A Bayesian approach implemented in the program Migrate-N (Beerli 2006, Beerli & Felsenstein 2001) was used to estimate relative rates of long-term migration between population clusters. An initial run of four long MCMC chains (10,000 steps) was used to estimate mutation scaled migration (*M*) prior values for the final analysis. For each chain, the first 1,000 steps were removed as burn-in, and a static heating scheme was implemented with a swapping interval of one. The final simulation was implemented with defined priors for *M* (as estimated in the first run), and scaled effective population size (Θ) priors were estimated from F_{ST} statistics. Four long MCMC chains (100,000 steps) were employed with a static heating scheme, and the first 50,000 were removed as burn-in.

Mitochondrial DNA Data

Each ND4-ND5 sequence contig was aligned manually in Sequencher 4.9 (GeneCodes, Ann Arbor, USA) based on the translated amino acid sequence (Beard *et*

al. 1993), and trimmed to 1161 bp. Estimates of mean evolutionary diversity were calculated for each sampled *An. melas* and *An. gambiae* population in the program MEGA ver. 5 (Tamura *et al.* 2011) using the method of Nei and Kumar (2000) under a Kimura 2-parameter substitution model (Kimura 1980) with a gamma distribution.

A 95% statistical parsimony approach was employed in the program TCS ver. 1.21 (Clement *et al.* 2000) to create three minimum spanning (haplotype) networks. The first haplotype network includes all sampled *An. melas* haplotypes, the second network includes all sampled *An. gambiae* haplotypes and the third includes the sequences from several additional species in the complex published by Besansky *et al.* (1994) as well as the most common haplotypes from the *An. melas* population clusters identified by the program STRUCTURE, and the most common haplotypes from the *An. gambiae* Mongola and Ukomba populations.

A maximum likelihood approach was also used to investigate evolutionary relationships between the mtDNA of members of the *An. gambiae* complex, and lineages within *An. melas*. We used the program RAxML ver 7.0.0, implemented in raxmlGUI (Silvestro & Michalak 2011), to conduct this analysis. All sampled *An. gambiae* complex individuals listed in Tables 2 and 4 were included in this analysis. The program Modeltest ver. 0.1.1 (Posada 2008, Guindon & Gascuel 2003) was used to determine the most appropriate model of nucleotide evolution (GTR+ Gamma+ Invariant Sites). This nucleotide mutation model was selected based upon the Akaike Information Criterion (Posada and Buckley 2004). This mutation model was implemented in the maximum likelihood analysis in RAxML, which utilized the thorough bootstrap approach with 1,000 bootstrap replicates. The resulting tree with the highest bootstrap support values was visualized using FigTree ver. 1.3.1 (Rambaut 2009).

Next, we used a Bayesian analysis implemented in the program BEAST ver. 1.6.1 (Drummond & Rambaut 2007), to estimate divergence times between lineages of *An. melas*, and between *An. melas* and *An. gambiae*. An initial starting tree was created using RAxML including the most commonly samples haplotypes from each of the *An. melas* population clusters, as well as the Tiko haplotype most similar to Bioko Island

haplotypes, and the two sampled *An. gambiae* populations (Table 4). Additionally, *An. gambiae* and *An. melas* samples from Besansky *et al.* (1994) (Table 2) were included. First, we performed a likelihood ratio test in MEGA 5 to examine if our nucleotide sequences evolved at a clock-like fashion based upon the topology of our starting tree. This hypothesis could not be rejected based upon a 95% confidence interval, therefore we used a strict molecular clock approach with a 2.3% My^{-1} insect mitochondrial DNA nucleotide substitution rate (Gaunt & Miles 2002) to estimate divergence dates. Nodes exceeding 50% bootstrap support in the starting tree were restricted to monophyly during the run, and the time to most recent common ancestor (tmrca) was estimated for each of these nodes. Four independent runs were performed in BEAST, each with a different random seed, and 500 million steps in the Monte Carlo Markov Chain. ND4-ND5 sequences were treated as a single partition and a GTR+ Gamma+ Invariant Sites mutation model was used. Log files of each run were analyzed in Tracer ver 1.4 (Rambaut and Drummond 2007a) to assess convergence. LogCombiner (Rambaut and Drummond 2007b) was used to combine the log and tree file outputs of each run and remove the first 50 million (10%) of each as a burn-in. A maximum clade credibility tree was created in the program TreeAnnotator (Rambaut and Drummond 2007c) by excluding trees that did not have a posterior probability above the mean value. The maximum clade credibility tree was visualized in FigTree ver. 1.3.1 (Rambaut 2009).

CHAPTER III
RESULTS

Primer Development and Cross-Species Amplification Analysis

An. melas Specific Microsatellite Markers

A total of 45 *An. gambiae* microsatellite loci were re-sequenced in 1-3 *An. melas* individuals. Out of 90 *An. gambiae* primers, 48 contained a mismatch with the obtained *An. melas* sequence. The number of mismatches between primer and annealing site ranged from 0-5 (Table 1), with 21 primers containing more than a single mismatch; 14 primers contained 2 mismatches, 2 primers contained 3 mismatches, 3 primers contained 4 mismatches, and 2 primers contained 5 mismatches. In two loci (AGXH100 and AGXH678) no microsatellite repeat was present in *An. melas*. Because the mutation rate of microsatellites is positively correlated with repeat number (Weber 1990), only loci containing more than five uninterrupted repeats in the sequenced individuals were selected for adaptation to *An. melas*. This was done to obtain markers that are likely to contain a suitable amount of variability for a population genetic study. Nine loci were discarded for this reason. Two loci with only 5 continuous repeats were examined in *An. melas* populations from Bome and/or Cacahual, Equatorial Guinea, but both proved to be monomorphic. Locus AGXH810 was discarded because it contained 36 uninterrupted repeats, which would likely have resulted in significant slippage during PCR amplification, making it difficult, if not impossible, to assign alleles to the correct size class. An additional five loci were discarded because single or tri-nucleotide repeats were present in the flanking regions of the microsatellite repeat. Finally, upon comparison of *An. melas* sequences, it was realized that loci AG3H312 and AG3H154 are identical.

Therefore, 18 out of 45 re-sequenced *An. gambiae* loci were not considered suitable as genetic markers, and *An. melas* specific primers were designed for those of the remaining 27, where mismatches were present. Loci for which new primers were designed were renamed AMXH##, AM2H## and AM3H##, depending on whether they

are present on the X, 2nd or 3rd chromosome respectively. The level of polymorphism of the remaining 27 microsatellite loci was initially examined in *An. melas* populations from Cacahual and Bomé, Equatorial Guinea (Table 5).

To that end, *An. melas*-specific forward and reverse primers were designed for 27 loci (Table 5). In all cases where *An. gambiae* primers were used to amplify loci in *An. melas*, no mismatch was present with the *An. melas* sequence. Several primers were redesigned because Primer3 identified more optimal priming sites. Three *An. melas*-specific markers did not amplify in test populations using standard conditions, and were excluded (Table 5). Seven of the examined loci contained a limited amount of polymorphism ($H_o < 0.37$); and two loci deviated significantly from HW, in one or both of the test populations. Therefore, our re-sequencing effort yielded 15 microsatellite markers for use in *An. melas* (Table 3).

To better characterize these 15 loci, they were subsequently amplified in 48 *An. melas* specimens from each of four populations representing a wider geographic region: Luba, Bioko Island, Equatorial Guinea; Ipono, Cameroon; Ballingho, The Gambia; and Ponta Abaca, Guinea Bissau (Table 6). The number of observed alleles per locus in mainland populations ranged from 2 to 22, with an average of 9.2. In Luba, Bioko Island, the number of alleles ranged from 1 to 5, the average being 3.3 (Table 6). A similar picture of lower genetic variability on Bioko Island emerges when comparing levels of polymorphism (H_o). In the three mainland populations H_o ranged from 0.042 to 0.98, the average being 0.69. In six cases mainland H_o was less than 0.5, and five of these observations were in two loci: AM2H603 and AM3H753, which therefore have a more limited variability in the populations examined. In the Bioko Island population, average H_o ranged from 0 to 0.75, the average being 0.42.

Out of 60 tests, six indicated a significant deviation from HW equilibrium before Bonferroni correction. However, there was an excess of homozygotes in only four of these; AM2H793 in Ipono, AM3H93 and AM2H157 in Luba, and AM2H143 in Ballingho. After Bonferroni correction, only locus AM3H93 in Luba showed a significant excess of homozygotes. Microchecker detected the presence of null alleles in

this locus/population combination, but also in locus AM2H143 in Luba and AM3H753 in Ipono, both of which did not deviate significantly from HW equilibrium.

Cross-Species Amplification Using An. Gambiae *Primers*

To examine if the original *An. gambiae* microsatellite markers produce a usable data set when applied to *An. melas* populations, five loci were amplified in 48 individuals in each of four *An. melas* populations using the original *An. gambiae* primers (denoted as AG2H#, AG3H# or AGXH# respectively). Loci with 1-3 mismatches between the *An. gambiae* primers and *An. melas* sequences were chosen (Table 7).

The use of *An. melas*-specific primers resulted in slightly higher levels of heterozygosity than the *An. gambiae*-primers; 65.92% of *An. melas*-primer individuals and 63.70% of *An. gambiae*-primer individuals amplified as heterozygotes. Most of this difference is due to 21 individuals that were homozygous in the AGXH127, but not the AMXH127 data set. Conversely, three individuals were heterozygous in the AGXH127 but not the AMXH127 data set. In the AGXH25 and AGXH38 data sets, four individuals were homozygous that were heterozygous in the respective *An. melas*-primer data sets. In the other two loci, the respective data sets were in complete congruence. This indicates that the use of non-specific primers caused 2.6% (25/960) of individuals to amplify as false homozygotes.

In the *An. gambiae*-primer data set, 2 out of 20 tests showed a significant excess of homozygotes before Bonferroni correction; AG2H157 in Luba, and AG3H127 in Ponta Abaca. Only the latter was significant after Bonferroni correction (Table 7). Microchecker analyses confirmed the presence of null alleles in AG3H127 in Ponta Abaca, but not in AG2H157 in Luba. Given that AM2H157 also showed a slightly significant deviation from HW equilibrium, factors besides null alleles are likely responsible. Microchecker analyses also detected null alleles in AG3H127 in Ballingho, The Gambia, for which the excess of homozygotes was just below significant ($p=0.0577$). Because no excess of homozygotes (or null alleles; Microchecker) were detected for AM3H127 in Ballingho ($p=0.633$) or Guinea Bissau ($p=0.967$), these results

indicate that the development of *An. melas*-specific primers for this locus circumvented the problem of null alleles in these two populations.

The *An. gambiae* loci chosen for cross amplification in *An. melas* populations showed a varying number of mismatches (1-3) between the primers and the obtained *An. melas* sequences (Table 7). Not surprisingly, the *An. gambiae* primers amplifying locus AG2H127, which has a null allele problem in two examined populations, had the largest number of mismatches; 2 and 3 in the forward and reverse primer respectively. Single bp mismatches in the other primers did not appear to lead to the occurrence of null alleles.

When comparing the F_{ST} values derived from the *An. gambiae*-primer and *An. melas*-primer data sets (Table 8), the former showed higher levels of genetic differentiation between Ipono and Ballingho, and Ipono and Ponta Anabaca. In all other comparisons, *An gambiae*-primer data based F_{ST} values underestimated genetic differentiation. F_{ST} values calculated from null-corrected allele frequencies still showed some disparity between the two data sets genotype frequencies (Table 8).

Population and Evolutionary Genetics

Microsatellite DNA Data

A total of 617 *An. melas* wild-caught specimens from 11 populations (Table 3 and Figure 1) were genotyped at 15 microsatellite loci (Table 3). Observed heterozygosity values (H_O) deviated significantly (p-val.<0.05) from the expected heterozygosity (H_E) in 23 out of 165 tests in our genotype data (Table 9) before Bonferroni correction. Homozygote excess was the cause of this deviation in 22 out of 23 instances, eight of which occurred in Bioko Island, E.G. populations. H_O deviated significantly from H_E in five out of 165 tests after Bonferroni correction (p-val.<0.0003). Homozygote excess was the explanation for this deviation in all tests, and four deviations were associated with the presence of null alleles in our dataset. Null alleles were detected in 10 other tests, but were not associated with significant deviation from HWE after Bonferonni correction. These results are similar to other studies in the *An. gambiae* complex where heterozygote deficiencies were observed (Donnelly and Townson 2000, Lehman 2003).

The lowest mean H_O was sampled at locus AM2H603 (17%), while locus AM2H46 had the highest proportion of observed heterozygotes across all populations (75%) (Table 9). The average H_O (across all 15 loci) on Bioko Island is considerably less (40%), compared to mainland populations (average = 66%). Total (mainland and Bioko Island) H_O is 59% (Table 9). The mean number of alleles observed per locus ranged from 2.36 (locus AM2H603) to 12.55 (locus AM3H93). The average number of sampled alleles (across all 15 loci) was also much lower in Bioko Island populations, despite comparable, and sometimes larger, sampled population sizes (*N*) (Table 4). In fact Tiko, Cameroon and Ada Foah, Ghana, which have the two smallest sample sizes (*N*=16 and *N*=6, respectively) have a higher average number of observed alleles than any Bioko Island populations (*N*=35-96). These data show that microsatellite genetic diversity on Bioko Island is much lower than in mainland West Africa populations.

Population pair-wise F_{ST} and G_{ST}' statistics are reported in Table 10. Significant genetic differentiation (F_{ST} with p-val.>0.05) was observed in all but two population pair-wise comparisons. No significant differentiation was found Between Ipono, Cameroon and Bome, Equatorial Guinea (F_{ST}=0.000). These two populations are in very close proximity to each other (approx 40 km). Between Tiko, Cameroon and Ada Foah, Ghana genetic differentiation was similar to between several other populations (F_{ST}=0.0296) and the lack of significance can no doubt be attributed to the extremely small sample size of our sample from Ada Foah, Ghana.

A neighbour-joining tree was constructed using G_{ST}' values to visualize the pattern of genetic differentiation between *An. melas* populations (Figure 2). *An. melas* populations are subdivided into three major clusters. One cluster consists of the Bioko Island populations, whereas the mainland populations group into a West and Central cluster. Despite the close proximity of Tiko and Ipono, in Cameroon, these populations are highly differentiated (F_{ST}=0.1063, G_{ST}'=0.3263). Genetic differentiation between Ipono and Tiko is 3.6 times higher than that between Tiko and Ada Foah, Ghana, which are geographically much more distant. This represents the barrier between the two

populations, with Tiko, Cameroon belonging to the Western group, whereas Ipono, Cameroon belongs to the Central group.

Not surprisingly, genetic differentiation between Bioko Island populations (F_{ST}= 0.08-0.10) is lower than between island and mainland populations. Tiko, Cameroon is geographically the closest mainland population to Bioko Island, and is also the most similar mainland population to the island. The results of a global analysis of molecular variance (AMOVA), with populations assigned to the Bioko Island, Western, and Central population groups, show that within population genetic variation represented the majority of the overall genetic variation observed (73.9%, Table 11), while differences between populations within groups explains 3.8%. On the other hand, difference between the three population clusters explains 22.24% of the genetic variation.

The Bayesian assignment test implemented in STRUCTURE strongly supports the presence of three populations (*K=3*) (Figure 3), corresponding to the Bioko Island, West and Central clusters. Likelihood bar plots for *K*=3 (Figure 4) show that Bioko Island represents a genetically highly distinct population, with very little evidence of migrants from the mainland, with only a single individual having derived a small portion of their genome from a mainland individual. In Tiko, Cameroon, on the other hand, several individuals show evidence of immigration from Bioko Island.

While the results of the Bayesian assignment test helps us to infer hybridization and migration events between population clusters, this method of analysis does not allow us to delineate between recent and historical migration and/or hybridization events. Individuals that have mixed genotypes from two different clusters could possess migrant ancestry, or could be evidence of remnant genetic similarity between two recently divergent populations. To investigate this further, we measured historical mutation scaled effective population size and migration between Bioko Island, West, and Central population clusters using the program MIGRATE-N. The results of this analysis indicate that historical mutation scaled effective population size (Θ) is much lower on Bioko Island (0.51) than either of the mainland populations (Northwest = 1.96 and Central= 1.97) (Figure 5). As expected, historical levels of bi-directional gene flow are higher

between Bioko Island and the West cluster than between Bioko Island and the Central cluster. The two highest historical gene flow parameters (*M*) are between Bioko Island and the West (M_{13}=18.405) and the Central and West population clusters (M_{23}=21.190). These levels correspond with the results from STRUCTURE, which indicate that the West cluster has the highest probability of containing migrants (Figure 4.B), especially Tiko, Cameroon and Ada Foah, Ghana. These individuals may be the result of a recent migration event, or may represent lingering genetic similarities between subdivided populations that have not yet been purged from these populations by genetic drift. In any case, it indicates that migration between Bioko Island and the mainland has been primarily in one direction; from Bioko to the mainland.

Mitochondrial DNA Data

Patterns of mitochondrial DNA (mtDNA) ND4-ND5 genetic diversity within, and between *An. melas* populations follow similar patterns as those observed with microsatellite DNA data. Bioko Island populations have extremely low levels of genetic diversity, as shown by low levels of average genetic divergence between haplotypes within populations (Table 12). Within population genetic divergence within Bioko Island is on average 2.5 times lower than genetic divergence within Mateba, Angola, which has the lowest mainland within population genetic divergence (0.00065). Tiko, Cameroon, had the highest within population genetic divergence (0.00706), which is 28.6 times more divergent that the mean Bioko Island genetic divergence (Table 12).

Only four unique mtDNA haplotypes were sampled on Bioko Island, where a total of 90 individuals were sequenced. Only three base-pair substitutions differentiate the most sampled Bioko Island haplotype from the most closely related Tiko, Cameroon haplotype (Figure 6), indicating that Tiko may represent the source for the Bioko Island populations. There are no shared haplotypes between any of the three major clusters identified using microsatellite markers. Tiko appears to represent an intermediate mtDNA haplotype form between Bioko Island and West populations, and does not share any haplotypes with any of the three population clusters. West and Central haplotypes differ by a minimum of nine base-pair substitutions, and Tiko, Cameroon haplotypes are

divergent from other West populations by a minimum of 13 base-pair substitutions.

This high level of differentiation between *An. melas* Bioko Island and mainland populations contrast strongly with that observed between *An. gambiae* (M-form) populations located on Bioko Island (Mongola), and mainland Equatorial Guinea (Ukomba) (Table 12). Both of these have levels of within population genetic divergence on par with those from mainland *An. melas* populations (Table 12). However, no geographic structure was detected between the *An. gambiae* populations on the mainland and Bioko Island as both populations shared the two most commonly sampled haplotypes, and haplotypes did not cluster by population (Figure 7).

A comparison of haplotypes from multiple members of the *An. gambiae* complex (*An. gambiae, An. arabiensis*, and *An. melas*), demonstrates that the most commonly sampled mtDNA haplotypes from the Central, West (excluding Tiko), and Bioko Island population clusters are as divergent as any one cluster is from *An. gambiae* (Figure 8). Not surprisingly, *An. gambiae* and *An. arabiensis* haplotypes are closely related, as introgression has been demonstrated between these two species.

These same patterns are observed in an unrooted Maximum Likelihood tree (Figure 9). This tree includes all unique haplotypes sampled from *An. melas* and *An. gambiae* populations (Table 4), in addition to previously described ND4-ND5 sequences from *An. melas, An. gambiae, An. arabiensis, An. quadriannulatus,* and *An. merus* lab strains (Besansky *et al.* 1994) (Table 2). The internal nodes of this Maximum Likelihood tree have low bootstrap support. Therefore, we are not able to infer phylogenetic relationships between these species and clusters. *An. melas* lab strains were original derived from The Gambia, and cluster with the West cluster. *An. gambiae* and *An. arabiensis* comprise a non-monophyletic clade. *An. quadriannulatus,* and *An. merus* individuals, respectively, are monophyletic, and have high bootstrap values (82 and 100) supporting their internal nodes. As expected, Tiko, Cameroon haplotypes cluster close to Bioko Island haplotypes. Individuals sampled from the Central cluster populations are monophyletic with a bootstrap support of 90 on their most basal node. West populations (excluding Tiko) are monophyletic with the exception of one individual sampled from

Guinea Bissau (GUIP905), which clusters with the *An. gambiae* & *An. arabiensis* cluster. The most closely related *An. melas* individual to GUIP905 is another that was sampled from Guinea Bissau, although they differ by 11 base-pair substitutions (Figure 6). Most importantly, there is absolutely no support for monophyly of the *An. melas* population groups.

Using a Bayesian approach implemented in BEAST, we estimated divergence times between *An. melas* population clusters, as well as between previously published *An. gambiae* + *An. arabiensis,* and *An. melas* sequences (Figure 10 and Table 2). For this analyses the Tiko haplotype most similar to the Bioko Island haplotypes was included. Bootstrap support was not high enough between BOMP108 (Central) and West samples (GAMP621, BREFET, and BAL) to calculate a divergence date between the Central and Northwest clusters. Median divergence estimates and 95% highest posterior density (HPD) statistics are reported in Table 13. This analysis dates the split of *An. gambiae* and *An. melas* lineages to 423,455 (95% HPD: 261,993 – 621,259) years before present, and the split between Bioko Island populations and Tiko, Cameroon to 54,518 (95% HPD: 10,214 – 120,777) years before present.

CHAPTER IV
DISCUSSION AND CONCLUSIONS

Primer Development and Cross-Species Amplification Analysis

The usefulness of 45 microsatellite loci developed for the malaria mosquito *An.gambiae* were re-sequenced and evaluated based on a standard set of criteria to determine their potential as polymorphic markers for *An. melas*. Out of 45 loci, 17 were not considered promising markers for use in *An. melas.* Of the remaining loci, seven contained very low polymorphism and two deviated significantly from HWE in one or both of two *An. melas* populations in Equatorial Guinea. Additionally, three loci failed to amplify consistently. The result is 15 loci with varying levels of polymorphism in populations from a wider geographic area. The availability of the microsatellite markers reported in this study will facilitate genetic studies of *An. melas*, a locally important malaria vector along the West-African coast.

In all, 24 out of 45 (53.33%) of the *An. gambiae* loci examined in this study are not readily suitable for use as polymorphic markers in its sibling species *An. melas* due to either a very low/high repeat number, low variability, or single or tri-nucleotide repeats in the flanking sequence. This does not include the three loci that failed to amplify, or the two loci that are in HW disequilibrium even when using *An. melas*-specific primers, as additional effort presumably could have led to the development of better working primers. The fact that two loci showed a deviation from HW equilibrium even after designing species-specific primers illustrates that null alleles are a potential problem even within a single species, especially between extremely divergent populations.

Based on the mtDNA ND4 and ND5 genes, the genetic distance between *An. gambiae* and *An. melas* is only approximately 1.5% (Table 14). The high number of microsatellite loci which had to be screened to yield sufficient polymorphic markers for a population genetic study demonstrates the difficulty of using microsatellite loci across

species boundaries beyond the occurrence of null alleles, and the necessity of extensive screening of loci even between closely related species.

Of the loci that were amplified in *An. melas* populations with *An. gambiae* specific primers, four out of five yielded reliable data sets without the presence of null alleles. In all of these cases the number of mismatches between the *An. gambiae* primer and the *An. melas* sequence was small, with single bp mismatches present in one (AGXH25, AGXH38 and AG2H157) or in both primers (AG2H215). In locus AG3H127, two and three mismatches were present in the forward and reverse primers respectively, and null alleles were detected when this locus were amplified in *An. melas* populations from Ballingho, The Gambia, and Ponta Anabaca, Guinea Bissau. The null alleles at this locus resulted in an overestimation of the Fst values between some populations and a underestimation of the Fst values between other populations. It has been reported based on a simulation study that null alleles lead to an overestimation of Fst values (Chapuis & Estoup 2007). Our limited data demonstrate that null alleles can lead to either an over- or underestimation of genetic distances between populations.

We also found that even between these very closely related species, 21 out of a total of 90 primers contained more than a single bp mismatch (2-5 bp). Although we did not examine the amplification success of all of these primers, we expect that the probability null alleles occurring will increase with a higher number of mismatches. A single bp mismatch between primers away from the 3' region did not cause any null allele problems in our data set. Other studies have examined the amplification success of microsatellite markers in species other than the one they were designed for. For example, Carreras-Carbonell *et al.* (2008), Hendrix *et al.* (2010), and Primmer *et al.* (2005) assessed the cross-species amplification efficiency of microsatellites in fish, true salamanders, and birds, respectively, with an emphasis on the correlation between mtDNA divergence and amplification success. All of these authors concluded, not surprisingly, that amplification success decreased with increased mtDNA genetic divergence. Both Primmer *et al.* (2005) and Carreras-Carbonell *et al.* (2008) report that microsatellite polymorphism also decreases with increased genetic distance from the

target species, though only Primmer *et al.* (2005) warn against the potential for null alleles.

However, in their 2004 paper, Chambers *et al.* screened 47 human microsatellite markers for use in *Hylobates lar*, the white-handed gibbon. Of the screened loci, only eight amplified well and were polymorphic when tested in 49 individuals from 12 social groups. Of the other 39 loci, 23 were excluded because they failed to amplify or amplified poorly, and the remaining 16 were mono- or dimorphic. These data also indicate that while a high proportion of loci may amplify across more distantly related species boundaries, this is not a good measure of their usefulness as many of such loci do not have the qualities that make them suitable genetic markers.

A number of *An. gambiae* microsatellite loci (Zheng *et al.* 1996) have been used successfully in various studies of another sibling species in the complex; *An. arabiensis* (E.g. Donnelly *et al.* 1999, Kamau *et al.* 1999, Simard *et al.* 2000, Onyabe *et al.* 2001, Wondji *et al.* 2005), even though Kent *et al.* (2007) found that only 12 out of 20 *An. gambiae* loci amplified well in this species. The level of mtDNA divergence between *An. gambiae* and *An. arabiensis* is less than half of that between *An. gambiae* and *An. melas* (Table 14). In addition, a history of introgression is well documented between the two former species (della Torre *et al.* 1997, Besansky *et al.* 2003, Slotman *et al.* 2005). This may account for some of the success in using *An. gambiae* microsatellite loci in *An. arabiensis,* although numerous *An. gambiae* markers are available and the number screened before the start of a study is not always reported.

Although a common practice, the main hazard of using microsatellite markers across species boundaries is generally considered the occurrence of null alleles in the data set. Our results highlight another substantial problem; the rapid evolution of these markers renders many of them useless even in closely related species due to the absence of the repeat, or due to a low repeat number/lack of polymorpishm. Researchers planning a study based on the cross-amplification of markers across species boundaries are well advised to plan for the inclusion of several times the number of markers needed to produce the desired data set. This is particularly problematic for taxa for which only a

limited number of microsatellite markers are available and researchers should evaluate the need to develop species-specific microsatellite markers for their research subject.

Population and Evolutionary Genetics

While the non-random assortment of paracentric inversions within *An. melas* has been noted (Coluzzi *et al.* 2002), this is the first genetic study on this locally important malaria vector and provides a major contribution to our understanding of this species. Patterns of genetic differentiation and structure between populations of *An. melas,* based upon results from microsatellite and mtDNA genetic markers, are similar. The highest level of microsatellite genetic differentiation was observed between Bioko Island and mainland *An. melas* populations (Table 10). This differentiation is no doubt the result of the geographical barrier to gene flow posed by the Gulf of Guinea, aided possibly by the low genetic diversity in Bioko Island populations. This low genetic diversity, observed both in the microsatellite data and the mtDA, could be the result of population bottlenecks caused by vector control measures on the island implemented by the BIMCP (Sharp *et al.* 2007), founder effects during initial colonization of the island, and/or by genetic drift in the smaller, isolated island populations resulting in the loss of genetic variation.

This high genetic isolation of Bioko Island is also supported by the lack of shared mtDNA haplotypes between Bioko and the mainland, as well as the Bayesian assignment test implemented in the program Structure. This clustering analyses indicated that Bioko Island represents the most genetically distinct population cluster, with the two mainland clusters, although also highly diverged, being somewhat more similar to each other (Figure 4). There is almost no evidence for recent migration to the island in the Bayesian cluster analyses, as observed by the almost complete lack of mainland ancestry found in Bioko Island individuals. Furthermore, our estimates of historical migration between study populations based on the microsatellite data indicate that historical gene flow from Bioko Island to the mainland has been greater than gene flow in the opposite direction (Figure 5). Presumably, the detected historic gene flow to the island actually

represents the original colonization of the island, with the higher historic migration rate from the island to the mainland resulting from occasional, more recent, migration events.

Bioko Island lies on the continental shelf in the Gulf of Guinea, and is hypothesized to have been connected to mainland Cameroon during the last glaciation, becoming isolated only as sea levels rose. Therefore, the fauna of Bioko Island is species-rich and is closely associated with that of coastal Cameroon. Endemism is low due to the island's recent isolation (Jones 1994), which may have occurred as recently as 10,000-11,000 years ago (Eisentraut 1965, Moreau 1966). Based on the recent isolation of Bioko Island from the mainland, combined with our genetic data, we infer that populations of *An. melas* on Bioko Island were connected to, or founded by, the Western mainland cluster, with little or no gene flow to the island following the isolation of Bioko Island. The lower range of the 95% confidence interval of our estimate of the divergence times between Bioko Island haplotypes and the closest Tiko haplotype, matches the hypothesized timing of the isolation of Bioko Island and the mainland. This supports the idea that Bioko Island *An. melas* populations have been largely isolated from the mainland since the isolation of the Island approximately 10,000 years ago.

Both our microsatellite and mtDNA data reveal that Tiko, Cameroon, populations are the least diverged from Bioko Island populations. Tiko is located approximately 50 km away, and is the population that is geographically closest to Bioko Island. The microsatellite data places Tiko, Cameroon populations firmly with the Western cluster of mainland *An. melas* populations, with some evidence of Bioko immigrants being present. The mtDNA haplotypes from Tiko are intermediate between Bioko Island and the Western mainland cluster on the haplotype network. This suggests that Tiko may be the source for Bioko Island *An. melas* populations. However, the lack of shared haplotypes between Tiko and the other Western *An. melas* populations is somewhat surprising.

The West and Central population clusters meet in Cameroon, between Tiko and Ipono. A very high level of genetic differentiation was observed between these two geographically close locations. No mtDNA haplotypes are shared between them and no

less than 10 substitutions separate the two closests sampled haplotypes. However, the Bayesian clustering analyses of the microsatellite DNA does suggest that multiple individuals in Tiko, Cameroon and Ada Foah, Ghana carry some proportion of Central population ancestry, whereas very little geneflow from West to Central population is evident. These results correspond with the results from Migrate, which show a higher level of historical migration from Central populations to West populations, than in the opposite direction. This influx of migrants from Central and Bioko Island populations into Tiko, and Ada Foah, Ghana, also explains the large number of alleles observed in these two populations.

When comparing the most commonly sampled *An. melas* mtDNA haplotypes from each population cluster to *An. gambiae* and *An. arabiensis* haplotypes (Figure 8), no clear ancestral state of *An. melas* haplotypes is evident in relation to these other species. That is, the number of substitutions between *An. melas* haplotypes from different clusters is on par with that between *An. melas* and other *An. gambiae* complex species. These patterns are illustrated in more detail in the maximum likelihood tree, which includes additional species from the complex. The internal nodes of this maximum likelihood tree have very low bootstrap support, but it is clear that *An. melas* clusters are as divergent from each other as they are from any species within the *An. gambiae* complex. Importantly, there is no support for monophyly of *An. melas*, which would be expected if the three population clusters represented a relatively recent isolation event within *An. melas.* This lack of support for monophyly of *An. melas* is consistent both with an independent origin of the mainland population clusters, or with isolation between them occurring very soon after *An. melas* differentiated from the other species in the complex.

Our estimates of divergence dates between *An. melas* clusters indicate that the Tiko + Bioko Island lineage (Figure 10 node E) diverged from other all other populations 342,986 (95% HPD: 204,171 – 512,188) years before present. The 95% HPD estimates of the split between *An. melas* clusters (Figure 10 node F) overlaps broadly with those of the split between *An. melas* and *An. gambiae* lineages (Figure 10

node D), suggesting that divergence times between *An. melas* clusters may be as old as those between two these two recognized species.

An. gambiae, one of the sibling species of *An. melas,* has been shown to consist of two molecular forms that are widely considered incipient species (Favia *et al.* 1997). Low levels of genetic differentiation are present within the S molecular form throughout its range across the African continent (Lehman *et al.* 2003), and are comparable to levels of genetic distance (F_{ST}) within the three population clusters of *An. melas*. A comparison of a microsatellite study of population differentiation within and between the M and S molecular forms of *An. gambiae* in Mali and Cameroon (Slotman *et al.* 2007b) indicates that genetic differentiation (F_{ST}) between West and Central populations is in most cases double (and in some cases 10-fold) that between sympatric M and S forms of *An. gambiae*. In fact, the level of microsatellite differentiation between the *An. melas* clusters is comparably or higher than levels of genetic differentiation between *An. gambiae* and *An. arabiensis* (Slotman *et al.* 2007c).

Therefore, the data presented here suggest that *An. melas* may in fact represent multiple, previously unrecognized species. Following the realization that *An. gambiae* s.l. consists of multiple species (Davidson 1962, Hunt *et al.* 1998) additional layers of incipient speciation and population isolation within *An. gambiae s.s.* were uncovered. The M and S molecular forms are widely considered incipient species and part of their genomes are reproductively isolated (Turner *et al.* 2005, Slotman *et al.* 2006) These M and S forms are widely sympatric and hybrids between them are rare, although recently high levels of hybridization have been observed in Guinea Bissau (Oliveira *et al.* 2008). An additional layer of population differentiation was uncovered by demonstrating that two chromosomal forms within the M molecular form are highly genetically differentiated (Slotman *et al*. 2007c), representing adaptations to different ecological conditions (Forest vs. Sahel Savanna) resulting in geographic isolation. Here we have uncovered additional complexity within the *An. gambiae* complex by showing that *An. melas* populations have a level of genetic structure similar to that found between other species in the complex. Questions on the origin of the *An. melas* mainland populations

remain. The fact that Tiko and Ipono, which belong to different clusters, are geographically very close with no apparent barrier to gene flow between them, suggests a possible independent origin of the two *An. melas* mainland clusters, followed by range expansion until the clusters met in Cameroon. The small amount of migration observed between the two mainland clusters, suggests that pre- or post-zygotic isolation mechanisms are in place between the two mainland clusters. Further study is necessary to provide more conclusive proof of this.

One goal of this study was to determine the population genetic structure of *An. melas*, a locally important malaria vector, to better inform malaria control programs, with a specific emphasis on Bioko Island, Equatorial Guinea. Bioko Island is a candidate for a potential malaria eradication effort, assessing the likelihood that malaria vectors would either re-colonize the island after a vector eradication campaign, or would re-introduce malaria after transmission has been halted, is an important component of such a campaign. Additionally, the potential of insecticide genes to spread through vector populations is of immediate relevance to control projects. Finally, the realization that *An. melas* consists of three genetically distinct and possibly reproductively isolated populations has a bearing on our understanding of the biology of this species. This means that studies on ecological preference, host preference, vector competence, indoor biting behavior, etc, in one population cannot be extrapolated to populations in another cluster.

Our data indicate a very low probability of *An. melas* migrants re-colonizing Bioko Island, or re-introduction malaria parasites after an eradication effort. *An. melas* populations on the island are highly isolated from those on the mainland. This is in stark contrast to the other major vector on the island, *An. gambiae*. A microsatellite study on *An. gambiae* found no evidence of isolation between the Bioko Island and mainland populations (Moreno *et al.* 2007). We confirmed this result by sequencing mtDNA haplotypes from Bioko and mainland *An. gambiae* populations. Haplotypes did not cluster by location, and several of the most common haplotypes were shared between the island and the mainland. This suggests that migration of *An. gambiae* to the island may

be on ongoing phenomenon and that this species is therefore much more likely to reinfest the island after an eradication event. This stark contrast between *An. melas* and *An. gambiae* may indicate that in this case migration is human-mediated, rather than through wind dispersal. It is difficult to imagine how a coastal mosquito such as *An. melas* would have a markedly different wind dispersal pattern than *An. gambiae.* It is conceivable however that the highly anthropophilic *An. gambiae,* which lives and breeds in close association with humans, has a much higher probability of ending up op a plane or boat heading for Bioko Island.

Only by gaining a full understanding of how populations of vector species interact with one another, and with closely related species, can we fully grasp the role insect vectors play in the transmission of human disease. In the case of malaria in sub-Sahara Africa, this problem is confounded by extremely high levels of genetic diversity found within, and between, member species of the *An. gambiae* complex. This study has revealed yet another layer of genetic complexity within one of the member species; *An. melas.* Now, armed with a better understanding of genetic diversity and population structure of this species, we can continue to work toward a thorough understanding of the evolutionary biology of these malaria vectors, and hopefully, use this knowledge to benefit those who suffer from one of the world's most devastating infectious diseases.

REFERENCES

Akogbeto M, Romano R (1999) Infectivité d'*Anopheles melas* vis-à-vis du *Plasmodium falciparum* dans le milieu côtier lagBull. *Soc. Pathol. Exot*, **92**, 3-5.

Barbará T, Palma-Silva C, Paggi GM, Bered F, Fay, MF, Lexer C (2007) Cross-species transfer of nuclear microsatellite markers: potential and limitations. *Molecular Ecology*, **16**, 3759-3767.

Beard CB, Hamm DM, Collins FH, (1993) The mitochondrial genome of the mosquito *Anopheles gambiae*: DNA sequence, genome organization, and comparisons with mitochondrial sequences of other insects. *Insect Molecular Biology*, **2**, 103-124.

Beerli P (2006) Comparison of Bayesian and maximum likelihood inference of population genetic parameters. *Bioinformatics*, **22**, 341-345.

Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences of the USA*, **98**, 4563-4568.

Belkhir K., Borsa P., Chikhi L., N. Raufast, Bonhomme F (1996-2004) GENETIX 4.05, Windows TM software for population genetics. Laboratoire Génome, Populations, Interactions, CNRS UMR 5171, Université de Montpellier II, Montpellier (France).

Besansky NJ, Krzywinski J, Lehmann T, Simard F, Kern M, Mukabayire O, Fontenille D, Touré Y, Sagnon N'F (2003) Semipermeable species boundaries between *Anopheles gambiae* and *Anopheles arabiensis*: evidence from multilocus DNA sequence variation. *Proceedings of the National Academy of Sciences USA*, **100**, 10818-10823.

Besansky NJ, Powell JR, Caccone A, Hamm DM, Scott JA, Collins FH (1994) Molecular phylogeny of the *Anopheles gambiae* complex suggests genetic introgression between principal malaria vectors. *Proceedings of the National Academy of Sciences USA*, **91**, 6885-6888.

Bøgh, C., S. W. Lindsay, S. E. Clarke, A. Dean, M. Jawara, M. Pinder, and C. J. Thomas (2007) High spatial resolution mapping of malaria transmission risk in The Gambia, West Africa, using landsat TM satellite imagery. *The American Journal of Tropical Medicine and Hygiene,* **76**, 875–881.

Brookfield JFY (1996) A simple new method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology*, **5**, 453-455.

Bryan JH (1983) *Anopheles gambiae* and *An. melas* at Brefet, The Gambia, and their role in malaria transmission. *Annals of Tropical Medicine and Parasitology*, **77**, 1-12.

Bryan JH, Di Deco MA, Petrarca V, Coluzzi M (1982) Inversion polymorphism and incipient speciation in *Anopheles gambiae* s.str. in The Gambiae, West Africa. *Genetica* , **59**, 167-176.

Bryan JH, Petrarca V, Di Deco MA, Coluzzi M (1987) Adult behavior of members of the *Anopheles gambiae* complex in the Gambia with special reference to *An. melas* and its chromosomal variants. *Parassitologia*, **29**, 221-249.

Callen DF, Thompson AD, Shen Y, Phillips HA, Richards RI, Mulley JC, Sutherland GR (1993) Incidence and origin of "null" alleles in the (AC)n microsatellite markers. *The American Journal of Human Genetics*, **52**, 922-927.

Carreras-Carbonell J, Macpherson E, Pascual M (2008) Utility of pairwise mtDNA genetic distances for predicting cross-species amplification and polymorphism success in fishes. *Conservation Genetics*, **9**, 181-190.

Chakraborty R, De Andrade M, Daiger SP, Budowle B (1992) Apparent heterozygote deficiencies observed in DNA typing data and their implications in forensic applications. *The Annals of Human Genetics*, **56**, 45-57.

Chambers KE, Reichard UH, Möller A, Nowak K, Vigilant L (2004) Cross-species amplification of human microsatellite markers using noninvasive samples from white-handed gibbons (*Hylobates lar*). *American Journal of Primatology*, **64**, 19-27.

Chandre F, Manguin S, Brengues C, Yovo JD, Darriet F, Diabate A, Carnevale P, Guillet P (1999) Current distribution of a pyrethroid resistance gene (kdr) in *Anopheles gambiae* complex from west Africa and further evidence for reproductive isolation of the Mopti form. *Parassitologia* , **41**, 319-322.

Chapuis MP, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution*, **24**, 621-631.

Clement M, Posada D, Crandall K, (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657-1660.

Coluzzi M, Petrarca V, Di Deco MA (1985) Chromosomal inversion intergradation and incipient speciation in *Anopheles gambiae*. *Bollenttino Di Zoologia,* **52**, 45-63.

Coluzzi M, Sabatini A, della Torre A, Di Deco MA, Petrarca V (2002) A Polytene Chromosome Analysis of the *Anopheles gambiae* Species Complex. *Science*, **298**, 1415-1418.

Coluzzi M, Sabatini A, Petrarca V, Di Deco MA (1979) Chromosomal differentiation and adaptation to human environments in the *Anopheles gambiae* complex. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* **73**, 483-497.

Coetzee M (2004) Distribution of the African malaria vectors of the *Anopheles gambiae* complex. *The American Journal of Tropical Medicine and Hygiene*, **70**, 103-104.

Corander J, Marttinen P, Mäntyniemi S (2006) Bayesian identification of stock mixtures from molecular marker data. *Fishery Bulletin*, **104**, 550–558.

Corander J, Marttinen P, Sirén J, Tang J (2008) Enhanced Bayesian modeling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics*, **9**, 539.

Dakin EE, Avise JC (2004) Microsatellite null alleles in parentage analysis. *Heredity*, **93**, 504-509.

Davidson G (1962) The *Anopheles gambiae* complex. *Nature,* **196**, 907.

Dawson DA, Horsburgh GJ, Küpper C, Stewart IRK, Ball AD, Durrant KL, Hansson B, Bacon I, Bird S, Klein Á, *et al.* (2010) New methods to identify conserved

microsatellite loci and develop primer sets of high cross-species utility – as demonstrated for birds. *Molecular Ecology Resource*s, **10**, 475-494.

della Torre A, Constantini C, Besansky NJ, Caccone A, Petrarca V, Powell JR, Coluzzi M (2002) Speciation within *Anopheles gambiae*- the glass is half full. *Science*, **298**, 115-117.

della Torre A, Fanello C, Akogbeto M, Dossou-Yovo J, Favia G, Petrarca V, Coluzzi M (2001) Molecular evidence of incipient speciation within *Anopheles gambiae s.s.* in West Africa. *Insect Molecular Biology*, **20**, 9-18.

della Torre A, Merzagora L, Powell JR, Coluzzi M (1997) Selective introgression of paracentric inversions between two sibling species of the *Anopheles gambiae* complex. *Genetics*, **146**, 239-244.

Donnelly MJ, Cuamba N, Charlwood JD, Collins FH, Townson H (1999) Population structure in the malaria vector, *Anopheles arabiensis* Patton, in East Africa. *Heredity*, **83**, 408-417.

Donnelly MJ, Townson H (2000) Evidence for extensive genetic differentiation among populations of the malaria vector *Anopheles arabiensis* in eastern Africa. *Insect Molecular Biology*, **9**, 357-367.

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214.

Earl DA (2011) Structure Harvester ver 0.6.8. See http://users.soe.ucsc.edu/~dearl/software/structureHarvester/

Eisentraut M (1965) Rassenbildung bei Säugetieren und Vögeln auf der Insel Fernando Poo. Zoologischer Anzeiger **174**:37-53.

Ekblom R, Galindo J (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, **107**, 1-15.

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

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to

perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564-567.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of Molecuar Variance Inferred from Metric Distances Among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data. *Genetics*, **131**, 479-491.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567-1587.

Favia G, della Torre A, Bagayoko M, Lanfrancotti A, Sagnon N, Touré YT, Coluzzi M (1997) Molecular identification of sympatric chromosomal forms of *Anopheles gambiae* and further evidence of their reproductive isolation. *Insect Molecular Biology*, **6**, 377-383.

Garrett-Jones C, Borehamn PFL, Plant CP (1980) Feeding habits of Anophelines (Diptera:Culicidae) in 1971-78, with reference to the human blood index: a review. *Bulletin of Entomological Research,* **70**, 165-85.

Gaunt MW, Miles MA (2002) An insect molecular clock dates to the origin of the insects and accords with palaeontological and biogeographic landmarks. *Molecular Biology and Evolution,* **19**, 748-761.

Gentile G, Slotman M, Ketmaier V, Powell JR, and Caccone A (2001) Attempts to molecularly distinguish cryptic taxa in Anopheles gambiae s.s. *Insect Molecular Biology*, **10**, 25–32.

Guichoux E, Lagache L, Wagner S, Chaumeil P, Léger P, Lepais O, Lepoittevin C, Malausa T, Revardel E, Salin F, Petit RJ (2011) Current trends in microsatellite genotyping. *Molecular Ecology Resources*, **11**,591-611.

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

Hedrick PW (2005) A standardized genetic differentiation measure. *Evolution,* **59**, 1633-1638.

Hendrix R, Hauswaldt JS, Veiths M, Steinfartz S (2010) Strong correlation between

cross-amplification success and genetic distance across all members of 'True Salamanders' (Amphibia: Salamandridae) revealed by *Salamandra salamandra*-specific microsatellite loci. *Molecular Ecology Resources*, **10**, 1038-1047.

Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, Wincker P, Clark AG, Ribeiro JM, Wides R, *et al.* (2002) The genome sequence of the malaria mosquito *Anopheles gambiae. Science*, **298**, 129-49.

Howe K, Bateman A, Durbin R (2002) QuickTree: building huge neighbour-joining trees of protein sequences. *Bioinformatics,* **18**, 1546-1547.

Hunt RH, Coetzee M, Fettene M (1998) The *Anopheles gambiae* complex: a new species from Ethiopia. *Transactions of The Royal Society of Tropical Medicine and Hygiene*, **92**, 231-235.

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

Jones PJ (1994) Biodiversity in th Gulf of Guinea: an overview. *Biodiversity and Conservation*, **3**, 772-784.

Kamau L, Mukabana WR, Hawley WA, Lehmann T, Irungu LW, Orago AAS, Collins FH (1999) Analysis of genetic variability in *Anopheles arabiensis* and *Anopheles gambiae* using microsatellite loci. *Insect Molecular Biology*, **8**, 287-297.

Kim KS, Min MS, An JH, Lee H (2004) Cross-species amplification of bovidae microsatellites and low diversity of the endangered Korean Goral. *Journal of Heredity*, **95**, 521-525.

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

Kent RJ, Mharakurwa S, Norris DE (2007) Spatial and temporal genetic structure of *Anopheles arabiensis* in Southern Zambia over consecutive wet and drought years. *The American Journal of Tropical Medicine and Hygiene*, **77**, 316-323.

Kleinschmidt I, Schwabe C, Benavente L, Torrez M, Ridl FC, *et al.* (2009) Marked increase in child survival after four years of intensive malaria control. *The American Journal of Tropical Medicine and Hygiene*, **80**, 882–88.

Lawson D, Arensburger P, Atkinson P, Besansky NJ, Bruggner RV, Butler R, Campbell KS, Christophides GK, Christley S, Dialynas E, *et al.* (2009) VectorBase: a data resource for invertebrate vector genomics. *Nucleic Acids Research,* **37**, D583-587.

Lehman T, Licht M, Elissa N, Maega BTA, Chimumbwa JM Watsenga FT, Wondji CS, Simard F, Hawley WA (2003) Population Structure of *Anopheles gambiae* in Africa*. Journal of Heredity*, **94**, 133-147.

Miles LG, Lance SL, Isberg SR, Moran C, Glenn TC (2009) Cross-species amplification of microsatellites in crocodilians: assessment and applications for the future. *Conservation Genetic,* **10**, 935-954.

Moreau RE (1966) *The Bird Faunas of Africa and its Islands*. Academic Press, London.

Moreno M, Salgueiro P, Vicente JL, Cano J, Berzosa PJ, de Lucio A, Simard F, Caccone A, Do Rosario VE, Pinto J, Benito A (2007) Genetic population structure of *Anopheles gambiae* in Equatorial Guinea. *Malaria Journal,* **6**, 137.

Mukabayire O, Caridi J, Wang X, Toure YT, Coluzzi M, and Besansky NJ (2001) Patterns of DNA sequence variation in chromosomally recognized taxa of Anopheles gambiae: evidence from rDNA and single-copy loci. *Insect Molecular Biology,* **10**, 33–46.

Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences, USA,* **70**, 3321-3323.

Nei M, Kumar S (2000) *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.

Oliveira E, Salgueiro P, Palsson K, Vicente JL, Arez AP, Jaenson TG, Caccone A, Pinto J (2008) High levels of hybridization between molecular forms of *Anopheles gambiae* from Guinea Bissau. *Journal of Medical Entomology*,**45**, 1057-1063.

Onyabe DY, Conn JE (2001) Population genetic structure of the malaria mosquito *Anopheles arabiensis* across Nigeria suggests range expansion. *Molecular Ecology,* **10**, 2577-2591.

Paetkau D, Strobeck C (1995) The molecular basis and evolutionary history of a microsatellite null allele in bears. *Molecular Ecology*, **4**, 519-520.

Panova M, Mäkinen T, Fokin M, André C, Johannesson K (2008) Microsatellite cross-species amplification in the genus *Littorina* and detection of null alleles in *Littorina saxatilis*. *Journal of Molluscan Studies*, **74**, 111-117.

Parmakelis A, Moustaka M, Poulakakis N, Louis C, Slotman MA, Marshall JC, Awono-Ambene PH, Antonio-Nkondjio F, Simard F, Caccone A, Powell JR (2010) *Anopheles* immune genes and amino acid sites evolving under the effect of positive selection. *PLoS One,* **5**, e8885.

Parmakelis A, Slotman MA, Marshall J, Awono-Ambene P, Antonio-Nkondjio C, Simard F, Caccone A, Powell JR (2008) The molecular evolution of four anti-malarial immune genes in the *Anopheles gambiae* complex. *BMC Evolutionary Biology*, **8**, 79.

Pinto J, Donnelly MJ, Sousa CA, Gill V, Ferreira C, Elissa N, Do Rosário VE, Charlwood JD (2002) Genetic structure of *Anopheles gambiae* (Diptera: Culicidae) in São Tomé and Príncipe (West Africa): implications for malaria control. *Molecular Ecology*, **11**, 2183-2187.

Pompanon F, Bonin A, Bellemain E, Taberlet P (2005) Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics*, **6**, 847-859.

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

Posada D, Buckley TR (2004) Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology,* **53**, 793-808.

Primmer CR, Painter JN, Koskinen MT, Palo JU, Merilä J (2005) Factors affecting avian cross-species microsatellite amplification. *Journal of Avian Biology*, **36**, 348-360.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945-959.

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

Rambaut A (2009) FigTree: Tree figure drawing tool, ver 1.3.1. http://tree.bio.ed.ac.uk/software/figtree/. Institute of Evolutionary Biology, University of Edinburgh.

Rambaut A, Drummond AJ (2007a) Tracer ver 1.4, See http://beast.bio.ed.ac.uk/Tracer/. Institute of Evolutionary Biology, University of Edinburgh.

Rambaut A, Drummond AJ (2007b) LogCombiner ver 1.4, See http://beast.bio.ed.ac.uk/LogCombiner/. Institute of Evolutionary Biology, University of Edinburgh.

Rambaut A, Drummond AJ (2007c) TreeAnnotator ver1.4, See http://beast.bio.ed.ac.uk/TreeAnnotator/. Institute of Evolutionary Biology, University of Edinburgh.

Reimer LJ, Tripet F, Slotman MA, Spielman A, Fondjo E, Lanzaro GC (2005) An unusual distribution of the *kdr* gene among populations of *Anopheles gambiae* on the island of Bioko, Equatorial Guinea. *Insect Molecular Biology,* **14**, 683-688.

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (eds Krawetz S, Misener S), pp. 365-386. Humana Press, Totowa, NJ.

Scott JA, Brogdon WG, Collins FH (1993) Identification of single specimens of the *Anopheles gambiae* complex by the polymerase chain reaction. *The American Journal of Tropical Medicine and Hygiene*, **49,** 520-529.

Sharp BL, Ridl FC, Govender D, Kuklinski J, Kleinschmidt I (2007) Malaria vector control by indoor redidual insecticide spraying on the tropical island of Bioko, Equatorial Guinea. *Malaria Journal*, **6**, 52.

Silvestro D, Michalak I (2011) raxmlGUI: a graphical front-end for RAxML. *Organisms Diversity and Evolution*. DOI: 10.1007/s13127-011-0056-0

Simard F, Lehmann T, Lemasson JJ, Diatta M, Fontenille D (2000) Persistence of *Anopheles arabiensis* during the severe dry season conditions in Senegal: an indirect approach using microsatellite loci. *Insect Molecular Biology*, **9**, 467-479.

Sinka ME, Bangs MJ, Manguin S, Coetzee M, Mbogo CM, Hemingway J, Patil AP, Temperley WH, Gething PW, Kabaria CW, *et al.* (2010) The dominant *Anopheles* vectors of human malaria in Africa, Europe, and the Middle East: occurrence data, distribution maps, and bionomic précis. *Parasites and Vectors*, **3**, 117.

Slotman MA, della Torre A, Calzetta M, Powell JR (2005) Differential introgression of chromosomal regions between *Anopheles gambiae* and *Anopheles arabiensis*. *The American Journal of Tropical Medicine and Hygiene,* **73**, 326-335.

Slotman MA, della Torre A, Powell JR (2004) The genetics of inviability and male sterility in hybrids between *Anopheles gambiae* and *Anopheles arabiensis.* *Genetics*, **167**, 275-287.

Slotman MA, della Torre A, Powell JR (2007a) Female sterility in hybrids between *Anopheles gambiae* and *A. arabiensis*, and the causes of Haldane's rule. *Evolution*, **59**, 1016-1026.

Slotman MA, Mendez MM, della Torre A, Dolo G, Toure YT, Caccone A (2006) Genetic differentiation between the BAMAKO and SAVANNA chromosomal forms of *Anopheles gambiae* as indicated by amplified fragment length polymorphism analysis. *The American Journal of Tropical Medicine and Hygiene,* **74**, 641-648.

Slotman MA, Parmakelis A, Marshall JC, Awono-Ambene PH, Antonio-Nkondjio C, Simard F, Caconne A, Powell JR (2007b) Patterns of selection in anti-malarial

immune genes in malaria vectors: evidence for adaptive evolution in LRIM1 in *Anopheles arabiensis*. *PLoS One,* **2**, e793.

Slotman MA, Tripet F, Cornel AJ, Meneses CR, Lee Y, Reimer LJ, Thiemann TC, Fondjo E, Fofana A, Traore SF, Lanzaro GC (2007c) Evidence for subdivision within the M molecular form of *Anopheles gambiae*. *Molecular Ecology*, **16**, 639-649.

Stinchcombe JR, Hoekstra HE (2008) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity,* **100**, 158-179.

Tabachnick WJ (2003) Reflections on the *Anopheles gambiae* genome sequence, transgenic mosquitoes and the prospect for controlling malaria and other vector borne diseases. *Journal of Medical Entomology*, **40**, 597-606.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731-2739.

Thomas MA, Klaper R (2004) Genomics for the ecological toolbox. *TRENDS in Ecology and Evolution*, **19**, 439-445.

Tripet F, Dolo G, Lanzaro GC (2005) Multilevel analyses of genetic differentiation in *Anopheles gambiae s.s.* reveal patterns of gene flow important for malaria-fighting mosquito projects. *Genetics*, **169**, 313-324.

Tripet F, Touré YT, Taylor CE, Norris DE, Dolod G, Lanzaro GC (2001) DNA analysis of transferred sperm reveals significant levels of gene flow between molecular forms of *Anopheles gambiae*. *Molecular Ecology,* **10**, 1725-1731.

Tuno N, Kjaerandsen J, Badu K, Kruppa T (2010) Blood-feeding behavior of *Anopheles gambiae* and *Anophelesmelas* in Ghana, western Africa. *Journal of Medical Entomology*, **47**, 28-31.

Turner TL, Hahn MW, Nuzhdin SV (2005) Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology*, **3**, e285.

van Oosterhout C, Hutchinson WF, Wills DPM , Shipley P (2004) Michro-Checker: software for identifying and correcting genotyping error is microsatellite data. *Molecular Ecology Notes*, **4**, 535-538.

Weber, JL (1990) Informativeness of human $(dC\text{-}dT)_n.(dG\text{-}dA)_n$ polymorphisms. *Genomics*, **7**, 524-530.

Weill M, Chandre F, Brengues C, Manguin S, Akogbeto M, Pasteur N, Guillet P, Raymond M (2000) The kdr mutation occurs in the Mopti form of *Anopheles gambiae s.s.* through introgression. *Insect Molecular Biology*, **9**, 451-455.

White GB (1974) The *Anopheles gambiae* complex and disease transmission in Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* **68**, 278-298.

White GB, Tessafaye F, Boreham PFL, Lemma G (1980) Malaria vector capacity of *Anopheles arabiensis* and *An. quadriannulatus* in Ethiopia: chromosomal interpretation after 6 years storage of field preparations. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* **74**, 683–684.

Wilson ACC, Massonnet B, Simon JC, Prunier-Leterme N, Dolatti L, Llewellyn KS, Figueroa CC, Ramirez CC, Blackman RL, Estoup A, Sunnucks P (2004) Cross-species amplification of microsatellite loci in aphids: assessment and application. *Molecular Ecology Notes*, **4**, 104-109.

Wondji C, Simard F, Fontenille D (2002) Evidence for genetic differentiation between the molecular forms M and S within the Forest chromosomal form of *Anopheles gambiae* in an area of sympatry. *Insect Molecular Biology*, **11**, 11-19.

Wondji C, Simard F, Lehmann T, Fondjo E, Samè-Ekobo A, Fontenille D (2005) Impact of insecticide-treated bed nets implementation on the genetic structure of *Anopheles arabiensis* in an area of irrigated rice fields in the Sahelian region of Cameroon. *Molecular Ecology*, **14**, 3683-3693.

Zheng L, Benedict MQ, Cornel AJ, Collins FH, Kafatos FC (1996) An integrated genetic map of the African human malaria vector mosquito, *Anopheles gambiae*. *Genetics*, **143**, 941-952.

APPENDIX A

FIGURES

Figure 1 *An. melas* sample locations throughout West Africa, including Bioko Island, Equatorial Guinea and neighboring Cameroon and mainland Equatorial Guinea (inset).

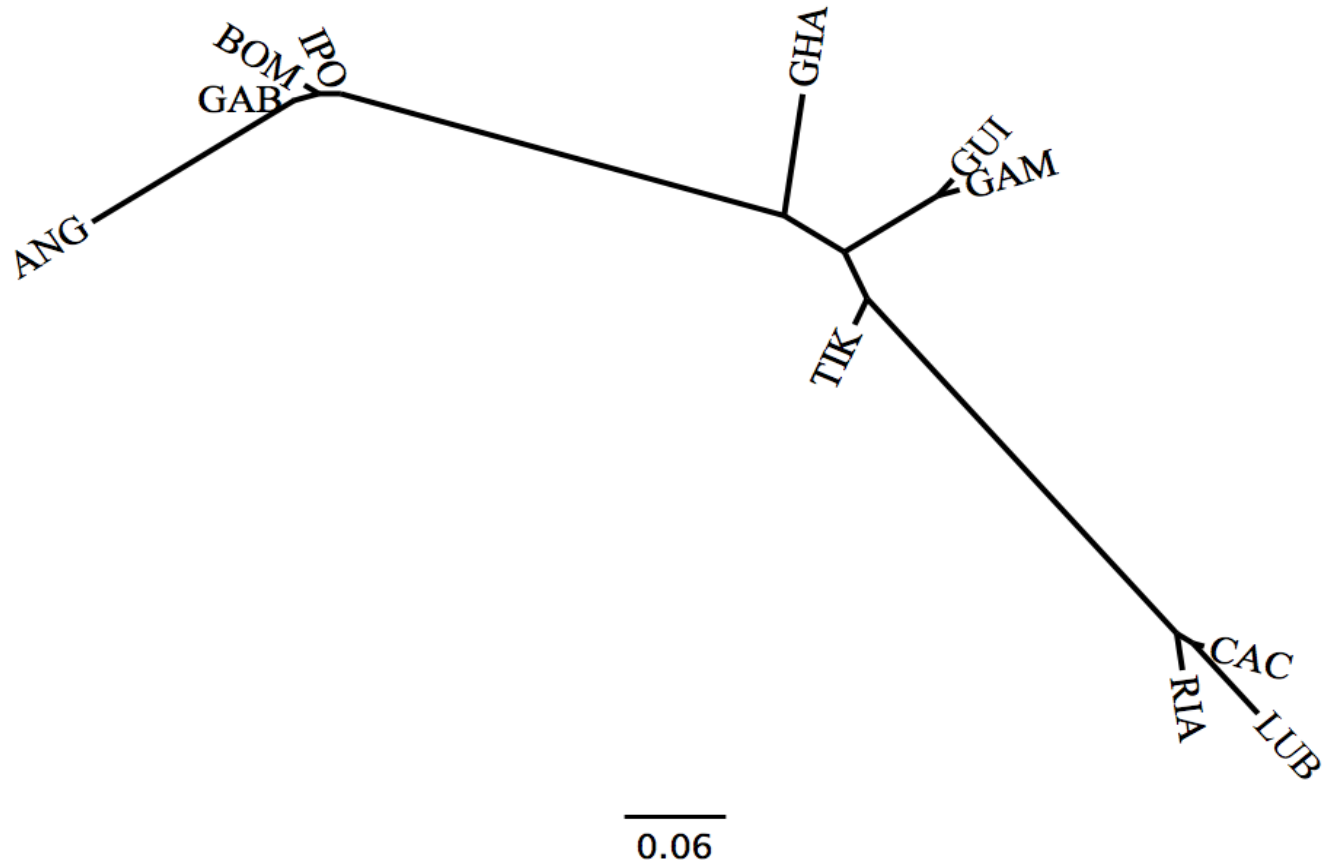

Figure 2 Dendrogram of populations of *An. melas* based upon pair-wise G_{ST}' values, constructed using a neighbor-joining cluster analysis. Population abbreviations correspond with those defined in Table 4.

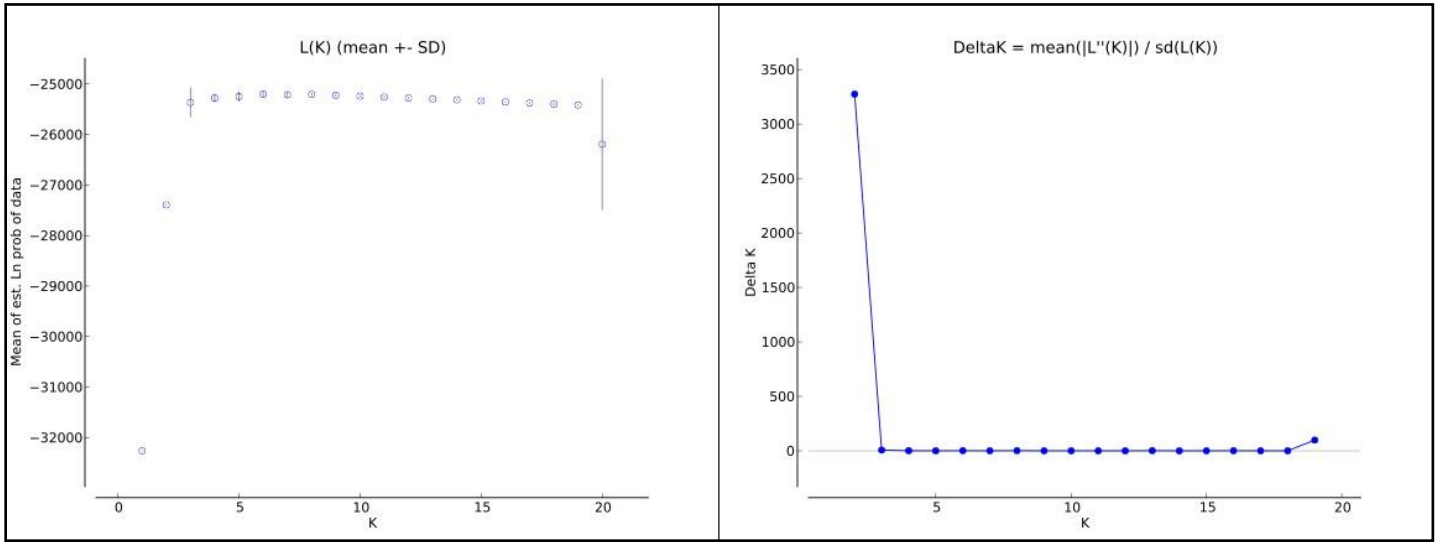

Figure 3 Results of Structure Harvester (Earl 2011), indicating likelihood scores of *K* populations (left) and Delta*K* values (right).

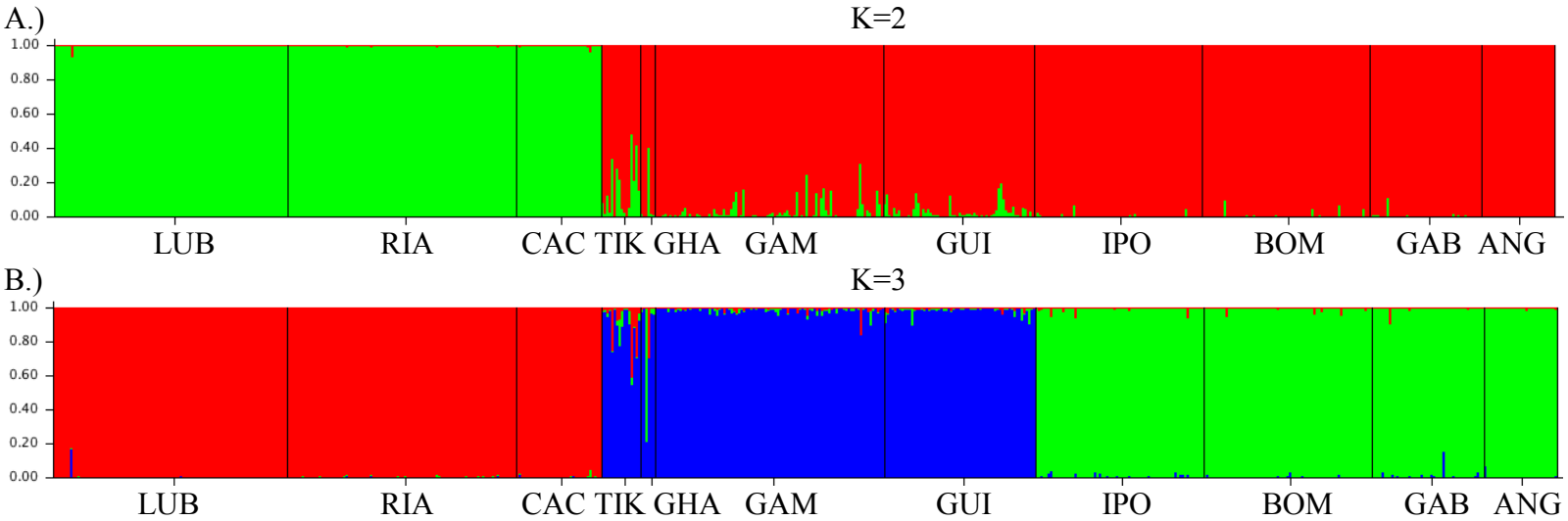

Figure 4 Results of the Bayesian assignment test for A.) *K*=2 and B.) *K*=3 based upon microsatellite DNA data implemented in the program Structure (Pritchard *et al.* 2000). Each vertical bar corresponds to a single individual, and colors represent the proportion of the genome that is assigned to a particular cluster based upon the admixture model. Sample populations are annotated according to population abbreviations defined in Table 4.

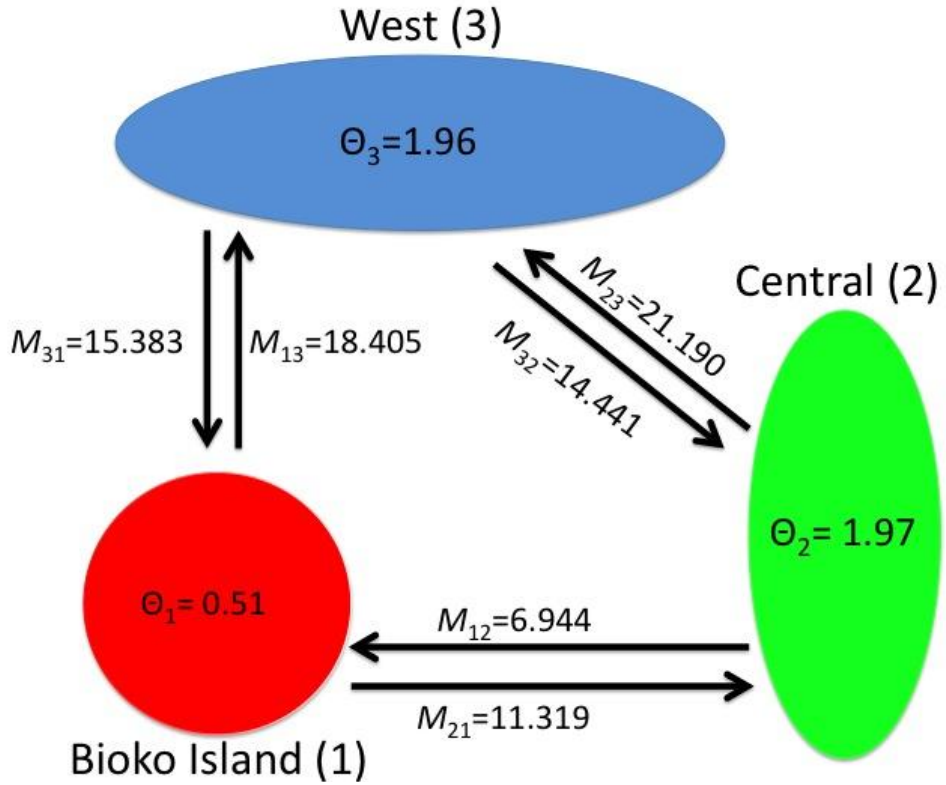

Figure 5 Estimations of historical migration between population clusters and effective population size within population clusters resulting from analysis performed in the program MIGRATE-N (Beerli 2006, Beerli and Felsenstein 2001). *M*= uni-directional mutation scaled migration. Θ= mutation scaled effective population size. Ovals represent the three *An. melas* population clusters, and are annotated as such. Directional *M* is annotated according to the direction of gene flow.

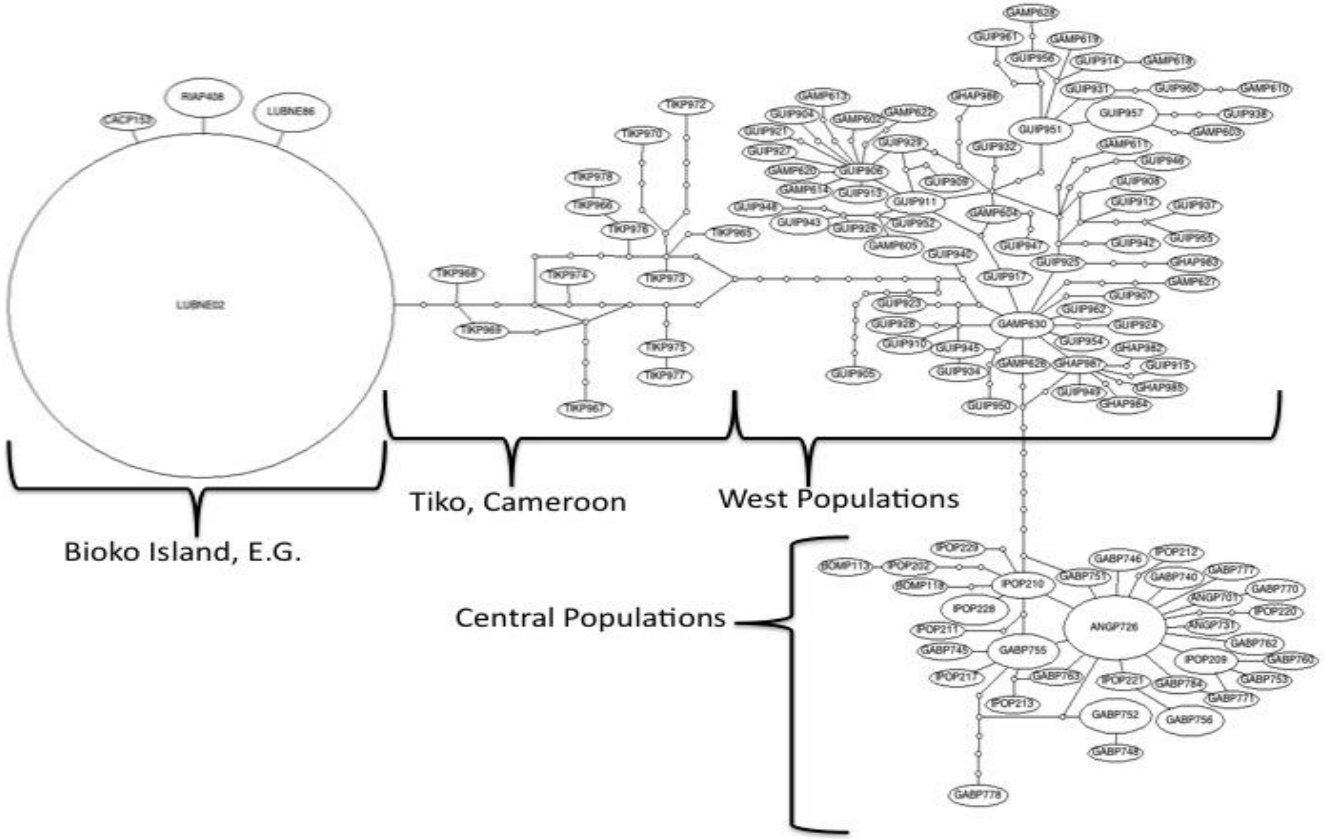

Figure 6 *An. melas* haplotype network constructed through statistical parsimony in the program TCS (Clement *et al.* 2000). Large ovals represent sampled haplotypes. Small intermediate circles represent ancestral or unsampled haplotypes. Population abbreviations refer to those defined in Table 4. Population and clusters are annotated according to *An. melas* population cluster names.

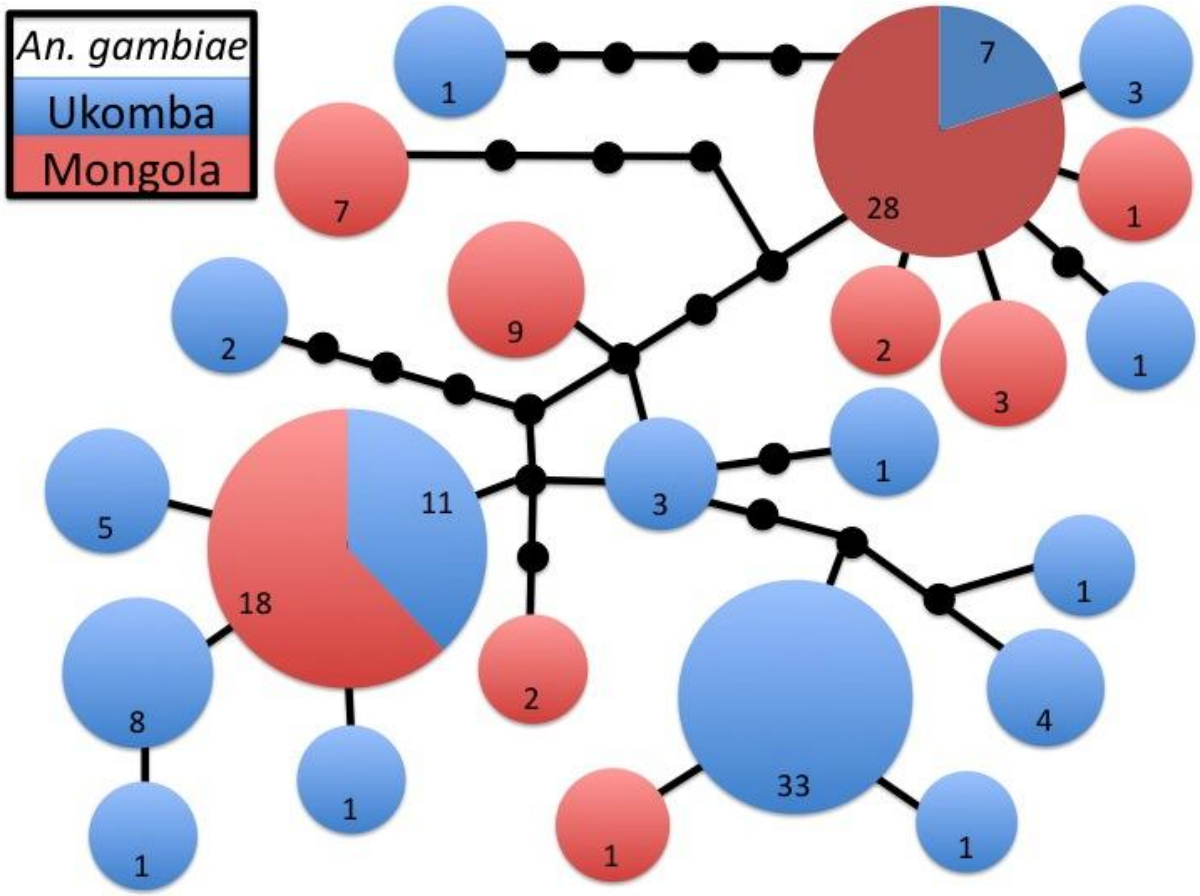

Figure 7 *An. gambiae* haplotype network constructed through statistical parsimony in the program TCS (Clement *et al.* 2000). Large colored circles represent sampled haplotypes. Small black intermediate circles represent ancestral or unsampled haplotypes. Haplotypes sampled from Mongola (Bioko Island, E.G.) are red, and haplotypes sampled from Ukomba (mainland E.G.) are blue. Haplotypes shared between both populations are shown as a pie chart. Numbers in each circle represent the number of times that haplotpye was sampled in the respective population.

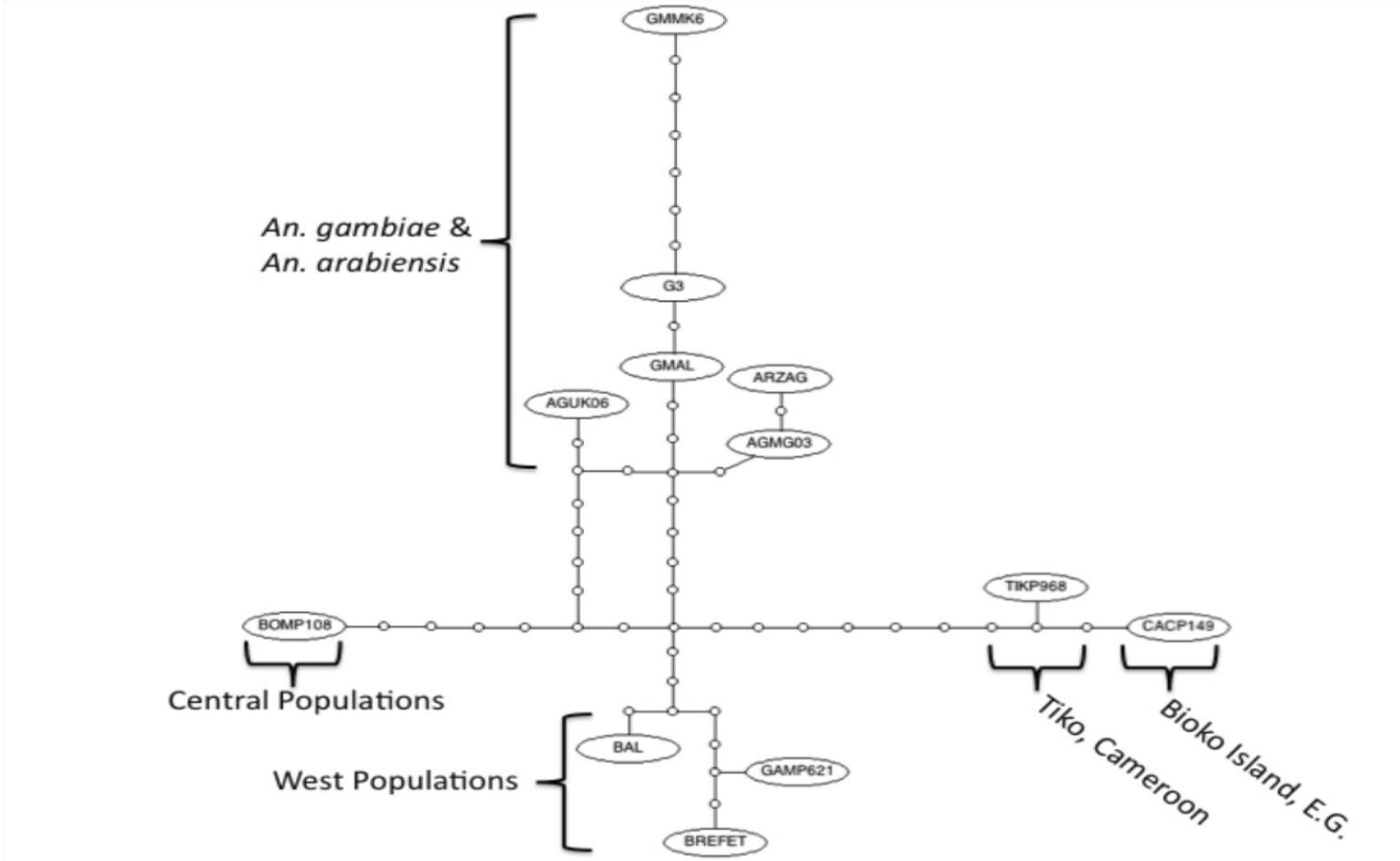

Figure 8 *An. gambiae* complex haplotype network constructed through statistical parsimony in the program TCS (Clement *et al.* 2000). Large ovals represent sampled haplotypes. Small intermediate circles represent ancestral or unsampled haplotypes. Population and strain abbreviations refer to those defined in Table 2 and Table 4. Population and species clusters are annotated according to *An. melas* population cluster or species (if different from *An. melas*).

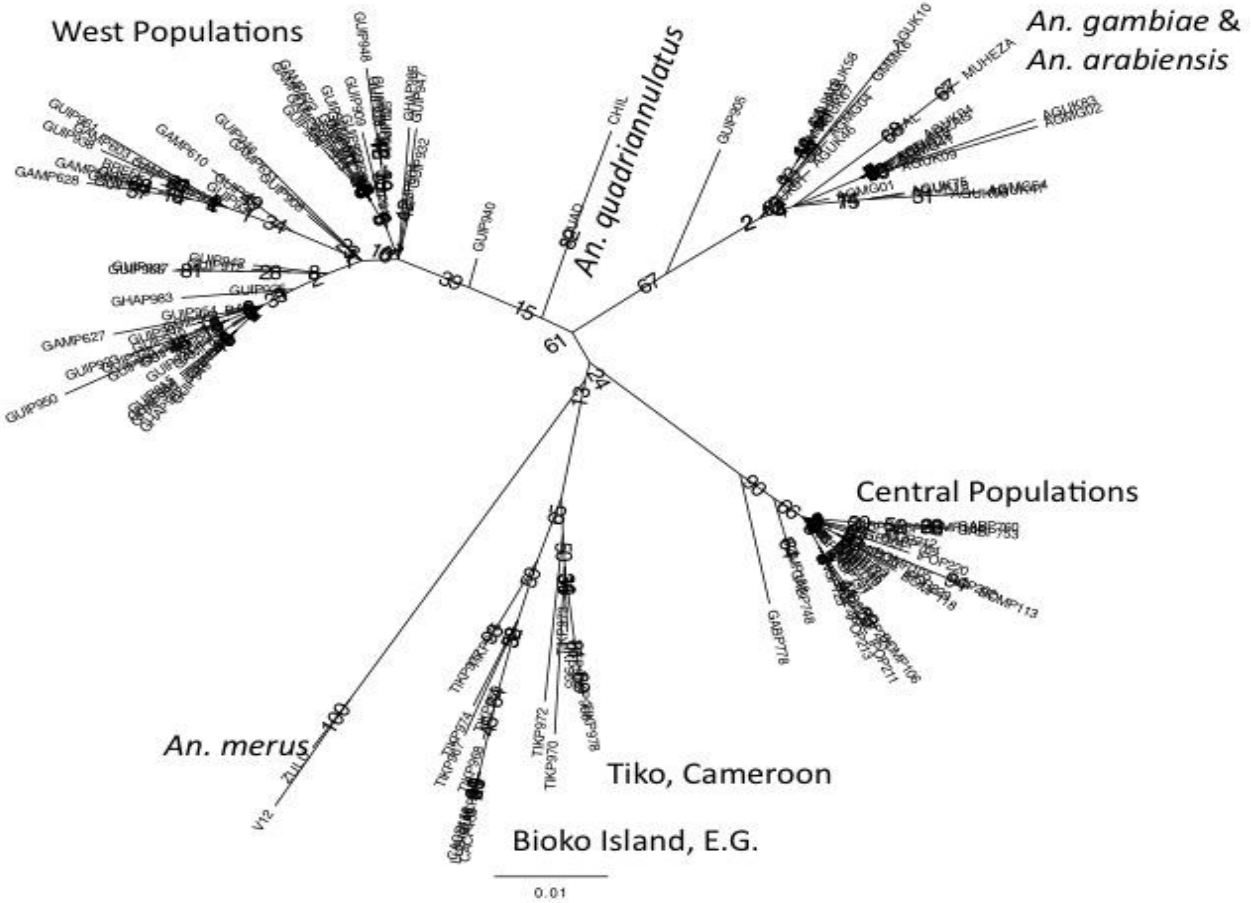

Figure 9 Unrooted maximum likelihood tree of *An. gambiae* complex member species including Central, West, and Bioko Island *An. melas* populations. Node annotations indicate bootstrap support resulting from 1000 replicates. Population and species clusters are annotated according to *An. melas* population cluster or species (if different from *An. melas*).

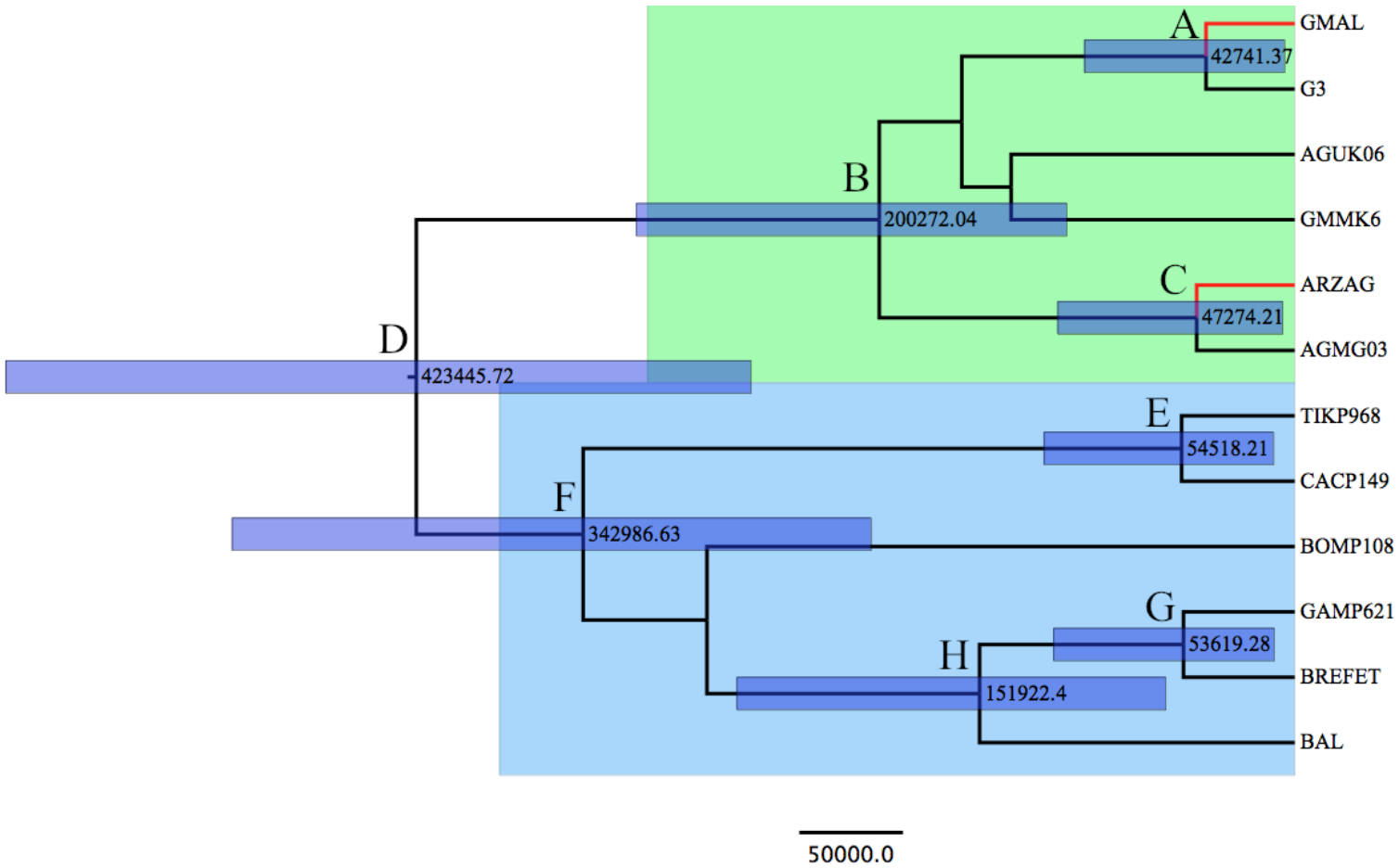

Figure 10 Unrooted phylogenetic tree of *An. gambiae* Complex member species including *An. melas* populations. Node annotations indicate bootstrap support resulting from 1000 replicates. Population and species clusters are annotated according to *An. melas* population cluster or species (if different from *An. melas*).

APPENDIX B

TABLES

Table 1 *An. gambiae* microsatellite loci sequenced in *An. melas.* NMF, Number mis-matches between Zheng *et al.* (1996) *An. gambiae* forward primer and *An. melas* sequence. NMR, Number mis-matches between Zheng *et al.* (1996) *An. gambiae* reverse primer and *An. melas* sequence. INC indicates that locus is included in final *An. melas* set. Reasons for exclusion include: BR, Bad repeat; HWD, Hardy-Weinberg Disequilibrium; FR, Repeats in flanking region; HR#, High repeat number; NA, Did not amplify in test populations; NR, No repeat at locus; and LV, Low variability.

Locus	NMF	NMR	*An. melas* Repeat Motif	Included / Reason Excluded
AG2H117	1	0	(GT)2+(GT)5+(GT)4	BR
AG2H175	1	2	(CA)2+(CA)4+(CA)3	BR
AG2H325	0	0	(GT)4+(CA)3	BR
AG3H059	0	0	(TG)5	BR
AG3H750	2	1	(GT)5	BR
AG3H811	1	1	(GT)5+(AT)2	BR
AGXH053	2	0	(TG)2+(TG)4+(TG)2+(TG)5+(TG)3	BR
AGXH099	0	1	(GT)3+(GT)4+(GT)2	BR
AGXH459	4	0	(GT)3+(GT)2+(GT)4	BR
AG2H147	2	2	(GT)11	FR
AG2H164	2	1	(TG)4+(TG)5	FR
AG3H746	1	0	(GT)15	FR
AG3H758	0	1	(TG)3+(TG)4+(GT)10	FR
AG3H817	1	0	(GT)2+(GT)7	FR
AGXH810	1	1	(GT)36+(GT)7+(GT)6+(GA)14	HR#
AG2H161	0	0	(TG)6	HWD
AGXH106	3	1	(GT)25	HWD
AG2H046	1	0	(GT)13	INC
AG2H143	0	0	(CT)6	INC
AG2H157	0	1	(TG)12	INC
AG2H215	1	1	(GT)10	INC
AG2H603	0	0	(GT)6	INC
AG2H793	4	1	(TG)10	INC

Table 1 (continued)

Locus	NMF	NMR	*An. melas* Repeat Motif	Included / Reason Excluded
AG3H093	0	0	(GT)26	INC
AG3H127	2	3	(GT)10	INC
AG3H555	5	0	(GT)13	INC
AG3H753	0	0	(AC)6+(AC)5	INC
AGXH025	0	1	(GT)3+(GT)5	INC
AGXH038	0	1	(GT)26	INC
AGXH293	0	4	(GA)7	INC
AGXH755	0	5	(GT)8	INC
AGXH808	0	0	(GT)15	INC
AG3H312	0	0	(AC)6	LV
AG2H417	0	2	(GT)18	LV
AG2H675	0	0	(AC)5+(CA)4+(CA)3	LV
AG3H088	2	1	(TG)2+(TG)6	LV
AG3H154	1	0	(AC)6	LV
AG3H249	0	1	(TG)6	LV
AG3H544	2	2	(GT)5+(GT)3	LV
AGXH019	1	2	(GT)2+(GT)7	LV
AG3H242	1	0	(GT)2+(GT)18	NA
AG3H341	0	2	(TG)10	NA
AGXH289	0	0	(GA)10+(TA)4	NA
AGXH100	2	1	N/A	NR
AGXH678	0	1	N/A	NR

Table 2 *An. gambiae* Complex ND4-ND5 mtDNA sequences that represent 5 different species from the *An. gambiae* complex and were originally published by Besansky *et al.* (1994).

Species	Strain	GenBank Accession No.	Geographical Origin
An. melas	BAL	U10123	The Gambia
An. arabiensis	GMAL	U10124	Sudan
An. arabiensis	ARZAG	U10125	Burkina Faso
An. gambiae	G3	U10126	The Gambia
An. gambiae	GMMK6	U10127	Burkina Faso
An. gambiae	MUHEZA	U10128	Tanzania
An. melas	BREFET	U10129	The Gambia
An. merus	V12	U10130	Kenya
An. merus	ZULU	U10131	Zululand
An. quadriannulatus	CHIL	U10132	Zimbabwe
An. quadriannulatus	SQUAD	U10133	Unknown

Table 3 *An. melas* specific primer information. Tm, *An. melas* specific primer melting temperature. As, allele Size of original *An. melas* clone. R,: allele Range in test populations. Mean Na, mean allelic richness, calculated across four test populations. Mean Ho, mean observed heterozygosity, calculated across four test populations. Location, location on *An. gambiae* chromosome.

Locus Name	*An. melas* specific primer (5' -> 3')	*An. melas* Repeat Motif	Tm (˚C)	As	RR (bp)	Mean Na	Mean Ho	Location
AMXH25	F: AAAAGGGAAGCCGAAAACAT	(GT)3+(GT)5	47.7	143	134-148	4.25	0.4688	X: 2C
	R: CAGTTATGCCGGCATGCTAC		53.8					
AMXH38	F: TCCAGTGACTACGCTTCTCG	(GT)26	53.8	236	180-244	11.25	0.7857	3R: 32A
	R: TCAGCGCTATCACAGCAAAC		51.8					
AM2H46	F: GCGCCCATAGACAAAGAGAG	(GT)13	53.8	124	104-136	10	0.7458	2R: 7A
	R: GAGGGTGCAGAACATTTACCA		52.4					
AM3H93	F: GTTGTCTCGCGCCGTCTA	(GT)26	52.6	306	276-332	15	0.7176	3R: 29A
	R: TCGTCATCGTACATCAATACCC		53.0					
AM3H127	F: CGCCGGTACGTCATTAAACT	(GT)10	51.8	129	119-131	5.25	0.6495	3L: 41C
	R: CTGGGAGTTCAGGGAATTGA		51.8					
AM2H143	F: TCTACGCACAAGGTCGTTTC	(CT)6	51.8	296	289-303	7.25	0.6802	2L: 25D
	R: CGCACGTCTCTGTTATGCTC		53.8					
AM2H157	F: TTAAAGTGTGCACGGGAAATC	(TG)12	50.5	173	158-186	7.25	0.6326	2R: 9A
	R: AGTTCGCGCAACTAGAAACG		51.8					
AM2H215	F: GGAACTGATTGTGGTGATCAAA	(GT)10	51.1	125	112-132	7.25	0.7712	2L: 24A
	R: ACGGTTTGGTCTGCAAGTGT		51.8					
AMXH293	F: ACATCTTTCAGCACCACTGG	(GA)7	51.8	145	138-170	7.25	0.6950	X: 4C
	R: GGTGCCACATTGTGTTACTGA		52.4					
AM3H555	F: GTGGAGCAGCTGACCTCATT	(GT)13	53.8	155	140-232	10.75	0.6946	3R: 32C
	R: TTGCCGTCTGATATGAATGC		49.7					
AM2H603	F: TGCACCGTTGATGCACATGC	(GT)6	53.8	111	107-111	2.5	0.1789	2L: 26D
	R: GTTGGTTGTGGACGATGTGA		51.8					
AM3H753	F: GGCAAAACAGGATGGTCGT	(AC)6+(AC)5	51.1	112	106-114	2.75	0.1952	3L: 44C
	R: CAGGCCAATGAGGTATCGAG		53.8					
AMXH755	F: CAGCAGCAGCTGAACGATATT	(GT)8	52.4	156	149-159	4.5	0.6605	X: 4A
	R: AGGCAGCGGGTTAAAAAGAT		49.7					
AM2H793	F: TTACGACGGAATGCAATGTT	(TG)10	47.7	198	191-249	12.75	0.7119	2R: 8B
	R: GTAATCGGCTCGTTTTCTGC		51.8					
AMXH808	F: CAGTGTGACCGAAGCTGTTG	(GT)15	53.8	177	152-178	7.5	0.7226	X: 3D
	R: AAACGGGTGGACACGATAAG		51.8					

Table 4 *An. melas* and *An. gambiae* population individual specimen collection location information. Microsatellite *N* defines the number of individuals that was genotyped for each of 15 microsatellite loci from each respective population. mtDNA *N* defines the number of individuals that was sequenced for the ND4-ND5 mtDNA locus.

Species	Geographical Origin	Abbreviation	Microsatellite *N*	mtDNA *N*
An. melas	Luba, Equatorial Guinea (Bioko)	LUB	96	61
An. melas	Riaba, Equatorial Guinea (Bioko)	RIA	94	16
An. melas	Cacaual, Equatorial Guinea (Bioko)	CAC	35	13
An. melas	Tiko, Cameroon	TIK	16	14
An. melas	Ada Foah, Ghana	GHA	6	6
An. melas	Ballingho, The Gambia	GAM	94	19
An. melas	Ponta Anabaca, Guinea Bissau	GUI	62	52
An. melas	Ipono, Cameroon	IPO	69	19
An. melas	Bome, Equatorial Guinea	BOM	69	11
An. melas	Port Gentil, Gabon	GAB	46	27
An. melas	Mateba, Angola	ANG	30	10
An. gambiae	Mongola, Equatorial Guinea (Bioko)	AGMG	N/A	71
An. gambiae	Ukomba, Equatorial Guinea	AGUK	N/A	83

Table 5 Population diversity estimates for *An. melas* specific loci in two initial test populations, Cacahual (CAC) and Bome (BOM). NMF, Number mis-matches between Zheng *et al.* (1996) *An. gambiae* forward primer and *An. melas* sequence. NMR, Number mis-matches between Zheng *et al.* (1996) *An. gambiae* reverse primer and *An. melas* sequence. N_a, number of observed alleles. H_O, observed heterozygosity. H_E, Expected heterozygosity. HWE p-val., Hardy-Weinberg Equilibrium test p-value. INC indicates that locus is included in final *An. melas* set. Reasons for exclusion include: HWD, Hardy-Weinberg Disequilibrium; NA, Did not amplify in test populations; and LV, Low variability.

	CAC				BOM						
Locus	N_a	H_O	H_E	HWE p-val.	N_a	H_O	H_E	HWE p-val.	Included / Reason Excluded	NMF	NMR
AM2H046	4	0.457	0.431	0.643	13	0.836	0.834	0.562	INC	1	0
AM2H143	3	0.543	0.575	0.233	9	0.738	0.730	0.380	INC	0	0
AM2H157	2	0.029	0.029	1.000	8	0.662	0.728	0.016	INC	0	1
AM2H215	3	0.567	0.590	0.466	5	0.629	0.671	0.385	INC	1	1
AM2H603	1	0.000	0.000	N/A	3	0.431	0.404	0.654	INC	0	0
AM2H793	4	0.324	0.505	0.000	15	0.643	0.827	0.000	INC	4	1
AM3H093	3	0.086	0.084	1.000	27	0.954	0.949	0.972	INC	0	0
AM3H127	6	0.686	0.723	0.769	7	0.470	0.673	0.000	INC	2	3
AM3H555	4	0.412	0.514	0.426	23	0.806	0.872	0.217	INC	5	0
AM3H753	1	0.000	0.000	N/A	3	0.609	0.607	0.813	INC	0	0
AMXH025	3	0.486	0.502	0.817	3	0.059	0.058	1.000	INC	0	1
AMXH038	3	0.229	0.364	0.046	13	0.772	0.900	0.010	INC	0	1
AMXH293	5	0.696	0.676	0.332	3	0.523	0.474	0.759	INC	0	4
AMXH755	3	0.484	0.516	0.468	6	0.750	0.767	0.508	INC	0	5
AMXH808	2	0.455	0.416	0.690	9	0.768	0.748	0.155	INC	0	0
AMXH106	1	0.000	0.000	N/A	15	0.949	0.926	0.004	HWD	3	1
AM2H161	4	0.000	0.267	0.000	3	0.279	0.515	0.000	HWD	0	0
AM3H249	2	0.200	0.183	1.000	2	0.220	0.198	1.000	LV	0	1
AM3H312	3	0.313	0.404	0.089	3	0.135	0.130	1.000	LV	0	0
AM2H417	6	0.371	0.333	1.000	20	0.821	0.843	0.883	LV	0	2
AM3H088	3	0.588	0.526	0.804	2	0.024	0.024	1.000	LV	2	1
AM3H544	1	0.000	0.000	N/A	1	0.000	0.000	N/A	LV	2	2
AM2H675	3	0.457	0.429	1.000	1	0.000	0.000	N/A	LV	0	0
AMXH019	1	0.000	0.000	N/A	3	0.581	0.456	0.125	LV	1	2
AG3H242	Excluded: No Analysis Performed								NA	1	0
AG3H341	Excluded: No Analysis Performed								NA	0	2
AGXH289	Excluded: No Analysis Performed								NA	0	0

Table 6 Population diversity estimates at each of 15 *An. melas* specific loci amplified in four *An. melas* populations. Bold indicates p-val.<0.05. Outlined HWE p-val. indicates that null alleles were detected by Microchecker for the denoted locus/population comparison. N_a, number of observed alleles. H_O, observed heterozygosity. H_E, expected heterozygosity. HWE p-val., Hardy-Weinberg Equilibrium test p-value. IPO: Ipono, Cameroon. LUB: Luba, Equatorial Guinea. GAM: Ballingho, The Gambia. GUI: Ponta Anabaca, Guinea Bissau.

	IPO				LUB				GAM				GUI			
Locus	N_a	H_O	H_E	HWE p-val.	N_a	H_O	H_E	HWE p-val.	N_a	H_O	H_E	HWE p-val.	N_a	H_O	H_E	HWE p-val.
AMXH25	2	0.042	0.041	1.000	3	0.625	0.641	0.252	7	0.625	0.645	0.611	5	0.583	0.607	0.314
AMXH38	17	0.938	0.914	0.149	3	0.500	0.441	0.419	15	0.851	0.888	0.219	10	0.854	0.801	0.691
AM2H46	9	0.875	0.816	0.209	5	0.468	0.492	0.372	10	0.723	0.802	0.285	16	0.917	0.873	0.611
AM3H93	22	0.979	0.940	0.157	5	0.208	0.399	**0.000**	15	0.896	0.859	0.313	18	0.787	0.830	0.622
AM3H127	5	0.500	0.591	0.455	4	0.750	0.723	0.423	6	0.646	0.714	0.621	6	0.702	0.745	0.967
AM2H143	6	0.711	0.734	0.170	3	0.404	0.556	0.084	11	0.763	0.817	**0.025**	9	0.842	0.840	0.812
AM2H157	8	0.792	0.805	0.380	2	0.021	0.061	**0.032**	12	0.913	0.827	0.412	7	0.805	0.790	0.575
AM2H215	7	0.681	0.752	0.752	4	0.723	0.612	**0.027**	9	0.792	0.787	0.941	9	0.889	0.833	0.382
AMXH293	4	0.563	0.527	0.513	4	0.604	0.652	0.786	11	0.787	0.829	0.438	10	0.826	0.843	0.183
AM3H555	21	0.875	0.873	0.684	4	0.565	0.510	0.841	12	0.692	0.707	0.585	6	0.646	0.738	0.422
AM2H603	3	0.375	0.428	0.554	1	0.000	0.000	N/A	3	0.191	0.178	1.000	3	0.149	0.141	1.000
AM3H753	4	0.500	0.649	0.092	1	0.000	0.000	N/A	3	0.213	0.198	1.000	3	0.068	0.067	1.000
AMXH755	5	0.813	0.763	**0.035**	3	0.646	0.577	0.321	4	0.638	0.585	0.857	6	0.545	0.606	0.444
AM2H793	11	0.745	0.818	**0.032**	5	0.313	0.315	0.499	18	0.936	0.881	0.979	17	0.854	0.879	0.674
AMXH808	8	0.750	0.758	0.581	2	0.438	0.442	1.000	10	0.870	0.814	0.213	10	0.833	0.795	0.873

Table 7 Population diversity estimates at each of five *An. gambiae* microsatellite loci amplified in four *An. melas* populations. Bold indicates p-val.<0.05. Outlined HWE p-val. indicates that null alleles were detected by Microchecker for the denoted locus/population comparison. NMF, Number mis-matches between Zheng *et al.* (1996) *An. gambiae* forward primer and *An. melas* sequence. NMR, Number mis-matches between Zheng *et al.* (1996) *An. gambiae* reverse primer and *An. melas* sequence. N_a, number of observed alleles. H_O, observed heterozygosity. H_E, expected heterozygosity. HWE p-val., Hardy-Weinberg Equilibrium test p-value. IPO: Ipono, Cameroon. LUB: Luba, Equatorial Guinea. GAM: Ballingho, The Gambia. GUI: Ponta Anabaca, Guinea Bissau.

			IPO				LUB				GAM				GUI			
An. gambiae Locus	NMF	NMR	N_a	H_O	H_E	HWE p-val.	N_a	H_O	H_E	HWE p-val.	N_a	H_O	H_E	HWE p-val.	N_a	H_O	H_E	HWE p-val.
AGXH25	0	1	2	0.042	0.041	1.000	3	0.617	0.637	0.235	7	0.660	0.658	0.718	5	0.644	0.624	0.565
AGXH38	0	1	16	0.936	0.913	0.163	3	0.479	0.429	0.340	9	0.813	0.833	0.612	6	0.778	0.759	0.894
AG3H127	2	3	5	0.542	0.587	0.843	4	0.745	0.725	0.490	6	0.521	0.699	0.058	5	0.372	0.691	**0.000**
AG2H157	0	1	7	0.792	0.803	0.337	2	0.021	0.062	**0.032**	11	0.915	0.826	0.459	7	0.795	0.794	0.940
AG2H215	1	1	7	0.688	0.751	0.700	4	0.723	0.638	0.135	9	0.787	0.786	0.943	9	0.911	0.830	0.319

Table 8 *An. melas*-primer amplified pairwise F_{ST} (lower diagonal), *An. gambiae*-primer amplified pairwise F_{ST} (upper diagonal), and null-corrected pairwise F_{ST} values (in parentheses). Italicized values are non-significant with p-value>0.05. Population abbreviations correspond with those defined in Table 4.

	IPO	LUB	GAM	GUI
IPO		0.33147 (0.33147)	0.17624 (0.17591)	0.20222 (0.20053)
LUB	0.33765 (0.31971)		0.23044 (0.23026)	0.24735 (0.24987)
GAM	0.17145 (0.17145)	0.23601 (0.21986)		0.00944 (*0.00510*)
GUI	0.1845 (0.18450)	0.25074 (0.23886)	0.00966 (0.00966)	

Table 9 Population diversity estimates at each of 15 *An. melas* specific loci amplified in 11 *An. melas* populations. Bold indicates p-val.<0.05. Asterisk (*) indicates significant deviation from HWE after Bonferonni correction (p-val.<0.0003). Outlined HWE p-val. indicates that null alleles were detected by Microchecker for the denoted locus/population comparison. N_a, number of observed alleles. H_O, observed heterozygosity. H_E, expected heterozygosity. HWE p-val., Hardy-Weinberg Equilibrium test p-value. s.d., standard deviation. Population abbreviations correspond with those defined in Table 4.

	LUB				RIA				CAC				TIK			
Locus	N_a	H_o	H_E	HWD p-val.	N_a	H_o	H_E	HWD p-val.	N_a	H_o	H_E	HWD P-val.	N_a	H_o	H_E	HWD p-val.
AMXH25	3	0.60	0.64	0.12	3	0.31	0.48	**0.00***	3	0.49	0.50	0.82	7	0.75	0.80	0.41
AMXH38	3	0.43	0.41	0.73	3	0.41	0.36	0.17	3	0.23	0.36	**0.05**	10	0.81	0.90	0.12
AM2H46	5	0.48	0.52	0.74	5	0.57	0.63	0.61	4	0.46	0.43	0.65	8	0.81	0.86	0.81
AM3H93	8	0.40	0.61	**0.00***	5	0.37	0.42	**0.00***	3	0.09	0.08	1.00	5	0.75	0.65	0.73
AM3H127	5	0.72	0.73	0.07	5	0.80	0.74	0.40	6	0.69	0.72	0.77	7	0.88	0.81	0.39
AM2H143	4	0.51	0.57	**0.03**	4	0.59	0.58	**0.04**	3	0.54	0.58	0.23	10	0.75	0.87	0.05
AM2H157	2	0.02	0.04	**0.03**	2	0.04	0.04	1.00	2	0.03	0.03	1.00	7	0.87	0.82	0.44
AM2H215	4	0.58	0.56	0.61	5	0.57	0.60	**0.00**	3	0.57	0.59	0.47	9	0.81	0.84	0.38
AMXH293	5	0.64	0.64	0.97	6	0.59	0.65	0.06	5	0.70	0.68	0.34	11	0.88	0.88	0.79
AM3H555	4	0.52	0.50	0.66	5	0.45	0.47	0.06	4	0.41	0.51	0.43	9	0.75	0.84	0.26
AM2H603	1	0.00	0.00	N/A	1	0.00	0.00	N/A	1	0.00	0.00	N/A	4	0.43	0.46	1.00
AM3H753	2	0.01	0.01	1.00	2	0.12	0.11	1.00	1	0.00	0.00	N/A	3	0.25	0.28	0.30
AMXH755	3	0.59	0.58	0.45	4	0.61	0.60	1.00	3	0.48	0.52	0.47	8	0.56	0.76	0.19
AM2H793	5	0.27	0.26	0.72	6	0.19	0.20	0.06	4	0.48	0.52	0.12	12	0.69	0.91	**0.04**
AMXH808	2	0.44	0.42	0.81	3	0.43	0.40	0.88	2	0.45	0.42	0.69	8	0.73	0.82	0.64

Table 9 (continued)

	GHA				GAM				GUI				IPO			
Locus	N_a	H_o	H_E	HWD p-val.	N_a	H_o	H_E	HWD p-val.	N_a	H_o	H_E	HWD p-val.	N_a	H_o	H_E	HWD p-val.
AMXH25	4	0.67	0.71	0.79	8	0.61	0.63	0.36	5	0.57	0.61	0.42	4	0.09	0.09	1.00
AMXH38	7	0.67	0.86	0.29	15	0.85	0.89	0.06	11	0.85	0.81	0.66	18	0.96	0.92	0.35
AM2H46	7	0.67	0.91	0.17	12	0.81	0.83	0.09	13	0.92	0.86	0.55	8	0.74	0.81	0.04
AM3H93	6	1.00	0.82	0.21	15	0.88	0.87	0.10	20	0.82	0.82	0.99	23	0.97	0.94	0.38
AM3H127	5	0.83	0.83	0.56	7	0.63	0.72	0.16	7	0.73	0.74	0.98	7	0.54	0.60	0.30
AM2H143	7	1.00	0.88	0.75	12	0.81	0.82	**0.00**	10	0.87	0.85	0.79	6	0.76	0.75	0.09
AM2H157	4	0.83	0.65	1.00	15	0.88	0.82	0.56	8	0.81	0.80	0.40	9	0.77	0.79	0.39
AM2H215	4	0.67	0.76	0.58	11	0.84	0.80	0.98	10	0.86	0.83	0.57	8	0.72	0.75	0.96
AMXH293	6	0.67	0.85	**0.02**	16	0.83	0.84	0.31	10	0.85	0.84	0.26	4	0.52	0.54	0.67
AM3H555	4	0.83	0.71	1.00	15	0.64	0.73	0.11	7	0.67	0.76	0.59	24	0.90	0.88	0.65
AM2H603	2	0.17	0.17	1.00	3	0.19	0.18	1.00	3	0.15	0.14	1.00	3	0.36	0.39	0.66
AM3H753	2	0.33	0.30	1.00	3	0.17	0.16	1.00	3	0.12	0.12	1.00	4	0.63	0.64	0.87
AMXH755	3	0.17	0.59	**0.03**	7	0.66	0.61	0.16	6	0.55	0.63	0.58	6	0.81	0.77	0.07
AM2H793	7	1.00	0.91	1.00	22	0.91	0.90	0.98	18	0.82	0.87	0.42	14	0.74	0.81	0.05
AMXH808	6	0.50	0.88	0.06	10	0.84	0.80	0.18	10	0.84	0.80	0.70	8	0.74	0.73	0.31

Table 9 (continued)

	BOM				GAB				ANG			
Locus	N_a	H_o	H_E	HWD p-val.	N_a	H_o	H_E	HWD p-val.	N_a	H_o	H_E	HWD p-val.
AMXH25	3	0.06	0.06	1.00	7	0.24	0.30	**0.01**	1	0.00	0.00	N/A
AMXH38	13	0.81	0.90	**0.04**	17	0.89	0.88	0.67	9	0.73	0.84	0.24
AM2H46	13	0.84	0.83	0.55	12	0.84	0.84	0.98	9	0.82	0.77	0.83
AM3H93	27	0.95	0.95	0.97	17	0.87	0.92	0.13	9	0.73	0.84	**0.01**
AM3H127	7	0.65	0.69	**0.04**	5	0.67	0.71	0.38	4	0.53	0.60	0.28
AM2H143	9	0.74	0.73	0.38	6	0.71	0.80	0.39	4	0.67	0.72	0.62
AM2H157	8	0.66	0.73	**0.02**	11	0.73	0.82	**0.00***	11	0.71	0.80	**0.01**
AM2H215	5	0.63	0.67	0.39	8	0.58	0.63	0.35	4	0.53	0.53	0.90
AMXH293	3	0.52	0.47	0.76	3	0.53	0.51	1.00	5	0.43	0.48	0.09
AM3H555	23	0.81	0.87	0.16	25	0.90	0.88	0.77	8	0.69	0.60	0.92
AM2H603	3	0.43	0.40	0.66	3	0.10	0.09	1.00	2	0.08	0.08	1.00
AM3H753	3	0.61	0.61	0.82	6	0.44	0.46	0.22	5	0.59	0.62	0.51
AMXH755	6	0.75	0.77	0.50	5	0.78	0.67	0.06	3	0.05	0.12	**0.02**
AM2H793	15	0.75	0.84	**0.00***	19	0.67	0.74	0.60	9	0.70	0.77	**0.04**
AMXH808	9	0.77	0.75	0.17	9	0.75	0.79	0.31	8	0.75	0.73	0.10

Table 9 (continued)

Locus	Mean N_a	Mean N_a s.d.	Mean H_O	Mean H_O s.d.	Mean H_E	Mean H_E s.d.
AMXH25	4.36	2.16	0.40	0.27	0.44	0.28
AMXH38	9.91	5.52	0.70	0.23	0.74	0.24
AM2H46	8.73	3.35	0.72	0.16	0.75	0.16
AM3H93	12.55	8.25	0.71	0.30	0.72	0.27
AM3H127	5.91	1.14	0.70	0.11	0.72	0.07
AM2H143	6.82	3.03	0.72	0.14	0.74	0.12
AM2H157	7.18	4.31	0.58	0.36	0.58	0.35
AM2H215	6.46	2.81	0.67	0.12	0.69	0.11
AMXH293	6.73	4.00	0.65	0.15	0.67	0.16
AM3H555	11.64	8.56	0.69	0.17	0.71	0.16
AM2H603	2.36	1.03	0.17	0.17	0.17	0.17
AM3H753	3.09	1.45	0.30	0.24	0.30	0.25
AMXH755	4.91	1.81	0.55	0.24	0.60	0.18
AM2H793	11.91	6.17	0.66	0.25	0.70	0.26
AMXH808	6.82	3.09	0.66	0.16	0.69	0.18

Table 10 Pair-wise estimates of genetic divergence (G_{ST}' and F_{ST}) values between 11 *An. melas* populations amplified at 15 microsatellite loci. Top diagonal: pair-wise G_{ST}' values. Bottom diagonal: pair-wise F_{ST} values (significant values (p-val.<0.05) in bold). Statistics computed using 10,000 permutations. Population abbreviations correspond with those defined in Table 4.

	Bioko Island			Northwest				Central			
G_{ST}'/F_{ST}	LUB	RIA	CAC	TIK	GHA	GAM	GUI	IPO	BOM	GAB	ANG
LUB		0.0796	0.0656	0.3685	0.4918	0.4403	0.4348	0.7143	0.7432	0.7127	1.0000
RIA	**0.0967**		0.0554	0.3294	0.4668	0.4231	0.4059	0.6542	0.6792	0.6526	0.9063
CAC	**0.0847**	**0.0788**		0.2974	0.4270	0.4161	0.3817	0.6659	0.6948	0.6581	0.9219
TIK	**0.2550**	**0.2468**	**0.2226**		0.1889	0.1311	0.1452	0.3263	0.3583	0.3786	0.4871
GHA	**0.3070**	**0.3138**	**0.3065**	0.0296		0.1850	0.1803	0.3513	0.3739	0.3661	0.5308
GAM	**0.2520**	**0.2454**	**0.2321**	**0.0405**	**0.0460**		0.0273	0.4012	0.4465	0.4395	0.5550
GUI	**0.2695**	**0.2604**	**0.2424**	**0.0470**	**0.0474**	**0.0060**		0.4114	0.4496	0.4473	0.5824
IPO	**0.3265**	**0.3242**	**0.3028**	**0.1063**	**0.1108**	**0.1435**	**0.1468**		0.0156	0.0601	0.1546
BOM	**0.3429**	**0.3405**	**0.3304**	**0.1188**	**0.1294**	**0.1551**	**0.1601**	0.0000		0.0461	0.1475
GAB	**0.3459**	**0.3484**	**0.3316**	**0.1261**	**0.1253**	**0.1553**	**0.1618**	**0.0218**	**0.0137**		0.1429
ANG	**0.4293**	**0.4360**	**0.4299**	**0.1849**	**0.2101**	**0.2044**	**0.2173**	**0.0723**	**0.0756**	**0.0767**	

Table 11 Results of the global analysis of molecular variance (AMOVA) as a weighted average over 15 *An. melas* microsatellite loci, with groups defined as West, Central, and Bioko Island population clusters.

Source of Variation	Sum of Squares	Variance Components	Percentage of Variation
Among Groups	1108.95	1.33 Va	22.24
Among Populations Within Groups	208.31	0.23 Vb	3.82
Within Populations	5139.17	4.42 Vc	73.94
Total	6456.43	5.98	

Table 12 Estimates of mean mtDNA sequence divergence within sampled *An. melas* and *An. gambiae* populations. The rate variation among sites was modeled with a gamma distribution (shape parameter = 6). Codon positions included were 1st+2nd+3rd+noncoding. All positions containing gaps and missing data were eliminated. There were a total of 1161 bp positions in the final dataset.

Species Population	Distance Estimate	Standard Error
An. melas		
Luba, Equatorial Guinea (Bioko)	0.00021	0.00014
Riaba, Equatorial Guinea (Bioko)	0.00031	0.00021
Cacaual, Equatorial Guinea (Bioko)	0.00027	0.00019
Tiko, Cameroon	0.00706	0.00143
Ada Foah, Ghana	0.00509	0.00122
Ballingho, The Gambia	0.00579	0.00115
Ponta Abanaca, Guinea Bissau	0.00576	0.00101
Ipono, Cameroon	0.00317	0.00073
Bome, Equatorial Guinea	0.00261	0.00085
Port Gentil, Gabon	0.00259	0.00059
Mateba, Angola	0.00065	0.00039
An gambiae		
Mongola, Equatorial Guinea (Bioko)	0.00312	0.00096
Ukomba, Equatorial Guinea	0.00366	0.0018

Table 13 Divergence date estimates between *An. melas* population clusters and *An. gambiae* and *An. arabiensis* individuals. Node names correspond with tree annotations in Figure 10. HPD, highest posterior density. BP, before present.

Node	Median Divergence Date (years BP)	Lower 95% HPD (years BP)	Upper 95% HPD (years BP)
A	42,741.37	4,801.66	101,272.76
B	200,272.04	109,769.74	317,389.87
C	47,274.21	5,784.25	114,177.44
D	423,445.72	261,993.99	621,259.42
E	54,518.21	10,214.47	120,777.02
F	342,986.63	204,171.86	512,187.90
G	53,619.28	9,744.75	116,168.92
H	151,922.40	62,089.65	268,870.96

Table 14 Genetic distances between member *An. melas*, *An. gambiae*, and *An. arabiensis* based upon ND4 and ND5 mtDNA sequences.

Species 1	Species 2	Genetic Distance	Std. Err of Mean
An. gambiae	*An. melas*	0.01404	0.00279
An. gambiae	*An. arabiensis*	0.00593	0.00152

VITA

Name:	Kevin Canning Deitz
Address:	Department of Entomology Texas A&M University TAMU 2475 Minnie Belle Heep Bldg. Rm. 412 College Station, TX 77843-2475
Email Address:	kcdeitz@tamu.edu
Education:	B.S., Conservation Biology, The State University of New York College of Environmental Science and Forestry, 2008 M.S., Entomology, Texas A&M University, 2011