VECTOR BIOLOGY AND GENOMICS OF *ANOPHELES* IN SOUTHERN AND CENTRAL AFRICA

by
Christine Michelle Jones

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

December, 2018

Abstract

Nchelenge District in Zambia has high malaria transmission despite intensive malaria control. The primary mosquito species contributing to transmission is *Anopheles funestus s.s.*, a species for which genetic and genomic data is lacking compared to other important vectors. To investigate the population dynamics of *An. funestus* that contribute to transmission and to eventually prepare a suitable set of genetic markers for future study, it was necessary to improve the existing baseline of genomic information. 43 field specimens of *An. funestus* were gathered from sites in central and southern Africa. Illumina shotgun sequencing was conducted and the first 43 complete high-quality *An. funestus* complete mitogenomes were assembled. The full-length mitogenome yielded a set of 567 polymorphic sites from which population genetic markers may be developed. Analyses conducted on the full set of 43 mitogenomes from three countries illuminated phylogenetic relationships and a complex demographic history.

To investigate outdoor transmission in Nchelenge, which may partially explain the refractoriness of transmission to standard indoor-based malaria interventions, Centers for Disease Control light traps (CDC LTs) were set outdoors in August of 2016. The anophelines collected were more diverse than seen from primarily indoor collections in Nchelenge. Sequencing, phylogenetics, and morphology revealed more than 12 phylogenetic groups of anophelines, some of which appear to be competent for malaria parasites. Approximately 1% of outdoor-collected *An. funestus* were parasite-positive,

which calls for further investigation of the involvement of these vectors in outdoor transmission.

The outdoor CDC LTs were set with three different schemes, including traps baited with a synthetic human odorant blend (BG-Lure®). Models indicated that the neither the abundance nor the species diversity of female anophelines differed between traps placed outdoors near humans, animals, or baited with synthetic attractant. Instead, for both the number of anophelines caught per trap as well as the number of species present in a trap, the best predictors were site (whether traps were lakeside or inland) and the numbers of people sleeping under or without bednets. Together, these studies represent significant steps forward in understanding the entomological drivers of malaria transmission in Nchelenge.

Thesis Advisory Committee

Thesis Advisor:	Douglas E. Norris, Ph.D.
Thesis Committee Chair:	Thomas A. Louis, Ph.D.
Thesis Committee Members:	William J. Moss, M.D.
	Conor McMeniman, Ph.D.
Thesis Committee Alternates:	Robert Gilman, M.D.
	Monica Mugnier, Ph.D.
	Timothy M. Shields, M.A.

Acknowledgements

I would like to thank first and foremost my family and friends, who were essential over the last six years to completion of this dissertation. Without them, I would be lost.

My advisor, Doug Norris, was a kind and excellent mentor throughout my time at Johns Hopkins. My colleagues provided a lovely work environment, which made it almost easy to get up and go to work every morning. I would like to thank especially Julia Pringle, who came into lab just a year after me and who has provided countless hours of assistance and positive energy. Giovanna Carpi, who was a postdoc and mentor, helped me learn some essential topics and rescued me from a number of mistakes. Ilinca Ciobotariu devoted a massive amount of her time and energy to helping me complete data processing, and was always sweet while doing so. Jordan Hoffman and Mary Gebhart represent the newest and endlessly cheerful iteration of the Norris lab, and I hope for nothing but good things for them. Smita Das, Tyler Henning, and Maureen Kessler were my original lab-mates, and they taught me a lot, while keeping me entertained. Hannah MacLeod always brought a unique and valuable perspective, and was rightfully considered an honorary member of our lab.

I'd also like to thank field team members in Zambia who expended time and effort to help me collect my specimens, as well as numerous other members of ICEMR who helped me throughout the years. Tim Shields, in particular, was an important mentor, instructor, and all-around nice guy. Other people in MMI who were incredibly helpful

and accommodating and who deserve a lot of recognition and thanks: Gail O'Connor, Anne Jedlicka, Amanda Dziedzic, and Thom Hitzelberger. And finally, I would like to thank my Thesis Committee for agreeing to be part of this whole process.

Table of Contents

Section

List of Tables

Table **Page**

List of Figures

Figure **Page**

Chapter I

Introduction

Malaria is among the most important infectious diseases worldwide. It is caused by *Plasmodium* parasites, which are transmitted in a cycle between humans and *Anopheles* mosquitoes (referred to as malaria vectors). Over 200 million cases of malaria occurred in 2016, with approximately 445,000 deaths.[1] The vast majority of cases and deaths (~90%) occurred in Africa, with 70% of deaths occurring in children younger than five. More than a third of cases and deaths worldwide occurred in Nigeria and the Democratic Republic of the Congo alone.[1] The social and economic burden of malaria cannot be overstated, with an estimated loss of over 550,000 disability-adjusted life years (DALYs) in 2015.[2] The complex ecology of the disease cycle and ubiquity of the mosquito vector globally throughout tropical and temperate zones have historically made malaria a difficult disease to combat.

The first push for global malaria eradication, which began in 1955, effectively shrank the distribution of malaria cases and transmission. Where cases had been widespread, ranging even into subpolar climates, active transmission became restricted to tropical regions. The enthusiasm that drove this program dwindled in the 1960's as funding declined and vectors and parasites acquired resistance to the novel chemicals which made up the bulk of the eradication strategy. Renewed interest in combating malaria at the turn of the century resulted in the Roll Back Malaria Partnership (RBMP). The RBMP established the Global Malaria Action Plan (GMAP), a detailed 2007-2015 strategy for

combating and eventually eliminating malaria. GMAP led to widespread deployment of standardized malaria control, resulting in reductions in malaria incidence and mortality of 20%-40% and 31%, respectively, between 2000 and 2015.[3,4] Despite significant progress towards reduction and regional elimination of malaria, malaria remains a leading global cause of morbidity and mortality.[3] Several phenomena contribute to the failure of malaria control efforts, including challenges in vector control that are further compounded by inadequate funding and poor public health infrastructure.[5]

Challenges in Malaria Vector Control

The primary malaria interventions are antiparasitic drugs and vector control.[1] Vector control is estimated to be responsible for approximately 80% of the reductions in malaria between 2000 and 2015, likely because of the relative ease of large-scale implementation as compared to antiparasitic drugs.[3] Vector species share some overarching characteristics that are exploited by the major vector control interventions. Namely, they tend to feed and rest inside homes at night while humans are sleeping. Therefore, the predominant tools for malaria vector control are long-lasting insecticide-treated bednets (LLINs) and indoor residual spraying (IRS).

Bednets are hung around sleeping humans and physically block the mosquito as it attempts to feed. LLINs provide an additional barrier in the form of imbedded insecticides (untreated nets may have reduced protective effect).[6–8] The insecticide component of LLINs provides a community level of protection, since the insecticide

reduces vector lifespan and therefore diminishes the number of infectious bites incurred by individuals in a community.[9,10] Delayed acquisition of a blood meal may also shorten the lifespan, fecundity, and the transmission capacity of a mosquito thwarted by a bed net.[11]

Insecticides are sprayed on the walls or eaves of a home during IRS campaigns. These chemicals are meant to kill mosquitoes that land on these surfaces to rest after feeding, though they can also provide a repellent effect.[12] Each insecticide comes with its own residual timeline, or the length of time it is effective against mosquitoes after application. This timeline is dependent on the surface on which it is sprayed, but ranges from 2 months to 6 months (or more, in some cases).[13] Since IRS is intended to kill resting mosquitoes that have theoretically already fed on a host, it serves primarily as a community protection by reducing the overall number of mosquitoes (specifically, those with a preference for feeding on humans indoors) in the sprayed region.

LLINs and IRS have been shown to reduce mortality and morbidity due to malaria, and they comprise a large portion of anti-malaria efforts.[14,15] However, vector control is facing many challenges, including insecticide resistance, insufficient knowledge of primary malaria vectors, residual transmission, and the rise of secondary vectors.

Insecticide Resistance

Insecticide resistance is a serious problem for public health because insecticides are an integral part of vector control. There are currently only four classes of insecticides used for malaria control: organochlorines (DDT), pyrethroids, carbamates and organophosphates. The widespread and prolonged application of these insecticides has selected for several types of insecticide resistance in vector populations: target-site resistance, metabolic resistance, behavioral resistance, and cuticular resistance.[16]

Most of the studies on insecticide resistance to date have focused on target-site and metabolic resistance. Target-site resistance is caused by mutation in a targeted molecule that reduces binding affinity with the insecticide, resulting in reduced insecticidal activity. Some classes of insecticides have the same target molecule (pyrethroids and DDT both target the sodium channel, while carbamates and organophosphates target acetylcholinesterase), so it is possible for a single point mutation in the coding sequence of the target molecule to confer cross-resistance to multiple insecticides and even multiple classes of insecticides.[17] These known mutations can be easily and quickly screened for in target populations using PCR or DNA sequencing technologies. A genetic resistance marker commonly screened for in some vector species called kdr (knock-down resistance) is responsible for conferring both DDT and pyrethroid resistance. In kdr, a single point mutation results in a single amino acid change which reduces nervous system sensitivity to the insecticide.[18] Similarly, cross-resistance to carbamates and organophosphates can be conferred by a single point mutation in the acetylcholinesterase gene (ACE-1).[19,20]

Metabolic resistance generally confers resistance to multiple types of insecticides, since it involves the increased expression of metabolic detoxification enzymes which may degrade or detoxify a variety of chemicals. This type of resistance is complex, and thus more difficult to screen for, since classically it required assessment of enzymatic activity. Recent studies have elucidated some key classes of enzymes that confer metabolic resistance, and up- and down-regulation can be evaluated using molecular approaches. Unfortunately, there are currently no quick or easy standardized tools available for routine screening in insecticide resistance/malaria control surveillance programs.[21]

Behavioral resistance and cuticular resistance are the least understood forms of insecticide resistance. Behavioral resistance (long-recognized in agricultural arthropod pests) occurs when vector populations exhibit altered behaviors that reduce their exposure to insecticides. The most important adaptations for malaria control are when populations begin feeding earlier and/or when they begin to feed outdoors. Current vector control strategies are likely to be ineffective against such populations.[11] Studies of behavioral resistance have lagged due to difficulty assessing behavioral phenotypes in the field, and studies of cuticular resistance in malaria are relatively new.[11,22–25] The cuticle of an arthropod is its protective outer 'skeleton', and in cuticular resistance the upregulation of particular proteins causes thickening of the outer cuticle, which in turn reduces insecticide penetration/uptake.[26,27]

The majority of studies into the molecular mechanisms of insecticide resistance have relied on gene and protein expression-level data to determine the genetic causes of

resistance phenotypes. In many cases, however, the ultimate cause of resistance remains obscure. Some newer studies have focused on the whole-genome in order to understand the complex combinations of mutations that confer many resistance phenotypes, especially complex metabolic resistance phenotypes.[28–33] Abundant genomic data is crucial to these efforts to find markers and mechanisms.[34–36]

The rise of resistance and cross-resistance across multiple classes drives the desperate need for new insecticides. LLINs are particularly vulnerable to resistance, as only pyrethroid-class insecticides have been considered both safe and effective enough for use in nets and pyrethroid resistance has become widespread in the last few decades.[37] IRS can be conducted with all classes of insecticides, but unfortunately resistance to all currently utilized classes of insecticides has been documented.[1,37] Often for IRS, the rapid spread of resistance phenotypes causes logistical and economic constraints. New insecticide classes and formulations are on the horizon, but these may require implementation of new handling and disposal protocols and can many times be much more expensive. In 2012, of the 80 countries reporting IRS programs, over half relied solely on pyrethroids.[1] By 2016, the number of countries reporting any IRS program declined alongside the proportion of programs that solely used pyrethroids.[1] As malaria control programs switch to these newer chemicals, cost has become a barrier and without additional funds, coverage in often sacrificed.[1]

With the most recent push for malaria eradication, the malaria control community has recognized the overwhelming importance of managing resistance and has responded by

developing novel kinds of vector control interventions. Some are still chemical-dependent, such as PBO. PBO is an organic compound with no inherent insecticidal activity that enhances the activity of pyrethroid and other insecticides by inhibiting insect detoxification enzymes; it provides hope for maintaining the utility of pyrethroid-treated bednets in the face of increasingly widespread resistance. Other strategies, like introduction of transgenic mosquitoes, could bypass the resistance phenomenon altogether. These methods rely on introducing genetically- or otherwise manipulated mosquitoes into a field setting to reduce the natural vector population or replace it with mosquitoes having decreased capacity for transmission. All approaches to vector control require an understanding of targeted mosquito populations. It is therefore extremely important to understand the dynamics within these critical vector populations by studying the biology and genomics of natural populations.

Anopheles funestus: lack of knowledge

Though there are over 40 species of *Anopheles* mosquitoes that are considered efficient at transmitting malaria parasites, only a few have been considered important for malaria control. Detailed knowledge of the anophelines present in an area is required for effective targeting of vector control, because these species exhibit distinct behaviors. Species complexes, in which distinct species are morphologically indistinguishable, are common among anophelines and often only one or a few members of species complex are important for transmission. In this case, molecular assays are required to speciate effectively. Genomic and genetic analyses over the last several decades have not only

improved the tools we use to differentiate anophelines, but have also been used to study anophelines and their population dynamics, which has led to a better understanding of the relationships between anopheline populations and malaria control.[31,38–46]

The major vectors in sub-Saharan Africa are *Anopheles gambiae sensu lato* and *An. funestus sensu stricto*. Much like *An. gambiae s.s.*, *An. funestus s.s.* is the titular member of a species complex. Unlike *An. gambiae s.s.*, *An. funestus s.s.* is the only member of its species complex thought to be meaningfully important for malaria transmission. There are now 13 species recognized in the *An. funestus* group (Table 1),[47] and studies have supported the existence of additional chromosomal forms in *An. funestus s.s.* Chromosomal forms are populations within a species that are distinguished by differing chromosome inversion polymorphisms, and are well known in studies of *An. gambiae s.s.* In *An. gambiae s.s.*, chromosomal forms constitute several ecological 'types' which demonstrate different environmental preferences.[48–50] One of these chromosomal forms eventually became classified as *An. coluzzii* – a species distinct from *An. gambiae s.s* in insecticide resistance profile and ecological preferences.[51–55] In *An. funestus s.s.*, a small body of literature has associated chromosomal forms with environmental features as well as indoor/outdoor resting preferences.[56–60]

Despite its importance in the transmission of malaria parasites in sub-Saharan Africa (see Figure 1 for geographical distribution), research regarding *An. funestus* has lagged behind that of the other major malaria vectors in Africa, largely due to technical issues in rearing this species in laboratory colonies. The bulk of information about *An. funestus* comes

from early malaria studies in the 1950s and further study stalled until advanced technologies in the 1990's made for easier study of field populations. Before molecular diagnostics were available to differentiate *An. funestus s.s.* from more benign members of the complex, the only other reliable method for discrimination relied on cytogenetics – the study of unique banding patterns on prepared chromosome specimens (otherwise known as karyotyping). Cytogenetics was first used to characterize the *An. funestus* species complex in 1980,[61–63] and subsequently used to illuminate population structure in *An. funestus s.s*.[59,64–68] Differential PCR and similar methods began to emerge in the late 1990's, when sequencing technologies became more advanced.[69–72] Around the same time, the first permanent laboratory colony of *An. funestus* was established, which has proven a great step forward in research on this species.[73] Though progress has been slow, baseline knowledge of *An. funestus* has been established.

In 2006, Krzywinski *et al.* published an *An. funestus* mitochondrial genome.[74] This reference genome was a significant improvement upon preexisting information available for the species, but still contained information for only ~70% of the total mitochondrial genome, with gaps in the sequence mainly in key coding regions. Newer studies have included sequencing of the transcriptome as well as limited whole-genome sequencing and high-depth targeted sequencing.[44,75] These are the kinds of studies which allow for investigation of the mechanisms of insecticide resistance and other functional correlations. In 2015, the whole nuclear genome sequence of *An. funestus* was released as part of an effort to characterize the genomics of 16 different anopheline species.[76] Despite being in the public domain for several years, this genome has yet to be assembled to

chromosome level and remains in 1,392 separate scaffolds of contigs and supercontigs, making many analyses difficult.

Nonetheless, the overall increase in *Anopheles* genetic data enabled study of mosquito populations across sub-Saharan Africa. Most of these studies have been limited in geographical scale, often to one country, and also to specific continental regions, many in western Africa. Country-level studies have generally found *An. funestus* to be a single population or have found weak signals of differentiation of populations as a function of distance.[64,65,77–81] One exception is a study in Kenya, which showed significant differentiation between populations across the Great Rift Valley, which is known to serve as a genetic barrier for populations of *An. gambiae*.[77,82] Across larger geographic areas, markers used in early studies indicated one large panmictic population of *An. funestus*, but more sophisticated markers have since shown a pattern of shallow structure across the continent.[44,72,83–86]

Though most molecular population genetics studies of *An. funestus* have used microsatellites from the nuclear genome, a few have used mitochondrial markers as well. One group showed two distinct lineages in comparisons of mitochondrial DNA.[85] Lineage I (referred to as clade I in that study) was found in all 11 countries included in the study, while Lineage II (referred to as clade II) was found only in Madagascar and Mozambique. The concomitant nuclear DNA studies did not show a similar phylogenetic grouping suggesting that this is an ancestral artifact. Follow-up studies have been sparse,

and it remains unclear whether these lineages have any biological relevance, especially for transmission of *Plasmodium*.

The impact of vector population genetic structure on malaria control is potentially large. High genetic diversity could contribute to the emergence of insecticide resistance genotypes and the failure of interventions. CRISPR-Cas9 systems of gene drive rely on a conserved target sequence to drive a genotype into a population; even a relatively uncommon single nucleotide polymorphism (SNP) in the targeted sequence could lead to a failed intervention.[87] Additionally, gene flow between populations or even species can lead to the rapid spread of the mutated target sequence or insecticide resistance. For example, adaptive introgression has been documented between *An. gambiae* and *An. coluzzii* in response to selective pressure induced by humans through the use of LLINs and IRS.[38] Understanding the population genetics and genomics of vector populations is essential to proper management of insecticide resistance as well as non-insecticide vector control. Inbreeding in or highly structured natural populations could also make successful introduction of a transgenic control strategy unlikely.[88]

Residual Transmission and Secondary Vectors

It is important to recognize that even with universal coverage of the commonly used vector control interventions, some transmission often still occurs. This is known as 'residual transmission,' and it has become an important issue in the malaria elimination community. Residual transmission often results from a combination of phenomena,

including shifts in vector species composition, the involvement of previously-unrecognized anopheline species in transmission, and altered behavior of primary vectors.

Several species-dependent mosquito behaviors can have an impact on vector control and malaria transmission. After a mosquito takes a blood meal and increases body mass several-fold, she often rests on a nearby surface to begin the digestion process and conversion of blood to eggs. Populations of mosquitoes preferring to rest on surfaces within a structure are termed endophilic. Endophilic mosquitoes tend to be more readily controlled by measures such as IRS, which aims to coat potential resting surfaces with insecticide. Exophilic mosquitoes in contrast, prefer resting outdoors – a behavior that may allow the mosquito to avoid IRS altogether. In addition to resting preference, the mosquito may prefer feeding either indoors (endophagic) or outdoors (exophagic). Like endo*philic* mosquitoes, endo*phagic* mosquitoes are more effectively controlled by indoor-based interventions such as LLINs. Additionally, different species or populations within a single species may have varying preference for host. Some mosquitoes are anthropophilic, meaning that they prefer to feed on humans, while zoophilic species prefer to feed on non-human animals. To make the situation even more complex, different species or populations will feed at different times of the day. While peak activity and feeding often occur during late night hours for anophelines, feeding may begin relatively early, at dusk, or extend relatively late, at dawn.

The most important primary vectors of malaria in sub-Saharan Africa are *An. gambiae s.s.*, *An. coluzzii*, *An. arabiensis*, and *An. funestus s.s.* These species have distinct preferences for breeding sites and have diverse feeding and resting behaviors, as well as different susceptibilities to insecticides. Each is important because each is highly competent for human-infecting *Plasmodium* and is generally highly anthropophilic. They also tend to be nocturnal feeders with a high preference for feeding and resting indoors, which are precisely the mosquito behaviors that LLINs and IRS target.[89] However, because these interventions are implemented inside of sleeping structures, universal coverage of LLINs/IRS will still leave people vulnerable to residual transmission from exophagic and exophilic vectors or vectors that feed while people are unprotected.[90] This includes evening, before individuals retire inside to sleep under LLINs or inside sprayed households, and pre-dawn, when many wake to start the day, especially in rural agricultural-based communities.

Control of these primary vectors in sub-Saharan Africa has often been sufficient for dramatic reductions of malaria transmission in endemic areas. However, in many cases, formerly low-abundance primary vectors that resisted control interventions, as well as ever-present 'secondary' vectors, have risen to fill the transmission niche left behind by the successfully reduced/eliminated vector populations.[89,91,92] Secondary vectors are traditionally considered to be of minor importance due to behaviors which make them less efficient at transmitting the parasite (e.g. primarily or exclusively zoophagic, exophagic, low population densities). However, in the last several years, in areas where the primary vectors have been reduced, investigators have increasingly detected parasites

in these ‘secondary vectors’ and a growing number of formerly-ignored or completely uncharacterized species are now recognized as critically important malaria vectors.[93–99]

Behavioral adaptation in some species, including both primary and secondary vectors, has increasingly been implicated as an important factor in transmission. These behavioral alterations may result from the interplay of genetic adaptation and phenotypic plasticity.[23] Studies have shown shifts to exophily and exophagy in populations of historically endophilic and/or endophagic vector mosquitoes.[22,23,100–103] In Benin, the proportion of *An. funestus* mosquitoes feeding outdoors increased significantly after multiple rounds of universal LLIN coverage, and peak feeding times shifted to times approaching dawn.[23] On Bioko Island, *An. gambiae* and *An. melas* exhibited a high degree of outdoor host-seeking behavior, despite limited display of such behavior before implementation of intense indoor vector control.[104] Because these changes happened over several generations of mosquitoes, they are more likely to be due to selection of those subpopulations that are better genetically adapted. Phenotypic plasticity in populations of the same species may also contribute to differences in degree of endophagy observed in different environments. Phenotypic plasticity refers to the varying expression of a single genotype’s phenotype as a function of circumstance or environment. For example, in a region with low access to humans, Lefèvre et al showed that although only 40% of the blood meals in captured *An. gambiae* were of human origin, when *An. gambiae* was given the choice between traps baited with either human or calf odor, 88% of feeds were on humans.[105]

Due to the complex differences in foraging, resting, and other ecological characteristics of even highly related species, speciation of mosquitoes is incredibly important in malaria control. High-throughput assays like PCR have been developed to quickly differentiate between common species based on primers differentially binding to regions of DNA based on the species- or population- specific sequences present. These assays can be inadequate when a single mosquito population or species is even moderately diverse (causing unanticipated changes in binding sites), and are limited to species for which genomic information exists. Not only are there limited numbers of species for which PCR discrimination exists, but there are a very limited number of species represented by genetic or genomic data in public databases like NCBI. When ambiguous morphological identifications arise, due to either limited identification experience or damage of the specimen, there may be no molecular recourse for speciation for rarer or neglected species. To effectively control mosquitoes responsible for malaria transmission, the repertoire for *Anopheles* genetic data must become larger, more diverse and highly specific.

ICEMR

The Southern and Central (formerly Southern) Africa International Centers for Excellence in Malaria Research (ICEMR) has worked to investigate the underlying dynamics responsible for malaria transmission since 2007 at several sites with differing epidemiologies in southern and central Africa. These sites include Mutasa District in Zimbabwe (resurgent malaria after successful malaria control), Choma District in

southern Zambia (low transmission, approaching elimination), and Nchelenge District in northern Zambia (ineffective malaria control). Zambia has made significant strides toward malaria reduction in the last two decades, with commitments to malaria control in the 2000s leading to increased funding from a number of sources, including the Bill and Melinda Gates Foundation, USAID, and the US President's Malaria Initiative (PMI).[106] This funding led to large rollouts of LLINs and insecticides as part of a massive malaria control scale-up. The number of households with at least one bednet has risen from 37.8% in 2006 to 77% in 2015 and the number of households that had received IRS within the last 12 months rose from 9.5% in 2006 to 28.5% in 2015.[107] Prevalence of malaria was reduced between 2006 and 2012, with a decline from 22.1% to 14.9%, but prevalence rose again in 2015 to 19.4%.[107] The impact of national malaria control programs since the early 2000s has also been uneven across the 9 provinces. Southern Zambia has seen near elimination levels of parasite prevalence in 2015, with 0.6% in Southern Province (where Choma resides) and 2.4% in Lusaka Province.[107] However, the northern and eastern provinces remain heavily burdened – Luapula Province having the highest parasite prevalence at 32.5%, not much lower than the 32.9% prevalence reported in 2006[107]. Nchelenge District, which resides in Luapula Province, has been the focus of studies for this dissertation (Figure 2).

Nchelenge District

Nchelenge District in northern Zambia lies along Lake Mweru which constitutes the border with the Democratic Republic of the Congo. Nchelenge is characterized by year-

round holoendemic malaria transmission; malaria prevalence has remained amongst the highest in the country at 50% in all age groups.[108] Rates in this area stand in stark contrast to many other regions of Zambia, which have seen dramatic reductions in disease burden since scale-up of malaria interventions began in 2006.[109] The refractory nature of malaria burden in this area, despite widespread vector control, calls into question whether currently employed interventions are appropriate.

Heretofore, there has been no comprehensive examination of foraging behavior in Nchelenge. Entomological collections conducted in 2011-2012 provided a first-glance characterization of vectors in the area and suggested that *An. funestus* and *An. gambiae* were respectively the primary and secondary vectors, maintaining a year-round transmission cycle (M. Muleba, unpublished).[110,111] A more thorough set of collections were conducted through the wet and dry seasons of 2012 and the wet season of 2013.[112] These data showed that *An. funestus* was the primary anopheline in the area, composing >80% of the total collected in all seasons.[44] During the dry season, *An. funestus* rose to 99% of the total vector population.[112] *Anopheles gambiae* composed a smaller proportion, with a dramatic seasonal abundance cycle; falling from 9% in wet season 2012 to 0.6% during the dry season 2012 and then peaked at 18.9% in the wet season of 2013.[112] Both vectors showed high preference for human blood, and while both were found with parasites, *An. funestus* was estimated to be a much higher driver of transmission, especially during the dry season when *An. gambiae* was found at very low numbers.[112] Some spatial association was also noted, with *An. gambiae* at higher proportions in households nearer the lake and *An. funestus* especially abundant in households around the

more inland Kenani stream. *Anopheles leesoni* was also collected (at much smaller numbers), with no evidence that it serves as a vector species in this region. A more comprehensive entomological analysis, utilizing cross-sectional and longitudinal data from the larger ICEMR project between April 2012 and September 2014 in households throughout the district, confirmed these figures and established an overall entomological inoculation rate (EIR) of ~81 infectious bites per person per year (ib/person/year), with *An. funestus* contributing more than *An. gambiae* (respective EIRs: 71 ib/person/year and 7 ib/person/year).[113] Abundance of both species was again statistically significantly associated with season and spatial correlations were noted.[113]

Additional studies from 2013 to 2015 characterized insecticide resistance. While *An. gambiae* was resistant to pyrethroids and DDT, *An. funestus* was shown to be resistant to pyrethroids and carbamates (Coetzee, unpublished).[108,114] It has been observed that *An. gambiae* from Nchelenge District are homozygous for kdr at high rates (Muleba and Coetzee, unpublished) and it is presumed that metabolic resistance, including P450 regulation, is contributing to resistance as well. Continued use of all three classes of these compromised insecticides partially explains the stubbornly high transmission intensity in the face of ramped up vector control interventions from 2006-2011. DDT use was discontinued in 2010. After 2012, Zambian IRS programs dropped pyrethroids in favor of a combination of carbamates and organophosphates. Subsequently, they adopted only organophosphates in the form of pirimiphos-methyl, to which there is currently no resistance in the two primary vector species in Nchelenge.[115]

Entomological collections in Nchelenge prior to the studies described in this dissertation have utilized indoor collection methods, which bias toward primarily endophagic and/or endophilic mosquitoes. This represents a major blind spot in understanding why Nchelenge remains a high transmission area in the face of widespread malaria control. One major reason that current strategies might be ineffective is if vector populations exhibit non-canonical foraging behaviors that evade current indoor interventions (i.e. biting outdoors). Another potential contributor is the presence of so-called "secondary" vector species in the area, which may be primarily exophagic/-philic and would not be captured using current collection methods.

Additionally, population genomics and genetic study of *An. funestus* in Nchelenge, as in many places, is lacking. This has hindered efforts to study population dynamics in Nchelenge that may be important to transmission. Choi et al, while studying insecticide resistance in Nchelenge district, documented for the first time in 2014 the presence of *An. funestus* mitochondrial lineage II outside of Mozambique and Madagascar.[114] There has also been little to no study regarding the comparative distribution and bionomics of the two mitochondrial lineages since this report, though the relative abundance of the lineages was similar in a recent study where lineage II made up 14% of the population (compared to 20% in Choi et al 2014).[85,114,116]

Summary

This dissertation explores vector behavior and genetics relevant to transmission in Nchelenge District, Zambia, and throughout southern and central Africa. Baseline genomic information is lacking for *An. funestus*, one of the three major vectors of malaria in Zambia and many areas of sub-Saharan Africa. Improved understanding of the diversity and dynamics of natural *An. funestus* populations would be of great benefit to researchers and public health workers aiming for vector control. In that vein, 43 *An. funestus* mitochondrial genomes from the Democratic Republic of the Congo, Zambia, and Tanzania were sequenced and assembled. Population genetic analyses of these mitogenomes revealed high levels of diversity and some structure. These latent population dynamics could hinder vector control efforts if polymorphism, mutation, and migration enable evasion of strategies like insecticides and gene drive technologies.

An additional study was conducted in Nchelenge District – an established ICEMR field site in northern Zambia. This study aimed to understand the activity of anopheline species outdoors, and whether or not that activity could contribute to the stubbornly high levels of transmission found in the area. In this study, CDC light traps (CDC LTs) baited with artificial human odorant were equally as effective as CDC LTs set near either human congregation areas or animal pens, and the most important contributor to either abundance or diversity of anopheline species was the general geographic location of a trap – near the lake or inland. Most strikingly, an unexpected diversity of anopheline species was observed, and led to additional phylogenetic studies aiming to place

unknown specimens. This analysis revealed some anophelines known elsewhere as vectors, while others represent as-yet unrecognized species with an unknown capacity for malaria transmission. Altogether, these studies represent a step forward in understanding the bionomics and genetics of malaria vectors in a high transmission setting and contribute a baseline of genomic information for *An. funestus,* a major malaria vector across the African continent.

Specific Aims

The aims of this dissertation were to 1) establish an *An. funestus* population mitogenomic baseline and appropriate informative markers especially for Lineage I and Lineage II, 2) compare populations of anophelines caught outdoors in a high-transmission setting using highly scalable trapping and baiting schemes, and 3) to phylogenetically characterize populations of recognized and identify unrecognized anophelines species from outdoor foraging collections.

Figures

Figure 1: Map showing the relative probability of *An. funestus* being found as a gradient of color from green to red throughout sub-Saharan Africa, with red indicating high probability. Data is courtesy of the Malaria Atlas Project (MAP)[117], and ArcMap v10.6 was used to generate the figure.[118]

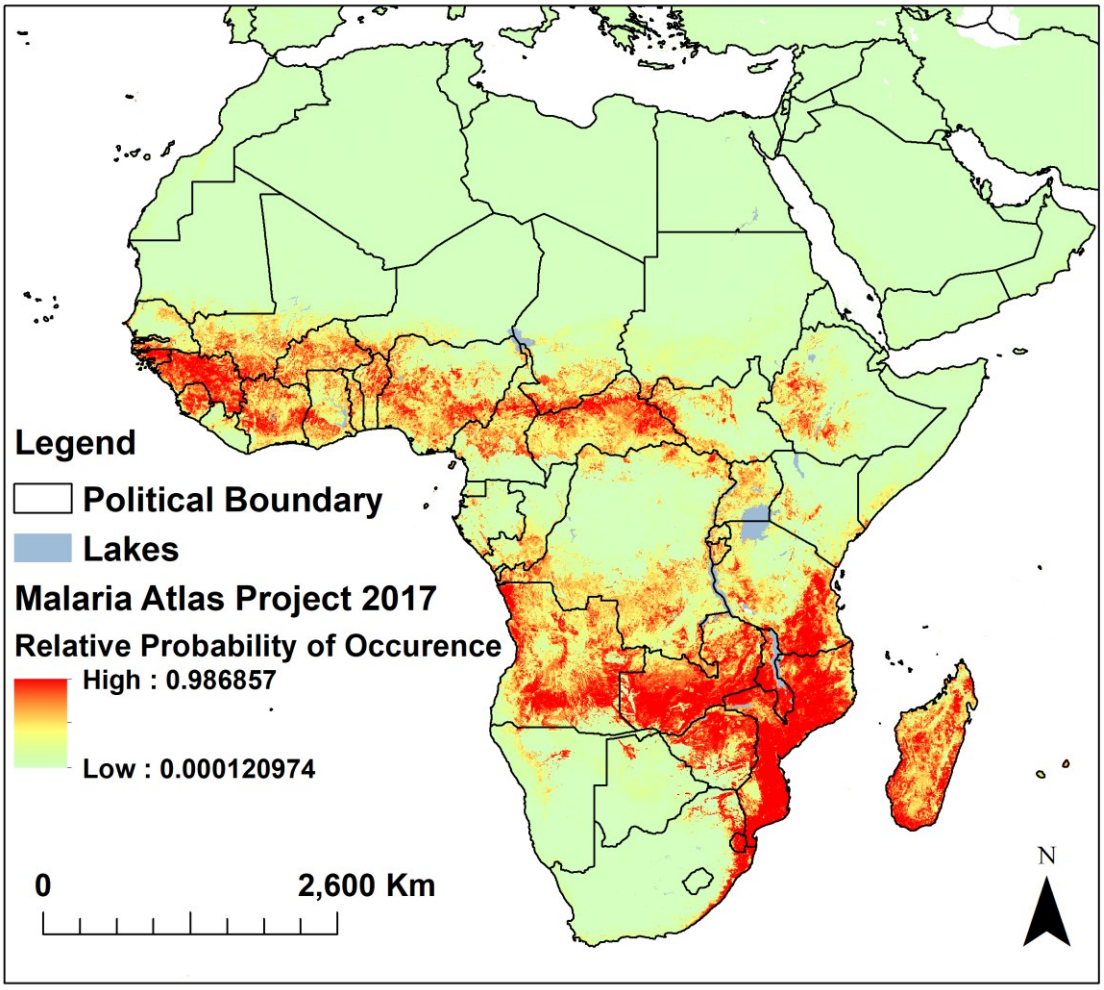
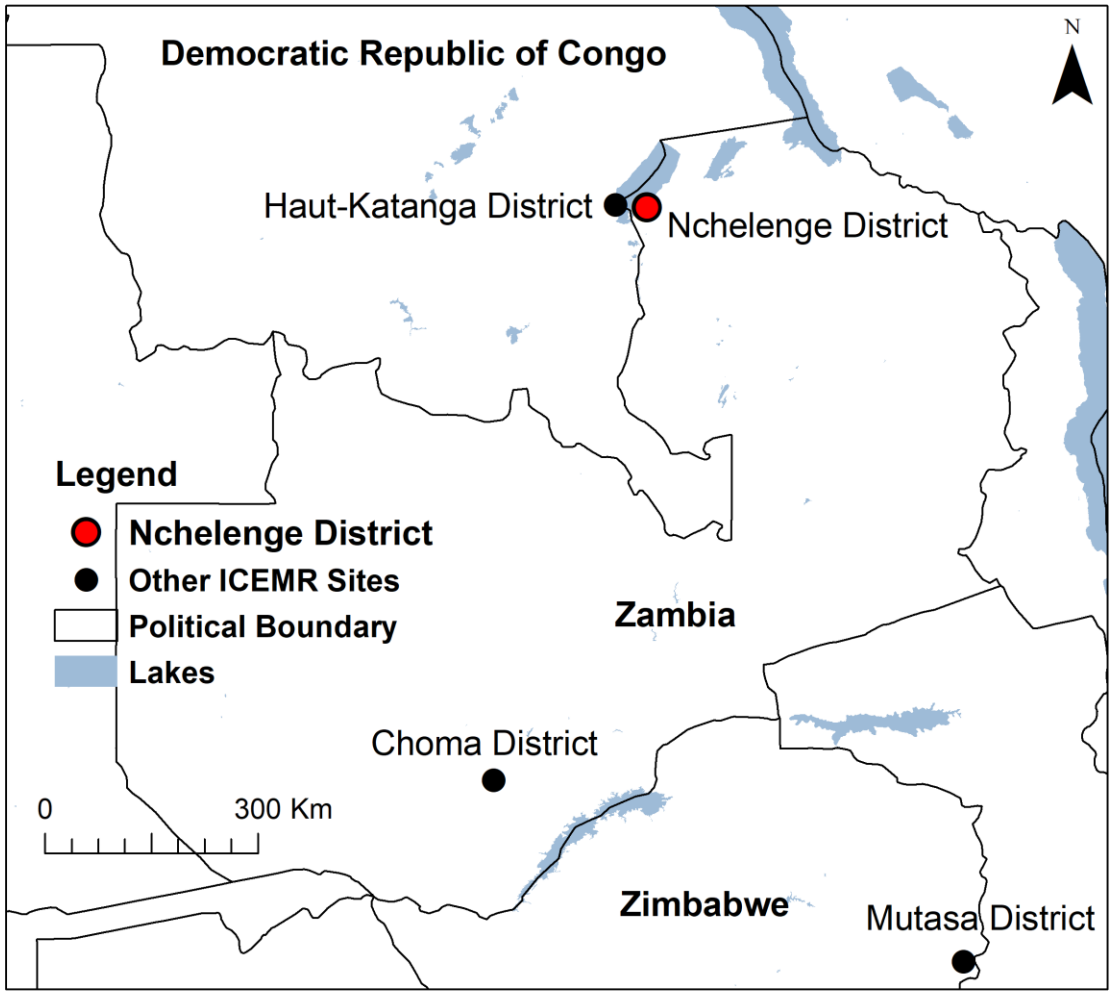


Figure 2: Map showing the geographic locations of Southern and Central Africa ICEMR study sites as circles within the Democratic Republic of Congo, Zambia, and Zimbabwe. Nchelenge District in northern Zambia is the focus of much of this dissertation and is highlighted in red. Figure was generated in ArcMap v10.6.[118]



Tables

Table 1: Table adapted from Ayala et al.[47] Summary of the species within the Funestus Group in terms of their geographical range and potential as vectors for malaria transmission.

Species	Distribution	Host preference	Vector role
An. funestus	continental	anthropophilic	major
An. funestus-like	local	unknown	unknown
An. aruni	local	unknown	unknown
An. confusus	regional	zoophilic	unknown
An. parensis	regional	unknown	minor
An. vaneedeni	local	unknown	unknown
An. longipalpis C	local	zoophilic	unknown
An. leesoni	continental	zoophilic	minor
An. longipalpis A	local	zoophilic	unknown
An. rivulorum	continental	zoophilic	minor
An. rivulorum-like	local	unknown	unknown
An. brucei	local	unknown	unknown
An. fuscivenosus	local	unknown	unknown

Chapter II

Complete *Anopheles funestus* mitogenomes reveal an ancient history of mitochondrial lineages and their distribution in southern and central Africa

The following chapter has been published:

Jones, C. M., Lee, Y., Kitchen, A., Collier, T., Pringle, J. C., Muleba, M., Irish, S., Stevenson, J. C., Coetzee, M., Cornel, A. J., Norris, D. E., Carpi, G. (2018). Complete Anopheles funestus mitogenomes reveal an ancient history of mitochondrial lineages and their distribution in southern and central Africa. Scientific Reports, 8(1), 9054.

Introduction

In 2016, there were approximately 216 million cases of malaria and approximately half a million deaths, most of which occurred in sub-Saharan Africa. [1] These data represent a drastically reduced incidence of malaria since 2000. [3] However, progress has plateaued in recent years and incidence and mortality have remained essentially the same in 2015 and 2016 [1]. The decline of malaria can be attributed to several causes, including a rise in coverage of vector control. [1,3] However, phenomena such as changing vector foraging and resting behaviors, [90,119] and the development of insecticide resistance, [38,120] have led to heterogeneity and stagnation in the success of malaria control worldwide. Population genetic and genomic methods, as a result of advances in sequencing strategies, are becoming useful tools for understanding and monitoring vector population diversity, [121–123] dispersal [124] and dynamics with the ultimate goal of informing malaria control. [39,40]

Anopheles funestus sensu stricto (hereafter "*An. funestus*") is a major malaria vector throughout sub-Saharan Africa and poses a significant threat to malaria control and elimination due to its high vectorial capacity, expansive distribution, and high rates of insecticide resistance. [125] While studies of the other major regional malaria vectors in the *An. gambiae sensu lato* species complex have been frequent over the past four decades, research on *An. funestus* has remained at a trickle, with very few investigations during recent decades. Notably, this dearth is largely due to the relative difficulty of rearing *An. funestus* in laboratories. With the advent of cytogenetic studies in the 1980's, as well as modern and more sophisticated molecular techniques, population studies of field-caught mosquitoes have become more common. [125] Additionally, the establishment of laboratory colonies of *An. funestus* within the last decade [73] has finally allowed for more complex genetic and genomic studies of this species. [125,126]

However, there are still only a limited number of genetic studies (using a variety of mitochondrial and nuclear markers) of *An. funestus* across its entire geographic range in sub-Saharan Africa. [60,66,77,84–86,114,127–129] These broad-scale studies largely agree that *An. funestus* populations can be split into major western and eastern groups. [84,85] Interestingly, there is compelling evidence for the hypothesis that the Great Rift Valley serves as an important barrier to gene flow between these populations, [77,84–86,127] which has similarly been documented for *An. gambiae.* [82] Additionally, Michel et al. [85] reported two mitochondrial lineages (I and II) of *An. funestus* based on partial mitochondrial gene sequences (COI and ND5), which are not reflected in parallel nuclear microsatellite analyses. While mitochondrial lineage I makes up the majority of samples found in

countries throughout sub-Saharan Africa, individuals belonging to lineage II have only been reported in the southeastern range of this species in Mozambique, Madagascar, northeastern Tanzania, and northern Zambia. [85,114,128,130]

To date, fine-scale population genetic studies of *An. funestus* have focused on only limited regions in western and eastern Africa, while *An. funestus* populations in central and southern Africa, such as Zambia and the Democratic Republic of the Congo, remain greatly understudied. More specifically, investigations of *An. funestus* maternal lineages have also been limited within southern Africa, and have solely relied on partial mitochondrial gene sequences (COI and ND5). [85,114,128,129] A key challenge to such studies is the unfinished nature of the published mitochondrial reference genome (GenBank: DQ146364.1), which is incomplete and lacking ~27.5% of the genome sequence, mainly in coding regions. [74] This incomplete reference represents a hurdle to future research in the field, as mitogenomes can serve as an important source of markers for population genetic studies, and also provide insight into evolutionary relationships within the *An. funestus* species complex. Further, the absence of large-scale mitochondrial genome (hereafter "mitogenome") and nuclear genome data of wild-caught *An. funestus* makes it difficult to catalogue genetic variation in natural populations and determine population structure and dispersal rates.

Complete mitogenomes are particularly useful for reconstructing phylogenies and inferring population history due to haploid maternal inheritance, [131,132] the rare occurrence of recombination, [133] and a higher mutation rate than the nuclear genome. [134]

Mitochondrial sequence polymorphisms may be particularly useful to study sex-biased dispersal known to occur in some anopheline mosquitoes (including *An. funestus*).[135,136] Mitogenomes have smaller effective population sizes than autosomal DNA, enabling better discrimination between populations due to the rapid effects of genetic drift. Additionally, mitochondria contain multiple genomic copies making mitogenomes amenable targets for sequencing at high coverage. Understanding historical gene flow and genetic structure via analysis of mitogenomes is a step toward revealing contemporary vector population dynamics and accurate discrimination between lineages and sub-populations. In turn, this information will contribute to an enhanced appreciation of malaria transmission dynamics, especially if vector genetic diversity reflects differences in biology, behavior, permissiveness to *Plasmodium* parasites, [137] or insecticide susceptibility, [39,138] all of which have consequences for malarial disease management, surveillance, and control measures.

To investigate the degree of genetic diversity in *An. funestus* across a critically understudied geographic region, and to further examine the evolutionary history and distribution of mitochondrial lineages, we performed shotgun mitogenome sequencing of *An. funestus* samples from northern Zambia, southeastern Democratic Republic of the Congo (hereafter "DRC"), and southern Tanzania. We first generated a new *An. funestus* mitochondrial reference genome that filled gaps in the existing incomplete reference (GenBank: DQ146364.1)[74] and then assembled and annotated 43 *An. funestus* mitogenome sequences from these regional collections. Bayesian phylogenetic and classical population genetic analyses were performed to characterize *An. funestus*

mitochondrial lineages, document their distribution in southern and central Africa, and assess their demographic history. Notably, the data generated in this study are part of an initial collection effort to build a digital repository of genomic data from field-caught *An. funestus* across southern and central Africa.

Methods

Study Sites and Sample Selection

An. funestus samples were chosen to geographically represent this species in southern and central Africa where we are actively conducting research on malaria transmission. In total, 43 *An. funestus* samples were selected for further sequencing (Table 1). Nchelenge District of northern Zambia was chosen as a long-standing site for malaria research in within the framework of "The Southern and Central Africa International Centers for Excellence in Malaria Research (ICEMR)", which is a research program designed to understand drivers of persistently high malaria transmission. Nchelenge District abuts the Democratic Republic of the Congo, with the border bisecting Lake Mweru. The sampling area lies 807 meters above sea level, with a marsh ecotype and three seasons: a single rainy season from November to May, a cool dry season from May to August, and a hot dry season from August to November. Malaria transmission occurs at high rates year-round, despite widespread use of long-lasting insecticide treated nets (LLINs) and indoor residual spraying (IRS). [108,111] Although both *An. funestus* and *An. gambiae* are present in this district, *An. funestus* is the primary vector in Nchelenge, with the population peaking during the dry season. [112]

Kapolowe is a town in Haut-Katanga Province, in southeastern Democratic Republic of the Congo. It is on the edge of Lake Tshangalele, an artificial lake created by the dam at nearby Mwadingusha. Kapolowe is at an elevation of 1,177 meters above sea level and has a rainy season lasting from November to April, with a dry season between May and October. Malaria prevalence is high despite widespread use of LLINs [139] and no IRS has been conducted in Kapolowe. *Anopheles gambiae*, *An. funestus*, and *An. coustani* group mosquitoes are the most commonly collected anopheline mosquitoes in Kapolowe. [140]

Lupiro is located within Kilombero Valley in southern Tanzania, a zone of intense perennial malaria transmission. [141–145] It is at an elevation of 300 meters above sea level and has a rainy season lasting from November to May. Epidemiological studies in this valley have revealed that malaria transmission intensities are very high, with 100-1000s of infective bites per person per annum. [141,145–147] A nation-wide LLIN distribution program is currently underway in Tanzania, through which net coverage has substantially increased in Kilombero Valley. [148] However, reduction in malaria transmission was not as great as anticipated based on the high LLIN coverage (75%) achieved. [148]

DNA extraction and sequencing

Field-caught mosquitoes were morphologically identified to species using standard keys at the time of collection. [149] Each identified mosquito was placed individually into a labelled 0.6 mL microcentrifuge tube containing silica gel desiccant and cotton wool and

stored either at room temperature or frozen at -20°C until laboratory processing. Genomic DNA extractions were performed on the head and thorax for each individual mosquito as previously described. [150] Quantitation of the genomic DNA was performed using a Qubit 2.0 Fluorometer (Life Technologies, Grand Island, NY) and genomic libraries were prepared as described, using an input of 10 ng of genomic DNA. [151] Indexed libraries were pooled and sequenced in a single lane on an Illumina HiSeq4000 to generate 150 bp paired-end reads. Sequencing was performed at the University of California-Davis DNA Technologies Core. Demultiplexed Illumina raw reads obtained from DNA Technologies Core were trimmed using Trimmomatic version 0.36. [152] We used the typical trimming parameters "ILLUMINACLIP:'{input.adapters_file}':2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36", which removes adapters, trims low quality or N bases below quality 3, scans the read with a 4-base sliding window cutting when the average quality per base drops below 15, and dropped reads below 36 bp long.

Mitochondrial genome assembly and variant detection

The incomplete *An. funestus* mitochondrial reference (GenBank: DQ146364.1) [74] was used as a 'seed' sequence to generate a new and complete mitogenome reference with MITObim v1.8 with 10 iterations, default parameters, and trimmed Illumina reads from sample AF13ICNC14-106. [153] Subsequently, raw Illumina sequence reads for each sample were aligned to the newly generated *An. funestus* mitogenome reference AF13ICNC14-106 (Figure S1), using BWA alignment tool v0.7.7 (bwa-mem, default parameters). [154,155] Duplicate sequences were identified and excluded from downstream

analysis using Picard Suite v1.117 MarkDuplicates. [156] Aligned reads were realigned around indels (insertions and deletions) using GATK v3.7 RealignerTargetCreator and IndelRealigner. Variants with respect to AF13ICNC14-106 were identified with GATK HaplotypeCaller (ploidy set to 1). [157] Indels and single nucleotide polymorphisms (SNPs) with signals of low mapping or genotyping quality were excluded with GATK VariantFiltration, using the following filters recommended by GATK: quality by depth (QD < 2.0), Fisher strand bias (FS > 200.0), mapping quality (MQ < 40.0), the Mann-Whitney rank sum test (ReadPosRankSum <− 20.0). [158] To create the consensus mitogenome sequence for each sample from the variant files, the GATK tool FastaAlternateReferenceMaker was used. The mitogenome coverage for each sample was calculated using the software GATK v3.7 (DepthOfCoverage with parameters mmq > 20 and mbq > 20). [157]

Phylogenetic analysis and divergence time estimation

The 43 newly generated *An. funestus* mitogenomes were aligned using MUSCLE with and without full mitogenomes from *An. gambiae* (GenBank: L20934.1) and *An. minimus* (GenBank: KT895423.1) as outgroups. [159–161] It is important to account for recombination when reconstructing evolutionary histories because homologous recombination has a profound impact on evolutionary trajectories and therefore the interpretation of inferred phylogenies. We used the 3SEQ software which implements a fast non-parametric recombination detection algorithm to infer recombinant tracts along the mitogenomes to rule out the possibility of recombination in our N=43 *An. funestus* mitogenome

alignment. [162] Maximum likelihood trees of the 43 samples were generated in SeaView v4, using PhyML and GTR substitution model, and default parameters with 1000 bootstrap replicates. [163] BEAST2 v2.4.5 was used to conduct phylogenetic analyses as well as generate estimates of divergence times and population size, and determine demographic history of southern and central African *An. funestus.* [164] Analyses were performed using a substitution rate of 1.2×10^{-8} mutations per site per year, following estimates from Brower. [165] Markov chains were run for 100 million generations or until convergence, with 10 million generations of each run discarded as burn in, and chains sampled every 10000 generations. Both HKY and GTR substitution models were used in combination with gamma site-specific rate variation (G) and a proportion of invariant sites (I) parameters, strict and relaxed log normal molecular clocks, as well as constant and Bayesian skyline population models. To compare models, the Path Sampler application from BEAST2 v2.4.5 was used to generate marginal likelihood estimates and the model with the highest estimate was used for demographic and population history inference. [164] Tracer v1.6 [166] was used to inspect convergence and confirm effective sample sizes were greater than 200 for parameters of interest. Tracer v1.6 was also used to generate Bayesian skyline plots. Because our evolutionary rate was in years, effective population size was confounded with generation time; we used a generation time of 3/52[167] to convert estimates of population diversity to N_e in our coalescent analyses. LogCombiner was used to resample 10000 trees from BEAST2 analysis and then TreeAnnotator was used to generate Maximum Clade Credibility (MCC) trees. [164] The multiple alignment of the 43 *An. funestus* mitogenomes was further analyzed using TCS

statistical methods as implemented in PopArt v1.7 to produce a mitochondrial haplotype network. [168,169]

To investigate how the genetic diversity of *An. funestus* samples sequenced in this study compared to previously known *An. funestus* diversity we extracted the NADH dehydrogenase, subunit 5 (ND5) sequences from our 43 mitogenome sequences and aligned them using MUSCLE to 400 published partial ND5 sequences (834 bp). [85] To further explore the diversity of our samples in the context of this large pan-African dataset, we used PhyML to generate a maximum likelihood tree of the ND5 alignment using the GTR nucleotide substitution model and 1000 bootstrap replicates. [170] tcsBU was used to visualize the TCS haplotype network generated by TCS v1.21 for partial ND5 sequences. [168,171]

Phylogenetic Analysis of Geographic Structure

To determine the extent of geographic structure in our *An. funestus* populations, we estimated the strength of association between phylogenetic relationships and sampling locations for the complete *An. funestus* mitogenome sequences using the software package BaTS.[172] BaTS generates a parsimony score (PS)[173] and association index (AI)[174] to assess the extent of geographical association with phylogenetic structure across the entire tree, as well as maximum monophyletic clade size statistics (MC)[172] to determine the association for particular sampling locations.

Estimation of demographic history

DnaSP v5 was used to generate general diversity statistics, conduct neutrality tests, and examine demography. [175] These statistics test the null hypothesis that populations are: neutral, of constant size, are in panmixia, and have no recombination. Arlequin v3.5 was used to calculate the mismatch distributions to test signal for population spatial expansion. [176] The raggedness index and SSD were used to evaluate how well the sample conforms to the null model of either demographic or spatial expansion. Mantel tests were used to evaluate for correlation between genetic distance and physical distance using the APE package in R v3.3.0. [177,178]

Annotation and data availability

Protein coding genes were identified and annotated manually by sequence similarity to the previous reference genome (GenBank: DQ146364.1) as well as the orthologous sequences of other anopheline species. [74,160,161] Transfer RNA (tRNAs) were identified by their putative secondary structures using tRNAscan-SE. [179] The ribosomal RNA genes (rrnL and rrnS) were identified by sequence similarity to the available homologous sequences using MITOS. [180]

The 43 newly generated *An. funestus* mitogenome sequences are available in the GenBank Database under the following accession numbers: MG742157-MG742199.

Results

An. funestus mitogenomes

A total of 43 female *An. funestus* from three regions across southern and central Africa (Zambia N=28; Tanzania N=10; DRC=6) were subjected to whole genome shotgun sequencing (Table 1). From these data, the first complete *An. funestus* reference mitochondrial genome (15,353 bp in length) was generated and the remaining 42 mitogenomes were assembled. On average, mitogenome coverage was 350x, ranging from 32x to 477x across the 43 samples (Figure S2). The nucleotide composition of the *An. funestus* mitogenome reference was heavily AT-skewed (average AT content = 78.2%), as is typical for the mitogenomes of many arthropod and anopheline taxa. [181] The mean nucleotide diversity (π) in the 43 *An. funestus* mitogenomes was 0.00625 (SD = 0.00054, Table 2), which is higher than the nucleotide diversity previously estimated using partial sequences of mitochondrial genes (π = 0.0042, SD = 0.007), [83,85] and that of other major anopheline malaria vectors, *An. gambiae* (π = 0.0038) and *An. arabiensis* (π = 0.0046). [182] The multiple alignment of the 43 mitogenomes revealed a total of 567 polymorphic sites. These variable sites defined a total of 41 mitogenome haplotypes from the 43 sampled individuals (mean haplotype diversity = 0.998, SD: = 0.006, Table 2), with only AF13ICNC14-155:AF15R31C10-A001 and AF15R35C07-B001:AF15R35C07-F002 sample pairs representing the same mitogenome haplotypes.

Phylogenetic analysis and divergence time estimation

To investigate the phylogenetic relationships of the 43 *An. funestus* mitogenomes sequenced in this study (Figure 1A), we constructed a maximum likelihood tree and identified two distinct lineages, herein defined as lineages I and II (Figure S3), which corresponded to clades 1 and 2 as described in Michel et al. (2005). [85] Our Bayesian coalescent analysis, implemented in BEAST2, produced a tree with concordant topology to the ML tree (Figure 1B). The most frequently sampled lineage in our study, lineage I, included mitogenomes from all sampled sites in the three countries. Lineage II, on the other hand, was absent from our DRC collection, which may be due to the small sample size.

We found 160 nucleotide differences between haplotypes in lineage I and II on average, with 47 fixed differences between the two mitochondrial lineages. Lineage II contained longer branch-lengths between samples than samples within lineage I, an observation which is also reflected in the diversity statistics (Table 2). The most recent common ancestor between the two lineages was estimated to have existed 504,016 years ago (95% Highest Posterior Density (HPD): 426,065 – 593,665 ya). This estimate is similar to that of the second-highest scoring model (not shown) from BEAST2 model selection. To validate the accuracy of divergence time estimations, we computed a separate BEAST2

analysis that included two outgroups, *An. gambiae* and *An. minimus*, in addition to our samples (see Figure S4). This analysis gave an approximate date of divergence between the two lineages of 528,336 years (95% HPD: 439,666 – 626,020). The divergence time of *An. gambiae* from all other anophelines in the analysis was 9.56 million years ago (95% HPD: 7.44 – 12.03 Mya), while *An. minimus* appears to have split from *An. funestus* approximately 5.36 million years ago (95% HPD: 4.06 – 6.68 Mya).

Within the two main lineages, we found several well-supported clades (Figure 1B). In lineage I, there appeared to be two well-defined clades (clusters A and B in Figure 1B, Figure 2), which diverged approximately 158,807 years ago (95% HPD: 128,864 – 192,908 ya). Cluster B is only found in our Zambia collections, while members of cluster A were found in all locations. In lineage II, there were also two well-defined smaller clades with an estimated divergence time of 276,074 years ago (95% HPD: 236,527 – 322,922 ya; Figure 1B). The smaller clade in lineage II, containing four individuals, lacked a SNP used as a diagnostic for lineage II in a recently developed high-throughput TaqMan assay. [183] The single-SNP-based TaqMan assay targets a SNP at position 405 in the sequence of COI, where two states are considered: a T or a C. [183] This definition of lineages misidentifies ~30% of our lineage II samples as lineage I.

We constructed a network to assess the genealogical relationships between the haplotypes and to gain insight into the population level phenomena that might have contributed to the maintenance of two mitochondrial lineages in *An. funestus* (Figure 2). Lineage I and II are very distinct, separated by ≥137 mutational steps. On average, the lineages differ

by ~160 nucleotides. Lineage II samples are separated by 77 nucleotide differences from each other on average, compared to 36 within lineage I (Table 2). Clusters A and B within lineage I are separated by ≥ 42 mutational steps. They correspond to well-supported inner clades (also clusters A and B) within lineage I in our phylogenetic analysis (Figure 1B).

Population demography and structure

We used several population genetic statistics to test for selection or historic changes in *An. funestus* population size (Table 2). For the full dataset (N=43), Tajima's D and Fu's Fs were not statistically significant, and Fu and Li's D was negative, but only moderately significant ($0.10 > p > 0.05$), suggesting population expansion. Neutrality analyses were also conducted for lineage I and lineage II samples separately. While lineage II did not produce statistically significant results for any neutrality tests, lineage I was either moderately or highly statistically significant for several statistics, again suggestive of population expansion (Table 2). Mismatch analysis (Table 2) was indicative of demographic expansion for the total sampled population, but not for lineage I alone. The Bayesian model selection suggested a complex demographic history, and when we analyzed the data under the Bayesian Skyline model we found a signature for population expansion in the total sampled population occurring approximately 76,400 years ago (Figure S5). Based on mismatch analysis, both lineage I and lineage II are consistent with models of spatial expansion, and lineage I is consistent with an increase in effective population size (demographic expansion).

A previous study based on partial mitochondrial gene sequences found no population structure within *An. funestus*.[85] Similarly, our analysis identified no clear and readily apparent geographic structure in the phylogeny of the *An. funestus* mitogenomes (Figure 1B). To more rigorously examine the strength of association between phylogenetic relationships and sampling locations, we used several statistical tests implemented in the BaTS package. This analysis revealed evidence for phylogenetic clustering (by country) using both the association index ($p = 0.02$) and the parsimony score ($p = 0.01$) (Table 3). The maximum clade size ($p = 0.02$) was statistically significant for Zambian sequences. This suggests that the samples from Zambia were not as interspersed with samples from Tanzania or the DRC as one would expect if geography and phylogeny were randomly associated.

Plots of geographic distance relative to nucleotide identity are shown in Figure S6 for the total sampled population and for each lineage. Mantel tests with 1000 permutations were conducted to determine whether there was a statistically significant relationship between genetic and geographic distance in these groups. [177,184] Both lineage I ($p = 0.029$*) and lineage II ($p=0.001$**) had statistically significantly related pairwise nucleotide identity and geographic distance matrices.

Phylogenetic analysis of partial mitochondrial genes

To examine how the potential ancient population structure identified from our samples relates to the larger context of known *An. funestus* diversity, we constructed a maximum likelihood tree and haplotype network from a large data set including published partial mitochondrial ND5 gene sequences available from GenBank [85] and the derived corresponding partial gene sequences from our 43 mitogenomes. The topology of the ML tree (Figure S7) as well as the haplotype network (Figure S8) again revealed a clear split between lineage I and lineage II samples. The haplotype network revealed a single, primary haplotype in lineage I containing a large number of samples from across Africa. A number of haplotypes were shared between Nigeria, Mali, and Kenya, which was reflected in the maximum likelihood tree (Figure S7). There was a large clade basal to the remainder of lineage I composed of mosquitoes from Kenya, Malawi, and Nigeria. None of our samples fell within this basal clade. Within lineage I as a whole, there was no obvious correlation of our samples with those from any other region in Africa, consistent with our results based on whole mitogenome sequences (Figure 1B). Samples from Mozambique fell basal to the rest of lineage II samples and tended to group apart from samples from Madagascar. Our lineage II samples fell into both groups, though samples from Madagascar appear to be more isolated within lineage II.

Discussion

To our knowledge, this is the first study to report the complete mitochondrial genome of *An. funestus* and to use complete mitogenomes to assess genetic diversity in southern and

central Africa. Our data revealed higher levels of genetic diversity than previously reported using single locus markers alone. Both the Bayesian and ML trees (Figures 1B, S2) supported the co-existence of two previously-described clades, herein defined as lineage I and lineage II, in Nchelenge District, northern Zambia, [128] as well as in southern Tanzania, indicating that these lineages are more widely distributed than previously appreciated. [85,128,130] This also represents the first study to examine the distribution of lineages in southeastern DRC and extends the known distribution of lineage II in Tanzania. We have described well-supported sub-structuring within the two lineages, which may reflect much higher diversity within *An. funestus s.s.* than has been reported to date. Notably, our data have been shown in the context of greater African diversity, using partial ND5 sequences in a haplotype network (Figure S8). This network showed that much of *Anopheles funestus* ND5 diversity was shared across distant sites, with limited clustering by region. However, our phylogenetic clustering analysis of full mitogenomes supported the inference of geographic structure in our sample. The differing conclusions from the two datasets may be the product of either decreased homoplasy and increased phylogenetic signal of full mitogenome data, or it may be the product of a small sample of mitogenomes.

The Bayesian coalescent analyses of the complete *An. funestus* mitogenomes provided an estimate of the divergence times for the two mitochondrial lineages and of the clusters within lineages. Our findings were consistent with these lineages having common ancestry dating back 500,000 years, which suggests that they have evolved independently since the Pleistocene (which extended from approximately 2.58 million to 12 thousand

years ago). Our divergence estimates fell on the low end of estimates from previous studies,[76,185–188] and specifically, our estimate of divergence between the two lineages is younger than that reported by Michel and colleagues, who used the same mutation rate (1.1 – 1.2% per million years) to generate an estimate of ~850,000 years. [85]

Our *An. funestus* samples harbored a genomic signature of historic population expansion for the total population as well as for lineage I, though not for lineage II (Table 2). A Bayesian Skyline reconstruction (Figure S5) indicated an expansion event in the total ancestral population (3.8 to 36 million in effective population size, $N_{e,}$) began approximately 80,000 years ago (Figure S5). Although the overall population did not reveal a signature of sudden spatial expansion (Table 2, Figure S6), mismatch analysis (Table 2) was consistent with spatial expansion for each lineage independently. Additionally, there was a statistically significant relationship between genetic and geographic distance for both lineages independently. However, this relationship became statistically insignificant when Tanzanian samples were removed from the analysis. Thus, these data suggested that a genetic barrier exists between our *An. funestus* samples, perhaps either the large physical distances between sampling sites or the Great Rift Valley, which separates our samples from Tanzania and Zambia/DRC. This latter possibility would be consistent with data from *An. funestus* and other related taxa across their range in sub-Saharan Africa, though our small sample size precludes eliminating the influence of extreme sampling distances. Importantly, both our identity-by-distance and Bayesian analysis of phylogenetic clustering by geography were indicative of statistical support for population structure. However, it was unclear if the weak population structure

identified here is associated with the maintenance of two divergent mitogenome lineages in structured *An. funestus* populations, or whether historical population sizes were sufficiently large for a panmictic *An. funestus* population to maintain two maternal lineages.

A TaqMan assay based on COI and developed for differentiation of lineage I from lineage II [183] based on a single SNP, failed to discriminate these lineages amongst our 43 samples. Four individuals that phylogenetically belong to lineage II (N=14) share a nucleotide polymorphism (a 'T') with lineage I instead of the diagnostic SNP used in the TaqMan assay to define lineage II (a 'C'). This finding reinforced the importance of complete mitochondrial sequences for accurate characterization of *An. funestus* diversity and/or revision of the assay to accurately reflect the new mitogenome data and diversity within *An. funestus*. We found 47 mitogenome-wide fixed SNP differences between the two lineages, thus a more specific marker set is required for accurate lineage discrimination and would benefit future studies that aim to describe *An. funestus* lineage composition, distribution and biology in sub-Saharan Africa. A maximum likelihood tree using the partial ND5 gene of our samples along with those from Michel et al. [85] (Figure S7) revealed a highly diverse sequence landscape for *An. funestus*, with no clear geographic clustering of our samples. Taken as a whole, our data indicate that caution must be taken when using single mitochondrial genes for intra-species and population studies, due to the highly variant mitogenome of *An. funestus.*

Conclusions

We have illustrated that *An. funestus* has a complex evolutionary history with previously undocumented levels of diversity in southern and central Africa. The diversity is ancient and geographically occurs throughout the region. We speculated that the two lineages split due to habitat partitioning in a changing African landscape during the Pleistocene, and then lineages underwent spatial expansion with consequent independent diversification. More recently the *An. funestus* population as a whole (predominately composed of lineage I) experienced a demographic expansion. At this time, data suggest that the lineages are at least partially sympatric. Preliminary analyses have indicated that lineages I and II may differ in habitat, insecticide resistance, and/or foraging preferences (unpublished observations); ongoing work is exploring these trends. Such a difference in behavior may have important implications for vector control. Though it is tempting to theorize that lineage I and II may represent reproductively isolated populations because of the strong separation of clades in mitogenome comparisons, these results will have to be interpreted within the context of future nuclear genomics and hybridization experiments between the two lineages. Given our findings, further investigations on whether the *An. funestus* mitochondrial lineages represent biologically meaningful populations are warranted.

Figures

Figure 1: Map and phylogenetic relationships of 43 *An. funestus* mitogenomes. A: Map indicating the collection sites for 43 *An. funestus* samples, created using ArcGIS v10.5.1 (www.esri.org). B: Bayesian maximum clade credibility phylogeny of complete mitogenomes from the 43 *An. funestus* samples of the best fitting model (GTR +G +I, Bayesian skyline plot, and a relaxed molecular clock) inferred using BEAST2. Samples are colored by geographic origin: blue indicates Zambia (N = 28), orange indicates Tanzania (N = 10), green indicates DRC (N = 5). Divergence dates (median estimates and 95 % HPD) are given in parenthesis for major nodes. Posterior probabilities > 0.5 are indicated at each node. The timescale is indicated below the tree and is in years before present.

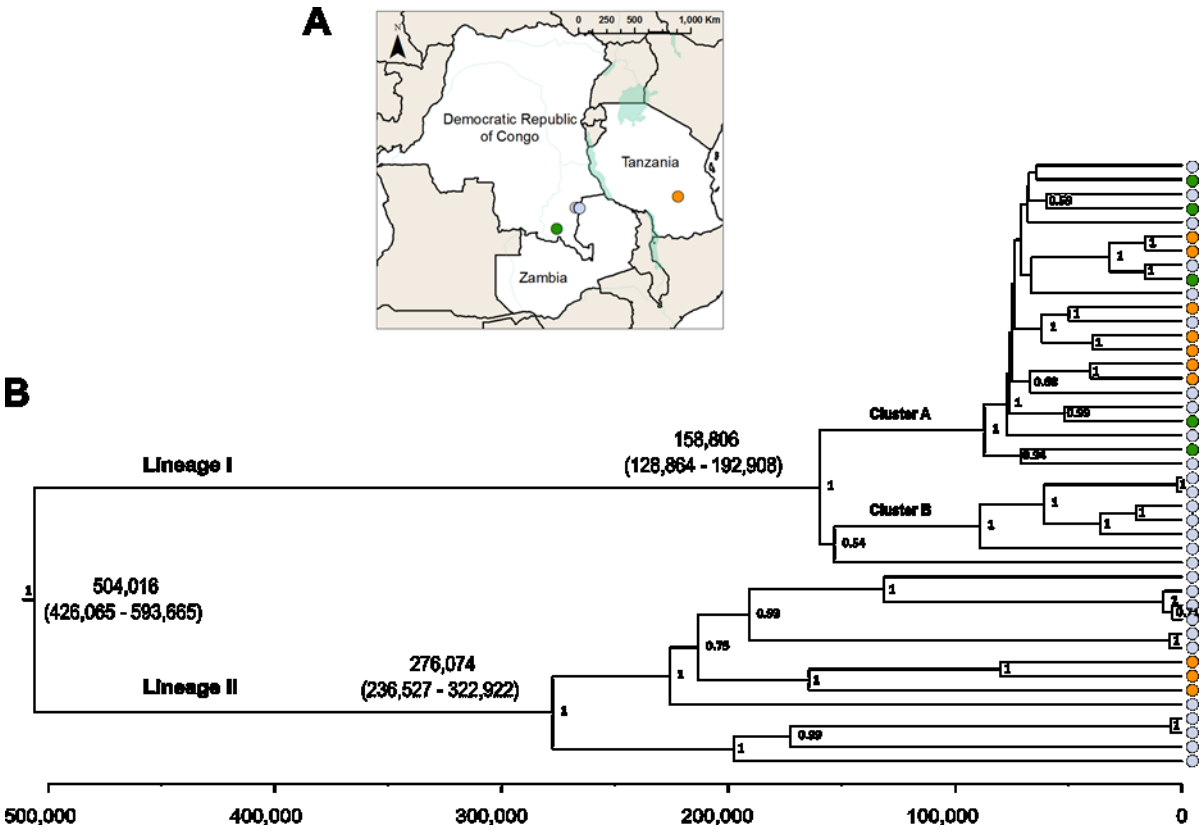


Figure 2: Haplotype network of 43 *An. funestus* mitogenomes. In this TCS network, each node indicates a haplotype, with nodes colored according to origin. The number of mutational steps between nodes are indicated in parentheses beside the line connecting one node to another. One group of samples (all lineage II) did not connect to the main cluster within 95 mutational steps (over a 95% confidence limit for connectivity): these are shown in the box in the lower right. There are two distinct groups within the main cluster (lineage I): one more highly clustered (cluster A), and another with fewer, more-distant nodes (cluster B). Cluster A and B in lineage I are separated by ≥ 42 mutations. The size of each node indicates the number of samples sharing a specific haplotype.

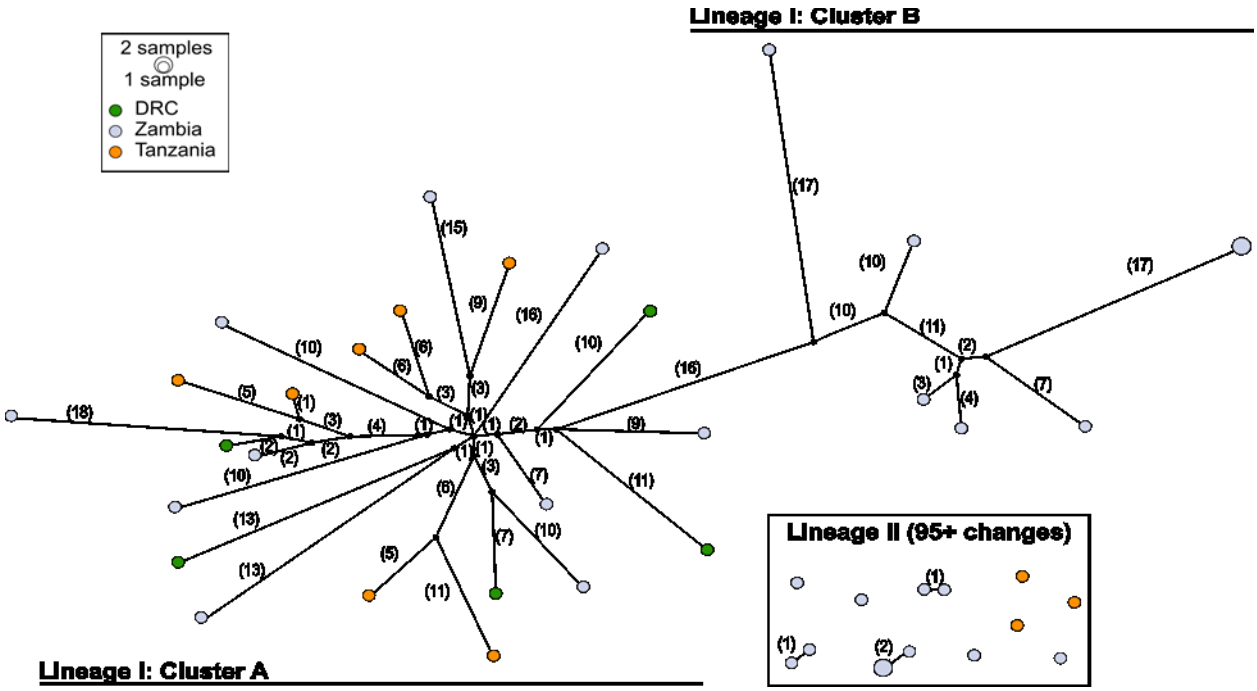


Figure S1: New complete mitochondrial genome reference of *Anopheles funestus* generated with MITObim v1.8,[153] using the incomplete mitogenome (GenBank: DQ146364.1)[74] as a 'seed' sequence. Coding regions are indicated and colored according to the sequence product (green for protein-coding, red for rRNA, and blue for tRNA). Reads from Zambia sample AF13ICNC14-106 (GenBank: MG742193.1).

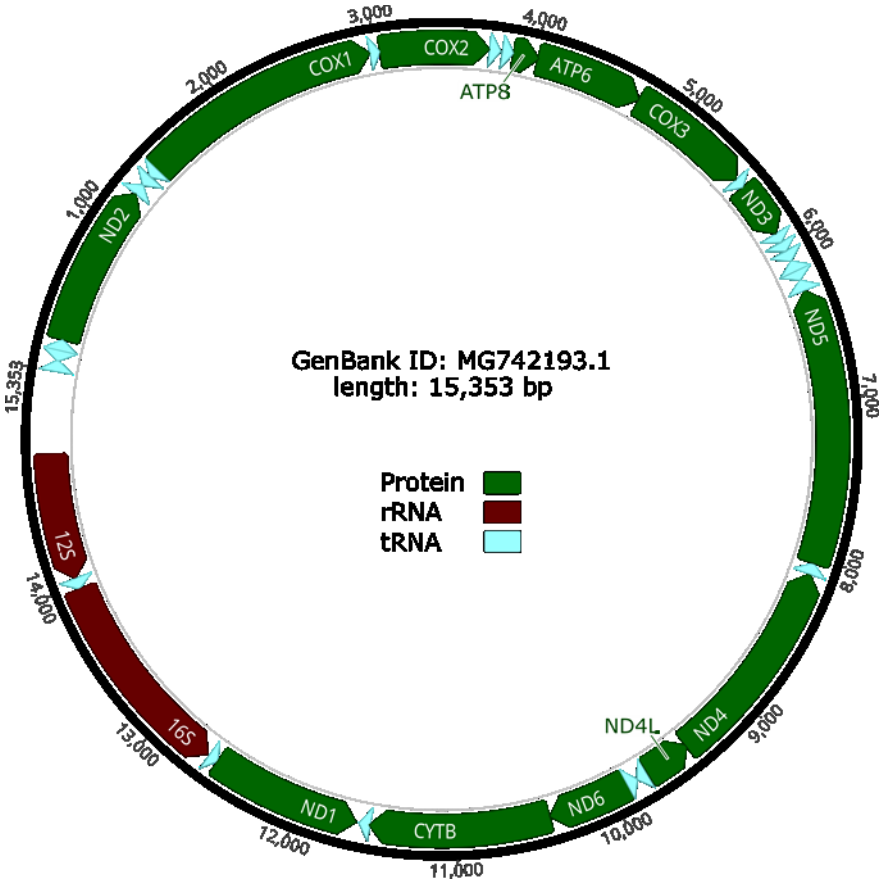


Figure S2: Average coverage along the mitochondrial genome for 43 *An. funestus* samples. The coverage depth is defined as the total number of sequenced bases which map to each nucleotide in the mitochondrial reference genome (AF13ICNC14-106) after removal of potential PCR duplicates and aligned reads with mapping quality below 20.

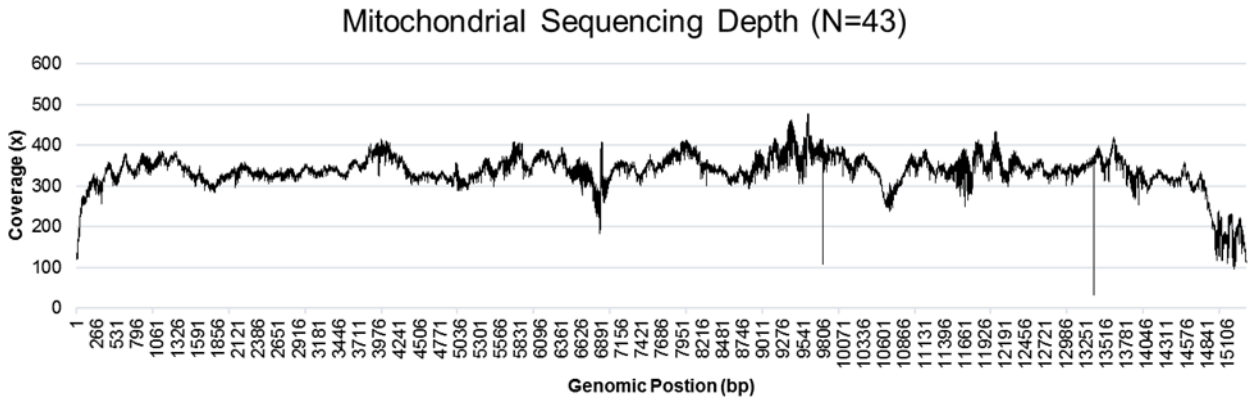


Figure S3: Maximum Likelihood tree of 43 *An. funestus* whole mitochondrial genomes. Maximum Likelihood tree of 43 newly sequenced *An. funestus* mitochondrial genomes using PhyML in SeaView v4 with GTR model and 1000 bootstrap replicates. Bootstrap support >0.5 shown next to appropriate nodes. Samples are colored by geographic origin, according to legend in the figure. Lineages are indicated.

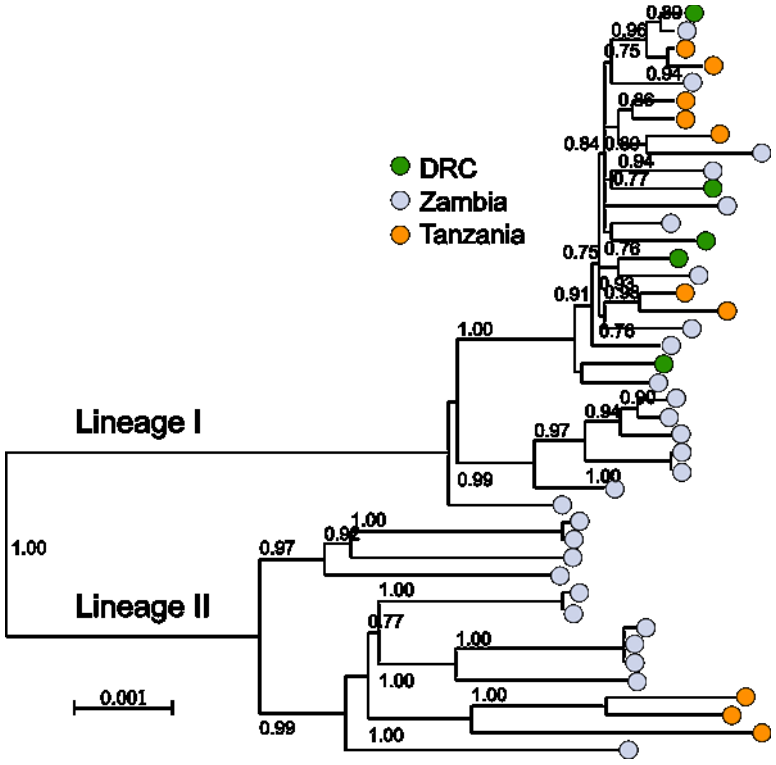


Figure S4: Bayesian tree with *An. funestus* samples and outgroups. Bayesian maximum clade credibility phylogeny from BEAST2 of complete mitochondrial genomes from the 43 *An. funestus* samples as well as 2 outgroups: *An. gambiae* and *An. minimus*. The model used GTR +G +I, constant population size, and a relaxed molecular clock. Samples are color-coded by geographic origin, according to the legend in the figure. Divergence dates (median estimates and 95% HPD) are given in parenthesis for major nodes. The timescale is indicated below the tree and is in year before present.

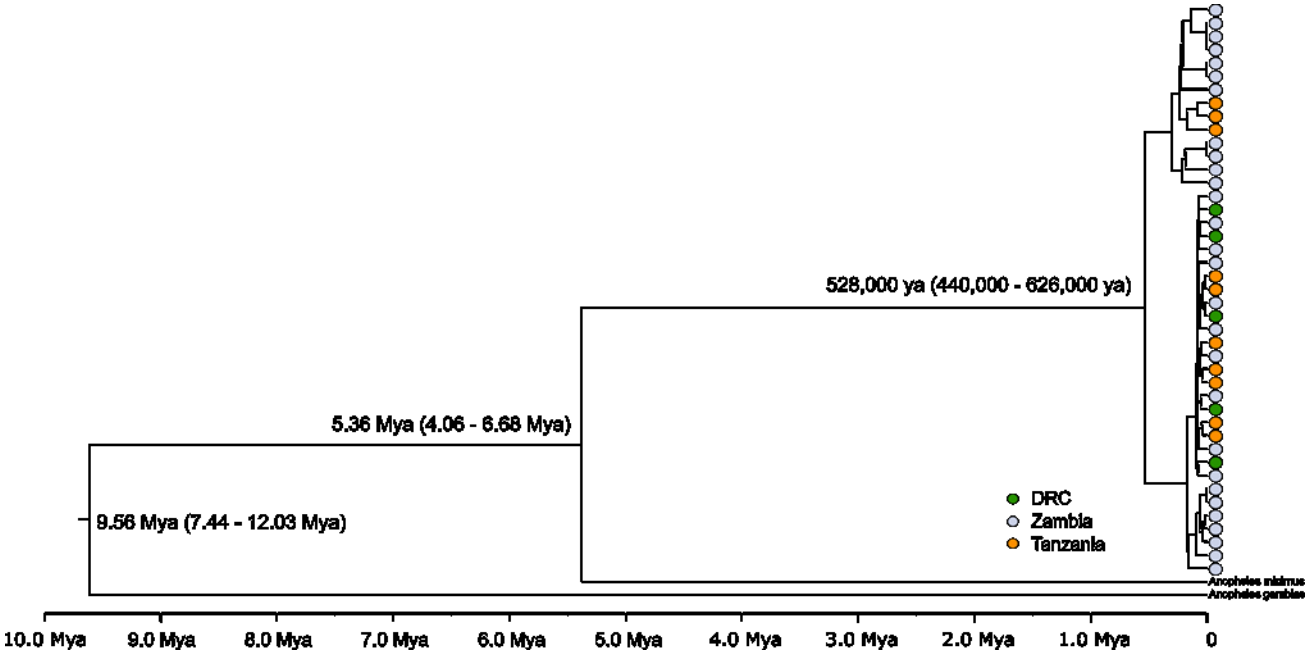


Figure S5: Bayesian skyline plots for the total population and lineages I and II. Bayesian skyline plot showing mean (red) and median (green) values of effective population size (95% HPD in blue). On the X axis is time going backward in years. Y axis is effective population size (log scale). Model used: GTR+G +I, relaxed clock. Top: Total population; Middle: Lineage I; Bottom: Lineage II.

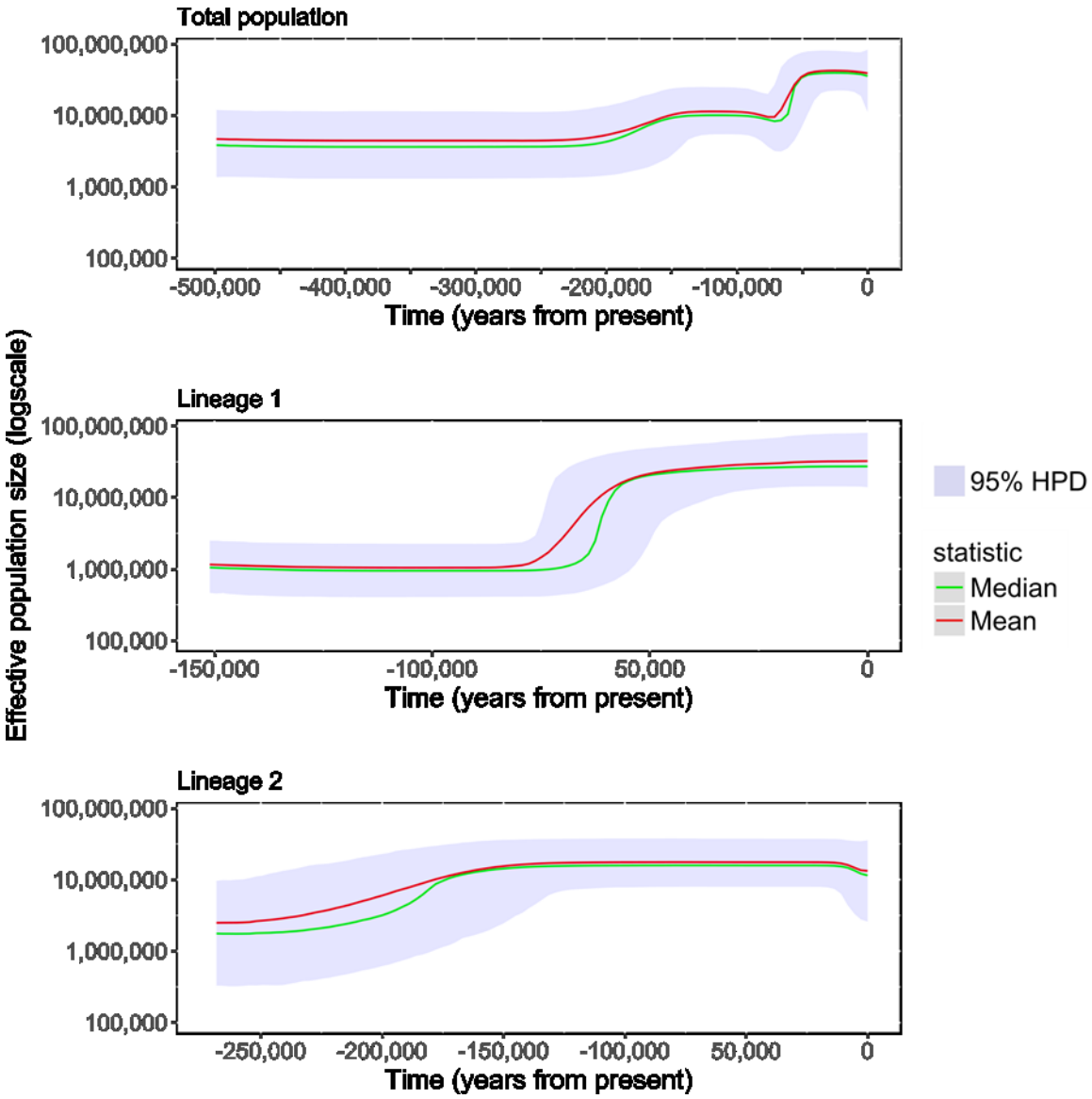


Figure S6: Correlation between genetic and physical distance. Pairwise genetic distance measured by nucleotide percentage identity was plotted against pairwise Euclidean distance between samples. Plots show the relationship between the genetic and physical distance for the total population and separately for each lineage. Mantel tests were conducted on each group. There was a statistically significant relationship between nucleotide identity and distance for lineage 1 ($p = 0.029$) and for lineage 2 ($p = 0.001$), though not for the total population. Not shown: when Tanzania samples (driving ~1 million-meter comparisons) were removed from pairwise analyses, no statistically significant correlation was seen for either lineage I ($p = 0.513$) or lineage II ($p = 0.549$).

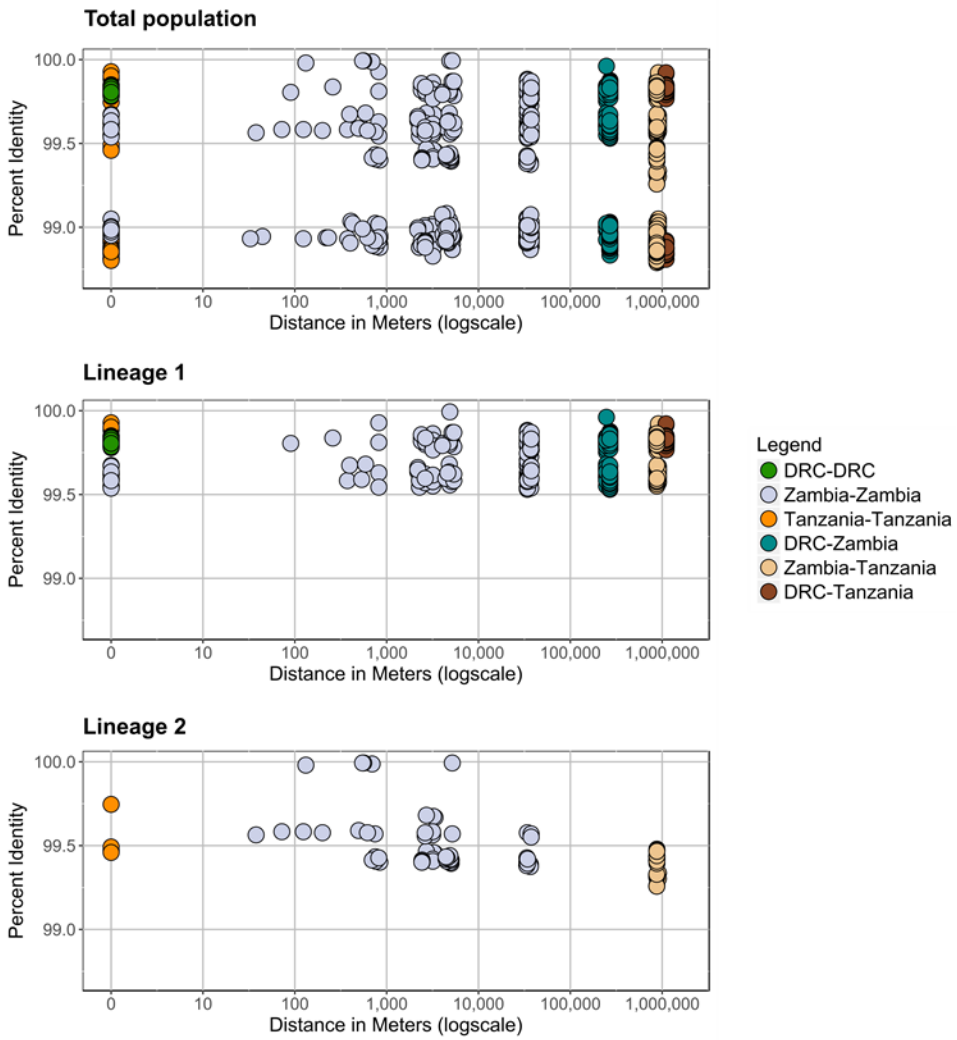


Figure S7: ML tree of partial ND5 sequences across Africa. Maximum Likelihood tree of the 43 *An. funestus* samples from the study in addition to 400 ND5 sequences from Michel *et al*. 2005 (GenBank DQ102854–DQ103253) generated using PhyML. GTR model with 1000 bootstrap replicates. Branches with bootstraps below 500 were collapsed to polytomies; remaining bootstrap support is shown in the figure. Samples are color-coded by geographic origin, as indicated in the legend on the bottom left.

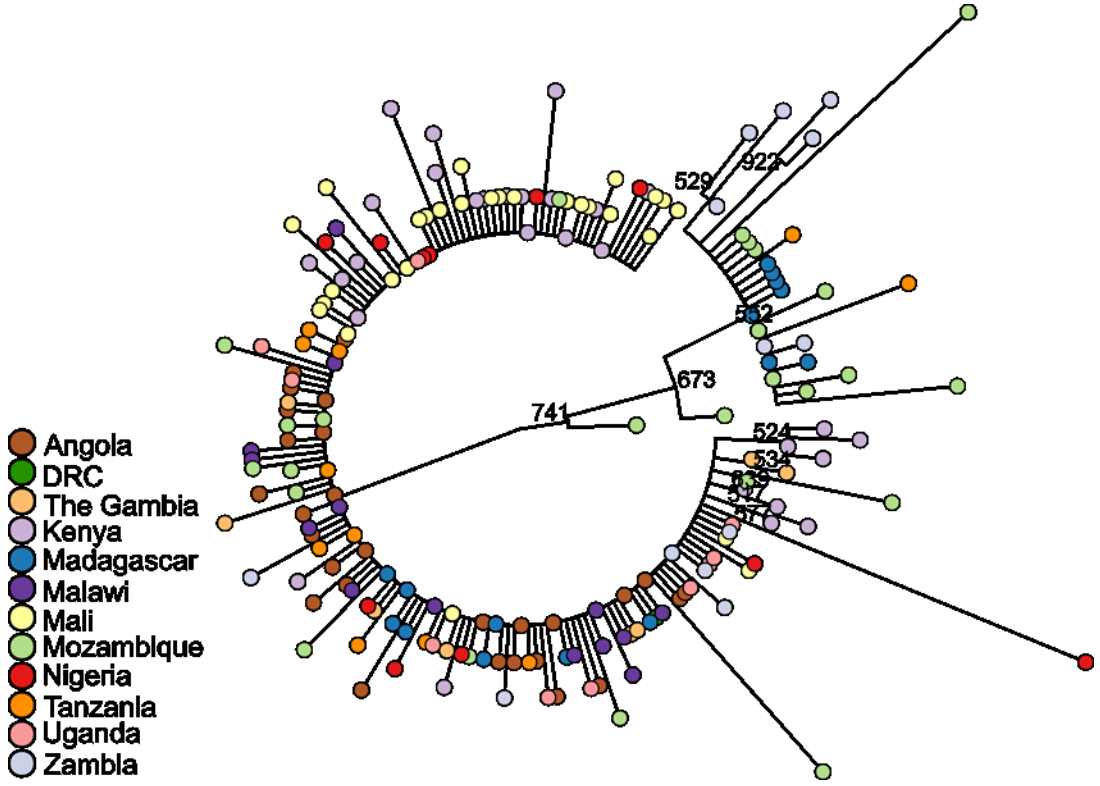
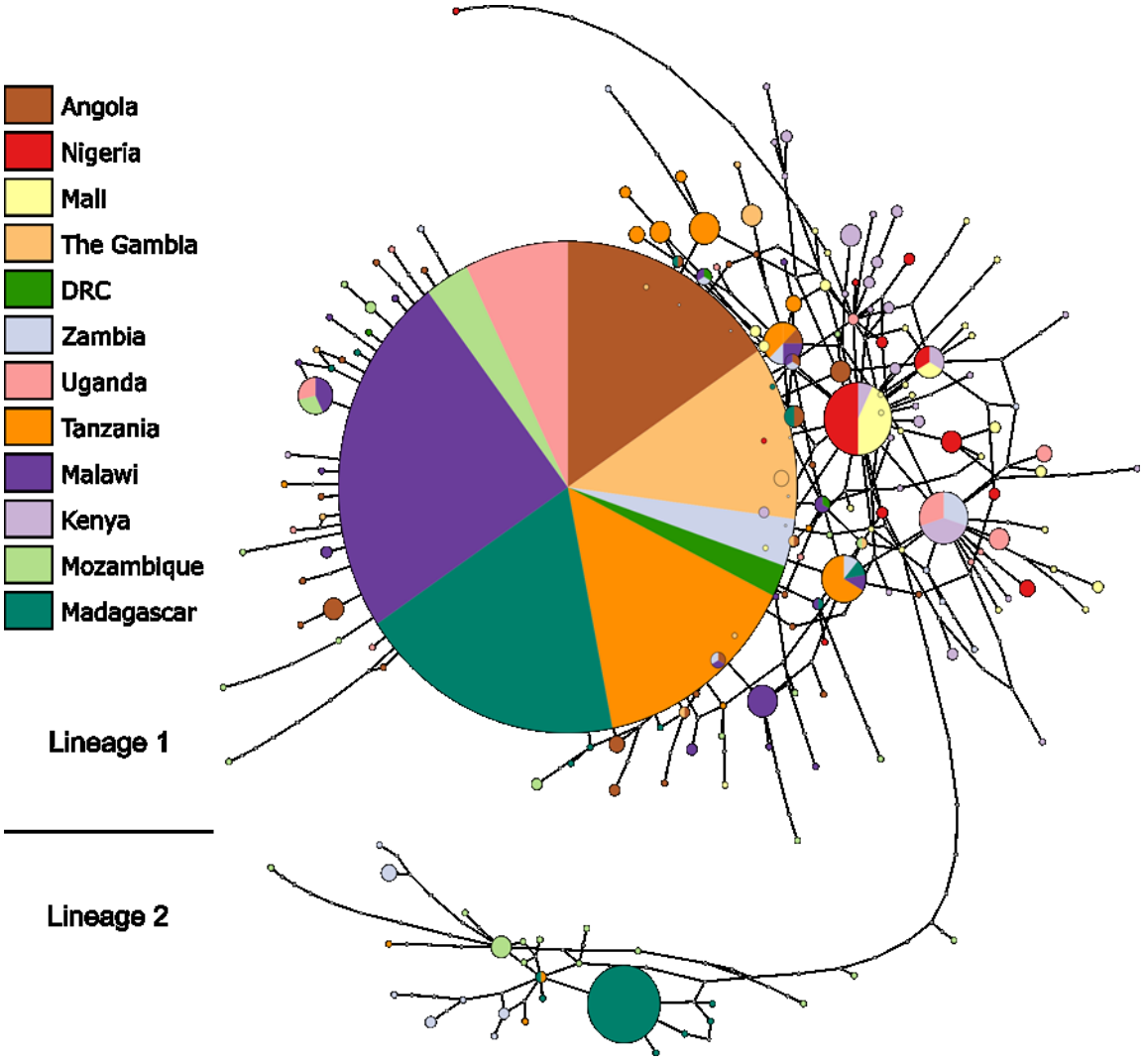


Figure S8: Haplotype network of N=443 partial ND5 sequences. TCS network of *An. funestus* partial ND5 sequences, including N = 43 from this study as well as N = 400 from Michel *et al*. 2005 (GenBank DQ102854–DQ103253). Each circle or node indicates a haplotype, with nodes segmented and colored proportionally to the number of sequences coming from 12 African countries (see legend on left). The size of each node indicates the total number of sequences sharing that haplotype. Haplotypes are separated by small uncolored circles, each of which indicates a single mutational step. The groups representing lineage I and II samples are indicated by a dividing line on the bottom left of the figure.



Tables

Table 1. Sampling sites, methods, numbers (N) and collection dates for whole genome sequenced specimens.

Site	N	Country	Coordinates (lat., long.)	Method	Collection Date
Nchelenge	6	Zambia	-9.2869, 28.7590	Indoor CDC-LT	Feb, 2015
Nchelenge	5	Zambia	-9.3247, 28.7819	Indoor PSC	Apr, 2015
Nchelenge	6	Zambia	-9.3042, 28.7822	Indoor CDC-LT	Feb, 2015
Nchelenge	6	Zambia	-9.2926, 28.7539	Indoor PSC	Apr, 2015
Kilwa Island	5	Zambia	-9.2675, 28.4500	Indoor backpack aspiration	Aug, 2014
Kapolowe	5	DRC	-10.9398, 26.9530	Indoor HLC	Apr, 2015
Lupiro	5	Tanzania	-8.383, 36.667	Indoor backpack aspiration	Jun, 2013
Lupiro	5	Tanzania	-8.383, 36.667	Animal-pen backpack aspiration	Jun, 2013

Table 2: Diversity statistics, neutrality tests, and demographic analysis. Samples have been split into two general comparisons: lineage I vs lineage II and DRC vs Tanzania vs Zambia. p-value is indicated by *: 0.10 > p > 0.05 or **: p < 0.05.

	Total	Lineage I	Lineage II	DRC	Zambia	Tanzania
# Samples	43	29	14	5	28	10
# Haplotypes	41	28	13	5	26	10
Hd (sd)	0.998 (0.006)	0.998 (0.010)	0.989 (0.031)	1.000 (0.126)	0.995 (0.011)	1.000 (0.045)
π (sd)	0.00625 (0.00054)	0.00237 (0.00019)	0.00500 (0.00033)	0.00163 (0.00022)	0.00668 (0.00046)	0.00611 (0.00139)
K	95.93	36.32	76.725	25.000	102.571	93.756
Neutrality Tests						
Tajima's D	-1.075	-1.841**	-1.040	-1.210	-0.514	0.023
Fu and Li's D*	-1.912*	-2.385*	-0.908	-1.210*	-1.098	0.076
Fu and Li's F*	-1.915	-2.606**	-1.086	-1.312	-1.069	0.071
Fu's Fs	-4.278	-6.543	0.904	0.778	-0.514	0.0525
Mismatch Distribution: Demographic Expansion						
SSD	0.00947	0.0123**	0.0284**	0.126**	0.013**	0.0365
Raggedness index	0.00309	0.00909*	0.0156	0.300	0.00638	0.0415
Mismatch Distribution: Spatial Expansion						
SSD	0.00939**	0.0135	0.0202	0.118**	0.0103**	0.034
Raggedness index	0.00309	0.00909	0.0156	0.300	0.00638	0.0415

Table 3: BaTS (Bayesian Tip-association Significance testing). MC = maximum clade size statistic; measures how closely particular sites are associated with monophyletic clade structure. P-value is indicated by *: $0.10 > p > 0.05$ or **: $p < 0.05$.

Statistic	**p-value**
Association Index	0.02**
Parsimony Score	0.01**
MC (Zambia)	0.02**
MC (Tanzania)	0.08*
MC (DRC)	1

Chapter III

Diversity and abundance of anophelines caught in outdoor CDC light traps in Nchelenge District, Zambia

Introduction

The Southern and Central African International Centers for Excellence in Malaria Research (ICEMR) engages in research in Nchelenge District in northern Zambia, among other locations, in an attempt to understand the dynamics of malaria transmission. In Nchelenge District, holoendemic malaria transmission continues despite years of malaria interventions, including artemisinin-combination therapies, targeted indoor residual spraying, and long-lasting insecticide treated bednets (LLINs). Nchelenge is approximately 800 meters above sea level, and the region's climate can be characterized as marshy. A river known as Kenani Stream runs through the study area in Nchelenge to empty into Lake Mweru, which forms the border with the Democratic Republic of the Congo. There are three seasons: a rainy season from November to May, a cool dry season from May to August, and a hot dry season from August to November. Malaria transmission is high throughout the year, and risk is especially associated with proximity to streams and marshland, which are characteristic of malaria vector breeding sites (M Hast, in preparation).[189]

Anopheles funestus sensu stricto (s.s.) and *An. gambiae s.s.* are the major malaria vectors in this area as determined by indoor vector collections, though the contribution of *An. funestus* to transmission is much higher than that of *An. gambiae* as estimated by their

respective entomological inoculation rates (EIRs) of 71 and 7 infectious bites per person per year.[113] An important driver of year-round transmission is the staggered peak abundancies of the two major vector species; while *An. gambiae* peaks during the wet season, *An. funestus* numbers are at their highest during the first part of the dry season.[112,113] Previous studies in the region have also established distinct spatial patterns for the two species. *An. gambiae* is generally found to be more abundant during the rainy season, while *An. funestus* is the most abundant vector overall and more prevalent inland alongside Kenani Stream rather than near Lake Mweru.[112,113]

Very few other anopheline species have been documented in Nchelenge, as sampling has been primarily conducted indoors. Only one outdoor collection has been done in the area, in the form of an adapted barrier screen collection in 2015 which resulted in 274 female anophelines (C Jones, unpublished).[190] The proportions of *An. gambiae* and *An. funestus* in that collection were approximately equal, with 2.6% of samples remaining molecularly unidentified due to either lack of molecular products or ambiguous results from standard molecular assays. Although barrier screen collections do not necessarily target host-seeking mosquitoes, the potential presence of outdoor foraging vectors may help to elucidate why the impact of current indoor-targeted vector control efforts have been limited in Nchelenge (J Stevenson, personal communication).[96,190]

Many researchers consider the human landing catch (HLC) the gold standard for collecting human-foraging anopheline vectors in the context of malaria transmission. HLCs require field workers to capture mosquitoes as they attempt to land on intentionally

exposed skin, but due to the potential risk for workers, the ethics of HLCs have long been debated. This, along with variability in attractiveness and skill of personnel and high labor and time costs has led to a need for alternative sampling approaches. Among other methods, Centers for Disease Control light traps (CDC LTs) placed near people protected and sleeping under bed nets are a common alternative to HLCs. CDC LTs require some form of bait (often in the form of protected people indoors or CO_2 outdoors), and because it's infeasible to acquire CO_2 in Nchelenge, another strategy for attracting mosquitoes is required. BG-Lure® (Biogents, Regensburg) is a chemical mosquito lure that has been tested for attractiveness to several vector species.[191–195] It is composed of a synthetic blend of chemicals meant to mimic human odors known to be attractive to arthropod vectors. Although designed for *Aedes* mosquitoes, at least one study found that traps baited with BG-Lure® had greater catches of *An. gambiae* than traps without (though BG-Lure® has been shown to perform poorly compared to other baits like CO_2 for malaria vectors).[191,193] Therefore, the BG-Lure® may serve as a replacement human-attractant for intensive outdoor collections where protected human bait sources are not readily available as they are for indoor overnight collections.

The primary purpose of this study was to determine the diversity and abundance of anophelines foraging outdoors that may contribute to residual *Plasmodium* transmission. The secondary objective was to compare catches using different baits with outdoor-deployed CDC LT collections, as CDC LTs serve as our primary vector collection methods throughout our ICEMR study sites. With these goals in mind, three trapping schemes were matched at each household; a standard CDC LT with incandescent light

placed outdoors where people gather in the evening, a CDC LT placed outdoors near livestock pens, and an outdoor CDC LT baited with the human analog BG-Lure®.

Methods

Study Sites and Sample Selection

This study was conducted in August 2016, during the dry season in Nchelenge District, when indoor anopheline population densities are at their highest. Thirteen households were selected from previous ICEMR data that indicated high abundance of anophelines caught by CDC LTs set indoors. From these thirteen households, eight households were selected for further collections based on their yield of anophelines after a single CDC LT collection. These households were divided between inland and lakeside locations as previously defined.[112,113] CDC LTs with three different bait schemes were randomly rotated through the households for eight consecutive days. The bait schemes were as follows: 1) CDC LTs set outdoor next to where people gather at night outside of the house (e.g. near cooking structures, shared common areas), 2) CDC LTs set outdoors at animal pens or shelters, and 3) CDC LTs set outdoors near high human-trafficked areas, such as washrooms, with an artificial human analog bait (BG-Lure®). Household metadata was recorded, including number of inhabitants, number and use of LLINs within the household, presence/absence of IRS in the household, location details of the trap, type and number of animals kept, as well as household socio-economic status including house construction characteristics, and household geolocation.

DNA extraction and species identification

Mosquitoes were morphologically identified to species using standard keys and then placed individually into labelled 0.6 mL microcentrifuge tubes containing silica gel desiccant and cotton wool and stored at room temperature until laboratory processing at Johns Hopkins Bloomberg School of Public Health. In the laboratory, abdomens and heads/thoraces were split and placed into separate tubes. DNA extractions were performed on the abdomens for each individual mosquito using a modified salt extraction protocol.[196] Abdominal DNA was subject to species complex-specific PCR assays, and an ITS2-based PCR when no specific PCR was available for a morphological ID or when species complex-specific results were unclear or null.[99,197–199] When samples repeatedly failed to amplify, pellets from the abdomen DNA extraction that had been saved at 20^{O}C from initial salt-extraction were re-extracted using the Qiagen DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany), and the re-extracted product was again subjected to the described molecular approach. Specimens from a single trap (which was a trap set in the human-attractant condition) were lost during transit to JHU and therefore did not undergo molecular analysis, although the collection count and morphological identifications were recorded. This brought the number of female anophelines that were molecularly processed to 790, though 1,087 were collected in total.

For samples giving ambiguous results from the ITS2 PCR as well as samples without *An. funestus* and *An. gambiae* morphological identifications, a Barcode of Life (BOL) portion of the cytochrome oxidase I gene (COI) was amplified from extracted DNA and sent to

the Johns Hopkins Medical Institutions (JHMI) Synthesis and Sequencing Facility for sequencing.[99] Forward and reverse sequences trimmed to remove ends with low quality and then high-quality trimmed sequences were aligned to generate a single consensus sequences. For samples where one read failed, the single high-quality trimmed read was used instead. Corresponding sequences of COI from NCBI (N=140) were aligned to BOL COI from outdoor samples and phylogenetic analyses were conducted (data not shown). Additional sequencing of internal transcribed spacer region 2 (ITS2), a nuclear locus, was conducted on at least one representative from each COI-based phylogenetic group. ITS2 consensus sequences were compared using BLASTN against the NCBI non-redundant nucleotide database for anopheline species identification and verification. Hits with a high percentage of query coverage (>70%) and a high percentage sequence identity (>90%) were considered good hits (data not shown). Molecular analyses were combined with morphological data to generate tentative species and other taxonomical identifications.

Parasite detection

Heads and thoraces of mosquitoes were homogenized in a grinding buffer of boiled casein and Nonidet P-40 (Carolina Barilla-Mury, personal communication). The homogenate was split into two portions for subsequent ELISA or PCR analysis. The Malaria Research and Reference Reagent Resource Center's (MR4) CSP ELISA protocol was used to detect *Plasmodium falciparum* sporozoites present in the first portion of the head/thorax homogenates. From the second portion of homogenate, genomic DNA was

extracted using the Qiagen DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany). *Pfldh* qPCR was conducted on the head and thorax genomic DNA extractions to confirm the presence of *Plasmodium falciparum* DNA in samples positive by CSP ELISA.[200]

Annotation and data availability

COI sequences are available from GenBank under the following accession numbers: MK016543-MK016657.

Statistical models and analysis

The R packages lme4 and glmmTMB were used to fit models for this study. Interval estimates for infection rate were conducted using base R, and plots were generated using ggplot2.[201–204] There were two phases for modeling both female anopheline counts and species diversity in traps: 1) testing *a priori* questions, and 2) data exploration. The basic structure of the models for both abundance/counts and diversity was decided based on likelihood ratio testing and comparing results from the R package DHARMa (which allows for fitted GLMM model diagnostics).[205]

For counts, the *a priori* question was: Does the expected number of mosquitoes in a given trap change as a function of the attractant scheme (animal, human, or BG-Lure®)? Remaining consistent with previous studies, location (lakeside or inland) was included as a fixed effect in tests as a potential confounding variable, and household was included as

a random effect to account for repeated measures at each household, which violate the assumptions of independence made by our regression models. The random effect intercept allows for variance of the baseline outcome (here, abundance) across households. Date was examined for potential confounding as a fixed effect as well as a random effect to examine the possibility of date-driven trap interdependence. Overall variance was much greater than the mean for counts for this study, and 31% of traps had zero captures. The negative binomial model is a generalization of the Poisson model often used for modelling count data, and it loosens the restriction of the Poisson model for variance being equal to the mean. A negative binomial distribution (Figure S2) fit the overall data better than an over-dispersed Poisson model, and no added benefits were seen with adding a zero-inflation term to the negative binomial model.

For diversity, the *a priori* question was: Does the diversity of populations of anophelines caught in any given trap differ as a function of the collection scheme/attractant (animal, human, or BG-Lure®)? Absolute number of species per trap was used as the response variable, and location was again added as a fixed effect and household as a random effect. For diversity, a Poisson model with overdispersion (Figure S3) was found to fit the data better than negative binomial. The overdispersion term in this Poisson model is another way of mathematically accounting for the variance of the data differing from the mean.

The trap that had specimens which were lost in transit was dropped in its entirety from diversity and species distribution analyses, due to lack of species confirmation for some

samples and potential bias resulting from the missing data. Thus, for the diversity models, the number of traps per treatment was 24 for the human-attractant context, 25 for the BG-Lure® traps, and 24 for the animal-attractant traps (N=73 traps total). Since abundance models rely simply on the count of female anophelines, which was reliably documented before transit, the full anopheline count was still used for abundance statistics. This brought the total number of human-attractant traps to N=25 and the overall total number of traps to N=74 for abundance models.

For both count and diversity, the final models were compared using cAIC values in the model.sel function in the R package MuMIn.[206] A version of R^2 developed for generalized linear mixed models (GLMM) was used to estimate the marginal and conditional pseudo-R^2 of each model, as implemented in MuMIn's r.squaredGLMM function using the trigamma function.[206] Marginal R^2 describes the proportion of variance explained by the fixed effects of a model, while the conditional R^2 describes the proportion variance explained by both fixed and random effects. The R package DHARMa was used for fitted GLMM model diagnostics.[205]

Please see supplementary figures and tables for all abundance and diversity models tested (Tables S1, S2) and the basic model specification diagnostics from DHARMa (Figures S1, S2).

Confidence intervals were estimated from mixed models using the bootstrap method in the confint.merMod function from lme4 and compared to R's confint function estimates

from corresponding models with household-level random effects excluded.[201,203] The confidence intervals reported in the results section represent the estimates from the GLMMs, and are consistent with and often more conservative than the estimates generated from excluding random effects (data not shown).

Specific distributions of female anophelines were analyzed for the five most abundant species. Chi square tests were used to compare the species-specific distribution among different attractants to an expected equal distribution of female anophelines. For species with low numbers (*An. gambiae* and *An. squamosus*), Fisher's exact test was used to compare counts from each type of attractant to the expected abundance (one third of the total abundance for that species). For each species, their proportion relative to the overall collection either lakeside or inland was calculated, and then Fisher's exact test was used to determine if the species-specific proportions differed between the two sites.

Map generation

Maps of anopheline species composition were depicted with ArcGIS v10.6 (ESRI, ArcGIS, Redlands, CA, USA) using geocoordinates from study households.[118]

Results

1,087 female anophelines were collected over the course of the 9-day study. Of the 790 which arrived at Johns Hopkins for processing, 747 (95%) were molecularly confirmed to species. The remaining samples did not amplify on any attempted PCRs (~5%). Of the

747 species-confirmed specimens, the majority of these were *An. funestus* (86%), with *An. gambiae* (2%) and other anopheline species (12%) making up the remainder.

Estimates of infection rate, blood-feeding rate, and human blood index

5/661 specimens tested by CSP ELISA (and confirmed by qPCR) were positive for *P. falciparum* sporozoites - all were *An. funestus* (5/524). The estimated infection rate for *An. funestus* in Nchelenge for this collection is therefore estimated to be 0.95% (95% CI: 0.12%, 1.8%). As described in Stevenson *et al*, mean catch per household per night from CDC LTs, unadjusted by household occupancy, was used as foraging rate in calculations of the entomological inoculation rate.[113] The mean confirmed *An. funestus* per household per night (13) was multiplied by the infection rate, which generated an estimate of 0.12 infectious bites/person/night (ib/p/night) in Nchelenge during mid-August 2016, or an estimated 44 ib/p/year.

Models of anopheline abundance

Models from hypothesis testing with the highest likelihood estimates (< 5 delta cAIC) indicate that location (lakeside vs inland) is the best (statistically significant) predictor of the count of anophelines expected in any given trap and the effect sizes for location in each model are highly concordant. The model with best fit by cAIC was:

$$\textbf{log (expected count)} \sim (\textit{location}) + (\textit{attractant}) + (1|\text{HH})$$

Note that that the term (1|HH) is the household-level random intercept term which was included in this and all other models discussed; the term allows for each household's baseline abundance to differ. This location + attractant model, accounting for household-level variation, predicted that 1) traps inland are expected to catch 15x (95% CI: 6.2x, 35x) more female anophelines than traps lakeside (p value = $7e^{-10}$), and 2) that neither the human attractant context traps (95% CI: 0.615, 3.07), nor the BG-Lure® traps (95% CI: 0.241, 1.28) had statistically significantly different abundance (p values = 0.41, 0.18 respectively) than the animal-attractant context traps. Additionally, attractant was not a statistically significant predictor in *any* of the tested models (Table S1). The location + attraction model had improved marginal and conditional R^2 0.33/0.36 versus the next-highest scoring model (delta cAIC = 0.3), which had marginal and conditional R^2 0.28/0.30 and included location as its only fixed effect covariate.

When abundance on was analyzed to examine the impact of date, the second day had statistically significantly ($p < 0.0003$) fewer female anophelines than the first study day, and the ninth day approached statistically significantly ($p = 0.06$) lower abundance. To partition the variance caused by date, a date-level random effect was added to and compared to the model with only the household level effect. It did not provide a significant improvement from the model with household-level random effects. Date also did not improve upon the location-only model when added as a fixed effect. Since Date was not an improvement and risked overparameterizing models, it was discarded from further analysis.

Exploratory analysis tested the impact of several additional household-level covariates on the location model, including: the total number of people sleeping in the household, the number of people sleeping under LLINs, and the number of LLINs in use, as well as roof- and eave-types (Table S1). The number of people sleeping *outside* of LLINs indoors in a household was a statistically significant predictor of abundance in many of the models tested, including the top model by cAIC after exploratory analysis:

$$\textbf{log (expected count)} \sim (\textit{location}) + (\textit{attractant}) + (\#\,\textit{people sleeping outside of nets}) + (1|HH)$$

In this model, accounting for household-level variation: 1) location was a statistically significant predictor (p value = $1e^{-7}$) of anopheline abundance, with 7.3x (95% CI: 3.5x, 16x) more mosquitoes expected per trap inland vs lakeside, 2) neither the human attractant traps (95% CI: 0.67x, 3.5x) nor the BG-Lure® traps (95% CI: 0.26x, 1.4x) were statistically significantly different than the traps in the animal attractant context (p values = 0.25 and 0.22, respectively), and 3) the number of people sleeping outside of nets was a statistically significant predictor (p value = 0.001) of abundance, with 0.68 (95% CI: 0.55x, 0.88x) fewer mosquitoes caught for each person outside the net. The marginal and conditional R^2 for this model were 0.37.

Models of anopheline diversity

There is no evidence to suggest that attractant statistically significantly impacted the diversity of anophelines caught in traps, as it was not a statistically significant predictor in any of the models tested (Table S2). Location was a statistically significant predictor (p value = $2e^{-5}$) of diversity by itself, with 2.9x (95% CI: 1.8x, 4.8x) more anopheline species seen in traps inland compared to traps lakeside. The location-only model of diversity had a conditional R^2 of 0.23 and a marginal R^2 of 0.38. After further exploratory analysis, the top model by cAIC was:

log (expected # species) ~ (*location*) + (*eaves*) + (1|HH)

Location was statistically significant (p = 0.001), with 2.2x (95% CI: 1.4x, 3.8x) more species inland than lakeside. This model showed that type of eaves approached significance (p < 0.1) as a predictor with approximately 2.5x (95% CI: 1.0x, 12x) more species per trap associated with open eaves. Marginal and conditional R^2 were 0.27/0.40. Since only one of the eight tested households had closed eaves, bias in this limited sample cannot be discounted. The next-best model by cAIC (dAIC = 1) was:

log (expected # species) ~ (*location*) + (*# people under bednet*) + (1|HH)

In this model, after accounting for household-level variation, 1) location was still a statistically significant predictor (p value = $1e^{-5}$) of diversity, with 3.1x (95% CI: 1.9x,

5.2x) as many distinct anopheline species present, and 2) the total number of people sleeping inside a household was *not* a statistically significant predictor (p value = 0.1; 95% CI: 0.846, 1.03). This model had marginal and conditional R^2 of 0.21/0.36.

Species Distribution

The five most abundant anopheline species were assessed in terms of distribution among both attractant and site, as shown in Figures 1-3. Of the five species, only *An. gambiae* had a statistically significant (p = 0.002) difference in abundance between attractant schemes. Most *An. gambiae* (N=11/14) were found in traps set near where people congregate. The primary species collected, *An. funestus,* was caught equally in all traps, regardless of trapping scheme/attractant.

Distinct spatial patterns for many anopheline species can be observed in Figure 1, which shows the relative proportion of each species in each trap. To account for the dramatic differences in overall abundance between lakeside and inland locations (Figure S1), the relative contribution of each species to either the lakeside or the inland collections were compared, as shown in Figure 2. *An. gambiae* was equally common in both inland and lakeside collections, as was *An. squamosus*. *An. funestus* comprised a statistically significantly higher proportion of inland collections than it was of lakeside collections. All *An. sp. 6* (N=29) were collected inland.

The overall population of *An. coustani* comprised a higher proportion of lakeside collections than inland, but the distinct clade structure of *An. coustani* (not shown) was further investigated. Figure 4 illustrates that there is a statistically significant difference between the proportions of each Clades B and C found inland versus lakeside. The low (N=5) number of specimens in clade A may mask a statistically significant difference to Clade C.

Discussion

This study found no statistically significant difference in mosquito count or diversity between outdoor collections made with CDC LTs placed in areas where humans congregate, near animal pens, or traps baited with BG-Lure®/human-analog attractant. This is encouraging for use of the BG-Lure® as an approximate outdoor equivalent of standard human-baited indoor CDC LT collection, especially in regions where CO_2 is unavailable as a mosquito attractant. However, since our study did not include a control CDC LT which was set far from high human traffic areas, we cannot be certain that our results indicate any additional benefit of BG-Lure® to the baseline attractiveness of CDC LTs. Additionally, the relatively small sample size and collection period (N=74 trap nights) of this study may have underpowered detection of smaller differences between the synthetic bait and attraction of either animals or humans in the proximity of the trap.

When split by species, collections began to indicate patterns of behavior that can be interpreted within the context of known species bionomics and general anopheline biology. *An. gambiae*, for example, was found mainly in CDC LTs placed near where

people congregate, which reflects its known specific attraction to human foraging cues. Other patterns were surprising. For example, *An. funestus* did not show the expected bias toward human attractants, and was found equally in each trapping scheme. This supports prior reports of animal blood meals associated with *An. funestus* and may also reflect the sheer abundance of this species across the collection area.[112] The overall abundance of *An. gambiae* was lower than that of other anopheline species, which is surprising considering that it makes up ~10% of indoor CDC LTs in this area.[113] In addition to skewing the expected proportions of species caught in Nchelenge, many of the anophelines in these outdoor traps had not been previously documented in this region of Zambia.

An. coustani was among the most abundant species in this study after *An. funestus*. *An. coustani* likely exists as a species complex, therefore our finding of distinct genetic structure amongst *An. coustani* specimens is not unexpected.[207,208] However, that the data supports biological differences among the *An. coustani* clades (in the form of spatial distribution differences) is surprising, especially in such a small sample set. The statistically significant difference in lakeside vs inland distributions of *An. coustani* clades reported here may be due to cryptic populations (within what is currently recognized as a single taxon) or due to additional biological constraints that are currently not understood. This finding is significant in the context of *An. coustani* as an emerging malaria vector, since understanding the biological limits of a species is important to vector control.[97,209]

Data analysis also revealed a seemingly counter-intuitive decrease in anophelines with more people outside of nets. There are two likely explanations for this phenomenon: 1) more people indoors and exposed without a bed net drew foraging mosquitoes away from traps placed outdoors, and 2) lower mosquito abundances in general led to lower net use. Our data are consistent with more nets in use in our inland households, where there is an overall higher abundance of foraging mosquitoes.

Limitations

Though it is likely that lakeside vs inland differences in abundance and diversity are due to ecological differences and biological suitability for mosquitoes, it is possible that socio-economic-driven changes in behavior and environment (such as types and number of animals kept, presence or absence of nearby open wells, and coverage of vector control) between the areas could contribute to these differences. A study focusing on such factors and designed with them in mind, would be required to tease apart these issues.

Households in this study were selected based on preliminary anopheline collections in areas previously identified as high-yielding. Though this non-random sampling strategy is unlikely to bias comparison results between the trapping schemes evaluated here, caution must be taken not to over-interpret the data. The non-random sampling in this study is a possible contributor/confounder of bionomics patterns, since low-yielding households may contain different anopheline species or mosquito foraging patterns.

Sampling may have also biased the estimated infection rates and could have inflated the average foraging rate in the area.

Conclusions

This study is the first in the region to target scalable collection methods for outdoor foraging anophelines, where commonly used baits such as CO_2 are not available. Trapping schemes utilizing easily-deployed and highly-scalable CDC LTs did not reveal a statistically significant difference in trap yield or female anopheline diversity between traps placed near animal pens, where people congregate, and traps fitted with the BG-Lure® human-analogue attractant. This is promising for the use of the lure as an alternative to using human volunteers to bait outdoor CDC LTs, at least when BG-Lure® CDC-LTs are set in areas likely to have lingering human odorants present.

There were an unexpectedly large number of distinct anopheline species collected outdoors in this study. Years of longitudinal and cross-sectional studies conducted in Nchelenge have included indoor CDC LTs, which have identified very little anopheline diversity, with the primary vectors *An. gambiae* and *An. funestus* dominating those collections. Although the additional species identified here were not found to harbor *P. falciparum* parasites, their low numbers ($N < 30$ for any particular species) preclude any assumption that they do not contribute to local transmission in Nchelenge, especially as some of these species have been implicated in transmission elsewhere.[96,149,209,210]

The spatial distributions of the anophelines in this study are suggestive of a highly complex ecology which likely contributes to the persistently high transmission rate in Nchelenge. A sufficient number of *P. falciparum* positive anophelines (N=5) were found in these outdoor foraging mosquitoes to warrant suspicion of outdoor transmission in the region. Further studies should rigorously examine the implication of outdoor transmission in Nchelenge District, as it would undermine the currently utilized strictly-indoor vector control.

Figures

Figure 1: Map showing relative proportion of each species to total female anophelines collected per trap. The upper cluster of traps are classified as 'lakeside' while the lower are considered 'inland.'

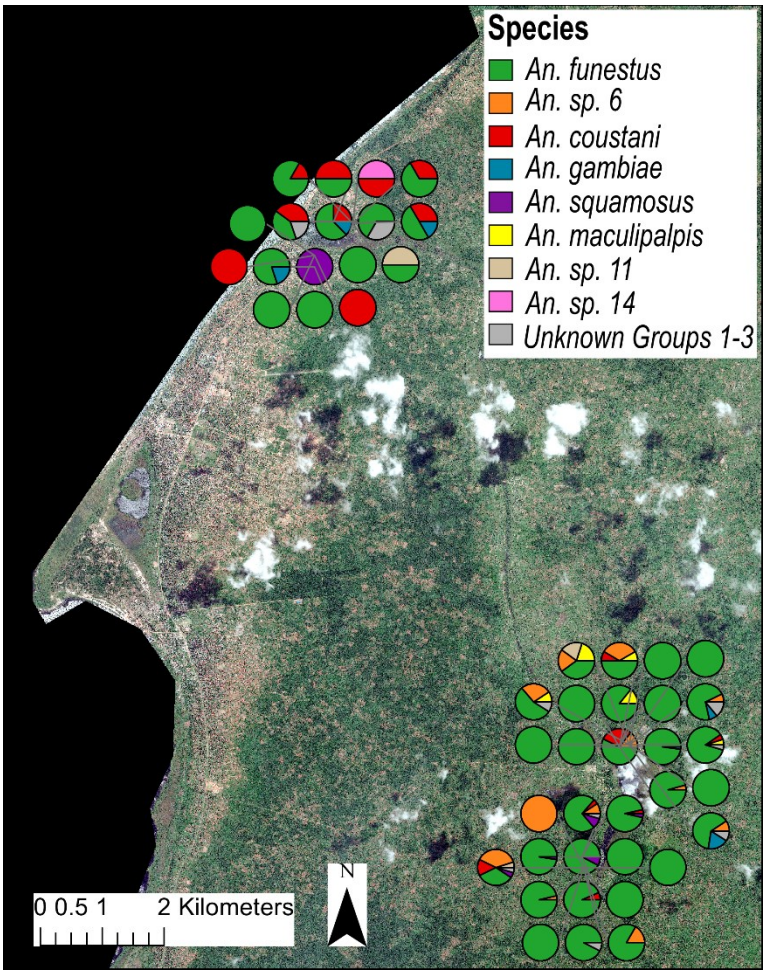


Figure 2: The proportion each species relative to the total collections either inland or lakeside. The five most abundant species are shown. ** indicates a statistically significant difference ($p < 0.01$) between lakeside and inland proportions. * indicates a moderate difference ($p < 0.1$). ns indicates no statistically significant difference.

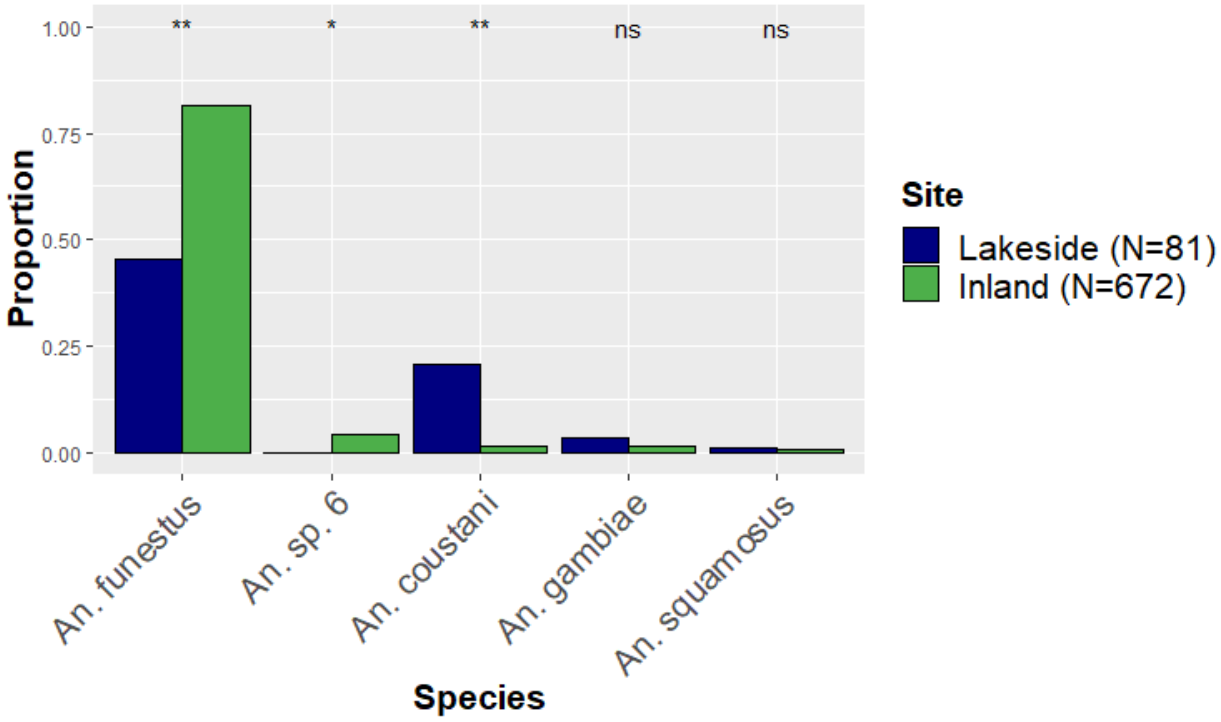


Figure 3: Species-specific proportion of specimens collected in traps among three different baiting schemes. Equal distribution among attractants is expected if attractant has no impact on species-specific abundance. The five most abundant species are shown. ** indicates that proportions statistically significantly deviate ($p < 0.01$) from the expected 1/3 in each attractant condition.

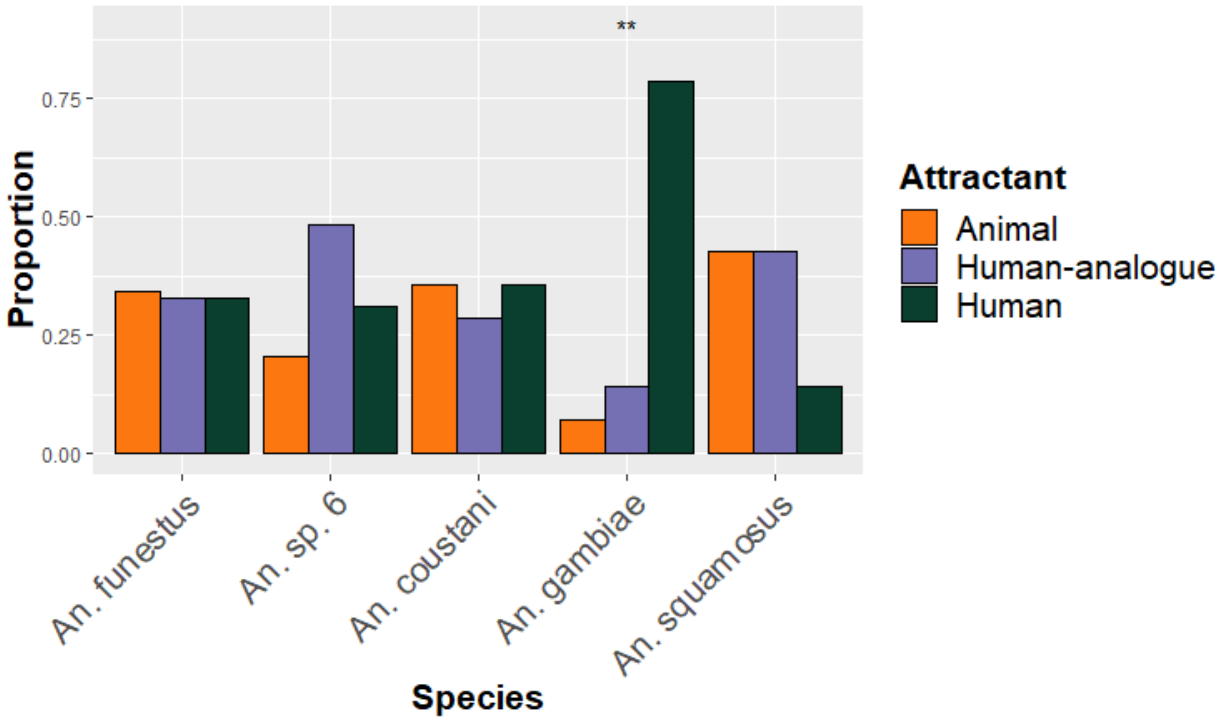


Figure 4: Comparison of the proportion inland vs lakeside of each *An. coustani* clade. ** indicates a statistically significant ($p < 0.01$) difference between clades.

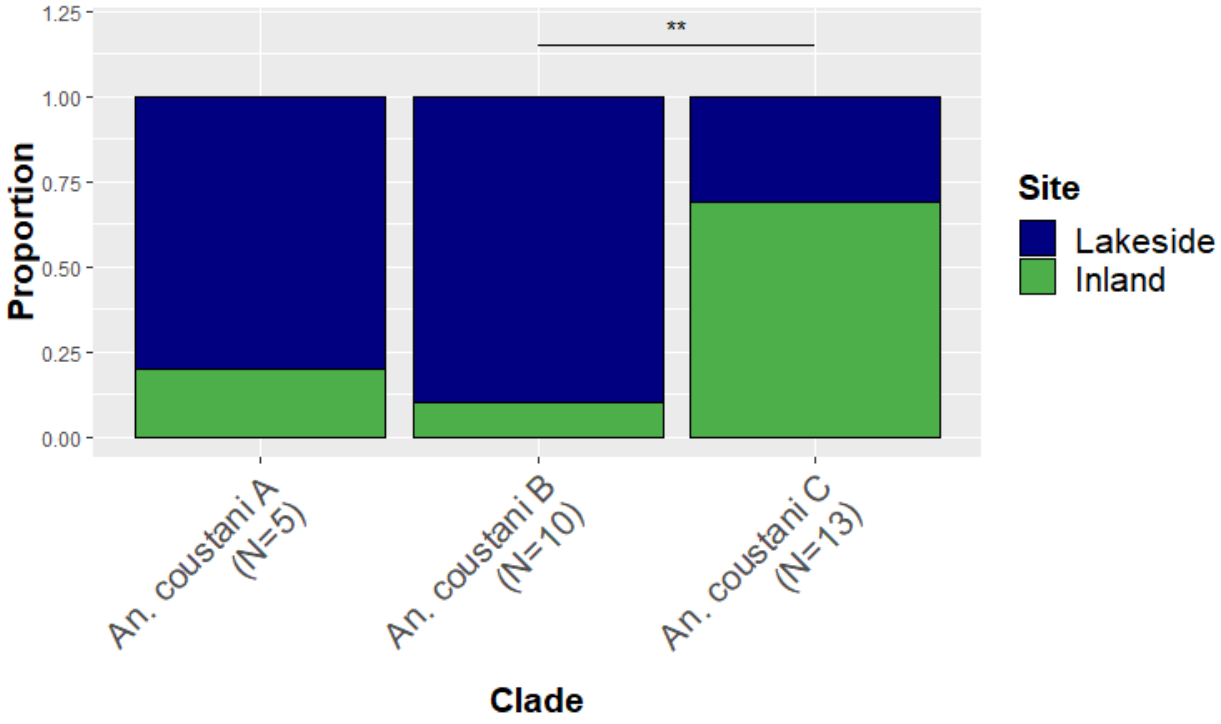


Figure S1: Map showing the relative abundance of female anophelines caught in each trap set during the study (N=73). Each circle represents an individual trap, and the size of the circle is relative to the number caught – a circle representing a trap with 7.1 female anophelines is shown in the legend.

Figure S2: The DHARMa diagnostics plots for the basic negative binomial GLMM for female anopheline abundance. On the left, plot shows no statistically significant deviation of the qqplot of the standardized residuals from the expected distribution. On the right, the distribution of standardized residuals against predicted values falls reasonably within expectations.

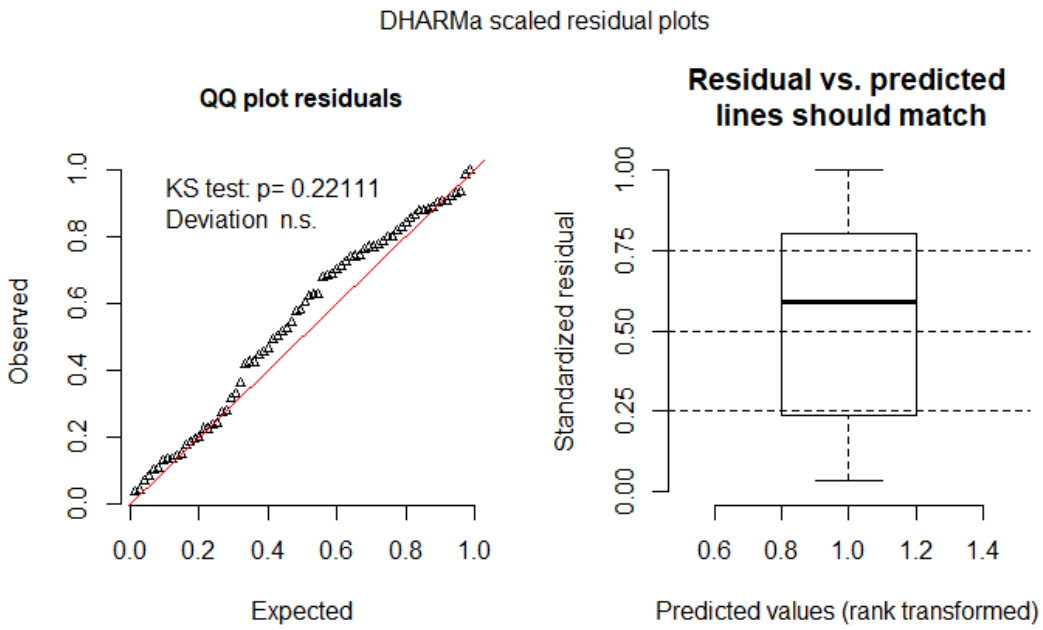
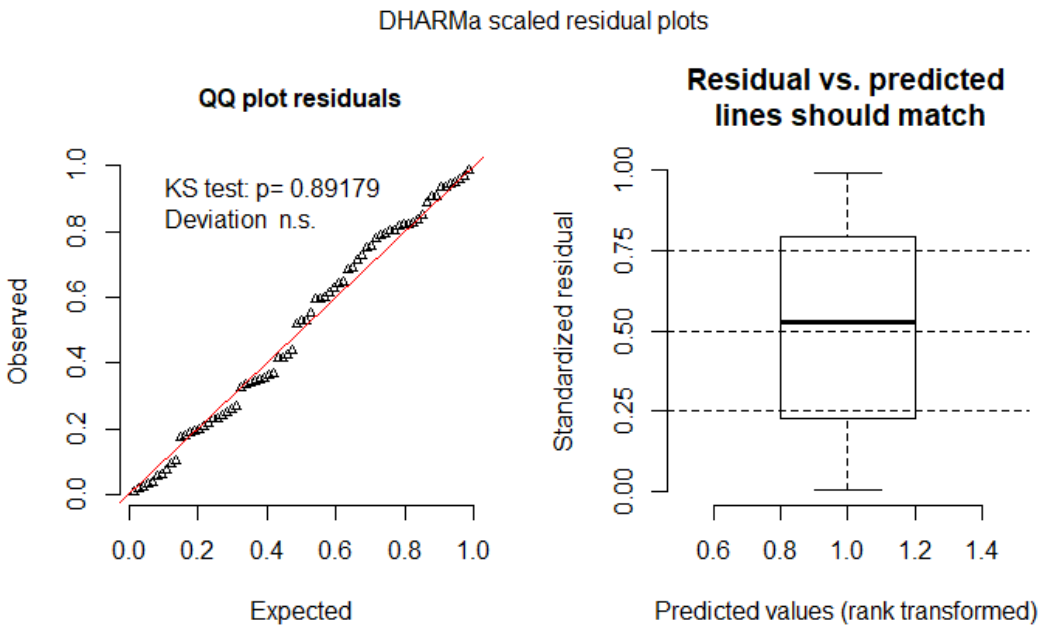


Figure S3: The DHARMa diagnostics plots for the basic over-dispersed Poisson GLMM for number of species of female anopheline present. On the left, plot shows no statistically significant deviation of the qqplot of the standardized residuals from the expected distribution. On the right, the distribution of standardized residuals against predicted values falls reasonably within expectations.



Tables

Table S1: Summary of models and their covariates for the N=20 tested negative binomial mixed models of female anopheline abundance per trap, which were ranked by cAIC with the model.sel function of the R package MuMIn.[206] Each '+' indicates that a covariate was used in the given model.

	roof	eave	#ITN	#ppl outside of net	#ppl in household	#ppl under net	date	attractant	on
				+				+	
				+					
			+					+	
					+			+	
					+				
						+		+	
						+			
				+				+	
			+						
								+	
								+	
		+							
		+						+	
	+						+	+	
	+							+	
	+						+	+	

Table S2: Summary of models and their covariates for the N=18 tested over-dispersed Poisson mixed models of the # species present in each trap, which were ranked by cAIC with the model.sel function of the R package MuMIn.[206] Each '+' indicates that a covariate was used in the given model.

eave	roof	#ITN	#ppl outside of net	#ppl under net	attractant	date	ion
+							
		+					
			+				
				+			
+					+		
+							
	+						
	+						
		+					
			+				
				+			
					+		
						+	
						+	
					+		
					+	+	
						+	

Chapter IV

Investigating the phylogenetic relationships between known and unknown *Anopheles* specimens from an outdoor collection in northern Zambia

Introduction

Human malaria is transmitted by species of mosquitoes in the genus *Anopheles*. There are >450 species of anopheline mosquitoes, which are divided into six main subgenera: *Stethomyia*, *Lophopodomyia*, *Kerteszia*, *Nyssorhynchus*, *Anopheles*, and *Cellia*. The largest of these subgenera by far are *Anopheles* (183 species) and *Cellia* (224 species).[211] *Cellia* has an old-world distribution, *Anopheles* is cosmopolitan, and the remaining subgenera are neotropical in distribution. There are fewer than 50 species within the entire *Anopheles* genus that are classically considered important to maintaining malaria transmission.

Subgenera of anophelines can be further divided into smaller taxonomic units, including Sections, Series, Groups, and species complexes. Members of species complexes cannot be differentiated solely by morphology, so a combination of morphological, behavioral/ecological, and molecular approaches must be used to identify species. Species-level identification is important, because even at the level of highly-related species within a complex, behavior and vector competence can be highly variable. In *An. funestus sensu lato (s.l.)*, for example, only 1 of the 13 species, *An. funestus sensu stricto* (*s.s.*), is considered to be a major malaria vector in sub-Saharan Africa. In addition, the species in this complex differ in terms of host preference, foraging preference, insecticide

resistance, and ecological niche.[47] For effective vector control efforts, it is important to be able to accurately identify species that sustain regional transmission. To date, most phylogenetic studies (especially molecular-based studies) have focused on classically-recognized malaria vectors and neglected the remaining ~400 species.[212] As newly-discovered species and largely disregarded existing secondary vector species are recognized as important for malaria transmission, the lack of genetic data and molecular diagnostics for a wider range of anopheline species becomes problematic.

When specimens from field collections remain unidentified using routine morphological and molecular methods, sequence comparison to a database (i.e. comparison of sequences to NCBI's non-redundant nucleotide database using the Basic Local Alignment Search Tool (BLASTN)) is often used.[213] Limitations of this database are the lack of reference sequences for neglected species and the inclusion of genetic information for misidentified or unverified specimens. The cytochrome oxidase c subunit I (COI) gene is often targeted for species-level identification as a so-called 'barcode' for many taxa, including many vector insects and anophelines.[214–217] The COI gene is targeted because 1) efforts to catalogue global species diversity using Barcode of Life (BOL) have led to many sequences being available for this locus,[218] and 2) the balance of sequence conservation and polymorphism in COI allow for comparison at the level of closely related species, such as within the complicated species complexes common to anophelines.

This study aims to classify a set of anopheline specimens from a set of outdoor Centers for Disease Control light trap (CDC LT) collections in Nchelenge District, Northern Zambia. The overall collection is part of a large effort to better understand malaria

transmission in this region of Zambia, where the disease remains largely uncontrolled. These anopheline specimens comprise both easily identified and commonly-recognized major vector species as well as many species for which species-level identification by morphology was not easily obtained. By sequencing a BOL portion of COI, it was possible to attain phylogenetic placement of some of these unassigned specimens with well-characterized anopheline species.[214] This allows for positive identification of some specimens and putative identification or clustering for many others for which well-referenced genetic data is not yet available. Such information will be critical for future malaria surveillance and interventions as these unassigned and potential vector species comprise increasing proportions of ongoing malaria investigations world-wide.

Methods

Study site and sample collection

Nchelenge District of northern Zambia is adjacent to the Democratic Republic of the Congo, with the border bisecting Lake Mweru. It is a marshy region lying ~800 meters above sea level. There are three seasons: a rainy season from November to May, a cool dry season from May to August, and a hot dry season from August to November. Malaria transmission occurs at high rates year-round, despite widespread use of long-lasting insecticide treated nets (LLINs) and indoor residual spraying (IRS).[111,219] *An. funestus* and *An. gambiae* are considered the primary vectors in the area, though *An. funestus* contributes more to transmission in Nchelenge due to its much larger abundance.[112,113] The *An. funestus* population peaks during the dry season when *An. gambiae* numbers are

at their lowest.[113] This study was conducted in August 2016 during the dry season. Centers for Disease Control light traps (CDC LTs) were set in a total of thirteen households in two different locations in Nchelenge – one close to the lake and one more inland. Traps were set outside overnight, adjacent to high trafficked areas for humans and near animal pens, for nine nights (a total of N=74 traps).

DNA extraction and species identification

Mosquitoes were morphologically identified to species using standard keys and then placed individually into labelled 0.6 mL microcentrifuge tubes containing silica gel desiccant and cotton wool and stored at room temperature until laboratory processing at Johns Hopkins Bloomberg School of Public Health. In the laboratory, abdomens and heads/thoraces were split and placed into separate tubes. DNA extractions were performed on the abdomens for each individual mosquito using a modified salt extraction protocol.[196]

For anopheline molecular identification, a series of PCR assays were conducted. For those specimens that would not amplify on species-complex-specific PCR diagnostics for *An. funestus* or *An. gambiae* species groups, a more general differential PCR based on the ITS2 region of rDNA was used.[69,214] For those samples that either did not amplify using the ITS2 assay, or that give an ambiguous fragment size, a COI-based BOL PCR protocol was used to amplify and then sequence generated fragments.[99] When samples repeatedly failed to amplify, pellets from the abdomen DNA extraction that had been saved at -20^{O}C

were re-extracted using the Qiagen DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany), and the product was again subjected to the described PCR assays.

Parasite detection

Heads and thoraces of mosquitoes were homogenized in a grinding buffer of boiled casein and Nonidet P-40 (Carolina Barilla-Mury, personal communication). The homogenate was split into two portions for subsequent ELISA or PCR analysis. The Malaria Research and Reference Reagent Resource Center's (MR4) CSP ELISA protocol was used to detect *Plasmodium falciparum* sporozoites present in the first portion of the head/thorax homogenates. From the second portion of homogenate, genomic DNA was extracted using the Qiagen DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany). For samples positive by CSP ELISA, *Pfldh* qPCR was conducted on the head and thorax extraction to confirm *Plasmodium falciparum* DNA.[200]

Sequencing and phylogenetic analysis

COI BOL amplicons were sent for Sanger sequencing at the Johns Hopkins Medical Institutions (JHMI) Synthesis and Sequencing Facility using the forward and reverse LCO1490 (5'-GGT CAA CAA ATC ATA AAG ATA TTG G-3') and HCO2198 (5'-TAA ACT TCA GGG TGA CCA AAA AAT CA-3') primers described in Hebert *et al* 2003.[214] Forward and reverse sequences trimmed to remove ends with low quality and then high-quality trimmed sequences were aligned to generate a single consensus

sequences. For samples where one read failed, the single high-quality trimmed read was used instead. Corresponding sequences of the COI from known anopheline species, as well as several taxonomically unassigned species (i.e. "*An. sp. 1*"), were downloaded from NCBI (N=140) to represent a spectrum of taxa in the genus *Anopheles* as well as several sequences from sister genera for outgroup analysis (Table S1). All sequences (N=256) were aligned in Geneious v11.1.5 (https://www.geneious.com) using the MUSCLE algorithm[159] and trimmed to a final length of 488 bp. Duplicate haplotypes were collapsed to a single haplotype for analysis using FaBox (see Table S2 for study haplotypes with >1 member), resulting in a final 196-member alignment.[220] A high-scoring model from jModelTest2 was used (GTR +G +I) for subsequent phylogenetic analysis.[221,222] Phylogenetic trees were built using both maximum likelihood (ML) and Bayesian statistics, implemented in phyML (Supplemental Figure 1) and BEAST2 (Figure 1), respectively .[164,170] 1,000 bootstraps were specified for the phyML tree. For BEAST2, 100,000,000 Markov chain Monte Carlo chains were used with 10,000,000 discarded as burn-in. LogCombiner was used to resample 1,000 states, and then a maximum clade credibility tree was generated with TreeAnnotator. Tracer v1.7 was used to ensure sufficient (>200) ESS values for all parameters of interest.[223] Trees were annotated and visualized using R packages ape, phytools, and ggtree.[177,223,224] Nodes with low support (fewer than 500 bootstrap trees for ML or lower than 50% posterior probability for the maximum clade credibility tree) were collapsed into polytomies, and trees were left unrooted.

To validate results from COI BOL sequencing, representatives of each phylogenetic group were sequenced using the internal transcribed spacer region (ITS2) in the nuclear genome. ITS2 is both easy to amplify and has relatively high variability within species, and is often used in conjunction with COI for molecular studies of species.[99,185,225–227] ITS2 amplicons were sent for Sanger sequencing at the JHMI Synthesis and Sequencing Facility using the forward and reverse ITS2A (5'- TGT GAA CTG CAG GAC ACA T -3') and ITS2B (5'- TAT GCT TAA ATT CAG GGG GT -3') primers from the protocol described by Koekemoer *et al* 2002.[69] Forward and reverse sequences were trimmed to remove ends with low quality and then high-quality trimmed sequences were aligned to generate single consensus sequences. For samples where one read failed, the single high-quality trimmed read was used instead. Final ITS2 sequences were compared using BLASTN against the NCBI non-redundant nucleotide database. Hits with a high percentage of query coverage (>70%) and a high percentage sequence identity (>90%) were considered good hits (Table 1).

Annotation and data availability

COI sequences are available from GenBank under the following accession numbers: MK016543-MK016657.

Map generation

Maps of anopheline species composition were depicted with ArcGIS v10.6 (ESRI, ArcGIS, Redlands, CA, USA) using geocoordinates from study households.

Results

A total of 790 female anophelines were molecularly processed for this study and 115 total COI BOL sequences were generated to assist with species verification. For 43/790 (~5%) samples, which repeatedly failed to amplify for any PCR attempted, species verification was not possible. The majority of the species-validated specimens (as determined through a combination of morphological and molecular analyses) were *An. funestus* (644/747), with only a few *An. gambiae* (14/747), and an unexpectedly high diversity of additional species (Table 1). Of the 5 samples positive by CSP ELISA, all were morphologically and molecularly *An. funestus s.s*., giving an estimated infection rate of *An. funestus* of 0.96% (95% CI: 0.12% – 1.8%).

Phylogenetic Analyses

The 488 bp multi-alignment for the COI BOL included 196 unique haplotypes and was used to construct both maximum likelihood (ML) and Bayesian trees (Supplemental Figure 1, Figure 1). The sequences from *Chagasia* (sister genus to *Anopheles*) and *Culex* (a more distant genus-level taxon) did not have a highly supported phylogenetic relationship with any other sequences in the alignment, as expected based on their

relatively distant taxonomical relationship. The overall 'backbone' of the full phylogeny suffered from very low support (Figure S1, Figure 1), as previously described for full genome-based analyses in Foster *et al*.[211] Therefore, many relationships at the level of genera and subgenera were not resolved. This is consistent with other studies where COI was not a good marker for this taxonomical level of discrimination.[188,211,228,229]

Some species groups and complexes were grouped into well-supported clades (i.e. *Gambiae* Complex and *Coustani* Group*,* not shown), which is consistent with widespread use of COI as a good discriminator at the approximate level of species.[214,218] However, some Group- and Series-level groupings were not fully resolved. In many cases, Groups or Series grouped in exclusive clades, but these clades did not have enough phylogenetic support to declare them monophyletic (i.e. *Funestus* Group or *Anopheles Laticorn Myzorhynchus* Series, data not shown).

An. funestus morphological identifications were confirmed by molecular and sequence analysis and fall in a single well-supported clade with NCBI sequences from *An. funestus s.s.* (Figure 2, top). Both of the two sequenced *An. gambiae* samples fell into a well-supported clade with sequences representing the *An. gambiae* complex (data not shown). The ML and Bayesian trees showed approximately 10 groups of sequenced study specimens outside of *An. funestus* and *An. gambiae* (Table 1, Figures 2-9). A relatively confident inference of species was made for a group of 5 specimens that formed a well-supported clade with *An. maculipalpis* (Figure 3). The additional 9 groups were more

difficult to place given the small fragment size of the COI BOL that was sequenced: these groups are illustrated in Figures 4-9, and are discussed below.

Subgenus Cellia

The second-most abundant group collected in the study (N=31) fell within a single highly-supported cluster (Figure 2, bottom). Due to the inclusion in this clade of the sequence of "*An. sp. 6*" from NCBI, these samples have been classified as *An. sp. 6*. One specimen, potentially related to *An. sp. 6*, was a highly supported partner to the NCBI sequence for "*An. sp. 14*" (Figure 2, bottom). *An. sp. 6* and *An. sp. 14* may be related to *An. theileri*, shown at the bottom of Figure 2. These three clades form a supported monophyletic group (posterior probability = 0.56) that, at a higher level, clustered with species from *Cellia Myzomyia* (posterior probability = 0.58), including members of the *Funestus* Group (Figure 2, top). The morphological identifications for *An. sp. 6* specimens were inconsistent and varied: 21/31 (67.7%) were identified as a member of *Cellia Myzomyia*, 7/31 (22.6%) were identified as *An. gambiae*, and 3/31 (9.68%) were marked as unidentified in the field. The single *An. sp. 14* specimen was morphologically identified as *An. gambiae*.

A group of 4 specimens clustered with *An. squamosus* sequences with high support (Figure 4, top). This clade formed a group with 3 additional specimens that more closely matched a NCBI sequence for "*An. sp. 15*," but remained monophyletically clustered with the *An. squamosus* sequences (Figure 4, bottom). All 7 samples were

morphologically identified as *An. squamosus* in the field, and corresponding ITS2 sequences returned BLAST matches to *An. squamosus*.

Three specimens in Unknown Group 1 (UG1), fell into a cluster (posterior probability = 0.62) containing sequences from the *Funestus* Group of *Cellia Myzomyia* that are found in southeast Asia (Figure 5). UG1 were morphologically identified as *An. funestus* (2) and unidentified (1).

Subgenus Anopheles

A large number of specimens (N=28) formed three distinct and well-supported clusters with the *Coustani* Group within the subgenus *Anopheles*. All specimens were morphologically identified as *An. coustani*. The clade that clustered most tightly with *Coustani* Group sequences was classified as *An. coustani* clade A. The other two clades were classified as *An. coustani* clade B and *An. coustani* clade C (Figure 6).

Other

Three specimens clustered tightly with *An. sp. 11*, and were classified as such (Figure 7). All three were morphologically identified as *An. squamosus*, though neither COI nor ITS2 sequences were matches to *An. squamosus*. Five specimens fell into each of Unknown Group 2 (UG2) and Unknown Group 3 (UG3), respectively (Figures 8 and 9). UG2 and UG3 had varied morphological identifications. 1/5 UG2 was morphologically

unidentified, 2/5 were identified as *An. brunnipes*, and 2/5 as *An. rhodesiensis*. 2/5 UG3 was morphologically unidentified, 2/5 as *An. funestus*, and 1/5 as *An. tchekedii*. Neither UG2 nor UG3 COI clustered with significant support with NCBI entries for *An. funestus* and *An. rhodesiensis*, and no COI sequences for *An. brunnipes* nor *An. tchekedii* were available. BLAST results for ITS2 sequences from both groups returned only poor matches to database sequences.

Discussion

Diversity

This collection represents a much higher diversity of anopheline species than has previously been documented in Nchelenge District, despite extensive sampling in the region spanning almost two decades. Several things may help to explain this unexpected diversity: 1) this is one of only two studies in the region in which collections were conducted outdoors, and many studies have documented higher species diversity of anophelines outdoors as compared to traditional indoor collections, 2) several consecutive years of IRS may have lowered numbers of *An. funestus* enough to open ecological niches for other anophelines, and 3) lack of extensive sequencing and rigorous identification of unidentified specimens in previous collections obscured an unrecognized level of species diversity. Members of both the *Gambiae* and *Funestus* Groups are highly recognized as vectors driving human malaria transmission in sub-Saharan Africa.[149] Though only *An. funestus* was found to be *P. falciparum*-positive in this collection, the

low sample size of non-*An. funestus* samples makes accurate estimation to their vector potential impossible.

The best match for 31 of the anophelines collected in this study was an NCBI database sequence for "*An. sp. 6*" (GenBank Accession: KJ522834), which was identified as *An. theileri* clade F by another group (Neil F Lobo, personal communication). *An. theileri* is an understudied species commonly collected outdoors in South Africa, though it has been observed as far north as Zambia, Tanzania, and Mozambique.[149] Not much is known about the adult biology of this species. *An. theileri* is morphologically placed within *Cellia Myzomyia* in the *Wellcomei* Group, for which only the COI sequence of *An. theileri* could be found. Though *An. theileri* was supported as a related sequence for the *An. sp. 6* and *An. sp. 14* specimens, the most conservative and confident conclusion from this analysis is that this unnamed species lies within the *Cellia Myzomyia* Series.

An. squamosus and its sister species *An. pharoensis* have long been considered secondary vectors of *Plasmodium* to humans.[149] Findings from southern Zambia indicate that *An. squamosus* may also serve as a vector of malaria parasites.[96] Seven specimens in this study have been tentatively classified as *An. squamosu*s based on morphological and molecular data. Three of them match closely with a sequence known as *An. sp. 15* (GenBank Accession: KJ522843), which we also tentatively classify as *An. squamosus* on the basis of both morphological identification and molecular analysis. Given the close phylogenetic relationship between *An. squamosus* and *An. sp. 15*, further investigations on the role of these species as secondary vectors of malaria are warranted.

An. coustani and closely-related species *An. ziemanni* and *An. paludis* have likewise increased in notoriety as potentially important malaria vectors in sub-Saharan Africa. In some cases, these species have served as major vectors in some areas, ranging from the Cameroon and the Central Africa Republic to Kenya and Madagascar.[97,209,210,230] One reason these species have been overlooked for so long is that they were originally known as primarily exophagic and zoophilic mosquitoes. Reports of high degrees of anthropophily indicate that in some regions or potentially in some cryptic populations, their vectorial capacity is much higher.[231] The *An. coustani* specimens within this study fell into three highly supported clusters. One subgroup may be *An. coustani s.s.*, while the other two clusters may represent either subpopulations or distinct but highly related species, perhaps within a single species complex.

An. maculipalpis, according to Gillies and DeMeillon, is generally a low-abundance species found throughout savannah- and tropical-type environments in Africa that tends to be zoophilic and rest outdoors.[149] It has never been implicated as a disease vector of any significant importance. Five samples from our study were identified as *An. maculipalpis* through molecular analysis, though only one of these six was morphologically identified as such. The others were morphologically identified as unidentified (2) and *An. coustani* (3).

Specimens which lack definitive species identification from this study (*An. sp. 11*, UG1-3) represent unknown anopheline populations which may possibly be important for

maintaining malaria transmission in unrecognized transmission cycles. Proper identification of specimens such as these require not only additional field material for more accurate morphology, but corresponding genetic data from taxonomically-verified specimens. For example, UG2 forms a relatively low support cluster with *An. darlingi*, a neotropical species and genus (*Nyssorhynchus*). This is likely an artifact of the relatively short sequence fragment used in our phylogenetic analyses and absence of corresponding genetic data from understudied African taxa in existing databases. Further study must be undertaken to properly document these populations to determine if they represent previously-undocumented species or are simply species for which we lack genomic information.

Limitations

The relationships of the anopheline subgenera to one another remain unclear and somewhat contentious.[211,229,232–234] Studies based on combinations of nuclear and mitochondrial DNA as well as amino acid sequence and morphological characters show that sections/series and even subgenera are para- or polyphyletic. This may be unsurprising, as the original taxonomic classification of anopheline mosquitoes was hypothesized based largely on morphological characters. As closely-related anopheline species can be morphologically distinct, and distantly-related species remarkably similar, morphological classification may suffer from some degree of inaccuracy.

There have been relatively few molecular phylogenetic studies of anopheline mosquitoes at a broad geographic scale. One recent study used full mitochondrial genomes to analyze the phylogenetics of *Anophelinae* below the genus level.[211] Even with the much more extensive sequence data that was used in this study, there was low support when using nucleotide sequences, and therefore the corresponding amino acid translations had to be utilized for analysis. Due to the relatively short sequence we used, amino acid translation was not a more discriminatory approach in our analyses.

Though the COI BOL is among the most common targets used for phylogenetic analysis in this group of organisms, its utility is likely limited to comparing relatively closely related species. Reports have been mixed with regard to the useful phylogenetic signal in COI for comparing subgenera within *Anopheles*.[226,228] To more accurately place ambiguous groups from this study, sequencing using a different or multiple targets would be helpful. ND5 from the mitochondrial genome, along with D2 from the nuclear genome, have been successfully used to resolve relationships at the subgenus level.[228,235] In addition, ITS2 is a very common locus that might be a useful addition and validation of COI-based phylogenetics.

Morphological misidentification remains a problem, even for experienced investigators. Misidentification of anophelines for *An. gambiae* specimens in this small sample set was common with 39% (9/23) of specimens morphologically identified as *An. gambiae* molecularly identified as something else. Comparatively, only 6.2% of morphologically identified *An. funestus* were misidentified. A high proportion of the remaining

populations and species were morphologically mis- or unidentified, which is likely due to inexperience with identification of relatively rare species of anophelines as well as damage to specimens during collection. More extensive documentation of rare species, including verified voucher specimens for comparison, would be of great benefit to malaria researchers and vector biologists.

Conclusions

By going beyond standard PCR assays for speciation of samples and conducting phylogenetic analysis, this study was able to show an unprecedented diversity of anophelines in Nchelenge District, Zambia. Several of these anophelines represent species known to be important for malaria transmission in other areas. At such low numbers in our collection, it is impossible to conclude that they represent more than potential minor contributors to transmission in Nchelenge District. Other specimens in this study remain unverified and represent either unnamed species or named species which have yet to be genetically characterized. Future taxonomic efforts are clearly needed to link anopheline morphology to genomic data.

Figures

Figure 1: Unrooted Bayesian tree (BEAST2) constructed from a 488 bp multi-alignment of 196 unique anopheline haplotypes of the BOL COI sequence. Tree shows subgenus-level taxonomy by tip color (as shown in legend). White circles indicate study samples while black circles indicate either *Bironella*, *Chagasia*, *Culex*, or unassigned taxonomy (ie "*An. sp.*"). Nodes under 50% posterior probability are collapsed to polytomies, with remaining posterior probabilities >50% indicated.

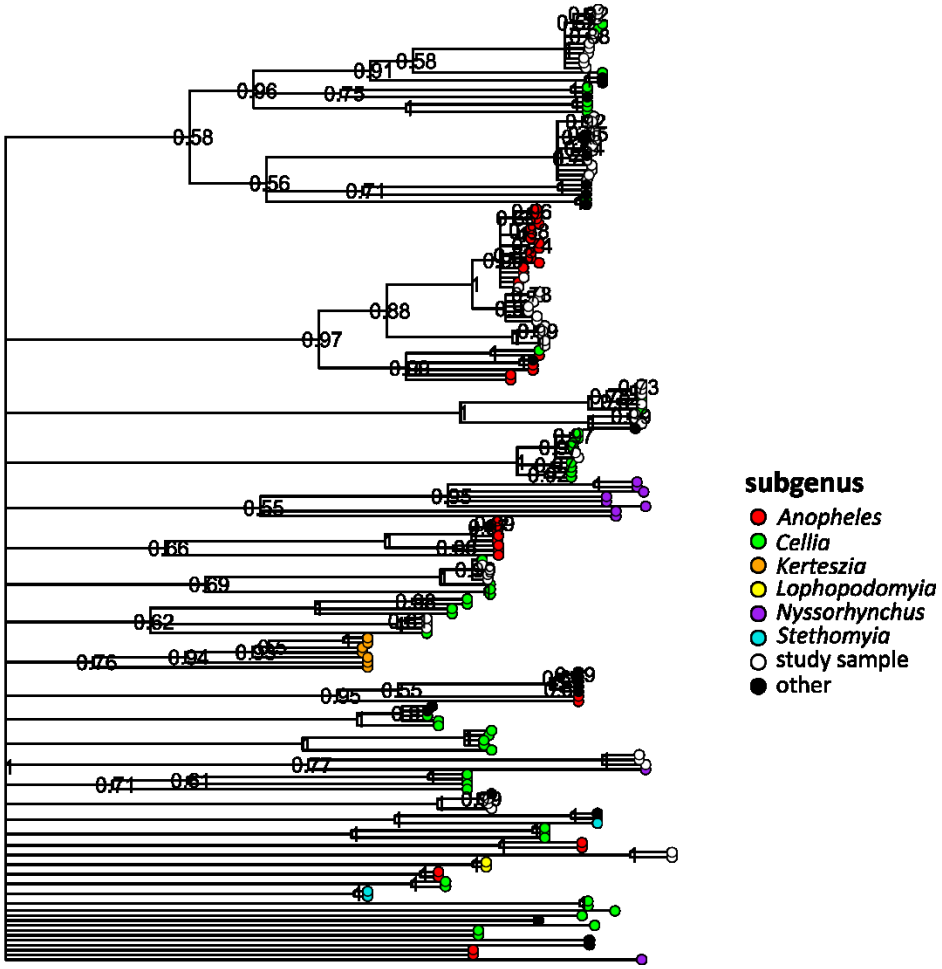


Figure 2: Highlighted clade detail of Bayesian tree (Figure 1 is the left of this figure), showing study specimens falling into a clade that includes *An. funestus*, *An. sp. 6*, and *An. sp. 14* groups, as well as other members of *Cellia Myzomyia*. Posterior probabilities above 0.5 are indicated. For study haplotypes having multiple members, the number of members is indicated in parenthesis.

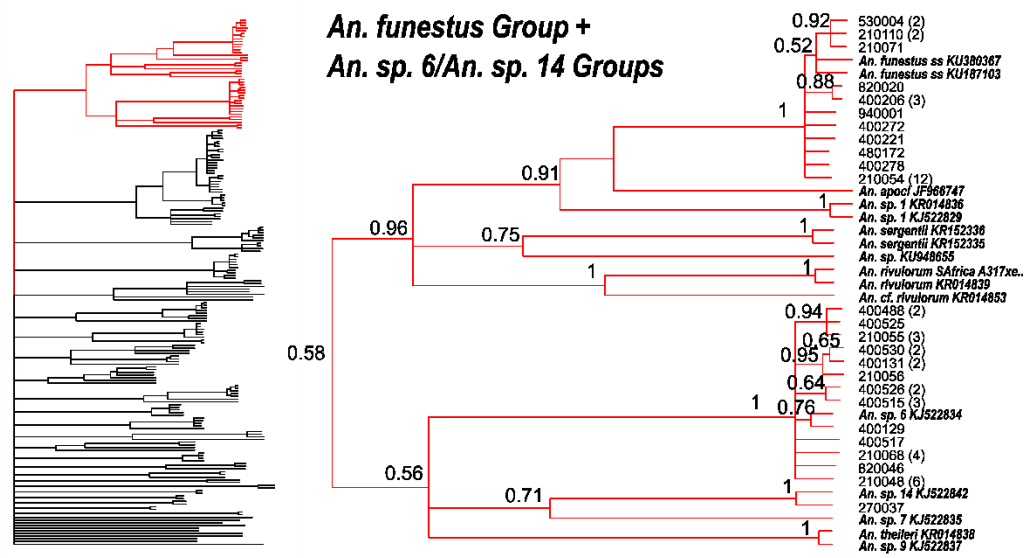


Figure 3: Highlighted clade detail of Bayesian tree (Figure 1 is the left of this figure), showing specimens falling into a clade that includes *An. maculipalpis*. Posterior probabilities above 0.5 are indicated. For study haplotypes having multiple members, the number of members is indicated in parenthesis.

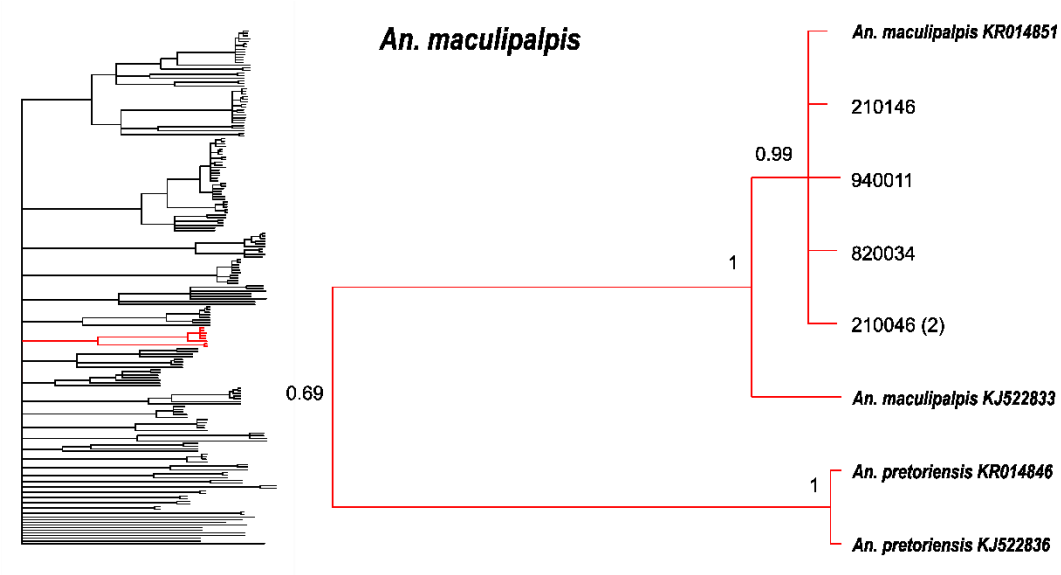


Figure 4: Highlighted clade detail of Bayesian tree (Figure 1 is the left of this figure), showing specimens falling into a clade that includes *An. squamosus* and *An. sp. 15*. Posterior probabilities above 0.5 are indicated. ** = NCBI sample considered "unverified"

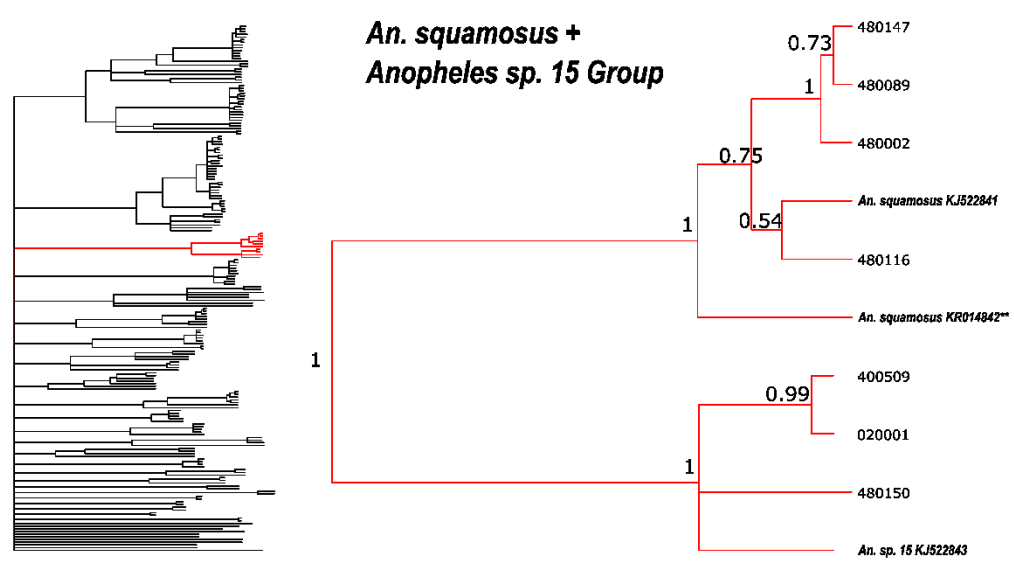


Figure 5: Highlighted clade detail of Bayesian tree (Figure 1 is the left of this figure), showing specimens falling into a clade that includes species from *Cellia Myzomyia*. Posterior probabilities above 0.5 are indicated.

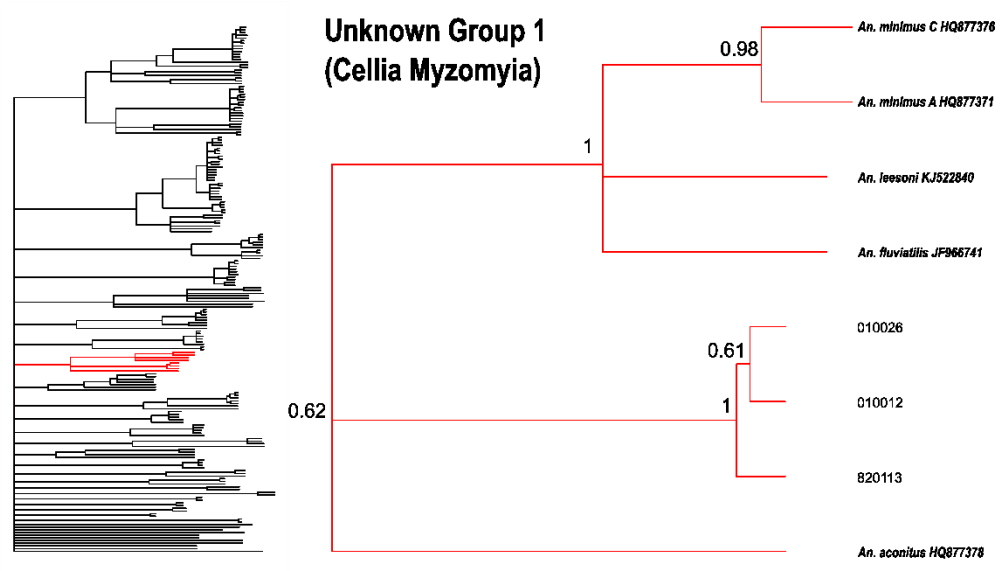


Figure 6: Highlighted clade detail of Bayesian tree (Figure 1 is the left of this figure), showing specimens falling into a clade that includes species from *Coustani* and *Hyrcanus* Groups. Posterior probabilities above 0.5 are indicated. For study haplotypes having multiple members, the number of members is indicated in parenthesis.

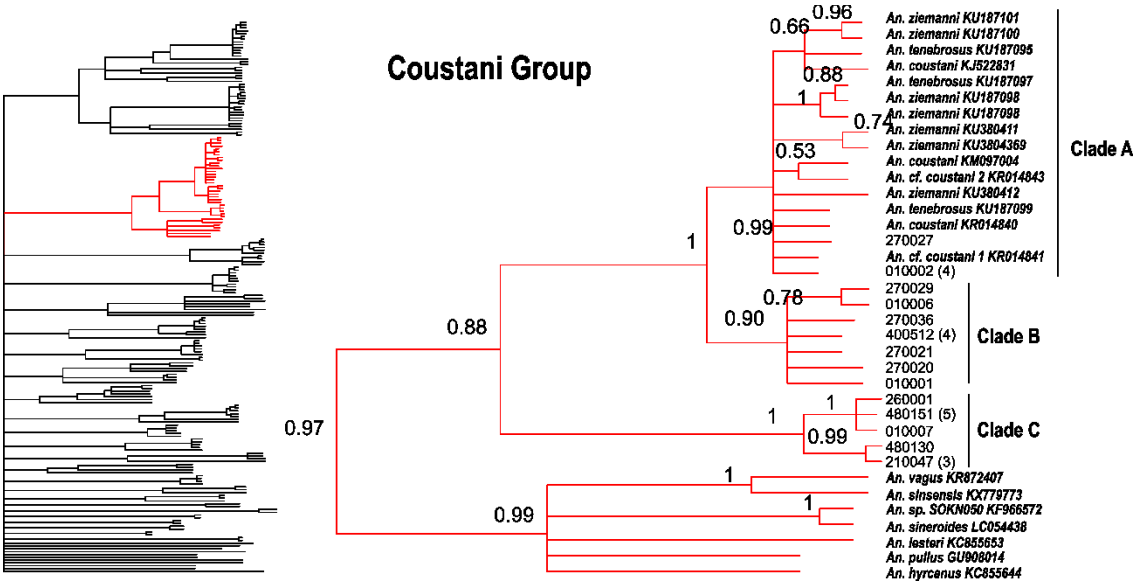


Figure 7: Highlighted clade detail of Bayesian tree (Figure 1 is the left of this figure), showing specimens falling into a clade that includes unassigned species *An. sp. 11*. Posterior probabilities above 0.5 are indicated.

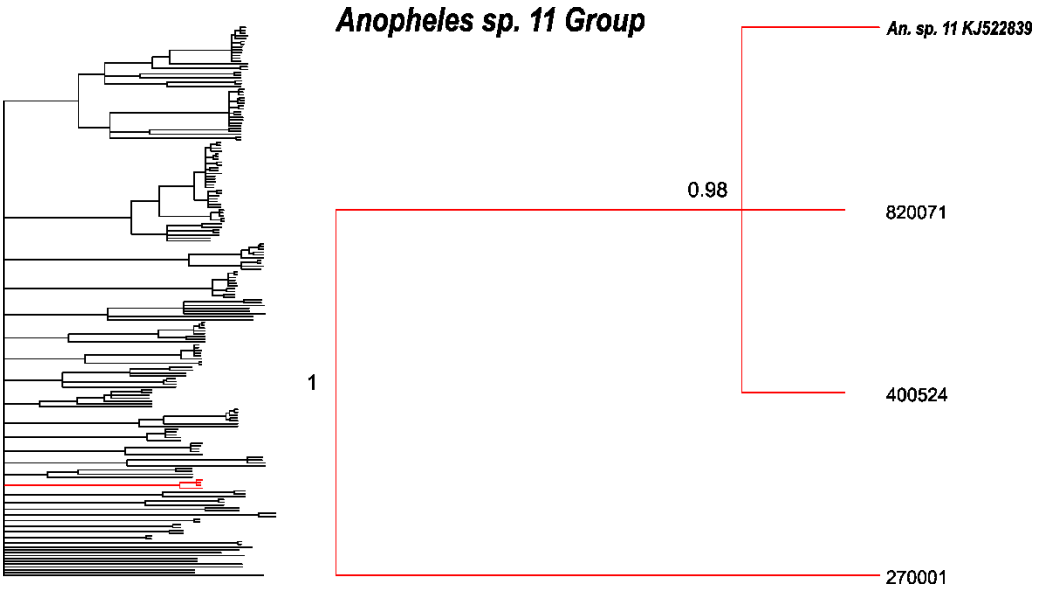


Figure 8: Highlighted clade detail of Bayesian tree (Figure 1 is the left of this figure), showing specimens falling into a clade that includes *An. darlingi*. Posterior probabilities above 0.5 are indicated. For study haplotypes having multiple members, the number of members is indicated in parenthesis.

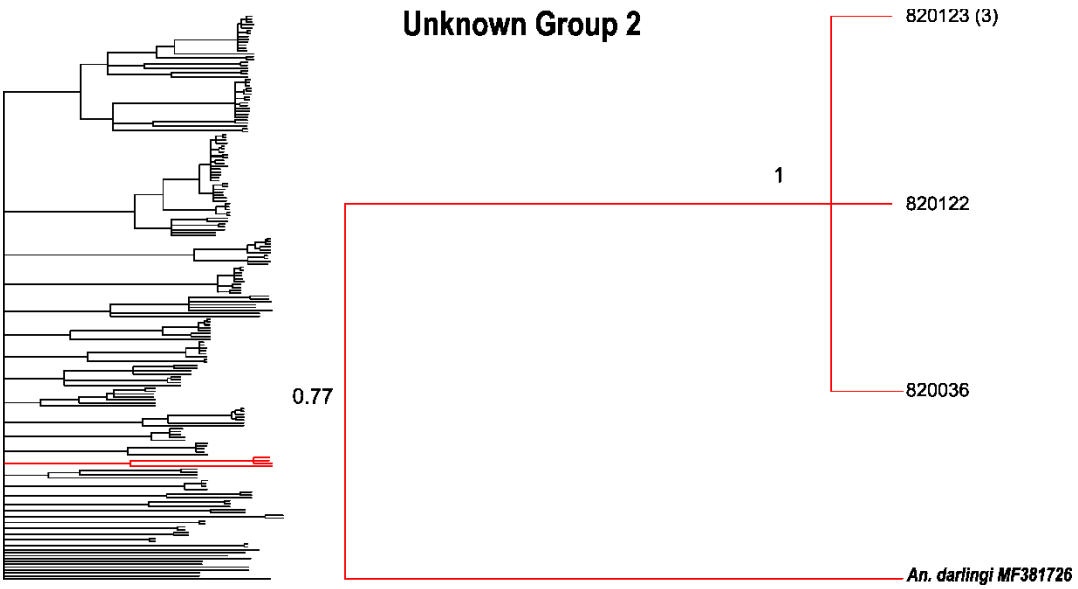


Figure 9: Highlighted clade detail of Bayesian tree (Figure 1 is the left of this figure), showing specimens that are not consistent with any other sequences from this analysis. Posterior probabilities above 0.5 are indicated. For study haplotypes having multiple members, the number of members is indicated in parenthesis.

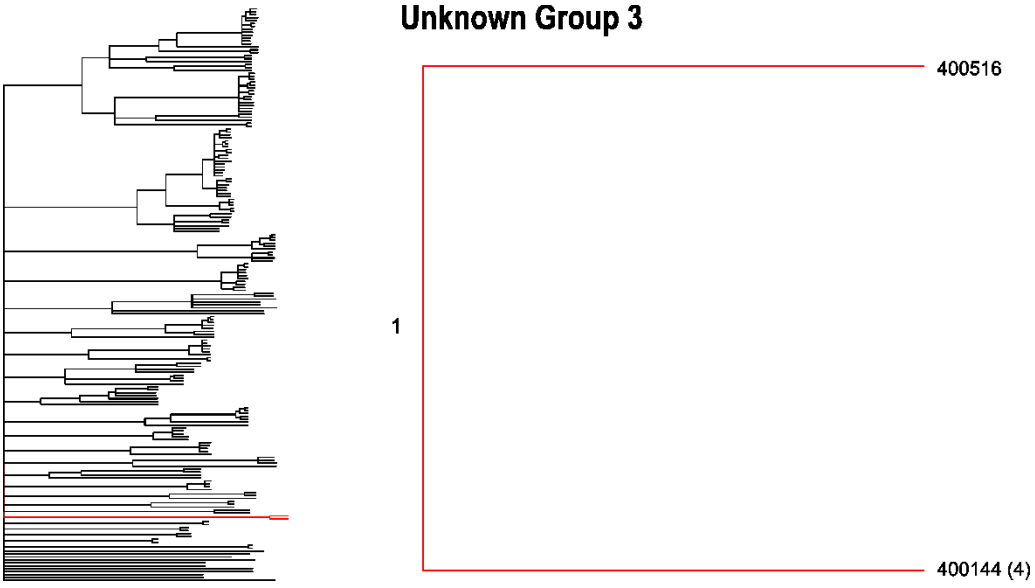
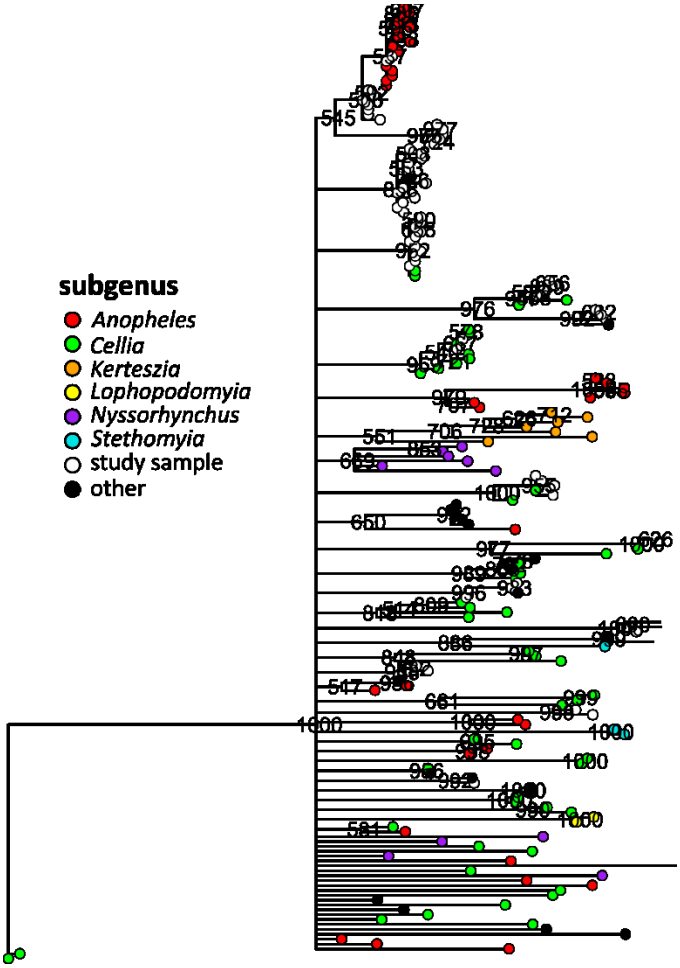


Figure S1: Unrooted Maximum Likelihood tree (phyML) showing subgenus-level taxonomy by tip color (as shown in legend on bottom right). White circles indicate study samples while black circles indicate either *Bironella*, *Chagasia*, *Culex*, or unassigned taxonomy (ie "*An. sp.*"). Nodes under 500 bootstraps collapsed to polytomies; the remaining bootstrap support values (out of 1000) are indicated.



Tables

Table 1: Phylogenetic groups confirmed through PCR and sequencing (N=747).

Phylogenetic Group (# PCR-confirmed members)	Primary Morphological ID (% of total)	# COI sequence	% COI Seq Identity (within group)	COI best match (% Sequence Identity)	ITS2 best match (# sequences)	Final Classification	Hypothesized Taxonomy?
A (644)	*An. funestus* (99%)	26	>98%	*An. funestus* (>99%)	*An. funestus* (1)	*An. funestus*	
B (31)	*An. funestus* (35%)	29	>98%	*An. sp. 6* (>98%); *An. theileri* (~94%)	*An. sp. 6* (11)	*An. sp. 6*	Subgenus *Cellia*, Series *Myzomyia*, *Wellcomei* Group
C (28)	*An. coustani* (100%)	28	>95%	*Coustani* Group (>95%)	*An. coustani/An. cf. coustani* (24)	An. coustani Group	
D (14)	*An. gambiae* (86%)	3	99.8% (2 haplotypes)	*An. gambiae s.l.* (>99%)	*An. gambiae* (2)	*An. gambiae*	
E (6)	*An. coustani* (50%)	5	>98%	*An. maculipalpis* (>97%)	*An. maculipalpis* (2)	*An. maculipalpis*	
F (5)	*An. brunnipes* (67%)	5	0.98	*An. darlingi* (~91%)/*An. tenebrosus* (~90%)	NA (4)	Unknown Group 2	Unknown (Anopheles sp.)
G (5)	*An. funestus* (50%)	5	98% (2 haplotypes)	*An. theileri*/*An. sp. 6* (~91%)	NA (3)	Unknown Group 3	
H (4)	*An. squamosus* (100%)	4	>98%	*An. squamosus* (>97%)	*An. sp. 16* (1)	*An. squamosus*	
I (3)	*An. squamosus* (100%)	3	>98%	*An. Squamosus*/*An. sp. 15* (>97%)	*An. sp. 16* (3)		
J (3)	*An. squamosus* (100%)	3	>98%	*An. sp. 11* (>98%)	*An. sp. 11* (2)	*An. sp. 11*	Unknown (*Anopheles sp.*)
K (3)	*An. funestus* (67%)	3	>99%	*An. aconitus*/*An. minimus C* (~94%)	NA (3)	Unknown Group 1	Subgenus *Cellia*, Series *Myzomyia*, *Funestus* Group
L (1)	*An. gambiae*	1	NA	*An. sp. 14* (99%)/*An. theileri* (94%)	*An. sp. 14* (1)	*An. sp. 14*	Subgenus *Cellia*, Series *Myzomyia*, *Wellcomei* Group

Table S1: Accession numbers and genus- and subgenus- level taxonomy for NCBI sequences used in analyses.

Genus	Subgenus	Species	Accession
Anopheles	-	*An. sp.*	KU948655
		An. sp.	KP045761
		An. sp.	MF821611
		An. sp.	KR519109
		An. sp.	KR774514
		An. sp.	MF828325
		An. sp. 1	KJ522829
		An. sp. 1	KR014836
		An. sp. 11	KJ522839
		An. sp. 14	KJ522842
		An. sp. 15	KJ522843
		An. sp. 6	KJ522834
		An. sp. 7	KJ522835
		An. sp. 9	KJ522837
		An. sp. M36YA	KU187107
		An. sp. MBI-14	KM097029
		An. sp. SOKN050	KF966572
		An. sp. TIP1	KF671003
		An. sp. TIP1	KF671004
	Anopheles	*An. labranchiae*	HQ860331
		An. aff. peryassui	MG701391
		An. costai	MF381614
		An. costai	MF381631
		An. costai	MF381607
		An. costai	NC_037794
		An. costai	KF698865
		An. forattinii	NC_037813
		An. malefactor	KF698839
		An. minor	NC_037802
		An. minor	MF381684
		An. nr. costai	NC_037821
		An. saperoi	AB738150
		An. saperoi	AB738176
		An. barbirostris	AB435998
		An. campestris	AB331588
		An. cf. coustani 1	KR014841
		An. cf. coustani 2	KR014843
		An. coustani	KR014840
		An. coustani	KM097004
		An. coustani	KJ522831

		An. tenebrosus	KU187096
		An. tenebrosus	KU187097
		An. tenebrosus	KU187099
		An. tenebrosus	KU187095
		An. ziemanni	KU187098
		An. ziemanni	KU380369
		An. ziemanni	KU380412
		An. ziemanni	KU380411
		An. ziemanni	KU187101
		An. ziemanni	KU187100
		An. hyrcanus	KC855644
		An. sineroides	LC054438
		An. sinsensis	KX779773
		An. pullus	GU908014
		An. lesteri	KC855653
	Cellia	*An. pharoensis*	KR014844
		An. pharoensis	KU380470
		An. pharoensis	KU380430
		An. pharoensis	KU380435
		An. squamosus	KU187110
		An. squamosus	KU187111
		An. squamosus	KJ522841
		An. squamosus	KR014842
		An. apoci	JF966747
		An. sergentii	KR152335
		An. sergentii	KR152336
		An. jeyporiensis	HQ877379
		An. aconitus	HQ877378
		An. pampanai	HQ877381
		An. varuna	HQ877380
		An. culicifacies	KF406658
		An. funestus ss	KU380367
		An. funestus ss	KJ522832
		An. funestus ss	KU187103
		An. funestus ss	KU380404
		An. rivulorum	KR014839
		An. rivulorum	MK044802
		An. cf. rivulorum	KR014853
		An. leesoni	KJ522840
		An. fluviatilis	JF966741
		An. minimus A	HQ877371
		An. minimus C	HQ877376
		An. theileri	KR014838
		An. maculipalpis	KJ522833
		An. maculipalpis	KR014851

		An. pretoriensis	KJ522836
		An. pretoriensis	KR014846
		An. rufipes	KR014845
		An. rufipes	KJ522838
		An. annularis	KF406653
		An. annularis	KF406654
		An. splendidus	KF406678
		An. nr. dravidicus	KF406679
		An. nili	KR014837
		An. cracens	JX219733
		An. dirus A	JX219731
		An. farauti	JX219741
		An. farauti	JX219735
		An. farauti	JX219736
		An. rhodesiensis	KU187106
		An. vagus	KR872407
		An. arabiensis	KY670610
		An. coluzzii	KR152322
		An. arabiensis	KJ522830
		An. gambiae ss	KP980693
		An. gambiae ss	DQ792578
		An. gambiae ss	KU187109
		An. melas	DQ792579
		An. melas	DQ792580
		An. quadriannulatus	KR014849
		An. quadriannulatus	DQ792581
		An. subpictus	KC970278
		An. subpictus	KF406734
	Kerteszia	*An. bellator*	KU551287
		An. cruzii	KU551284
		An. cruzii	KU551285
		An. cruzii	KU551286
		An. cruzii	KU551289
		An. homunculus	MF381605
		An. laneanus	MF381613
		An. laneanus	NC_030250
	Lophopodomyia	*An. squamifemur*	KU900771
		An. squamifemur	KU900772
	Nyssorhynchus	*An. nr. konderi*	KF809141
		An. nr. konderi	KF809142
		An. oswaldoi	KF809131
		An. albertoi	MF381634
		An. benarrochi	MF381588
		An. albitarsis	KJ492432
		An. argyritarsis	NC_037807

		An. darlingi	MF381726
		An. guarani	NC_037816
		An. pristinus	NC_037821
	Stethomyia	*An. kompi*	NC_037827
		An. nimbus	NC_037811
		An. nimbus	HM022409
Bironella		*Bironella hollandi*	NC_03307796
Chagasia		*Chagasia. sp.* SP54_C	MF381717
Culex		*Culex. aff. dolosus*	MF381620

Table S2: Haplotypes occurring more than once. Includes both study specimens and sequences downloaded from NCBI.

Haplotype 'Name'	# sequences	Sequences belonging to haplotype
210046	2	210046 820127
210110	2	210110 480156
400131	2	400131 820130
400488	2	400488 820129
400526	2	400526 940003
400530	2	400530 820128
530004	2	530004 940006
An_funestus_ss_KU187103	2	An_funestus_ss_KU187103 An_funestus_ss_KU380404
An_laneanus_MF381613	2	An_laneanus_MF381613 An_laneanus_NC_030250
An_squamosus_KU187110	2	An_squamosus_KU187110 An_squamosus_KU187111
210047	3	210047 480090 480152
210055	3	210055 400518 400522
400206	3	400206 400254 400519
400515	3	400515 400520 820072

480129	3	480129 480148 820121
820123	3	820123 820125 820126
10002	4	010002 260003 260005 400527
10016	4	010016 020007 An_arabiensis_KY670610 An_gambiae_ss_KU187109
210068	4	210068 400187 400523 820054
400144	4	400144 480149 820041 820042
400512	4	400512 260002 260006 270003
480151	5	480151 010014 010015 400513 820015
210048	6	210048 210057 210120 400036 400180 820124

210054	13	210054 210137 400008 400034 400065 400216 400248 400256 400266 480171 820114 820119 An_funestus_ss_KJ522832

Chapter V

Conclusions

This dissertation is the product of the first intensive studies of outdoor behavior and anopheline genetics and genomics in Nchelenge District of northern Zambia. *Anopheles funestus* mosquitoes were collected from Nchelenge and additional sites in the Democratic Republic of the Congo (DRC) and Tanzania and subjected to whole genome sequencing and subsequent population genomics analysis. Additional studies in Nchelenge examined populations of anopheline mosquitoes caught outdoors, with a secondary objective of examining the relative efficacy of Centers for Disease Control Light Trap placement/attractant schemes for sampling outdoor-foraging populations of anophelines that may be involved in residual malaria transmission. This study characterized the anophelines caught in these outdoor collections in terms of their distribution, abundance, infection rates, and species diversity. A subset of anophelines from this outdoor collection was subject to phylogenetic analysis for the purposes helping to describe taxonomical placement and species identification.

Nchelenge lies within Luapula Province, which remains one of the most malarious provinces in Zambia, despite vector control and other malaria control efforts having been instituted in the region since at least 2007. Though other studies have examined the impact of human movement and risk factors, only relatively recently have the basic vector bionomics in Nchelenge been investigated. Starting in 2011, entomological collections conducted by the Southern Africa International Centers for Excellence in

Malaria Research (ICEMR) showed that *An. funestus* and *An. gambiae* are the primary vectors that drive transmission in this region, with *An. funestus* having a much higher relative abundance and therefore a higher contribution to the malaria burden. Only a handful of other anopheline species have ever been documented in the area.

Additional studies revealed different ecological preferences for the two species. *An. gambiae* being more common during the wet season while *An. funestus* was found year-round, and especially prevalent in inland collections. Both species showed a preference for human blood meals, and both had significant rates of infection with *P. falciparum* parasites.[112,113] Studies of vector biology in Nchelenge have not ranged much beyond establishing basic spatio-temporal trends and confirming the vector status of these two species. Important questions remain unanswered, including whether outdoor vector activity has partially nullified the protection of indoor-focused vector control, if there are distinct populations within what is recognized as *An. funestus s.s.* which are relevant to transmission, and whether population genetics and genomics can be useful as tools for ongoing surveillance and identification of Nchelenge vector populations.

Pursuit of population genetics as a tool to examine dynamics within the primary vector *An. funestus* revealed a scarcity of appropriate markers for central and southern African populations of *An. funestus*. In an effort to provide the baseline data necessary for selecting and developing such markers, 43 *An. funestus* specimens from Nchelenge (N=28), a site in southeastern DRC (N=5), and a site in southern Tanzania (N=10), were whole genome sequenced. We hypothesized that the three geographically distant sites

would represent distinct populations of *An. funestus*, but analysis found no significant genetic difference between the DRC and Zambia samples (~ 270 km distant), and only mild differentiation between Zambia and Tanzania (~ 875 km distant). This shallow differentiation was accompanied by larger than expected sequence diversity in the overall collection including within sites. These results could be both potentially good and bad news for malaria control: on the one hand, a single panmictic population is theoretically more tractable in terms of transgenic release and population replacement strategies (though the high diversity despite low differentiation might nullify the benefit), but on the other, one large population with high diversity is an ideal situation for adaptation, including broad population acquisition and spread of novel insecticide resistance polymorphisms.

Population genomics analysis confirmed the presence of two distinct mitochondrial lineages in *An. funestus* in Nchelenge, at a steady proportion of approximately 80% lineage I and 20% lineage II. Demographic and phylogenetic analyses determined that these two populations likely diverged half a million years ago, and that the total *An. funestus* population may have experienced a minor population burst around 80,000 years ago. An established high-throughput TaqMan assay for detection of lineages (based on a single SNP) was shown to have lower sensitivity than expected due to a clade of lineage II retaining the lineage I allele. Approximately 50 other single nucleotide polymorphisms had fixed differences between the two lineages in the mitogenomics study, which provides possibilities for a more sensitive assay to be developed. 567 polymorphic sites

were detected among the mitochondrial genomes sequenced; these can serve as a pool of potential population genetic markers for future studies in southern and central Africa.

An additional question aimed at residual malaria focused on the use of alternative placement/baiting strategies for outdoor collection of anophelines in Nchelenge. Human landing catches (HLCs) have long been considered the gold standard for outdoor collections of human-foraging mosquitoes, but are logistically demanding and suffer from ethical concerns. Though CDC light traps (CDC LTs) are a common alternative method, they must be baited. In Nchelenge, the CO_2 generally used for baiting purposes is not available. The purpose of our study was twofold: 1) characterize outdoor-caught anopheline populations, and 2) determine if a human analogue bait would serve as a standardized bait when added to CDC LTs

Promisingly, we caught similar numbers of female anophelines in each of our attractant schemes. However, the results from our relatively small-scale study do not necessarily indicate that the artificial bait alone would provide significant additional attraction of outdoor CDC LTs for anophelines, as we lacked a control BG-Lure® CDC LT set in a context without lingering human odorants present. The average count of anophelines in any given trap was most highly related to the location of capture (inland vs lakeside households). Exploratory analysis of additional covariates showed smaller counts with each additional person outside of a bed net during the night, which is consistent with the idea of declining incentives to use LLINs (the less often you're bitten, the less you care to sleep under a net).

It was also shown that the overall diversity of anophelines measured as the raw number of species per trap did not statistically significantly differ when comparing attractant scheme. However, distributions of particular species did show varying trends, including an apparent preference for human-baited trap for *An. gambiae*, exclusively inland capture of *An. sp. 6*, and primarily inland capture for *An. squamosus*. *An. funestus* had higher absolute numbers in inland communities, though it was found everywhere, with no apparent preference for trap placement. *Plasmodium* positive samples were found in inland households in *An. funestus* exclusively. This partially reflects the dominance of *An. funestus* specimens in all collections, as well as the much higher yield of anophelines inland. With low absolute numbers of the other species, including *An. gambiae*, it is unclear what role these species serve in transmission. Clearly, however, this is a community burdened by a large number of mosquito bites (some infectious), and a diverse array of anopheline species, despite widespread use of standard vector control.

Unidentified anophelines from the aforementioned outdoor study were investigated using molecular phylogenetics tools. Knowledge of which species are present regionally is important for implementation of effective malaria interventions, and this study unveiled a surprisingly large proportion of anophelines which were neither *An. funestus* nor *An. gambiae*. This high diversity may be attributed to a mixture of a few factors, the most likely of which are 1) all prior collections had been focused indoors, 2) prior misidentification of rarer anophelines, and 3) three consecutive indoor residual spray

campaigns inducing enough suppression of the *An. funestus* population that rarer species are now visible.

Sequencing of a widely-used speciation target, the Barcode of Life (BOL) fragment of the cytochrome oxidase c subunit I (COI) gene, allowed for phylogenetic placement and verification or putative identification of some of these diverse specimens. *Anopheles coustani* and *An. squamosus*, which have been implicated in malaria transmission in some areas, were confirmed molecularly and shown to have internal phylogenetic clade structure of unknown biological significance, which will require additional investigation. Other specimens remain somewhat mysterious, either because they represent undocumented species, or because they represent species for which there is little to no genomic information. One of these groups, *An. sp. 6*, was the second-most commonly collected anopheline in our outdoor collection study (N=29), present in higher numbers than *An. gambiae* (N=10) and eclipsed only by the dominant *An. funestus* (N=623). Our morphological identifications of *An. sp. 6* show it was commonly misidentified as either *An. funestus* or a related species, and it lies phylogenetically within the same taxonomical Series as *An. funestus,* which likely supports the morphological similarity. Further studies should include extensive photographic/morphological documentation of unusual specimens and an archive of voucher specimens.

Altogether the works of this dissertation build toward a better understanding of the entomological drivers of transmission in highly endemic Nchelenge District, and represent a large step forward in our understanding of the genomics of *An. funestus,* one

of the most important vectors of human malaria in sub-Saharan Africa. It also builds towards a better understanding of African anopheline diversity, distribution and genetics, which could lead to more effective targeting of malaria interventions.

References

1. World Health Organization (WHO). *World malaria report*. (2017).
2. Kassebaum, N. J. *et al.* Global, regional, and national disability-adjusted life-years (DALYs) for 315 diseases and injuries and healthy life expectancy (HALE), 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* **388,** 1603–1658 (2016).
3. Bhatt, S. *et al.* The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. *Nature* (2015). doi:10.1038/nature15535
4. World Health Organization (WHO). *World malaria report*. (2016).
5. Mendis, K. *et al.* From malaria control to eradication: The WHO perspective. *Trop. Med. Int. Heal.* **14,** 802–809 (2009).
6. Guyatt, H. L. & Snow, R. W. The cost of not treating bednets. *Trends Parasitol.* **18,** 12–6 (2002).
7. Kleinschmidt, I. *et al.* Implications of insecticide resistance for malaria vector control with long-lasting insecticidal nets: a WHO-coordinated, prospective, international, observational cohort study. *Lancet Infect. Dis.* **18,** 640–649 (2018).
8. Toé, K. H. *et al.* Increased pyrethroid resistance in malaria vectors and decreased bed net effectiveness, Burkina Faso. *Emerg. Infect. Dis.* **20,** 1691–6 (2014).
9. Killeen, G. F. *et al.* Preventing Childhood Malaria in Africa by Protecting Adults from Mosquitoes with Insecticide-Treated Nets. *PLoS Med.* **4,** e229 (2007).
10. Killeen, G. F. *et al.* The importance of considering community-level effects when selecting insecticidal malaria vector products. *Parasit. Vectors* **4,** 160 (2011).
11. Gatton, M. L. *et al.* The importance of mosquito behavioural adaptations to malaria control in Africa. *Evolution* **67,** 1218–30 (2013).
12. Global Malaria Programme. *Indoor residual spraying: use of indoor residual spraying for scaling up global malaria control and elimination*. (2006).
13. Walker, K. Cost-comparison of DDT and alternative insecticides for malaria control. *Med. Vet. Entomol.* **14,** 345–354 (2000).
14. Pluess, B., Tanser, F. C., Lengeler, C. & Sharp, B. L. Indoor residual spraying for preventing malaria. *Cochrane Database Syst. Rev.* CD006657 (2010). doi:10.1002/14651858.CD006657.pub2
15. Lengeler, C. Insecticide-treated bed nets and curtains for preventing malaria. in *Cochrane Database of Systematic Reviews* (2004). doi:10.1002/14651858.CD000363.pub2
16. Corbel, V. & N'guessan, R. Distribution, mechanisms, impact and management of insecticide resistance in malaria vectors : A pragmatic review. in *Anopheles mosquitoes - New insights into malaria vectors* (ed. Sylvie Manguin) 579–633

(InTech, 2013). doi:10.5772/56117/3392

17. Hemingway, J. & Ranson, H. Insecticide Resistance in Insect Vectors of Human Disease. *Annu. Rev. Entomol.* **45,** 371–391 (2000).

18. Donnelly, M. J. *et al.* Does kdr genotype predict insecticide-resistance phenotype in mosquitoes? *Trends Parasitol.* **25,** 213–219 (2009).

19. Weill, M. *et al.* The unique mutation in ace-1 giving high insecticide resistance is easily detectable in mosquito vectors. *Insect Mol. Biol.* **13,** 1–7 (2004).

20. N'Guessan, R. *et al.* Resistance to carbosulfan in *Anopheles gambiae* from Ivory Coast, based on reduced sensitivity of acetylcholinesterase. *Med. Vet. Entomol.* **17,** 19–25 (2003).

21. Donnelly, M. J., Isaacs, A. T. & Weetman, D. Identification, Validation, and Application of Molecular Diagnostics for Insecticide Resistance in Malaria Vectors. *Trends Parasitol.* **32,** 197–206 (2016).

22. Sougoufara, S. *et al.* Biting by *Anopheles funestus* in broad daylight after use of long-lasting insecticidal nets: a new challenge to malaria elimination. *Malar. J.* **13,** 125 (2014).

23. Moiroux, N. *et al.* Changes in *Anopheles funestus* biting behavior following universal coverage of long-lasting insecticidal nets in Benin. *J. Infect. Dis.* **206,** 1622–9 (2012).

24. Russell, T. L. *et al.* Increased proportions of outdoor feeding among residual malaria vector populations following increased use of insecticide-treated nets in rural Tanzania. *Malar. J.* **10,** 80 (2011).

25. Russell, T. L. *et al.* Anopheles farauti is a homogeneous population that blood feeds early and outdoors in the Solomon Islands. *Malar. J.* **15,** 151 (2016).

26. Yahouédo, G. A. *et al.* Contributions of cuticle permeability and enzyme detoxification to pyrethroid resistance in the major malaria vector *Anopheles gambiae*. *Sci. Rep.* **7,** 11091 (2017).

27. Wood, O., Hanrahan, S., Coetzee, M., Koekemoer, L. & Brooke, B. Cuticle thickening associated with pyrethroid resistance in the major malaria vector *Anopheles funestus*. *Parasit. Vectors* **3,** 67 (2010).

28. Müller, P. *et al.* Field-Caught Permethrin-Resistant *Anopheles gambiae* Overexpress CYP6P3, a P450 That Metabolises Pyrethroids. *PLoS Genet.* **4,** e1000286 (2008).

29. Müller, P., Donnelly, M. J. & Ranson, H. Transcription profiling of a recently colonised pyrethroid resistant *Anopheles gambiae* strain from Ghana. *BMC Genomics* **8,** 36 (2007).

30. David, J. P. *et al.* The *Anopheles gambiae* detoxification chip: A highly specific microarray to study metabolic-based insecticide resistance in malaria vectors.

Proc. Natl. Acad. Sci. **102,** 4080–4084 (2005).

31. Riveron, J. M. *et al.* Directionally selected cytochrome P450 alleles are driving the spread of pyrethroid resistance in the major malaria vector *Anopheles funestus*. *Proc. Natl. Acad. Sci.* **110,** 252–257 (2012).

32. Wondji, C. S. *et al.* Mapping a quantitative trait locus (QTL) conferring pyrethroid resistance in the African malaria vector *Anopheles funestus*. *BMC Genomics* **8,** 34 (2007).

33. Liu, N. Insecticide resistance in mosquitoes: impact, mechanisms, and research directions. *Annu. Rev. Entomol.* **60,** 537–59 (2015).

34. Ranson, H. *et al.* Evolution of supergene families associated with insecticide resistance. *Science* **298,** 179–81 (2002).

35. Hemingway, J., Hawkes, N. J., McCarroll, L. & Ranson, H. The molecular basis of insecticide resistance in mosquitoes. *Insect Biochem. Mol. Biol.* **34,** 653–665 (2004).

36. Denholm, I., Devine, G. J. & Williamson, M. S. Evolutionary genetics: Insecticide resistance on the move. *Science* **297,** 2222–2223 (2002).

37. Ranson, H. & Lissenden, N. Insecticide Resistance in African *Anopheles* Mosquitoes: A Worsening Situation that Needs Urgent Action to Maintain Malaria Control. *Trends Parasitol.* **32,** 187–196 (2016).

38. Norris, L. C. *et al.* Adaptive introgression in an African malaria mosquito coincident with the increased usage of insecticide-treated bed nets. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 815–20 (2015).

39. Main, B. J. *et al.* Complex genome evolution in *Anopheles coluzzii* associated with increased insecticide usage in Mali. *Mol. Ecol.* **24,** 5145–5157 (2015).

40. Main, B. J. *et al.* The genetic basis of host preference and resting behavior in the major african malaria vector, *Anopheles arabiensis*. *PLoS Genet.* **12,** e1006303 (2016).

41. Kamdem, C., Fouet, C., Gamez, S. & White, B. J. Pollutants and Insecticides Drive Local Adaptation in African Malaria Mosquitoes. *Mol. Biol. Evol.* **34,** 1261–1275 (2017).

42. Alout, H. *et al.* Insecticide resistance alleles affect vector competence of *Anopheles gambiae s.s.* for *Plasmodium falciparum* field isolates. *PLoS One* **8,** e63849 (2013).

43. Alout, H. *et al.* Interplay between *Plasmodium* infection and resistance to insecticides in vector mosquitoes. *J. Infect. Dis.* **210,** 1464–70 (2014).

44. Barnes, K. G. *et al.* Genomic Footprints of Selective Sweeps from Metabolic Resistance to Pyrethroids in African Malaria Vectors Are Driven by Scale up of Insecticide-Based Vector Control. *PLoS Genet.* **13,** e1006539 (2017).

45. Ibrahim, S. S. *et al.* Allelic Variation of Cytochrome P450s Drives Resistance to Bednet Insecticides in a Major Malaria Vector. *PLoS Genet.* **11,** e1005618 (2015).

46. Pinto, J. *et al.* Multiple origins of knockdown resistance mutations in the Afrotropical mosquito vector *Anopheles gambiae*. *PLoS One* **2,** e1243 (2007).

47. Ayala, D., Dia, I. & Guelbeogo, M. Advances and Perspectives in the Study of the Malaria Mosquito *Anopheles funestus*. in *Anopheles mosquitoes* (ed. Manguin, S.) (InTech, 2013). doi:10.5772/55389

48. Lehmann, T. & Diabate, A. The molecular forms of *Anopheles gambiae*: a phenotypic perspective. *Infect. Genet. Evol.* **8,** 737–46 (2008).

49. Bryan, J. H., Di Deco, M. A., Petrarca, V. & Coluzzi, M. Inversion polymorphism and incipient speciation in *Anopheles gambiae* s.str. in The Gambia, West Africa. *Genetica* **59,** 167–176 (1982).

50. Coluzzi, M., Petrarca, V. & di Deco, M. A. Chromosomal inversion intergradation and incipient speciation in *Anopheles gambiae*. *Bolletino di Zool.* **52,** 45–63 (1985).

51. Coetzee, M. *et al. Anopheles coluzzii* and *Anopheles amharicus*, new members of the *Anopheles gambiae* complex. *Zootaxa* **3619,** 246–74 (2013).

52. Costantini, C. *et al.* Living at the edge: biogeographic patterns of habitat segregation conform to speciation by niche expansion in *Anopheles gambiae*. *BMC Ecol.* **9,** 16 (2009).

53. Gimonneau, G. *et al.* Behavioural responses of *Anopheles gambiae sensu stricto* M and S molecular form larvae to an aquatic predator in Burkina Faso. *Parasit. Vectors* **5,** 65 (2012).

54. Simard, F. *et al.* Ecological niche partitioning between *Anopheles gambiae* molecular forms in Cameroon: the ecological side of speciation. *BMC Ecol.* **9,** 17 (2009).

55. Kerah-Hinzoumbé, C. *et al.* Insecticide resistance in *Anopheles gambiae* from south-western Chad, Central Africa. *Malar. J.* **7,** 192 (2008).

56. Guelbeogo, W. M., Sagnon, N., Liu, F., Besansky, N. J. & Costantini, C. Behavioural divergence of sympatric *Anopheles funestus* populations in Burkina Faso. *Malar. J.* **13,** 65 (2014).

57. Guelbeogo, W. M. *et al.* Seasonal distribution of *Anopheles funestus* chromosomal forms from Burkina Faso. *Malar. J.* **8,** 239 (2009).

58. Dia, I., Boccolini, D., Antonio-Nkondjio, C., Costantini, C. & Fontenille, D. Chromosomal inversion polymorphism of *Anopheles funestus* from forest villages of South Cameroon. *Parassitologia* **42,** 227–9 (2000).

59. Dia, I., Lochouarn, L., Boccolini, D., Costantini, C. & Fontenille, D. Spatial and temporal variations of the chromosomal inversion polymorphism of *Anopheles*

funestus in Senegal. *Parasite* **7,** 179–184 (2000).

60. Costantini, C., Sagnon, N., Ilboudo-Sanogo, E., Coluzzi, M. & Boccolini, D. Chromosomal and bionomic heterogeneities suggest incipient speciation in *Anopheles funestus* from Burkina Faso. *Parassitologia* **41,** 595–611 (1999).

61. Green, C. A. Cladistic analysis of mosquito chromosome data (*Anopheles* (*Cellia*) *Myzomyia*. *J. Hered.* **73,** 2–11 (1982).

62. Pape, T. Cladistic analysis of mosquito chromosome data in *Anopheles* subgenus *Cellia* (Diptera: Culicidae). *Mosq. Syst.* **24,** (1992).

63. Green, C. A. & Hunt, R. H. Interpretation of variation in ovarian polytene chromosomes of *Anopheles funestus* Giles, *A. parensis* Gillies, and *A. aruni*? *Genetica* **51,** 187–195 (1980).

64. Cohuet, A. *et al.* Gene flow between chromosomal forms of the malaria vector *Anopheles funestus* in Cameroon, Central Africa, and its relevance in malaria fighting. *Genetics* **169,** 301–11 (2005).

65. Cohuet, A., Dia, I., Simard, F., Raymond, M. & Fontenille, D. Population structure of the malaria vector *Anopheles funestus* in Senegal based on microsatellite and cytogenetic data. *Insect Mol. Biol.* **13,** 251–8 (2004).

66. Lochouarn, L., Dia, I., Boccolini, D., Coluzzi, M. & Fontenille, D. Bionomical and cytogenetic heterogeneities of *Anopheles funestus* in Senegal. *Trans. R. Soc. Trop. Med. Hyg.* **92,** 607–12 (1998).

67. Boccolini, D. *et al.* Chromosomal differentiation of *Anopheles funestus* from Luanda and Huambo Provinces, western and central Angola. *Am. J. Trop. Med. Hyg.* **73,** 1071–6 (2005).

68. Michel, A. P. *et al.* Effective population size of *Anopheles funestus* chromosomal forms in Burkina Faso. *Malar. J.* **5,** 115 (2006).

69. Koekemoer, L. L., Kamau, L., Hunt, R. H. & Coetzee, M. A cocktail polymerase chain reaction assay to identify members of the *Anopheles funestus* (Diptera: Culicidae) group. *Am. J. Trop. Med. Hyg.* **66,** 804–11 (2002).

70. Koekemoer, L. L., Coetzee, M. & Hunt, R. H. HpaII endonuclease distinguishes between two species in the *Anopheles funestus* group. **7,** 273–277 (1998).

71. Hackett, B. J. *et al.* Ribosomal DNA internal transcribed spacer (ITS2) sequences differentiate *Anopheles funestus* and *An. rivulorum*, and uncover a cryptic taxon. *Insect Mol. Biol.* **9,** 369–74 (2000).

72. Garros, C. *et al.* Restriction fragment length polymorphism method for the identification of major African and Asian malaria vectors within the *Anopheles funestus* and *An. minimus* groups. *Am. J. Trop. Med. Hyg.* **70,** 260–5 (2004).

73. Hunt, R. H., Brooke, B. D., Pillay, C., Koekemoer, L. L. & Coetzee, M. Laboratory selection for and characteristics of pyrethroid resistance in the malaria

vector *Anopheles funestus*. *Med. Vet. Entomol.* **19,** 271–5 (2005).

74. Krzywinski, J., Grushko, O. G. & Besansky, N. J. Analysis of the complete mitochondrial DNA from *Anopheles funestus*: an improved dipteran mitochondrial genome annotation and a temporal dimension of mosquito evolution. *Mol. Phylogenet. Evol.* **39,** 417–23 (2006).

75. Gregory, R. *et al.* A De Novo Expression Profiling of *Anopheles funestus*, Malaria Vector in Africa, Using 454 Pyrosequencing. *PLoS One* **6,** e17418 (2011).

76. Neafsey, D. E. *et al.* Mosquito genomics. Highly evolvable malaria vectors: the genomes of 16 *Anopheles* mosquitoes. *Science* **347,** 1258522 (2015).

77. Braginets, O. P., Minakawa, N., Mbogo, C. M. & Yan, G. Population genetic structure of the African malaria mosquito *Anopheles funestus* in Kenya. *Am. J. Trop. Med. Hyg.* **69,** 303–8 (2003).

78. Ayala, D., Goff, G. Le, Robert, V., de Jong, P. & Takken, W. Population structure of the malaria vector *Anopheles funestus* (Diptera: Culicidae) in Madagascar and Comoros. *Acta Trop.* **97,** 292–300 (2006).

79. Ayala, D. *et al.* Chromosomal inversions, natural selection and adaptation in the malaria vector *Anopheles funestus*. *Mol. Biol. Evol.* **28,** 745–58 (2011).

80. Kamdem, C., Fouet, C. & White, B. J. Chromosome arm-specific patterns of polymorphism associated with chromosomal inversions in the major African malaria vector, *Anopheles funestus*. *Mol. Ecol.* **26,** 5552–5566 (2017).

81. Michel, A. P. *et al.* Molecular differentiation between chromosomally defined incipient species of Anopheles funestus. *Insect Mol. Biol.* **14,** 375–387 (2005).

82. Lehmann, T. *et al.* The Rift Valley Complex as a Barrier to Gene Flow for *Anopheles gambiae* in Kenya. *J. Hered.* **91,** 165–168 (1999).

83. Mukabayire, O., Boccolini, D., Lochouarn, L., Fontenille, D. & Besansky, N. J. Mitochondrial and ribosomal internal transcribed spacer (ITS2) diversity of the African malaria vector *Anopheles funestus*. *Mol. Ecol.* **8,** 289–97 (1999).

84. Temu, E. A., Hunt, R. H. & Coetzee, M. Microsatellite DNA polymorphism and heterozygosity in the malaria vector mosquito *Anopheles funestus* (Diptera: Culicidae) in east and southern Africa. *Acta Trop.* **90,** 39–49 (2004).

85. Michel, A. P. *et al.* Rangewide population genetic structure of the African malaria vector *Anopheles funestus*. *Mol. Ecol.* **14,** 4235–48 (2005).

86. Koekemoer, L. L. *et al.* Impact of the Rift Valley on restriction fragment length polymorphism typing of the major African malaria vector *Anopheles funestus* (Diptera: Culicidae). *J. Med. Entomol.* **43,** 1178–84 (2006).

87. Drury, D. W., Dapper, A. L., Siniard, D. J., Zentner, G. E. & Wade, M. J. CRISPR/Cas9 gene drives in genetically variable and nonrandomly mating wild populations. *Sci. Adv.* **3,** e1601910 (2017).

88. Zentner, G. E. & Wade, M. J. The promise and peril of CRISPR gene drives. *BioEssays* **39,** 1700109 (2017).

89. Gillies, M. T. & De Meillon, B. *The Anophelinae of Africa South of the Sahara (Ethiopian Zoogeographical Region).* (The South African Institute for Medical Research, 1968).

90. Killeen, G. F. Characterizing, controlling and eliminating residual malaria transmission. *Malar. J.* **13,** 330 (2014).

91. Gillies, M. T. & Smith, A. The effect of a residual house-spraying campaign in East Africa on species balance in the *Anopheles funestus* group. the replacement of *A. funestus* Giles by *A. rivulorum* Leeson. *Bull. Entomol. Res.* **51,** 243–252 (1960).

92. Afrane, Y. A., Bonizzoni, M. & Yan, G. Secondary Malaria Vectors of Sub-Saharan Africa: Threat to Malaria Elimination on the Continent? in *Current Topics in Malaria* (InTech, 2016).

93. Wilkes, T. J., Matola, Y. G. & Charlwood, J. D. *Anopheles rivulorum*, a vector of human malaria in Africa. *Med. Vet. Entomol.* **10,** 108–110 (1996).

94. Temu, E. A., Minjas, J. N., Tuno, N., Kawada, H. & Takagi, M. Identification of four members of the *Anopheles funestus* (Diptera: Culicidae) group and their role in *Plasmodium falciparum* transmission in Bagamoyo coastal Tanzania. *Acta Trop.* **102,** 119–125 (2007).

95. Mouatcho, J. *et al.* Detection of *Anopheles rivulorum-like*, a member of the Anopheles funestus group, in South Africa. *Malar. J.* **17,** 195 (2018).

96. Stevenson, J. C. *et al.* Detection of *Plasmodium falciparum* Infection in *Anopheles squamosus* (Diptera: Culicidae) in an Area Targeted for Malaria Elimination, Southern Zambia. *J. Med. Entomol.* **53,** 1482–1487 (2016).

97. Antonio-Nkondjio, C. *et al.* Complexity of the Malaria Vectorial System in Cameroon: Contribution of Secondary Vectors to Malaria Transmission. *J. Med. Entomol* **43,** 1215–1221 (2006).

98. St Laurent, B. *et al.* Molecular Characterization Reveals Diverse and Unknown Malaria Vectors in the Western Kenyan Highlands. *Am. J. Trop. Med. Hyg.* **94,** 327–35 (2016).

99. Lobo, N. F. *et al.* Unexpected diversity of *Anopheles* species in Eastern Zambia: Implications for evaluating vector behavior and interventions using molecular tools. *Sci. Rep.* **5,** 17952 (2015).

100. Meyers, J. I. *et al.* Increasing outdoor host-seeking in *Anopheles gambiae* over 6 years of vector control on Bioko Island. *Malar. J.* **15,** 239 (2016).

101. Thomsen, E. K. *et al.* Mosquito behaviour change after distribution of bednets results in decreased protection against malaria exposure. *J. Infect. Dis.* **215,** jiw615 (2016).

102. Reimer, L. J. *et al.* Malaria transmission dynamics surrounding the first nationwide long-lasting insecticidal net distribution in Papua New Guinea. *Malar. J.* **15,** 25 (2016).

103. Russell, T. L., Beebe, N. W., Cooper, R. D., Lobo, N. F. & Burkot, T. R. Successful malaria elimination strategies require interventions that target changing vector behaviours. *Malar. J.* **12,** 56 (2013).

104. Reddy, M. R. *et al.* Outdoor host seeking behaviour of *Anopheles gambiae* mosquitoes following initiation of malaria vector control on Bioko Island, Equatorial Guinea. *Malar. J.* **10,** 184 (2011).

105. Lefèvre, T. *et al.* Beyond nature and nurture: Phenotypic plasticity in blood-feeding behavior of *Anopheles gambiae s.s.* when humans are not readily accessible. *Am. J. Trop. Med. Hyg.* **81,** 1023–1029 (2009).

106. Steketee, R. W. *et al.* National malaria control and scaling up for impact: the Zambia experience through 2006. *Am. J. Trop. Med. Hyg.* **79,** 45–52 (2008).

107. Government of the Republic of Zambia Ministry of Health. *National Malaria Indicator Survey*. (2015).

108. Moss, W. J. *et al.* Malaria epidemiology and control within the International Centers of Excellence for Malaria Research. *Am. J. Trop. Med. Hyg.* **93,** 5–15 (2015).

109. Kamuliwo, M. *et al.* The changing burden of malaria and association with vector control interventions in Zambia using district-level surveillance data, 2006-2011. *Malar. J.* **12,** 437 (2013).

110. Das, S. *et al.* Underestimation of foraging behaviour by standard field methods in malaria vector mosquitoes in southern Africa. *Malar. J.* **14,** 12 (2015).

111. Mukonka, V. M. *et al.* High burden of malaria following scale-up of control interventions in Nchelenge District, Luapula Province, Zambia. *Malar. J.* **13,** 153 (2014).

112. Das, S., Muleba, M., Stevenson, J. C., Norris, D. E. & Southern Africa International Centers of Excellence for Malaria Research Team. Habitat partitioning of malaria vectors in Nchelenge District, Zambia. *Am. J. Trop. Med. Hyg.* **94,** 1234–1244 (2016).

113. Stevenson, J. C. *et al.* Spatio-temporal heterogeneity of malaria vectors in northern Zambia: implications for vector control. *Parasit. Vectors* **9,** 510 (2016).

114. Choi, K. S. *et al.* Insecticide resistance and role in malaria transmission of *Anopheles funestus* populations from Zambia and Zimbabwe. *Parasit. Vectors* **7,** 464 (2014).

115. The President's Malaria Initiative & PMI Africa IRS (AIRS). *Zambia Supplemental Environmental Assessment for Indoor Residual Spraying for Malaria Control: 2015-2020*. (2015).

116. Zawada, J. W. *et al.* Molecular and physiological analysis of *Anopheles funestus* swarms in Nchelenge, Zambia. *Malar. J.* **17,** 49 (2018).

117. Hay, S. I. & Snow, R. W. The Malaria Atlas Project: Developing Global Maps of Malaria Risk. *PLoS Med.* **3,** e473 (2006).

118. ESRI. ArcGIS Desktop. (2018).

119. Charlwood, J. D. & Graves, P. M. The effect of permethrin-impregnated bednets on a population of *Anopheles farauti* in coastal Papua New Guinea. *Med. Vet. Entomol.* **1,** 319–27 (1987).

120. Brown, A. W. Insecticide resistance in mosquitoes: a pragmatic review. *J. Am. Mosq. Control Assoc.* **2,** 123–40 (1986).

121. Emerson, K. J., Conn, J. E., Bergo, E. S., Randel, M. A. & Sallum, M. A. M. Brazilian Anopheles darlingi Root (Diptera: Culicidae) Clusters by Major Biogeographical Region. *PLoS One* **10,** e0130773 (2015).

122. Lee, Y. *et al.* Spatial and temporal distribution of genome divergence among California populations of *Aedes aegypti*. *bioRxiv* 166629 (2017).

123. Pless, E. *et al.* Multiple introductions of the dengue vector, *Aedes aegypti*, into California. *PLoS Negl. Trop. Dis.* **11,** e0005718 (2017).

124. Marsden, C. D. *et al.* An analysis of two island groups as potential sites for trials of transgenic mosquitoes for malaria control. *Evol. Appl.* **6,** 706–20 (2013).

125. Coetzee, M. & Koekemoer, L. L. Molecular systematics and insecticide resistance in the major African malaria vector *Anopheles funestus*. *Annu. Rev. Entomol.* **58,** 393–412 (2013).

126. Crawford, J. E. *et al.* De novo transcriptome sequencing in *Anopheles funestus* using illumina RNA-seq technology. *PLoS One* **5,** e14202 (2010).

127. Kamau, L., Hunt, R. & Coetzee, M. Analysis of the population structure of Anopheles funestus (Diptera: Culicidae) from western and coastal Kenya using paracentric chromosomal inversion frequencies. *J. Med. Entomol.* **39,** 78–83 (2002).

128. Choi, K. S., Koekemoer, L. L. & Coetzee, M. Population genetic structure of the major malaria vector *Anopheles funestus s.s.* and allied species in southern Africa. *Parasit. Vectors* **5,** 283 (2012).

129. Choi, K. S., Spillings, B. L., Coetzee, M., Hunt, R. H. & Koekemoer, L. L. A comparison of DNA sequencing and the hydrolysis probe analysis (TaqMan assay) for knockdown resistance (kdr) mutations in *Anopheles gambiae* from the Republic of the Congo. *Malar. J.* **9,** 278 (2010).

130. Kweka, E. J. *et al.* Application of hydrolysis probe analysis to identify clade types of the malaria vector mosquito *Anopheles funestus sensu stricto* from Muheza, northeastern Tanzania. *Med. Vet. Entomol.* (2017).

131. Reilly, J. G. & Thomas, C. A. Length polymorphisms, restriction site variation, and maternal inheritance of mitochondrial DNA of *Drosophila melanogaster*. *Plasmid* **3,** 109–115 (1980).

132. Stewart, J. B. & Larsson, N.-G. Keeping mtDNA in shape between generations. *PLoS Genet.* **10,** e1004670 (2014).

133. Ladoukakis, E. D. & Zouros, E. Evolution and inheritance of animal mitochondrial DNA: rules and exceptions. *J. Biol. Res. (Thessalonike, Greece)* **24,** 2 (2017).

134. Brown, W. M., George, M. & Wilson, A. C. Rapid evolution of animal mitochondrial DNA. *Proc. Natl. Acad. Sci. U. S. A.* **76,** 1967–71 (1979).

135. Weathersbee, A. A. & Meisch, M. V. Dispersal of *Anopheles quadrimaculatus* (Diptera: Culicidae) in Arkansas Ricelands. *Environ. Entomol.* **19,** 961–965 (1990).

136. Charlwood, J. D. Studies on the bionomics of male *Anopheles gambiae* Giles and male *Anopheles funestus* Giles from southern Mozambique. *J. Vector Ecol.* **36,** 382–394 (2011).

137. Fryxell, R. T. T. *et al.* Differential *Plasmodium falciparum* infection of *Anopheles gambiae* s.s. molecular and chromosomal forms in Mali. *Malar. J.* **11,** 133 (2012).

138. Tripet, F. *et al.* Longitudinal survey of knockdown resistance to pyrethroid (kdr) in Mali, West Africa, and evidence of its emergence in the Bamako form of *Anopheles gambiae* s.s. *Am. J. Trop. Med. Hyg.* **76,** 81–7 (2007).

139. Mvumbi, D. M. *et al.* High prevalence of *Plasmodium falciparum* infection in asymptomatic individuals from the Democratic Republic of the Congo. *Malar. Res. Treat.* **2016,** 1–4 (2016).

140. *The President's Malaria Initiative (PMI)/Africa Indoor Residual Spraying Project. November 2016. The Democratic Republic of Congo Entomological Monitoring*. (2016).

141. Smith, T. *et al.* Absence of seasonal variation in malaria parasitaemia in an area of intense seasonal transmission. *Acta Trop.* **54,** 55–72 (1993).

142. Kitua, A. Y. *et al. Plasmodium falciparum* malaria in the first year of life in an area of intense and perennial transmission. *Trop. Med. Int. Health* **1,** 475–84 (1996).

143. Charlwood, J. D. *et al.* Incidence of *Plasmodium falciparum* infection in infants in relation to exposure to sporozoite-infected anophelines. *Am. J. Trop. Med. Hyg.* **59,** 243–51 (1998).

144. Drakeley, C. *et al.* An estimation of the entomological inoculation rate for Ifakara: a semi-urban area in a region of intense malaria transmission in Tanzania. *Trop. Med. Int. Health* **8,** 767–74 (2003).

145. Killeen, G. F., Ross, A. & Smith, T. Infectiousness of malaria-endemic human

populations to vectors. *Am. J. Trop. Med. Hyg.* **75,** 38–45 (2006).

146. Smith, T., Charlwood, J. D., Takken, W., Tanner, M. & Spiegelhalter, D. J. Mapping the densities of malaria vectors within a single village. *Acta Trop.* **59,** 1–18 (1995).

147. Killeen, G. *et al.* Cost-sharing strategies combining targeted public subsidies with private-sector delivery achieve high bednet coverage and reduced malaria transmission in Kilombero Valley, southern Tanzania. *BMC Infect. Dis.* **7,** 121 (2007).

148. Killeen, G. F. *et al.* Quantifying behavioural interactions between humans and mosquitoes: evaluating the protective efficacy of insecticidal nets against malaria transmission in rural Tanzania. *BMC Infect. Dis.* **6,** 161 (2006).

149. Gillies, M. T. & Coetzee, M. A supplement to the Anophelinae of Africa south of the Sahara (Afrotropical Region). *Publ. South African Inst. Med. Res.* **55,** 1–143 (1987).

150. Nieman, C. C., Yamasaki, Y., Collier, T. C. & Lee, Y. A DNA extraction protocol for improved DNA yield from individual mosquitoes. *F1000Research* (2015).

151. Yamasaki YK, Nieman CC, Chang AN, Collier TC, Main BJ, L. Y. Improved tools for genomic DNA library construction of small insects. (2016).

152. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114–2120 (2014).

153. Hahn, C., Bachmann, L. & Chevreux, B. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads--a baiting and iterative mapping approach. *Nucleic Acids Res.* **41,** e129–e129 (2013).

154. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. (2013).

155. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).

156. Broad Institute. Picard Tools. Available at: http://broadinstitute.github.io/picard/. (Accessed: 14th September 2017)

157. McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20,** 1297–1303 (2010).

158. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43,** 491–8 (2011).

159. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32,** 1792–1797 (2004).

160. Beard, C. B., Hamm, D. M. & Collins, F. H. The mitochondrial genome of the mosquito *Anopheles gambiae*: DNA sequence, genome organization, and

comparisons with mitochondrial sequences of other insects. *Insect Mol. Biol.* **2,** 103–24 (1993).

161. Hua, Y. Q. *et al.* The complete mitochondrial genome of *Anopheles minimus* (Diptera: Culicidae) and the phylogenetics of known *Anopheles* mitogenomes. *Insect Sci.* **23,** 353–365 (2016).

162. Lam, H. M., Ratmann, O. & Boni, M. F. Improved Algorithmic Complexity for the 3SEQ Recombination Detection Algorithm. *Mol. Biol. Evol.* **35,** 247–251 (2018).

163. Gouy, M., Guindon, S. & Gascuel, O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27,** 221–224 (2010).

164. Bouckaert, R. *et al.* BEAST 2: a software platform for bayesian evolutionary analysis. *PLoS Comput. Biol.* **10,** e1003537 (2014).

165. Brower, A. V. Rapid morphological radiation and convergence among races of the butterfly Heliconius erato inferred from patterns of mitochondrial DNA evolution. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 6491–5 (1994).

166. Rambaut, A., Suchard, M. & Drummond, A. Tracer v1.6. Available at: http://tree.bio.ed.ac.uk/software/tracer. (Accessed: 27th April 2017)

167. Gillies, M. T. & De Meillon, B. *The Anophelinae of Africa South of the Sahara*. (South African Institute for Medical Research, 1968).

168. Clement, M., Posada, D. & Crandall, K. A. TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* **9,** 1657–9 (2000).

169. Leigh, J. W. & Bryant, D. popart : full-feature software for haplotype network construction. *Methods Ecol. Evol.* **6,** 1110–1116 (2015).

170. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59,** 307–321 (2010).

171. Múrias dos Santos, A., Cabezas, M. P., Tavares, A. I., Xavier, R. & Branco, M. tcsBU: a tool to extend TCS network layout and visualization. *Bioinformatics* **32,** 627–628 (2016).

172. Parker, J., Rambaut, A. & Pybus, O. G. Correlating viral phenotypes with phylogeny: Accounting for phylogenetic uncertainty. *Infect. Genet. Evol.* **8,** 239–246 (2008).

173. Slatkin, M. & Maddison, W. P. A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics* **123,** 603–13 (1989).

174. Wang, R., Zheng, L., Touré, Y. T., Dandekar, T. & Kafatos, F. C. When genetic distance matters: measuring genetic differentiation at microsatellite loci in whole-genome scans of recent and incipient mosquito species. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 10769–74 (2001).

175. Librado, P. & Rozas, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25,** 1451–1452 (2009).

176. Excoffier, L. & Lischer, H. E. L. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10,** 564–567 (2010).

177. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20,** 289–90 (2004).

178. Team, R. C. R: A Language and Environment for Statistical Computing. (2013).

179. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25,** 955–64 (1997).

180. Bernt, M. *et al.* MITOS: Improved de novo metazoan mitochondrial genome annotation. *Mol. Phylogenet. Evol.* **69,** 313–319 (2013).

181. Cameron, S. L. Insect mitochondrial genomics: implications for evolution and phylogeny. *Annu. Rev. Entomol.* **59,** 95–117 (2014).

182. Besansky, N. J. *et al.* Patterns of mitochondrial variation within and between African malaria vectors, *Anopheles gambiae* and *An. arabiensis*, suggest extensive gene flow. *Genetics* **147,** 1817–1828 (1997).

183. Choi, K., Coetzee, M. & Koekemoer, L. L. Detection of clade types (clades I and II) within *Anopheles funestus* sensu stricto by the hydrolysis probe analysis (Taqman assay). *Parasit. Vectors* **6,** 173 (2013).

184. Mantel, N. The detection of disease clustering and a generalized regression approach. *Cancer Res.* **27,** 209–20 (1967).

185. Garros, C. *et al.* Systematics and biogeographical implications of the phylogenetic relationships between members of the Funestus and Minimus Groups of *Anopheles* (Diptera: Culicidae). *J. Med. Entomol.* **42,** 7–18 (2005).

186. Kamali, M. *et al.* Multigene phylogenetics reveals temporal diversification of major African malaria vectors. *PLoS One* **9,** e93580 (2014).

187. Sharakhov, I. V *et al.* Inversions and gene order shuffling in *Anopheles gambiae* and *A. funestus*. *Science* **298,** 182–5 (2002).

188. Krzywinski, J. *et al.* Analysis of the evolutionary forces shaping mitochondrial genomes of a Neotropical malaria vector complex. *Mol. Phylogenet. Evol.* **58,** 469–477 (2011).

189. Pinchoff, J. *et al.* Predictive Malaria Risk and Uncertainty Mapping in Nchelenge District, Zambia: Evidence of Widespread, Persistent Risk and Implications for Targeted Interventions. *Am. J. Trop. Med. Hyg.* **93,** 1260–1267 (2015).

190. Burkot, T. R. *et al.* Barrier screens: a method to sample blood-fed and host-seeking exophilic mosquitoes. *Malar. J.* **12,** 49 (2013).

191. Pombi, M. *et al.* Evaluation of a protocol for remote identification of mosquito vector species reveals BG-Sentinel trap as an efficient tool for *Anopheles gambiae* outdoor collection in Burkina Faso. *Malar. J.* **14,** 161 (2015).

192. Batista, E. P. A. *et al.* Semi-field assessment of the BG-Malaria trap for monitoring the African malaria vector, *Anopheles arabiensis*. *PLoS One* **12,** e0186696 (2017).

193. Hoel, D. F. *et al.* Optimizing Collection of *Anopheles gambiae s.s.* (Diptera: Culicidae) in Biogents Sentinel Traps. *J. Med. Entomol.* **51,** 1268–1275 (2014).

194. Cilek, J. E., Weston, J. R. & Richardson, A. G. Comparison of Adult Mosquito Abundance From Biogents-2 Sentinel and Biogents Gravid *Aedes* Traps In Northeastern Florida. *J. Am. Mosq. Control Assoc.* **33,** 358–360 (2017).

195. Urquhart, C., Paulsen, D., Moncayo, A. & Trout Fryxell, R. T. Evaluating Surveillance Methods for Arboviral Vectors of La Crosse Virus and West Nile Virus of Southern Appalachia. *J. Am. Mosq. Control Assoc.* **32,** 24–33 (2016).

196. Post, R. J., Flook, P. K. & Millest, A. L. Methods for the preservation of insects for DNA studies. *Biochem. Syst. Ecol.* **21,** 85–92 (1993).

197. Scott, J. A., Brogdon, W. G. & Collins, F. H. Identification of single specimens of the *Anopheles gambiae* complex by the polymerase chain reaction. *Am. J. Trop. Med. Hyg.* **49,** 520–9 (1993).

198. Cohuet, A. *et al.* Species identification within the *Anopheles funestus* group of malaria vectors in Cameroon and evidence for a new species. *Am. J. Trop. Med. Hyg.* **69,** 200–5 (2003).

199. Beebe, N. W. & Saul, A. Discrimination of all members of the *Anopheles punctulatus* complex by polymerase chain reaction--restriction fragment length polymorphism analysis. *Am. J. Trop. Med. Hyg.* **53,** 478–81 (1995).

200. Hoffman, I. F. *et al.* Estimation of *Plasmodium falciparum* Transmission Intensity in Lilongwe, Malawi, by Microscopy, Rapid Diagnostic Testing, and Nucleic Acid Detection. *Am. J. Trop. Med. Hyg.* **95,** 373–377 (2016).

201. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models Using **lme4**. *J. Stat. Softw.* **67,** 1–48 (2015).

202. Brooks, M. E. *et al.* glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *R J.* **9,** 378–400 (2017).

203. Team, R. C. R: A language environment for statistical computing. (2018).

204. Wickham, H. *ggplot2: Elegant Graphic for Data Analysis.* (Springer-Verlag New York, 2016).

205. Hartig, F. DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. (2018).

206. Barton, K. MuMIn: Multi-Model Inference. (2018).

207. Coetzee, M. Chromosomal and cross-mating evidence for two species within *Anopheles (A.) coustani* (Diptera: Culicidae). *Syst. Entomol.* **8,** 137–141 (1983).

208. Coetzee, M. *Anopheles crypticus*, new species from South Africa is distinguished from *Anopheles coustani* (Diptera: Culicidae). *Mosq. Syst.* **26,** 125–131 (1994).

209. Ndiath, M. O. *et al.* Composition and genetics of malaria vector populations in the Central African Republic. *Malar. J.* **15,** 387 (2016).

210. Ogola, E. *et al.* Composition of *Anopheles* mosquitoes, their blood-meal hosts, and *Plasmodium falciparum* infection rates in three islands with disparate bed net coverage in Lake Victoria, Kenya. *Malar. J.* **16,** 360 (2017).

211. Foster, P. G. *et al.* Phylogeny of Anophelinae using mitochondrial protein coding genes. *R. Soc. open Sci.* **4,** 170758 (2017).

212. Harbach, R. E. The classification of genus *Anopheles* (Diptera: Culicidae): a working hypothesis of phylogenetic relationships. *Bull. Entomol. Res.* **94,** 537–53 (2004).

213. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–402 (1997).

214. Hebert, P. D. N., Cywinska, A., Ball, S. L. & deWaard, J. R. Biological identifications through DNA barcodes. *Proc. R. Soc. B Biol. Sci.* **270,** 313–321 (2003).

215. Cywinska, A., Hunter, F. F. & Hebert, P. D. N. Identifying Canadian mosquito species through DNA barcodes. *Med. Vet. Entomol.* **20,** 413–424 (2006).

216. Batovska, J., Blacket, M. J., Brown, K. & Lynch, S. E. Molecular identification of mosquitoes (Diptera: Culicidae) in southeastern Australia. *Ecol. Evol.* **6,** 3001–3011 (2016).

217. Chan, A. *et al.* DNA barcoding: complementing morphological identification of mosquito species in Singapore. *Parasit. Vectors* **7,** 569 (2014).

218. Ratnasingham, S. & Hebert, P. D. N. BOLD: The Barcode of Life Data System (www.barcodinglife.org). *Mol. Ecol. Notes* **7,** 355–364 (2007).

219. Moss, W. J. *et al.* Challenges and prospects for malaria elimination in the Southern Africa region. *Acta Trop.* **121,** 207–11 (2012).

220. Villesen, P. FaBox: an online toolbox for fasta sequences. *Mol. Ecol. Notes* **7,** 965–968 (2007).

221. Guindon, S. & Gascuel, O. A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. *Syst. Biol.* **52,** 696–704 (2003).

222. Darriba, D., Taboada, G., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* **9,** 772

223. Rambaut, A., Drummond, A., Xie, D., Baele, G. & Suchard, M. Posterior

summarisation in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **67,** 901–904 (2018).

224. Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3,** 217–223 (2012).

225. Yao, H. *et al.* Use of ITS2 region as the universal DNA barcode for plants and animals. *PLoS One* **5,** (2010).

226. Norris, L. C. *et al.* Phylogeny of anopheline (Diptera : Culicidae) species in southern Africa , based on nuclear and mitochondrial genes. 16–27 (2015).

227. Bandibabone, J. *et al.* Identification of *Anopheles* species in Sud Kivu, Democratic Republic of Congo, using molecular tools. *Trans. R. Soc. Trop. Med. Hyg.* (2018).

228. Sallum, M. A. M. *et al.* Phylogeny of Anophelinae (Diptera: Culicidae) based on nuclear ribosomal and mitochondrial DNA sequences. *Syst. Entomol.* **27,** 361–382 (2002).

229. Krzywinski, J. & Besansky, N. J. Molecular systematics of *Anopheles*: from subgenera to subpopulations. *Annu. Rev. Entomol.* **48,** 111–39 (2003).

230. Nepomichene, T. N. J. J., Tata, E. & Boyer, S. Malaria case in Madagascar, probable implication of a new vector, *Anopheles coustani*. *Malar. J.* **14,** 475 (2015).

231. Fornadel, C. M., Norris, L. C., Franco, V. & Norris, D. E. Unexpected anthropophily in the potential secondary malaria vectors *Anopheles coustani s.l.* and *Anopheles squamosus* in Macha, Zambia. *Vector Borne Zoonotic Dis.* **11,** 1173–9 (2011).

232. Freitas, L. A. *et al.* Diversification of the Genus Anopheles and a Neotropical Clade from the Late Cretaceous. *PLoS One* **10,** e0134462 (2015).

233. Foley, D. H., Bryan, J. H., Yeates, D. & Saul, A. Evolution and Systematics ofAnopheles:Insights from a Molecular Phylogeny of Australasian Mosquitoes. *Mol. Phylogenet. Evol.* **9,** 262–275 (1998).

234. Krzywinski, J., Wilkerson, R. C. & Besansky, N. J. Toward understanding Anophelinae (Diptera, Culicidae) phylogeny: insights from nuclear single-copy genes and the weight of evidence. *Syst. Biol.* **50,** 540–56 (2001).

235. Krzywinski, J., Wilkerson, R. C. & Besansky, N. J. Evolution of Mitochondrial and Ribosomal Gene Sequences in Anophelinae (Diptera: Culicidae): Implications for Phylogeny Reconstruction. *Mol. Phylogenet. Evol.* **18,** 479–487 (2001).

Appendices

Appendix A
ITS2 rDNA PCR

This PCR is very robust and therefore can be used to check the quality of DNA extractions. It targets the ITS2 region of nuclear rDNA and produces amplicons of varying sizes depending on mosquito species. It can be used in tandem with the Funestus PCR to identify ambiguous samples. Because ITSA binds to the conserved 5.8S rDNA and ITS2B binds to the 28S rDNA, this PCR can be used to sequence samples from almost any anopheline mosquito for species identification. ITS2B1, a novel, alternate primer, binds slightly downstream from ITS2B and produces a slightly larger amplicon that can be used to sequence through the entire ITS2.

Expected product sizes for different mosquito species:
Funestus group:

An. leesoni ~520 bp
An. parensis ~ 620 bp
An. vaneedeni ~ 830 bp
An. rivulorum and *rivulorum*-like ~520 bp
An. longipalpis ~620 bp and ~900 bp
An. funestus and *funestus*-like ~850 bp

Other species:

An. rufipes, maculipalpis, and *pretoriensis* ~500 bp
An. theileri ~ 520 bp
An. coustani ~620 bp
An. gambiae complex ~600 bp
An. squamosus with SQFor/Rev ~300 bp

Primers:

ITS2A: 5'- TGT GAA CTG CAG GAC ACA T -3'
ITS2B: 5'- TAT GCT TAA ATT CAG GGG GT -3'
ITS2B1: 5'- GTC CCT ACG TGC TGA GCT TC -3'
SQFor405: 5'- CCA TTT CCA TTA TGT CCT ATC TAT AGG -3'
SQRev707: 5'- GGG AAA GCA GGA GTT CGT TGA G- 3'

Note: Only the ITS2B and ITS2B1 primers work well for sequencing.

PCR Program: (ITS2)

1.	94ºC	2 min
2.	94ºC	30 sec
3.	50ºC	30 sec
4.	72ºC	40 sec
5.	Go to step 2	39x
6.	72ºC	10 min
7.	4ºC	forever

(Continued on next page)

Reaction Mixture:	**25 µL**
10X	2.5 µL
dNTPs 2.5 mM	2.0 µL (final conc. 200 µM each)
ITS2A	0.3 µL (30 pmol)
ITS2B	0.3 µL (30 pmol)
Taq	2.0 U
dH_20	fill to 25 µL

Use 1 µL of template DNA.

Reference:

Koekemoer, L.L., L. Kamau, R.H. Hunt, M. Coetzee. 2002. A cocktail polymerase chain reaction assay to identify members of the *Anopheles funestus* (*Diptera: Culicidae*) group. Am. J. Trop. Med. Hyg. 6(6): 804-811.

Appendix B
Cytochrome Oxidase subunit I (COI) mitochondrial PCR

This PCR targets the 3' portion of the cytochrome oxidase I gene (bp 2121-2998), and can be used for sequencing most anopheline mosquito species for phylogeny building. Because it is a mitochondrial gene and has high copy number, it is fairly robust. The amplicon target size is 877 bp. Adapted from Lobo et al. 2015.

Primers:
LCO 1490: 5'-GGT CAA CAA ATC ATA AAG ATA TTG G-3'
HCO 2198: 5'-TAA ACT TCA GGG TGA CCA AAA AAT CA-3'

PCR Program:		(COILobo)	
1.	94ºC	5 min	
2.	94ºC	40 sec	5 cycles
3.	45ºC	1 min	
4.	72ºC	1.5 min	
5.	94ºC	40 sec	30 cycles
6.	51ºC	1 min	
7.	72ºC	1.5 min	
8.	72ºC	5 min	
9.	4ºC	forever	

Reaction Mixture:	**25** µL
10X	2.5 µL
dNTPs 2.5 mM	2.0 µL (final conc. 200 µM each)
LCO 1490	0.3 µL (30 pmol)
HCO 2198	0.3 µL (30 pmol)
Taq	2.0 U
dH_20	fill to 25 µl

Use 1.0 µl DNA template.

Reference:
Lobo, N. F., B. S. Laurent, C. H. Sikaala, B. Hamainza, J. Chanda, D. Chinula, S. M. Krishnankutty, J. D. Mueller, N. A. Deason, Q. T. Hoang, H. L. Boldt, J. Thumloup, J. Stevenson, A. Seyoum, and F. H. Collins. 2015. Unexpected diversity of Anopheles species in Eastern Zambia: implications for evaluating vector behavior and interventions using molecular tools. *Scientific Reports* **5**:17952.

Appendix C
Differentiation of the *Anopheles gambiae* complex by PCR

This PCR uses 4 primers that in combination produce three differentially-sized amplicons of the ribosomal DNA spacer region of *An. gambiae* complex mosquitoes. The expected product sizes are as follows: *An. gambiae* s.s. (~390 bp), *An. arabiensis* (~315 bp), and *An. quadriannulaus* (~150 bp).

Primers:

UN: 5'- GTG TGC CCC TTC CTC GAT GT -3'
GA: 5'- CTG GTT TGG TCG GCA CGT TT -3'
AR: 5'- AAG TGT CCT TCT CCA TCC TA -3'
QD: 5'- CAG ACC AAG ATG GTT AGT AT -3'

PCR Program: (SCOTT)

1.	94ºC	2 min
2.	94ºC	30 sec
3.	50ºC	30 sec
4.	72ºC	30 sec
5.	Go to step 2	29x
6.	72ºC	7 min
7.	4ºC	forever

Reaction Mixture:	**25 µL**
10X	2.5 µL
dNTPs 2.5 mM	2.0 µL (final conc. 200 µM each)
AR	3.0 µL (150 pmol)
QD	3.0 µL (150 pmol)
GA	0.5 µL (25 pmol)
UN	1.0 µL (50 pmol)
Taq	1.5 U
dH_20	fill to 25µL

Use between 0.5 and 1 µL of template DNA.

Reference:

Scott, J.A., W.G. Brogdon and F.H. Collins. 1993. Identification of single specimens of the *Anopheles gambiae* complex by the polymerase chain reaction. Am. J. Trop. Med. Hyg. 49(4): 520-529.

Appendix D
Differentiation of the *Anopheles funestus* complex by PCR

This PCR differentiates species of the *An. funestus* complex based on variation in the ITS2 region of nuclear rDNA. There is a universal forward primer and seven species-specific primers. The expected product sizes are as follows: *An. funestus* (505 bp)*, An. leesoni* (146 bp), *An. vaneedeni* (587 bp)*, An. parensis* (252 bp), *An. rivulorum* (411 bp), *An. rivulorum*-like (313 bp), and *An. funestus*-like (390 bp). Because the expected amplicons from *An. rivulorum* and *An. funestus*-like are too close in size to be effectively visualized on an agarose gel, only one of these primers should be used at a time in the reaction mixture, OR, if both are added and a band appears, remove each primer to determine which species is present. This is an appropriate strategy only if both species are uncommon in your study area.

Primers:

UV: 5'- TGT GAA CTG CAG GAC ACA T -3'
FUN: 5'- GCA TCG ATG GGT TAA TCA TG -3'
VAN: 5'- TGT CGA CTT GGT AGC CGA AC -3'
RIV: 5'- CAA GCC GTT CGA CCC TGA TT -3'
PAR: 5'- TGC GGT CCC AAG CTA GGT TC -3'
LEES: 5'- TAC ACG GGC GCC ATG TAG TT -3'
RIVLIKE: 5'- CCG CCT CCC GTG GAG TGG GGG -3'
FUNLIKE (MalaFB) 5′- GTT TTC AAT TGA ATT CAC CAT T -3′

PCR Program: (FUNESTUS)

1.	94ºC	2 min
2.	94ºC	30 sec
3.	45ºC	30 sec
4.	72ºC	40 sec
5.	Go to step 2	29x
6.	72ºC	5 min
7.	4ºC	forever

(Continued on next page)

Reaction Mixture:	**25 µL**
10X	2.5 µL
dNTPs 2.5 mM	2.0 µL (final conc. 200 µM each)
UV	0.3 µL (33 pmol each primer)
FUN	0.3 µL
VAN	0.3 µL
RIV	0.3 µL
FUNLIKE	0.3 µL
PAR	0.3 µL
LEES	0.3 µL
RIVLIKE	0.3 µL
Taq	1.6 U
dH_20	fill to 25 µl

Use 1 µL of template DNA.

References:

Cohuet, A. F. Simard, J.C. Toto, P. Kengne, M. Coetzee, D. Fontenille. 2003. Species identification within the *Anopheles funestus* group of malaria vectors in Cameroon and evidence for a new species. Am. J. Trop. Med. Hyg. 69(2): 200-5.

Koekemoer, L.L., L. Kamau, R.H. Hunt, M. Coetzee. 2002. A cocktail polymerase chain reaction assay to identify members of the *Anopheles funestus* (Diptera: Culicidae) group. Am. J. Trop. Med. Hyg. 6(6): 804-811.

Spillings, B.L., B.D. Brooke, L.L. Koekemoer, J. Chiphwanya, M. Coetzee, R.H. Hunt. 2009. A new species concealed by *Anopheles funestus* Giles, a major malaria vector in Africa. Am. J. Trop. Med. Hyg. 81(3): 510-5.

Appendix E
M/S Form Differentiation of *Anopheles gambiae* s.s. by PCR

This PCR diagnostic differentiates the M-form and S-form of An. gambiae s.s. by amplifying the a portion at the 5' end of the rDNA intergenic spacer region. The S-form will have a band at 475 bp and the M-form will have a band at 727 bp. Hybrid M/S form will have two bands at 475 bp and 727 bp.

Primers

R5: 5'- CGA ATT CTA GGG AGC TCC AG - 3'
R3: 5' - GCC AAT CCG AGC TGA TAG CGC - 3'
Mopint: 5' - GCC CCT TCC TCG ATG GCA T - 3'
B/Sint: 5' - ACC AAG ATG GTT CGT TGC - 3'

PCR Program (MSDIFF)

1. 94°C 10 min
2. 94°C 30 s
3. 63°C 30 s
4. 72°C 30 s
5. Go to Step 2 x24
6. 72°C 7 min
7. 4°C forever

Reaction Mixture:	**25 µL**
10X	2.5 µL
dNTPs 2.5 mM	1.0 µL
R5	0.5 µL
R3	0.5 µL
Mopint	0.4 µL
B/Sint	0.25 µL
Taq	2 U
dH2O	fill to 25 µL

Use 2 µL DNA (from abdomen extraction eluted in 50 µL dH2O).

Reference

Favia, G et al., 2001. Molecular characterization of ribosomal DNA polymorphisms discriminating among chromosomal forms of *Anopheles gambiae* s.s. *Insect Molecular Biology* 10(1): 19-23.

Appendix F
Marriott DNA Extraction Procedure

Materials:

Bender Buffer

- 0.1 M NaCl (5 mL from a 1M stock solution you may need to make)
- 0.2 M sucrose (3.42 grams)
- 0.1 M Tris-HCl (5 mL from a 1M stock)
- 0.05 M EDTA pH 9.1 (5 mL from a 0.5M stock)
- 0.5% SDS in DEPC water (0.25 mL from a 0.1M stock)

For 50 mL Bender Buffer: Add 3.42 grams dry sucrose to a 50 mL conical tube. Add the other ingredients, listed above in parentheses. Fill to a final volume with DEPC water. (Be sure to add SDS last, after mixing, otherwise the detergent will foam). Filter-sterilize with a 0.2-micron filter before using. Store at room temperature.

To make 1 M NaCl stock solution, add 2.9 grams dry NaCl into 50 mL HPLC H_2O and vortex. Stock solutions of the liquid reagents should come in the molar concentrations listed.

8M Potassium acetate

To make an 8 molar stock solution, add 19.63 grams into 25 mL HPLC H_2O. Store at 4°C.

Extraction Protocol:

1. If specimens are dry, rehydrate them in a 1.5 mL microfuge tube containing 20 µl HPLC H_2O for 10 minutes. If specimens are frozen, begin the procedure from Step 2.
2. Add 100 µl Bender Buffer directly into the tube with the specimen and homogenize until there are no recognizable mosquito parts. Place used pestle in 1M NaOH.
3. Incubate homogenized samples at 65°C for 1 hour.
4. Add 15 µl cold 8M potassium acetate to each sample. Mix gently and incubate on ice for 45 minutes. (Procedure may be stopped here overnight.)
5. Spin samples in a microcentrifuge (14,000 rpm) for 10 minutes, and then transfer the supernatant to a new 1.5 mL microfuge tube.
6. Add 300 µl 100% ethanol (2X volume) to each supernatant to precipitate DNA. Mix well by inverting the tube. Incubate samples at room temperature for 5 minutes.
7. Centrifuge samples (14,000 rpm) for 15 minutes. Following this spin there should be a small pellet of DNA at the bottom of the tube.
8. Carefully remove the supernatant and discard it, leaving the pellet behind in the tube. Let the pellets dry completely before resuspending—residual ethanol can interfere with PCR later.
9. Resuspend pellets in 50 µl HPLC H_2O for head/thorax or abdomen extractions (100 µl for whole mosquitoes). Ideally, store overnight at 4°C before use. Store DNA permanently at -20°C.

Pestle washing: To prevent DNA contamination in PCR-based analyses, pestles should be soaked in 1M NaOH after use. They should then be washed in soapy water, rinsed off in distilled water, and autoclaved before they are used again.

Appendix G
CSP (Circumsporozoite protein) ELISA

This assay detects *Plasmodium falciparum* CSP protein in mosquito samples. CSP is only expressed during the sporozoite stage of malaria development, so this assay detects only sporozoite-positive mosquitoes, which are capable of transmitting malaria. The monoclonal capture antibody nonspecifically binds to the ELISA plate, after which the addition of blocking buffer prevents nonspecific binding of other proteins. After the addition of mosquito homogenate, the capture antibody binds to CSP and holds it during subsequent wash steps. After the monoclonal antibody is added, it also binds CSP and remains after washing. This antibody is conjugated to a peroxidase which catalyzes ABTS indicator solution, turning the solution green, while negative samples remain uncolored.

Adapted from the Malaria Research and Reference Reagent Resource Center (MR4) Methods in Anopheles Research Manual, available at: http://www.mr4.org/Publications/MethodsinAnophelesResearch/tabid/336/Default.aspx. Limited amounts of Plasmodium falciparum positive controls, capture antibodies, and conjugated antibodies are available free of cost through the MR4 website ((MR #890)).

Materials

PBS (phosphate buffered saline, available from MMI Dept.)
BSA (bovine serum albumin) (A7906)
Casein (Sigma C7078)
Phenol red (Sigma P4758)
IGEPAL CA-630 (Sigma I3021)
Nonidet P-40 Substitute (VWR E109-50ML)
NaOH
HCl
Tween (Fisher BP337)
P.f. capture MAb (MR #890)
P.f. conjugate MAb (MR #890)
P.f. CSP positive control (MR #890)
Glycerol (Sigma G6279)
ABTS solution (Kirkegaard Perry)
10% (w/v) SDS (sodium dodecyl sulfate) (Gibco #15553-035)
96-well U bottom vinyl ELISA plates (Corning #2797)

(Continued next page)

Solutions

BSA Blocking Buffer (Buffer B): 250 mL

250 mL PBS
2.5 g BSA
1.25 g casein
50 µl 0.1 g/mL phenol red stock

Add 250 mL PBS to casein, BSA, and phenol red along with a stir bar. Stir gently (avoid introducing air/foam into the solution) for several hours on stir plate until dissolved (**suggested**: prepare the night before). Store overnight or up to 5 days at 4°C or freeze for future use. For best results, stir for ~20min right before use. Store BSA at 4°C.

Boiled Casein Blocking Buffer (Buffer C): 250 mL

1.25 g casein
25 mL NaOH (0.1M)
225 mL PBS
~5 mL HCl (1 M)

Suspend 1.25 g casein in 25 mL 0.1M of NaOH and bring to a boil while stirring on a hot plate. After casein dissolves, slowly add 225 mL PBS, allow to cool, and then adjust pH to 7.4 with HCl (**suggested**: prepare beforehand). Store aliquots at -20C

Buffer C-Nonidet P-40 (BUFFER C-NP40)

5 uL NP-40 per each 100 uL BUFFER C

Add NP-40 to BUFFER C and mix thoroughly by vortexing. Make fresh daily.

PBS: Tween (wash buffer): 500 mL

500 mL PBS
0.25 mL Tween

MAb (monoclonal antibody) stock

Dissolve lyophilized antibody in 1:1 dH_2O: glycerol, following instructions on the bottle. Store antibody at -20°C. Make the following antibody dilutions immediately prior to use:
Capture antibody: 40 µl stock in 5 mL PBS—this is enough for one 96-well plate
Conjugated antibody: 10 µl stock in 5 mL BUFFER B—this is enough for one 96-well plate

(Continued next page)

P.f. positive control stock
Resuspend *Plasmodium falciparum* CSP protein in 250 µl BUFFER B (vial I)
Take 10 µl from vial I, dissolve in 990 µl BUFFER B (vial II, 100x dilution)
Take 10 µl from vial II, dissolve in 990 µl BUFFER B for working stock (vial III, 10,000x dilution)
For the positive control serial dilution, add 100 µl from vial III to a plate well. Transfer 50 µl of this to the next well down, mix well with 50 µl BUFFER B. Using a new pipet (tip, transfer 50 µl to the next well down, mix well with 50 µl BUFFER B, etc., resulting in 1X, 2X, 4X, 8X, 16X, 32X, 64X, and 128X positive control dilutions.

Mosquito homogenate
Grind each whole mosquito (or head thorax) in 50 µl BUFFER C:NP40 with sterile pestle. Rinse pestle with 125 µl BUFFER C. Transfer 100 uL of this homogenate to a new labeled tube for gDNA extraction (store at -20C or -70C for long-term). To the remaining ELISA homogenate, add 75 uL of BUFFER C-NP40 for a total of 150 uL of ELISA homogenate. ELISA homogenates can be prepared in advance and stored at -20°C (or -70C for long-term).

Negative controls
Homogenize uninfected colony mosquitoes as above for negative controls.

ABTS solution
Pour 1-component solution into 15 mL conical, ~10 mL per 96-well plate. Store at 4°C

Stop Solution
1% SDS (1 mL 10% w/v SDS in 9 mL dH_2O for one 96-well plate). Can heat gently to dissolve.

(Continued next page)

Plate Setup

	1	2	3	4	5	6	7	8	9	10	11	12
A	neg	(+) 1x	(+) 1x	3	7	11	15	19	23	27	31	35
B	neg	(+) 2x	(+) 2x	3	7	11	15	19	23	27	31	35
C	neg	(+) 4x	(+) 4x	4	8	12	16	20	24	28	32	36
D	neg	(+) 8x	(+) 8x	4	8	12	16	20	24	28	32	36
E	1	(+) 16x	(+) 16x	5	9	13	17	21	25	29	33	37
F	1	(+) 32x	(+) 32x	5	9	13	17	21	25	29	33	37
G	2	(+) 64x	(+) 64x	6	10	14	18	22	26	30	34	38
H	2	(+) 128x	(+) 128x	6	10	14	18	22	26	30	34	38

ELISA Protocol

Note: All incubations are carried out at room temperature.

1. Add 50 µl capture MAb solution to each well (40 µl MAb in 5 mL PBS). Cover and incubate overnight.

2. Remove solution by knocking plates upside-down. Fill wells with BUFFER B (~200 µl) and incubate for 1 hour.

3. Remove solution and add positive controls and negative controls to their respective wells. Add 10 µl of mosquito homogenate from each the first five samples to the first well (this is the **pooled** protocol; please add 50 uL of a single sample when using the **unpooled** protocol). Repeat in a second well as a duplicate (refer to the plate setup diagram). Add 50 µl BUFFER C to any empty wells. Incubate for 2 hours.

4. During the 2-hour incubation:
   - Prepare the ABTS solution
   - Dilute the conjugate MAb in BUFFER B as described above (10 µl MAb in 5 mL BUFFER B).
   - Confirm enzyme activity by mixing 5 µl of the conjugate Mab/BUFFER B solution with 100 µl ABTS. A dark green color should begin developing within a few minutes.

5. Remove mosquito homogenate. Wash plate 7 times with PBS-Tween using a plate washer. If no plate washer is available, hand wash by: pipette 20uL PBS-Tween to each well and then tap out the solution. Do this 3 times.

6. Add 50 µl conjugate MAb/BUFFER B to each well, incubate for 1 hour.

7. Remove conjugate MAb/BUFFER B, wash 7 times with PBS-Tween. You can now start rinse/storage for the plate washer equipment. If no plate washer is available, hand wash by: pipette 20uL PBS-Tween to each well and then tap out the solution. Do this 4 times.

8. Add 100 µl ABTS solution to each well and incubate for 60 minutes.
(Continued next page)

9. Add 100 µl Stop Solution to each well and read plate absorbance at 405 nm.

10. The absorbance cut-off for positive samples in 2X the average absorbance of the negative controls.

(Continued next page)

DNA Extraction from CSP ELISA Homogenate

1. 100 µl of CSP ELISA homogenate was transferred to a new 1.5 mL microcentrifuge tube during homogenization for ELISA and stored at -20C or -70C for short- or long-term storage.

2. Using these homogenates, follow the Qiagen DNeasy Blood and Tissue Kit instructions for tissue samples (or the insect-specific protocol), with the following specifications:
    a. Step 1a: Incubate at 56C for 2.5 hr.
    b. Step 2: Incubate samples at 56C for 10 min after adding 200 uL Buffer AL (even though they are technically not blood samples).
    c. Add 30 sec to all centrifugation steps (i.e. 1.5 min instead of 1 min) as the centrifuge takes a little while to get up to speed and also to ramp down.
    d. Elute DNA twice from the column in 30 uL of Buffer AE, storing the second elution in case the first fails. Final centrifugation at 8,000 rpm.

Curriculum Vitae

Christine M. Jones

DOB: June 6, 1991
Email: pbacjones@gmail.com
Phone Number: 443-287-4490 (Business)/940-465-3344 (Personal)
Business Address: Johns Hopkins Bloomberg School of Public Health
Department of MMI, Room E3402
615 N. Wolfe St., Baltimore MD, 21205
Personal Address: 1007 Little Baer Court, Glen Burnie MD, 21061

Education

PhD (current): Molecular Microbiology and Immunology, Johns Hopkins Bloomberg School of Public Health
BS (2013): Genetics, Texas A&M University

Research

Johns Hopkins Bloomberg School of Public Health — Doctoral Research
Advisor: Douglas E Norris — June 2014 - present

- **Mitochondrial genome generation and analysis of field-caught *An. funestus*.** Conducted library preparation for next-generation Illumina sequencing of 43 *An. funestus*. Generated complete mitochondrial genomes from sequence data. Used ArcGIS to select sampling sites, to display sample sites, and to generate pairwise distance matrices for further analysis in R. Analyzed population genetics and genomics to determine the relatedness of *An. funestus* individuals across large distances, in different ecological settings, and between different mitochondrial lineages.
- **Outdoor collection of Anopheles and analysis of the efficacy of an artificial human odorant bait.** Led a field team of 3-5 in a month-long collection series of resting and foraging mosquitoes in Zambia. Identified species by morphological characters. Conducted ELISA and PCR to characterize the human blood-feeding rates and *Plasmodium* prevalence in mosquito populations. Modelled the abundance and diversity of anophelines in traps as a function of covariates including type of attractant
- **Phylogenetic analysis of highly diverse outdoor-captured anophelines.** Aligned conserved sequence region of >100 unidentified study specimens to specimens from NCBI. Constructed Bayesian and Maximum Likelihood phylogenetic trees to generate hypotheses of taxonomy.

Johns Hopkins Bloomberg School of Public Health — PhD Rotations
Advisors: Jay Bream and Andrew Pekosz — Sept 2013 – June 2014

- Investigated the response of human nasal primary epithelial cells and mouse tracheal primary epithelial

cells to influenza A virus infection.
- Assessed the role of single nucleotide polymorphisms in the mRNA expression of a human immune protein

Department of Biochemistry and Biophysics, Texas A&M University — Undergraduate Research
Advisors: Margaret E Glasner and James Hu — Oct 2011 – May 2013
- Evaluated the role of specific residues in substrate specificity/catalytic efficiency and catalytic promiscuity of OSBS family members.
- Generated functional annotations for proteins and gene products for online databases

Teaching and Mentoring Experience

Mentor, PhD Laboratory — June 2014 - present
Johns Hopkins Bloomberg School of Public Health
Supervisor: Douglas E Norris (410-614-2710), May Contact
Provide training in basic molecular biology theory, technique, and phylogenetic analysis to students/trainees

Teaching Assistant, *Spatial Analysis I&II: ArcGIS* — Sept 2015 – Oct 2017
Johns Hopkins Bloomberg School of Public Health
Supervisor: Timothy Shields (410-502-9077), May Contact
Provide technical assistance to students during class time and office hours and grade assignments

Teaching Assistant, *Vector Biology and Vector-Borne Diseases* — Jan 2015 – March 2018
Johns Hopkins Bloomberg School of Public Health
Supervisor: Douglas E Norris (410-614-2710), May Contact

Write and grade exams, organize review sessions, and mentor students.

Job Related Training

Summer Institute in Statistical Genetics — July 2017
University of Washington, Seattle, WA
Courses: Population Genetic Data Analysis; Introduction to Advanced Population Genetics; Introduction to Molecular Epidemiology and Infectious Disease Phylodynamics

Skills

Computer Skills: ArcGIS *(advanced)*; QGIS *(basic)*; Google Earth Pro (*basic);* R *(intermediate)*; STATA *(basic);*
Language Skills: Russian (Spoken: Intermediate, Written: Intermediate, Read: Intermediate)

Publications

Jones CM, Lee Y, Kitchen A, Collier C, Pringle JC, Muleba M, Irish S, Stevenson JC, Coetzee M, Cornel AJ, Norris DE, Carpi G. Complete *Anopheles funestus* mitogenomes reveal an ancient history of mitochondrial lineages and their distribution in southern and central Africa. *Scientific Reports*, *8*(1), 9054.

Stevenson JC, Pinchoff J, Muleba M, Lupiya J, Chilusu H, Mwelwa I, Mbewe D, Simubali L, **Jones CM,** Chaponda M, Coetzee M, Mulenga M, Pringle JC, Shields T, Curriero F, Norris DE. Spatio-temporal heterogeneity of malaria vectors in northern Zambia: implications for vector control. *Parasit Vectors* 9, 510 (2016).

Stevenson JC, Simubali L, Mbambara S, Musonda M, Mweetwa S, Mudenda T, Pringle JC, **Jones CM,** Norris DE. Detection of *Plasmodium falciparum* Infection in *Anopheles squamosus* (Diptera: Culicidae) in an Area Targeted for Malaria Elimination, Southern Zambia. *J. Med. Entomol*. 53, 1482–1487 (2016).

Poster Presentations

Jones CM, Lee Y, Collier T, Pringle JC, Kitchen A, Stevenson JC, Coetzee M, Muleba M, Cornel AJ, Carpi G. *Anopheles funestus* in southern and central Africa: analysis of mitochondrial diversity. *American Society of Tropical Medicine and Hygiene*. 2017 Nov. 5-9: Baltimore, MD.

Jones CM, Stevenson JC, Ciubotariu IL, Lupiya J, Mbewe D, Muleba M, Norris DE. Outdoor barrier collection in Nchelenge District, Zambia. *American Society of Tropical Medicine and Hygiene*. 2017 Nov. 5-9: Baltimore, MD.

Jones CM, Lee Y, Collier T, Pringle JC, Kitchen A, Stevenson JC, Coetzee M, Muleba M, Cornel AJ, Carpi G. Mitochondrial population structure of *Anopheles funestus* in southern and central Africa. *The Future of Malaria Research Symposium*. 2017 Nov. 3: Baltimore, MD.

Pringle JC, Searle KM, Carpi G, **Jones CM**, Kobayashi T, Norris DE, Moss WJ. A genomics approach to track malaria transmission in different epidemiologic settings in Zambia. *Genetics Research Day*. 2017 Feb. 17: Baltimore, MD.

Jones CM, Lee Y, Collier T, Pringle JC, Carpi G, Stevenson JC, Coetzee M, Muleba M, Yamasaki Y, Cornel AJ, Norris DE. Population structure of *An. funestus* in southern Africa. *American Society of Tropical Medicine and Hygiene*. 2016 Nov. 13-17: Atlanta, GA.

Jones CM, Pringle JC, Carpi G, Stevenson JC, Norris DE. Vector biology, ecology, and genetics: Malaria transmission and control. *World Malaria Day Symposium*. 2016 Apr. 25: Baltimore, MD.

Jones CM, Muleba M, Kessler MK, Das S, Moss WJ, and Norris DE. Changes in malaria vector dynamics post-IRS in Nchelenge District, Zambia. *American Society of Tropical Medicine and Hygiene*. 2015 Oct. 25-29: Philadelphia, PA.

Das S, **Jones CM**, and Norris DE. Spatial and temporal dynamics of malaria vectors in Nchelenge district, Zambia. *American Society of Tropical Medicine and Hygiene 63rd Annual Meeting*. 2014 Nov. 2-6: New Orleans, LA.

Jones CM, McMillan A, Odokonyero D, Glasner ME. Evolution of enzyme specificity in the OSBS family. *Texas A&M Student Research Week*, March 2013: College Station, TX.

Affiliations

American Society of Tropical Medicine and Hygiene – Member