Universidade de São Paulo
Faculdade de Saúde Pública

SUSAN ELAINE GRENI

A multi-gene analysis and proposed distribution of species of the Strodei Subgroup of *Anopheles* (*Nyssorhynchus*) (Diptera: Culicidae)

Tese apresentada ao Programa de Pós-Graduação em Saúde Pública para obetenção do título de Mestrado em Saúde Pública

Área de concentração: Epidemiologia

Orientadora: Profa. Dra. Maria Anice Mureb Sallum

São Paulo
2016

A multi-gene analysis and proposed distribution of species of the Strodei Subgroup of *Anopheles* (*Nyssorhynchus*) (Diptera: Culicidae)

SUSAN ELAINE GRENI

Tese apresentada ao Programa de Pós-Graduação em Saúde Pública para obetenção do título de Mestrado em Saúde Pública

Área de concentração: Epidemiologia

Orientadora: Profa. Dra. Maria Anice Mureb Sallum

São Paulo
2016

Catalogação da Publicação
Biblioteca/CIR: Centro de Informação e Referência em Saúde Pública
Faculdade de Saúde Pública da Universidade de São Paulo
Dados fornecidos pelo(a) autor(a)

Elaine Greni, Susan
A multi-gene analysis and proposed distribution of species of the Strodei Subgroup of Anopheles (Nyssorhynchus) (Diptera: Culicidae) / Susan Elaine Greni; orientadora Maria Anice Mureb Sallum. -- São Paulo, 2016.
74 p.

Dissertação (Mestrado) -- Faculdade de Saúde Pública da Universidade de São Paulo, 2016.

1. Epidemiology. I. Mureb Sallum, Maria Anice, orient. II. Título.

To my light, Rodrigo, whose absolute and unconditional love and support has gotten me through times of utter darkness, to my dearest Jonah, who reminds me, each day, of how wonderful life can be and to my parents, for the years of and care.

ACKNOWLEDGEMENTS

To the following people, I owe my sincerest gratitude in helping me finish my project.

To my advisor, Professor Maria Anice Mureb Sallum, thank you for taking me under your wing, for helping me when I was most lost, for all the knowledge you have passed on to me, for all the patience, kindness, support, suggestions and endless corrections to my myriad drafts.

I am in debt with Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP Grant no. 2011/20397-7 and no. 2014/26229-7 to Maria Anice Mureb Sallum) for the financial support that allowed all field collections and my lab work.

To Professor Lincoln Suesdek for the CAPES fellowships Process #23038.005274/2011-24 over the years of my research.

To Bruna Silva, who has been my diligent mentor, throughout this process. You have taught me tirelessly about phylogenetics and academic writing. I truly could not have done it without you. Thank you for all the patience, support and hard work.

To Tatiane Oliveira, thank you for all your help in the lab, for your friendship and for your support.

To everyone in the Laboratório de Entomologia de Saúde Pública – Sistemática de Culicidae, thank you for accepting me when I did not even speak a word of Portuguese. You were all so kind and patient with me and I will never forget my time at FSP because of you all.

SUMMARY

Introduction *Anopheles strodei* sensu lato is an understudied subgroup of potential epidemiological importance, having been found naturally infected in Brazil with *Plasmodium falciparum*, *Plasmodium vivax* and *Plasmodium malariae*. *An. strodei* s.l. is currently composed on 8 species: *An. albertoi* Unti, *An.* CP Form, *An. rondoni (*Neiva & Pinto), *An. strodei* Root, *An. arthuri* Unti and three other unnamed species that have been proposed by Bourke et al. (2013): *An. arthuri* B, *An. arthuri* C and *An. arthuri* D. **Objectives** As delineating species accurately is an essential goal of public health entomology, the objectives of this study were to: 1) Determine the phylogenetic relationships within the Strodei Subgroup and reaffirm or reject the hypothesis of the 3 new species (*An. arthuri* B, *An. arthuri* C and *An. arthuri* D) 2) Address the potential spatial distribution of species of the *An. strodei* subgroup to provide support for the candidate species in the Strodei Subgroup **Methods** Bayesian inference, which included DNA sequences of one mitochondrial and three nuclear protein coding genes: *CO1*, *white*, *CAD* and *CAT*, was used to determine the phylogenetic relationship within the group. To propose a species distribution, collection localities, along with climatic and geographic data were input into MAXENT. **Results** When analyzing the four molecular markers employed, support was found for allopatry in the Strodei Subgroup. The paraphyletic clade of *An. arthuri* was supported. **Conclusion** Potential species distributions of the Strodei Subgroup were addressed for the first time. Fifty-five unique *CAT* sequences and 46 unique *CAD* sequences were newly characterized.

Key words: Anopheles strodei, MAXENT, Bayesian analysis, phylogenetics, diptera, culicidae

SUMMARY

1. INTRODUCTION

Mosquito-borne diseases, including Zika virus, malaria, dengue and yellow fever are posing a serious health threat across the world. The economic and clinical costs to society due to mosquito-borne diseases have been estimated in the billions, not to mention the prohibitive personal expenses (lost wages, travel costs associated with doctor visits) to some of the world's poorest and most vulnerable populations, such as low socioeconomic status pregnant women and children, who also happen to be some of the most affected by these diseases (Bôtto-Menezes et al. 2016, Shepard et al. 2014, Suaya et al. 2009).

In 2014 alone, malaria, a disease caused by at least six species of the *Plasmodium* parasite (Calderaro et al, 2013), transmitted by approximately 60 species of *Anopheles* mosquitoes (Neafsey et al. 2015, Warrell & Gilles, 2002), was responsible for a staggering 207 million cases worldwide, and 438.000 deaths. An estimated 36 of those deaths occurred in Brazil (Bôtto-Menezes et al. 2016).

Although there has been a pronounced decrease in the number of cases of malaria in Brazil from the year 2000 to the year 2014, down 84%, the number is still unacceptably large at 143,145 cases reported in Brazil in 2014 (WHO, 2014).

Current studies suggest that the changing climate and global warming trends will have a direct effect on arthropod vector population size, density and distribution and, consequently, vector-borne disease transmission rates (Laporta et al. 2015, Beard et al. 2016, Githeko et al. 2000). Because of this and because of the immense effect that these diseases have, it is of the upmost importance to understand the ecological dynamics and mosquito vector species involved in the transmission (Bôtto-Menezes, et al. 2016).

1.1 FAMILY CULICIDAE

Mosquitoes, as they are commonly known, are the most important arthropod vector of zoonotic disease affecting both humans and livestock (Gubler, 1998). They predominantly transmit parasitic diseases, such as malaria, and viral diseases, such as dengue fever, Zika virus and yellow fever, among many others pathogens (Gubler, 2008).

Mosquitoes are easily identifiable due to their slender, mostly scale-covered, two-winged bodies and their long proboscis used for feeding (Harbach, 2013). The females of some species are well-known irritants within the human population for their hematophagy before oviposition while the males do not take blood meals and feed mostly on fruits and plant nectar (Bradshaw, 1979).

Mosquitoes belong to the extremely diverse family Culicidae of the Diptera Order. The family is comprised of over 3,000 currently recognized species (Harbach, 2015). The family Culicidae is broken into two subfamilies Anophelinae and Culicinae (Service, 2004). The Anophelinae subfamily is further broken down into three genera, *Anopheles, Bironella* and *Chagasia (*Krzywinski et al., 2001; Sallum et al., 2005).

The *Anopheles* genus was first described by a German entomologist named Johann Wilhelm Meigen in 1818. The anophelines are the only known competent vectors for the five *Plasmodium* parasites that cause malaria in humans, *Plasmodium falciparum, Plasmodium knowlesi, Plasmodium ovale, Plasmodium vivax and Plasmodium malariae*, (WHO, 2015). The genus can be found in nearly all habitats across the globe from shaded forests in North America to cosmopolitan areas in southern Brazil (Warrell & Gilles, 2002). It is the largest and most diverse of the genera and is comprised of some 450 recognized species with an estimated 60 of those being competent *Plasmodium* vectors (Neafsey et al.

2015, Warrell & Gilles, 2002). Its wide reach and capacity as a competent disease vector of malaria and some types of the encephalitis virus make *Anopheles* an important target for further studies. The *Anopheles* genus is subdivided into eight subgenera: *Anopheles*, *Baimaia*, *Cellia*, *Chrystia*, *Kertezia*, *Lophopodomyia*, *Nyssorhynchus* and *Stethomyia* (Harbach and Kitching, 2015).

The *Nyssorhynchus* subgenus is currently comprised of 39 formally recognized species, which includes the main malaria vector in Brazil *An. darlingi* (Harbach, 2014; Moroni et al., 2010; Martins-Campos et al.). The *Nyssorhynchus* subgenus is further divided into 3 sections: the Albimanus Section (Faran, 1980), Argyritarsis Section (Linthicum, 1988) and the Myzorhynchella Section (Galvão, 1941).

The Albimanus section is divided into two groups: Albimanus (monotypic) and Oswaldoi (Harbach, 2014). Of the Albimanus Section very little is known of the Strodei Subgroup. The Strodei Subgroup includes *An. albertoi* Unti, *An.* CP Form (Sallum et al., 2010), *An. rondoni* (Neiva & Pinto), *An. strodei* Root, *An. arthuri* Unti and three other unnamed species that have been proposed by Bourke et al. (2013): *An. arthuri* B, *An. arthuri* C and *An. arthuri* D. *Anopheles arthuri* and *Anopheles albertoi* were described by Unti in 1941, and later were transferred to the synonym of *An. strodei* by Faran (1980).

1.2 STRODEI SUBGROUP

The Strodei Subgroup is comprised of *An. albertoi* Unti, *An.* CP Form, *An. rondoni (*Neiva & Pinto), *An. strodei* Root, *An. arthuri* Unti and three other unnamed species that have been proposed by Bourke et al. (2013): *An. arthuri* B, An. *arthuri* C and *An. arthuri* D

Anopheles strodei is currently known by possessing three synonyms, *An. strodei variety*

ramosi Unti, *An. strodei variety artigasi* Unti, e *An. strodei variety lloydi* Unti (Sallum, 2010). In 1926, Root described *An. strodei* based on external morphological characteristics of adult male and fourth instar larvae collected in the district of Agua Limpa, Juiz de Fora municipality, Minas Gerais state, Brazil. The author included in the description specimens from Panama, Paraguay and Argentina. Subsequently, Galvão and Lane (1936) described the eggs of *An. strodei* and noted morphological similarities in the mosquitoes but distinct differences in the egg morphology. In 1938, Galvão separated eggs from females of *An. strodei* that were collected over three years in different climatic regions in two groups based on major morphological differences. The Group I had large floats while the Group II had small floats. These descriptions, which noted morphological similarity between the adults but physical differences in the eggs, plus Root's original descriptions in 1926 was later used by Galvão in 1940 to describe *An. strodei* as a species complex.

Unti in 1940 and in 1941 described five species that were morphologically similar to *An. strodei*: *An. strodei ramosi, An. strodei arthuri, An. strodei artigasi, An. strodei albertoi,* and *An. strodei lloydi* based on the observed variations in their eggs and palmate setae (abdominal seta 1) on the abdomen. Faran (1980), due to lack of vouchers from Unti's study, transferred all five species to the synonymy of *An. strodei.* Mosquitoes of the *Nyssorhynchus* genus are notoriously difficult to identify using morphology alone due to few differences in external features (Sallum, 2010). Because of the difficulties, Fritz (1998) tried to characterize *An. strodei* morphologically using the DNA sequence of the ITS2 rDNA, however, he was unable to find any significant differences. Posteriorly, Fritz (2004) tried to identify *An. strodei, An. trinkae* Faran, A*n. rangeli* Gabaldón, Cova Garcia & Lopez and *An. triannulatus* using the ITS2 rDNA again and was able to differentiate

between these four and describe a protocol for molecular identification using the ITS2 region of the rDNA. Since molecular techniques have become more widely available, more sequencing and genetic analyses have been done and the evidence supporting the different species in the Strodei Subgroup has become even more overwhelming.

In 2010, Sallum and collaborators revised the subgroup and resurrected two species from the synonymy of *An. strodei*, thus added two species to the Strodei Subgroup, *An. albertoi* and *An. arthuri*. *Anopheles strodei albertoi*, unlike most other Anophelinae, can be distinguished easily based on its egg, which is much different due to the lack of floats. *Anopheles strodei strodei* has egg that is broader than the other species in the complex and that has floats (Sallum, 2010).

Subsequently, Bourke et al. (2013) performed a Bayesian phylogenetic analysis using the *ITS2, white* and *COI* with 61 samples from the Strodei Subgroup collected from various localities throughout Brazil. The gene tree produced using the concatenated DNA sequences of the three concatenated molecular markers supported seven monophyletic clades. Thus, Bourke and collaborators found evidence to suggest 4 species under the name An. *arthuri* sensu lato. They are: *An. arthuri* A, *An arthuri* B, *An arthuri* C and *An. arthuri* D (Bourke et al. 2013). The last three are likely undescribed new species, whereas *An. arthuri* A is likely the sensu stricto.

The Strodei Subgroup has a wide spatial distribution having been identified as far north as Panama and as far south as Argentina. The *An. strodei strodei* species has been observed as being highly zoophilic showing a clear preference for blood feeding on pigs and cattle, however, apart from this, very little is known about the feeding habits of the species (Tadei and Thatcher, 2000) and requires further investigations.

The involvement of species of the Strodei Group in the dynamics of malaria transmission remains unclear, however, mosquitoes identified as the species *An. strodei* have been incriminated as potential vectors. In a study that involved over 12,000 field-collected Anophelinae in the state of Rondônia in the Brazilian Amazon, Oliveira-Ferreira et al (1990) found *An. strodei* naturally infected with *Plasmodium vivax*. More recently Duarte et al. (2013) found *An. strodei* to be naturally infected with *Plasmodium malariae* in the Atlantic rain forest in the state of Sao Paulo in Brazil (Duarte et al, 2013). Additionally, *An. strodei* was also found naturally infected with *P. falciparum* in the Southeastern Atlantic Forest, thus providing more evidence to implicate *An. strodei* as a potential competent *Plasmodium* vector (Laporta, et al. 2015).

Whether species of the Strodei Subgroup play a role as either primary or secondary vectors of *Plasmodium* is still questionable. In light of such uncertainties, further studies are required to accurately delineate the species in the Strodei Subgroup, ascertain their feeding habits, larval habits, spatial distribution and then distinguish which species are involved in the epidemiology of malaria transmission.

1.3 WHAT DEFINES A SPECIES?

There are, of course, obvious characteristics that define a species. The oldest and easiest being external morphological differences (Hoef-Emden n.d.). When talking about different species of mosquitoes, sometimes it can be fairly easy to recognize two different species when their morphology is quite different. Comparing the beautiful *Sabethes cyaneus* to the rather plain looking *Aedes aegypti,* one could quite readily identify the physical differences. Their size, coloration, scale patterns, length of proboscis, length of segments of antennae might all be different, for example. These physical differences, however, can only

differentiate between very morphologically distinct species and not so for cryptic or sister species, which may be morphologically similar but genetically distinct. Another problem with the classical morphospecies concept is intraspecies variation or sexual dimorphisms (The University of California Museum of Paleontology, Berkeley, and the Regents of the University of California n.d.). Similar looking, but genetically different mosquitoes may display completely different feeding and larval habits and may not even occupy the same geographical area.

With that in mind, biologists set out to define species in a different, more complete way. One defining characteristic of a species, as cited by Poulton (1903), Dobzhansky (1937), and Mayr (1942), when he put forth the biological species concept, is the lack of interbreeding. Put another way, members of the same species can mate and produce viable offspring while members of different species cannot. Lack of gene flow, due to either pre-mating isolation (i.e.: physical separation, different larval habits, season separation, and so on) or post-mating isolation (i.e.: gametic incompatibility, hybrid unviability, and so on) may be helpful in defining real relationships in different populations of organisms (Mallet n.d.).

As mentioned previously, species distribution or ecological niche is another facet of defining a species. Organisms evolve to occupy a specific niche. Ecological and geographical data about a certain niche such as temperature, relative humidity, precipitation and altitude can be used in a model with occurrence data to predict the range of suitable habitat (Warren and Seifert 2011). Such an approach is called ecological niche modeling and can be used not only to help support a species classification but can also be used to predict the distribution of a species.

One final way to define a species is through the use of DNA methods. DNA barcoding in taxonomy has become quite standard as it is fast, cost effective, efficient and more accurate than classical taxonomical approaches based solely on morphological characters (Jörger et al., 2012). The DNA barcoding usually relies on utilizing and comparing a short fragment of 658 base pairs long of the protein-coding cytochrome c oxidase I gene (*COI*) mitochondrial gene being the most commonly used (Monaghan et al., 2005; Hebert et al., 2004). The *COI* gene has become widely used as DNA barcoding because it is easy to work with, is present in many organisms and is a multiple copy gene (The Ideal barcoding gene, 2009).

Notwithstanding, it is important to note that when examining and investigating phylogenetic relationships of closely related species, many have found that it is important to use DNA sequences from more than one gene to elucidate accurate relationships (Dayrat, 2005). Barcoding is great for molecular identification of species that already have published sequences however, when trying to reconstruct evolutionary relationships; it is ideal to use additional genetic markers.

For instance, Foster et al. (2013) found that the use of CO1 barcoding region alone was a poor choice for species delineation, however, when used in conjunction with the carbamoylphosphate synthetase-aspartate transcarbamylase-dihydroorotase (*CAD*) and *white* genes, in a Bayesian analysis the trio was quite effective at confirming species (Foster et al., 2013; Krzywinski et al., 2001).

Bourke et al. (2013) used the *CO1* gene, the Internal Transcribed Spacer 2 (ITS2) and the *white* gene to address the relationships within the Strodei Group (Bourke et al., 2013).

Surendran et al. (2013) used both the *ITS2* as well as the *CO1* genes in the identification of members of the *An. subpictus* complex.

Sum et al. used the ITS2 gene alone in their phylogenetic study of members of the *Anopheles* genus in Malaysia (Sum et al., 2014).

1.4 SPECIES DELINEATION AND ITS BIOLOGICAL IMPORTANCE

Accurate species delineation is valuable for examining earth's mostly unrecognized biological diversity (Hebert et al., 2004; Monaghan et al., 2005). It is estimated that there are at least 10 million undiscovered species suggesting an unimaginably biodiverse world. Unfortunately, however, biodiversity is decreasing at an unprecedented rate largely because of anthropogenic actions (deforestation, human-induced global warming, etc.). It follows then, that delineating, naming and describing species in an accurate and timely manner is now more important than ever (Wilson, 2004).

One confounding factor in delineating similar species, especially in insects, is the existence of cryptic species, also known as sibling species or species complexes (Saez & Lozano, 2005). Cryptic species (CS) are two or more species that seem to be a single species based on their morphological similarities but have been reclassified as distinct species based on the results of an integrative taxonomical analysis including morphological studies, phylogenetic studies and ecological niche modeling (Bickford et al., 2006). While species complexes are often morphologically indistinguishable from one another, they may demonstrate behavior differences and a wide genetic variation (Bourke et al., 2013).

Knowledge of CS is deepening rapidly thanks to the increased usage of DNA-based phylogenetic studies (Padial et al., 2010). Cryptic species have been discovered in many

different habitats across the world; from the neotropical skipper butterfly *Astraptes fulgerator* to tropical water beetles in Madagascar to complexes of mosquitoes in the *Nyssorhynchus* subgenus (Hebert et al., 2004).

In 2014 Wang et al used *ITS2* and *CO1* to investigate possible sister species in the Barbirostris Subgroup of *Anopheles* in China. They proposed two cryptic species in the subgroup *(*Wang et al, 2014).

In a 2012 study in Kenya, 422 female Anophelinae were caught. Of these 422 anopheline mosquitoes, 216 could not be identified using current taxonomic keys and had *ITS2* and *CO1* sequences that were different from any other published sequences. Fifteen of these were found to be new species with several proposed cryptic species (Stevenson et al, 2012).

A recent study in 2015 in Eastern Zambia found a number of cryptic species similar in taxonomy and sequences to *An. coustani and An. crypticus (Lobo et al, 2015).*

1.5 SPECIES DELINEATION AND ITS PUBLIC HEALTH IMPORTANCE

Important still, describing and naming possible disease vectors is vital goal of public health entomology. Different species demonstrate different feeding and breeding habits, which can affect vector control efforts and public health initiatives (Bourke et al., 2013).

A classic example of this is the *An. gambiae* sensu lato that is arguably the deadliest mosquito in the human world. The prime vector of *Plasmodium* in sub-Saharan Africa, *An. gambiae* has been attributed to transmitting the *Plasmodium* parasite to millions. The complex is currently comprised of eight species: *An. amharicus, An. arabiensis, An. bwambae, An. gambiae, An. coluzzii, An. melas* and *An. merus (*White et al. 2014). The complex shows the importance of species delineation in the context of public health

because of the difference in behaviors and biting habits, which affect the vector competence of each of these mosquitoes. *An. arabiensis,* a major contributor to malaria transmission rates in sub-Saharan Africa, shows much more exophilic and exophagic behaviors than other members of the complex (Kent et al 2007). *An. arabiensis,* when compared to *An. gambiae,* shows increased zoophilic feeding behaviors as well (Kent et al 2014). Being able to distinguish between these two species, clearly, would have an effect on public health measures taken to combat them.

1.6 SPECIES DELINEATION IN THE PRESENT STUDY

1.6.1 COI Gene

Cytochrome C oxidase (*COX)* is a mitochondrial gene that encodes for the complex IV that is the terminal component of the electron transfer chain in oxidative phosphorylation (Youfen, et al. 2006). *Cytochrome C oxidase 1* (COI) encodes for the 1 subunit of the complex IV. Oxidative phosphorylation (OP), which produces ATP, the primary energy source, is an absolutely fundamental need of the cell. Because of this, the genes that encode the components of OP are highly conserved.

Because some regions of CO1 are so highly conserved, it is an ideal target for interspecific genetic analysis but due to its high rate of substitution, it is also an ideal choice for intraspecific genetic analysis (The Ideal Barcoding Gene 2009). Also, because it is a mitochondrial gene, which means each cell has multiple copies of the gene, it is a less complicated target for amplification via polymerase chain reaction (PCR) (The Ideal Barcoding Gene 2009). It is a quick and efficient choice for genetics and is very widely used for DNA barcoding (Bourke et al., 2014).

1.6.2 *CAD* Gene

The *CAD* gene, carbamoylphosphate synthetase-aspartate transcarbamylase-dihydroorotase, also known as rudimentary, is a single-copy, protein encoding nuclear gene complex. The *CAD* gene is found in anopheles on the X chromosome. *CAD* encodes a protein that is involved with enzymatic activity in the pyrimidine synthesis pathway (Danforth, 2009; NCBI, 2016).

CAD, which has a relatively high rate of substitution, coupled with the fact that it is a single copy gene, makes it somewhat difficult to work with (Danforth et al., 2006). However, CAD has been used successfully before in phylogenetic analyses of mosquitoes as well as in other non-vector arthropods (Reidenbach et al., 2009; Danforth et al., 2006; Moulton et al., 2004; Wild et al., 2008).

1.6.3 *White* Gene

The *white* gene is a 16kb single-copy protein coding nuclear gene (Krzywinski et al. 2001) that codes for proteins that control eye color (Besansky and Fahey 1997). Both the *CAD* gene and the *white* genes as well are practical to use in genetic studies because of the ease of alignment and high phylogenetic resolution (Besansky and Fahey, 1997; Zhang et al., 2012).

Foster et al used the *white* gene to investigate relationships in the *Anopheles* subgenus and found that it was not informative enough to delineate species when used alone but when used in conjunction with other genes, such as *COI* and *CAD* could be used more efficiently for phylogenetic analysis (Foster et al., 2013).

1.6.4 *Catalase* Gene

The *catalase* gene (*CAT*) is a single copy, nuclear gene that codes for the catalase enzyme, an enzyme that is omnipresent in all living organisms. Hydrogen peroxide is a byproduct of cellular respiration and is produced in all living cells. Hydrogen peroxide is an acid that can cause significant damage to cells if not removed. The catalase enzyme is responsible for this removal by breaking up hydrogen peroxide into oxygen and water (Corona and Robinson 2006). Because of its biological importance, the *CAT* gene is highly conserved making it ideal for addressing intraspecific genetic differences. Catalase has been used previously to assess relationships in the Anophelinae and Culicinae subfamilies (Reidenbach et al., 2009).

1.7 ECOLOGICAL NICHE MODELING

An ecological niche is the relationship a species has with its environment in relation to the resources it needs and the services that it provides; basically the ecological niche is the species´ specific role in its environment. The niche that a certain species inhabits and its characteristics is a serious factor in determining its actual or possible distribution (Peterson, 2001). Knowing about the potential spatial distribution of a species, especially possible disease vector, like the Strodei Subgroup, is an important goal in medical entomology. To control disease vectors, knowledge of their feeding and breeding habits is as important as knowledge about their spatial distribution. Also, when trying to resolve a complex, such as the Strodei Subgroup, one might look at whether the populations are separate. If they are separate and occupy specific niches, it could be an important clue in determining their conspecific relationships.

The Genetic Algorithm for Rule Set Prediction (GARP) is a computer-based ecological niche model (ENM) developed by David Stockwell (Stockwell, 2004; Natural History

Museum of University of Kansas). The GARP ENM uses the collection localities as an input, along with information about geographic parameters that may hinder or promote the survival of a species (temperature, humidity, elevation, etc.) to predict its niche model or possible distribution (NHM of KU).

Another important model used in ecological niche modeling, and the model used in the present study, is maximum entropy (MAXENT) (Philips et al. 2006). Maxent, similar to other distribution models, is a computer-based algorithm that produces a potential distribution based on limited input information. The program proposes a potential spatial distribution of species based on presence only data points using a maximum entropy approach coupled with environmental factors for each cell on the landscape (Proosdij et al., 2015; Merow et al., 2013). Maxent has a number of advantages over other distribution models. First, it has been proposed as more accurate in predicting distributions based on small sample sizes (Merow et al., 2013; Folley et al. 2010). Furthermore, it is a presence point only model, as opposed to other distribution models, which use both presence and absence points. Being a presence only model becomes especially important when considering two points, one, absence points can be difficult to attain and secondly, they can be inaccurate as absence of a species could really be failure to observe said species and not true absence (Baldwin, 2009). A final advantage to consider is that Maxent accepts the input of both categorical as well as continuous variables (Baldwin, 2009). When Maxent is run, it creates a map with a probability of a species distribution as well as various statistics including the AUC, area under the curve. The AUC value is the gold standard to evaluate the performance of the model. The values of the AUC range from 0 to 1 with the closer to 1 being associated with a better discrimination between suitable habitat and

unsuitable habitat for the species. An AUC value of less than 0.5 is associated with a worse-than-random fit model. The AUC is similar to the Whitney-Mann U statistic and can be interpreted as the likelihood that when one absence point and one presence point are chosen at random, the presence point will have the higher predictive value than the absence point (Elith et al. 2006).

2. JUSTIFICATION

The Strodei Subgroup of the *Anopheles* genus is an understudied complex of mosquitoes of potential public health, biological and taxonomical importance. The subgroup could contain one or more vector of public health importance and is in need of revision. The subgroup is currently comprised of *An. strodei, An.* CP form and *An. albertoi* and could contain other four phylogenetic clades, and likely potential species, named *An. arthuri* A-D (Bourke et al., 2013).

This complex not only needs a revision but further study of its distribution and feeding and breeding habits are of paramount importance. The current study intends to use ecological niche models to predict the spatial distribution of the Strodei Subgroup, an important piece of information necessary to predict where potential vectors and humans or livestock may overlap (Folley et al., 2010)

3. OBJECTIVES

Determine the phylogenetic relationship within the *Anopheles strodei* subgroup and reaffirm or reject the hypothesis of the 7 species proposed by Bourke et al.

Propose the potential geographical distribution of species of the *An. strodei* subgroup to provide clues to better define species and to augment knowledge of the Strodei Subgroup, a subgroup of possible epidemiological importance.

4. HYPOTHESIS

Anopheles strodei subgroup comprises 7 distinct species of mosquitoes.

Anopheles arthuri A-D are valid species.

5. MATERIALS AND METHODS

5.1 COLLECTION

All collections were done by Sallum, Bergo et al. in the Brazilian municipalities and states listed in Table 1. The specimens collected were either adults or larvae and pupae. The adults were collected using a Shannon trap and the larvae and pupae were collected from habitats and then kept alive in the laboratory to obtain the adults linked with larval and pupal exuviae in the Laboratório de Sistemática e Ecologia de Culicidae (LASEC) at the Faculdade de Saude Publica of the University of Sao Paulo. The identification of all of the specimens was done by Dr. Maria Anice Mureb Sallum based on the male genitalia and/or scanning electron micrographs of the eggs, confirmed by ITS2 and *COI* sequence data.

Table 1. Information about the samples used including code used, species, municipality and state of collection, and geographical latitude and longitude of collection location.

Code	Species	Locality (state)	Latitude	Longitude
MG07124	*An. albertoi*	Frutal (MG)	-20.025278	-49.076500
MG0734	*An. albertoi*	Frutal (MG)	-20.025278	-49.076500
MG07710	*An. albertoi*	Frutal (MG)	-20.025278	-49.076500
CE1211	*An. arthuri* B	Ubajara (CE)	-3.8867500	-41.001250
CE1246	*An. arthuri* B	Ubajara (CE)	-3.8867500	-41.001250
CE17152	*An. arthuri* B	Ubajara (CE)	-3.8442220	-40.897778
CE1751	*An. arthuri* B	Ubajara (CE)	-3.8442220	-40.897778
CE20104	*An. arthuri* B	São Benedito (CE)	-4.0964170	-40.896361
CE20243	*An. arthuri* B	São Benedito (CE)	-4.0964170	-40.896361
GO713	*An. arthuri* A	Itarumã (GO)	-18.906128	-51.024917
GO72102	*An. arthuri* A	Itarumã (GO)	-18.906128	-51.024917
GO73105	*An. arthuri* A	Itarumã (GO)	-18.906128	-51.024917
GO76101	*An. arthuri* A	Itarumã (GO)	-18.906128	-51.024917
MG0312	*An. arthuri* A	Frutal (MG)	-19.981278	-49.096028
MG0412	*An. arthuri* A	Frutal (MG)	-19.988472	-49.093361
MG071100	*An. arthuri* A	Frutal (MG)	-20.025278	-49.076500
MG0710 106	*An. arthuri* A	Frutal (MG)	-20.025278	-49.076500
MG0718100	*An. arthuri* A	Frutal (MG)	-20.025278	-49.076500
MG07202	*An. arthuri* A	Frutal (MG)	-20.025278	-49.076500
MG0763	*An. arthuri* A	Frutal (MG)	-20.025278	-49.076500
MG241	*An. arthuri* A	Goianá (MG)	-21.538836	-43.200856
MG324	*An. arthuri* A	Oliveira (MG)	-20.746389	-44.915278
MG33112	*An. arthuri* A	Oliveira (MG)	-20.745598	-44.915613

Code	Species	Locality (state)	Latitude	Longitude
MG3312 6	*An. arthuri* A	Oliveira (MG)	-20.745598	-44.915613
MG3313 7	*An. arthuri* D	Oliveira (MG)	-20.508785	-44.770600
MG342	*An. arthuri* A	Oliveira (MG)	-20.712500	-44.974444
MG349	*An. arthuri* A	Oliveira (MG)	-20.712500	-44.974444
MG3511	*An. arthuri* D	São Franscisco de Paula (MG)	-20.754444	-44.917222
MG4414 2	*An. arthuri* A	Oliveira (MG)	-20.768428	-44.878209
RO2918	*An. arthuri* C	Campo Novo de Rondonia (RO)	-10.637639	-65.499833
RO31103	*An. arthuri* C	Campo Novo de Rondonia (RO)	-10.637639	-65.499833
RO81	*An. arthuri* C	Monte Negro (RO)	-10.268639	-63.555389
RO8104	*An. arthuri* C	Monte Negro (RO)	-10.268639	-63.555389
RO8109	*An. arthuri* C	Monte Negro (RO)	-10.268639	-63.555389
SP31120	*An. arthuri* A	Inubia Paulista (SP)	-21.681417	-50.919889
MG1511	*An.*CP Form	Coronel Pacheco (MG)	-21.635819	-43.319267
MG15612	*An.*CP Form	Coronel Pacheco (MG)	-21.635819	-43.319267
MG1596	*An.*CP Form	Coronel Pacheco (MG)	-21.635819	-43.319267
PR21110	*An.*CP Form	Foz do Iguaçu (PA)	-54.546528	-25.454583
BA233	*An. strodei*	São José da Vitória (Bahia)	-15.087060	-39.341560
BA254	*An. strodei*	São José da Vitória (Bahia)	-15.090910	-39.343700

Code	Species	Locality (state)	Latitude	Longitude
ES091	*An. strodei*	Santa Teresa (ES)	-19.916667	-40.600000
ES093	*An. strodei*	Santa Teresa (ES)	-19.916667	-40.600000
MG27108	*An. strodei*	Coronel Pacheco (MG)	-21.587778	-43.265833
MG30102	*An. strodei*	Coronel Pacheco (MG)	-21.587778	-43.265834
MG3391	*An. strodei*	Oliveira (MG)	-20.745598	-44.915613
PR2043	*An. strodei*	São Miguel do Iguaçu (PA)	-25.265361	-54.309583
PR2923 3	*An. strodei*	Foz do Iguaçu (PA)	-25.480556	-54.586667
RS3798	*An. strodei*	Maquiné (RS)	-29.589556	-50.262639
SP076	*An. strodei*	Buri (SP)	-23.800000	-48.566670
SP104181	*An. strodei*	Pindamonhangaba (SP)	-22.960472	-45.452083
SP10510 12	*An. strodei*	Pindamonhangaba (SP)	-22.999333	-45.495361
SP271	*An. strodei*	Lucélia (SP)	-21.618861	-50.940000
SP29121	*An. strodei*	Lucélia (SP)	-21.618861	-50.940000
SP31101	*An. strodei*	Inubia Paulista (SP)	-21.681417	-50.919889
SP5633	*An. strodei*	Mairiporã (SP)	-23.318889	-46.586944
SP568	*An. strodei*	Mairiporã (SP)	-23.318889	-46.586944
SP6615 1	*An. strodei*	Dourado (SP)	-22.134694	-48.391722
VP0511A	*An. strodei*	Pindamonhangaba (SP)	-22.959750	-45.452389
VP0652	*An. strodei*	Pindamonhangaba (SP)	-22.959750	-45.452389
VP0664	*An. strodei*	Pindamonhangaba (SP)	-22.959750	-45.452389

Source: Bourke et al. 2013

5.2 DNA SAMPLES

The samples have been collected throughout Brazil and can be seen in Table 1. The DNA has been previously extracted following the Qiagen protocol. DNA master solutions were diluted to 200 µl and were stored at -80°C in the Laboratório de Sistemática Molecular – Culicidae of the Department of Epidemiology at the School of Public Heath of the University of São Paulo. The stock genomic DNA is kept in a separate -80°C storage from DNA samples that were in use. The samples are from offspring of wild-caught females or wild-caught pupae, which were raised to maturity in the laboratory.

5.3 PCR PROTOCOL

Catalase – The catalase gene was amplified using the CAT-F Primer (5´- GAY GGY TWY CGI TTC ATG AAC G -3') and the CAT-R Primer (5´- GRC GKC CRA ART CRG CAT CAA G-3') (Table 2). The PCR will be carried out in a mixture of 25 uL – 2.50 µl of 1X buffer (Invitrogen), 0.75 µl of MgCl2, 0.5 µl of dNTP, 2.0 µl of both the primer CAT-F and the primer CAT-R, 2.5 µl of dimethyl sulfoxide, 13.625 µl of water, 0.25 µl of hot start Platinum® Taq enzyme and 1 µl DNA extract. The PCR was performed using the following temperatures: one cycle of 94°C for 2 mins, 40 cycles of: 94°C for 30 seconds for denaturation, 55.1°C for 30 seconds for annealing and 72°C for one minute for the final extension, and then one cycle of 72°C for 5 minutes.

CAD -- The *CAD* gene was amplified using the CAD-F Primer (5´- CCM RSC GST GCT ACA TGA C- 3') and the CAD-R Primer (5´- GAT GAT GAG CTG RGY CGA GTG -3'). The PCR was carried out in a mixture of 25 µl – 2.50 µl of 1X buffer (Invitrogen), 0.75 µl of MgCl2, 0.5 µl of dNTP, 2.0 µl of both the primer CAD-F and the primer CAD-R, 2.5

μl of dimethyl sulfoxide (DMSO), 13.625 μl of water, 0.25 μl of hot start Platinum® Taq enzyme and 1 μl DNA extract. The PCR was performed using the following temperatures: one cycle of 94°C for 2 mins, 35 cycles of: 94°C for 30 seconds for denaturation, 57.5°C for 30 seconds for annealing and 72°C for one minute for the final extension, and then one cycle of 72°C for 5 minutes.

White and CO1- The *white* and *CO1* genes were both previously sequenced in the LASEC at the University of Sao Paulo at the School of Public Health (Bourke et al., 2013) and were accessed from the NCBI GenBank site using the accession codes found in table 5.

Table 2. Primers used for amplification for Polymerase Chain Reaction (PCR) and sequencing reactions.

Gene	**Primer F**	**Primer R**
CAT	5´- GAY GGY TWY CGI TTC ATG AAC G -3’	5´- GRC GKC CRA ART CRG CAT CAA G-3’
CAD	5´- CCM RSC GST GCT ACA TGA C- 3’	5´- GAT GAT GAG CTG RGY CGA GTG-3’

5.4 PURIFICATION

The PCR product of *CAT* and *CAD* were purified using the polyethylene glycol (PEG) (20% PEG, 80% 2.5 M NaCl) precipitation protocol. The purified products were then electrophoresed on 1% agarose gel with GelRed™ (Biotium) to confirm that only the desired gene was amplified.

5.5 SEQUENCES AND SEQUENCING

CO1- The *CO1* gene was previously sequenced by Bourke et al. (2013) using the same DNA and sequences were accessed via the NCBI website.

White – The *white* gene was previously sequenced by Bourke et al. using the same DNA samples and were accessed via the NCBI website.

CAT -- The sequencing reaction was carried out in both directions using the Big Dye® Terminator V3.1 Cycle sequencing kit. A 10 µl mixture was used including: 0.5 µl of Big Dye®, 2.0 µl of 1X buffer (Invitrogen), 1.0 µl of DMSO, 2.0 µl of Primer CAT-F or 2.0µl of Primer CAT-R, 4.5 µl of ultra pure water and 1.0 µl of purified PCR product. The product was then dried and re-suspended using 10 µl of water. The DNA was sequenced following the Sanger method.

CAD - The sequencing reaction was carried out in both directions using the Big Dye® Terminator V3.1 Cycle sequencing kit. A 10 µl mixture was used including: 0.5 µl of Big Dye®, 2.0 µl of 1X buffer (Invitrogen), 1.0 µl of DMSO, 2.0 µl of Primer CAD-F or 2.0µl of Primer CAD-R, 4.5 µl of water and 1.0 µl of purified PCR product. The product was then dried and re-suspended using 10 µl of water.

The PCR products generated for all genes were sequenced following the Sanger method (Sanger et al., 1977).

5.6 PHYLOGENETIC ANALYSIS

The sequences were viewed and edited in Sequencher® (Sequencher version 5.4.1 sequence analysis software, Gene Codes Corporation) and the multiple sequence alignments (MSA) were made with Mesquite software using the ClustalW algorithm (Maddison & Maddison, 2015).

Bayesian analysis was performed on all the sequence data to create the evolutionary models using MrBayes version 3.1.2 (Ronquist and Huelsenbeck, 2003). Evolutionary models to be used in the Bayesian analysis were attained using Akaike Information Criterion (AIC) in jModelTest 2 (Posada, 2003).

Bayesian analysis was performed on both *CAD* and *CAT* genes individually and on the concatenated gene matrix of the four genes (*CAD, CAT, white* and *CO1)* to examine evolutionary relationships in the Strodei Subgroup.

All Bayesian analyses were carried out in two simultaneous runs, employing four Markov Chains Monte Carlo each (being three heated and one cold chain). The analyses were run in 10 million generations long, saving one sample every 1000 generation. The first 25% of the samples of the cold chain were discarded as burnin.

For the *CAD* analysis *An. konderi, An. oswaldoi* and *An. benarrochi* were used as outgroup (accession numbers: KC167248, KC167271 and KC167208, respectively). The outgroup used in the *CAT* analysis was *An. konderi, An. oswaldoi* and *An. benarrochi.*

5.7 ECOLOGICAL NICHE MODELING

The environmental variables used can be seen in Table 3. These variables were downloaded in 30 arc second, ESRI format from worldclim.org (Hijmans et al. 2005) and were input into qGIS (qGIS, 2015) to format equal cell size and geographical area. The rasters were cut to the following extent xmin -80.57499, xmax -27.02499, ymin -56.65000, ymax 11.69167. They were then converted into ASCII format to be used in Maxent.

Table 3. Information about the variables used for the niche modeling using the MAXENT software.

Variable name	Variable meaning
Bio1	Annual mean temperature
Bio2	Mean Diurnal Range (Mean of monthly (max temp – min temp)
Bio3	Isothermality (Bio2/Bio7)*(100)
Bio4	Temperature Seasonality (standard deviation*100)
Bio5	Max Temperature of Warmest Month
Bio6	Min Temperature of Coldest Month
Bio7	Temperature Annual Range (BIO5-BIO6)
Bio8	Mean Temperature of Wettest Quarter
Bio9	Mean Temperature of Driest Quarter
Bio10	Mean Temperature of Warmest Quarter
Bio11	Mean Temperature of Coldest Quarter
Bio12	Annual Precipitation
Bio13	Precipitation of Wettest Month
Bio14	Precipitation of Driest Month
Bio15	Precipitation Seasonality (Coefficient of Variation)
Bio16	Precipitation of Wettest Quarter
Bio17	Precipitation of Driest Quarter
Bio18	Precipitation of Warmest Quarter
Bio19	Precipitation of Coldest Quarter

Source: http://www.worldclim.org/current

The localities of the mosquitoes used in this analysis are in Table 4.

The mosquitoes were collected previously throughout other projects done at the School of Public Health of the University of Sao Paulo and were identified based on morphological characters by Dr. Maria Anice Mureb Sallum. Not all of the specimens collected at these collection points were used in the phylogenetic analysis. All mosquitoes that were previously identified as *An. strodei*, *An. arthuri* A-D or *An.* CP Form, however, were included in the analysis. The model was run individually for each species with all 19

variables and was re-run eliminating variables that were not significant in strengthening the model. The general rule followed for the analysis was that for one layer could be used for each 3-collection points that mean that for *Anopheles* CP Form, only one layer was used in the analysis. The species, longitude and latitude of each collection point were input into Microsoft excel and were saved in a CSV file format to be input into Maxent.

Table 4. Species information and geographical coordinates of the localities used in the niche modeling analyses in Maxent software

Species	**Longitude**	**Latitude**	**Datum**
Anopheles strodei	-40.6000005	-19.9166667	WGS84
Anopheles strodei	-40.6000005	-19.9166668	WGS84
Anopheles strodei	-40.6000005	-19.9166669	WGS84
Anopheles strodei	-40.6000005	-19.916667	WGS84
Anopheles strodei	-40.6000005	-19.9166671	WGS84
Anopheles strodei	-40.6000005	-19.9166672	WGS84
Anopheles strodei	-40.6000006	-19.9166673	WGS84
Anopheles strodei	-48.5833333	-22.95975	WGS84
Anopheles strodei	-48.5833333	-22.95975	WGS84
Anopheles strodei	-45.4523889	-22.95975	WGS84
Anopheles strodei	-45.4523889	-22.95975	WGS84
Anopheles strodei	-45.4523889	-22.95975	WGS84
Anopheles strodei	-45.4523889	-22.95975	WGS84
Anopheles strodei	-45.4523889	-22.95975	WGS84
Anopheles strodei	-45.4523889	-22.95975	WGS84

Species	Longitude	Latitude	Datum
Anopheles strodei	-54.3095833	-25.2653611	WGS84
Anopheles strodei	-54.3095833	-25.2653611	WGS84
Anopheles strodei	-54.3095833	-25.2653611	WGS84
Anopheles strodei	-54.3095833	-25.2653611	WGS84
Anopheles strodei	-54.3530278	-24.8629722	WGS84
Anopheles strodei	-50.9436944	-21.6188611	WGS84
Anopheles strodei	-50.9436944	-21.6188611	WGS84
Anopheles strodei	-50.9436944	-21.6188611	WGS84
Anopheles strodei	-50.9436944	-21.6188611	WGS84
Anopheles strodei	-50.9436944	-21.6188611	WGS84
Anopheles strodei	-50.9436944	-21.6188611	WGS84
Anopheles strodei	-50.9436944	-21.6188611	WGS84
Anopheles strodei	-54.5866667	-25.4805556	WGS84
Anopheles strodei	-50.9436944	-21.6188611	WGS84
Anopheles strodei	-50.9436944	-21.6215278	WGS84
Anopheles strodei	-50.9198889	-21.6814167	WGS84
Anopheles strodei	-43.3192667	-21.6358194	WGS84
Anopheles strodei	-48.391722	-22.1346944	WGS84
Anopheles strodei	-48.391722	-22.1346944	WGS84
Anopheles strodei	-47.046312	-22.873985	WGS84
Anopheles strodei	-45.452083	-22.9604722	WGS84
Anopheles strodei	-45.452083	-22.9604722	WGS84

Species	Longitude	Latitude	Datum
Anopheles strodei	-45.452083	-22.9604722	WGS84
Anopheles strodei	-45.452083	-22.9604722	WGS84
Anopheles strodei	-45.452083	-22.9604722	WGS84
Anopheles strodei	-45.452083	-22.9604722	WGS84
Anopheles strodei	-43.1826	-21.529742	WGS84
Anopheles strodei	-50.2626389	-29.5895556	WGS84
Anopheles strodei	-50.2626389	-29.5895556	WGS84
Anopheles strodei	-46.5869444	-23.3188889	WGS84
Anopheles strodei	-46.5869444	-23.3188889	WGS84
Anopheles strodei	-46.5869444	-23.3188889	WGS84
Anopheles strodei	-46.5869444	-23.3188889	WGS84
Anopheles strodei	-46.5869444	-23.3188889	WGS84
Anopheles strodei	-46.5869444	-23.3188889	WGS84
Anopheles strodei	-46.5869444	-23.3188889	WGS84
Anopheles strodei	-46.5869444	-23.3188889	WGS84
Anopheles strodei	-46.5869444	-23.3188889	WGS84
Anopheles strodei	-50.2175	-29.6598333	WGS84
Anopheles strodei	-50.2175	-29.6598333	WGS84
Anopheles strodei	-50.2175	-29.6598333	WGS84
Anopheles strodei	-50.2175	-29.6598333	WGS84
Anopheles strodei	-50.2175	-29.6598333	WGS84
Anopheles strodei	-50.2175	-29.6598333	WGS84

Species	Longitude	Latitude	Datum
Anopheles strodei	-50.2175	-29.6598333	WGS84
Anopheles strodei	-44.8836111	-20.4647222	WGS84
Anopheles strodei	-44.8836111	-20.4647222	WGS84
Anopheles strodei	-44.7188889	-20.4430556	WGS84
Anopheles strodei	-44.7188889	-20.4430556	WGS84
Anopheles strodei	-44.7188889	-20.4430556	WGS84
Anopheles strodei	-45.2483	-20.88	WGS84
Anopheles strodei	-45.2483	-20.88	WGS84
Anopheles arthuri B	-41.00125	-3.88675	WGS84
Anopheles arthuri B	-41.00125	-3.88675	WGS84
Anopheles arthuri B	-41.00125	-3.88675	WGS84
Anopheles arthuri B	-40.8977778	-3.8442222	WGS84
Anopheles arthuri B	-40.8977778	-3.8442222	WGS84
Anopheles arthuri B	-40.8977778	-3.8442222	WGS84
Anopheles arthuri B	-40.8977778	-3.8442222	WGS84
Anopheles arthuri B	-40.8977778	-3.8442222	WGS84
Anopheles arthuri B	-40.8977778	-3.8442222	WGS84
Anopheles arthuri B	-40.8977778	-3.8442222	WGS84
Anopheles arthuri B	-40.8977778	-3.8442222	WGS84
Anopheles arthuri B	-40.8977778	-3.8442222	WGS84
Anopheles arthuri B	-40.8977778	-3.8442222	WGS84
Anopheles arthuri B	-40.8977778	-3.8442222	WGS84

Species	**Longitude**	**Latitude**	**Datum**
Anopheles arthuri B	-40.8977778	-3.8442222	WGS84
Anopheles arthuri B	-40.8977778	-3.8442222	WGS84
Anopheles arthuri B	-40.8977778	-3.8442222	WGS84
Anopheles arthuri B	-40.8963611	-4.0214167	WGS84
Anopheles arthuri B	-40.8963611	-4.0214167	WGS84
Anopheles arthuri B	-40.8963611	-4.0214167	WGS84
Anopheles arthuri B	-40.8963611	-4.0214167	WGS84
Anopheles arthuri B	-40.8963611	-4.0214167	WGS84
Anopheles arthuri B	-40.8963611	-4.0214167	WGS84
Anopheles arthuri B	-40.8963611	-4.0214167	WGS84
Anopheles arthuri B	-40.8963611	-4.0214167	WGS84
Anopheles arthuri B	-40.8963611	-4.0214167	WGS84
Anopheles arthuri B	-40.8963611	-4.0214167	WGS84
Anopheles arthuri B	-40.8963611	-4.0214167	WGS84
Anopheles arthuri B	-40.8963611	-4.0214167	WGS84
Anopheles arthuri B	-40.8963611	-4.0214167	WGS84
Anopheles arthuri A	-49.0960278	-20.0252778	WGS84
Anopheles arthuri A	-49.0960278	-20.0252778	WGS84
Anopheles arthuri A	-49.0960278	-20.0252778	WGS84
Anopheles arthuri A	-49.0960278	-20.0252778	WGS84
Anopheles arthuri A	-49.0960278	-20.0252778	WGS84
Anopheles arthuri A	-49.0960278	-20.0252778	WGS84

Species	Longitude	Latitude	Datum
Anopheles arthuri A	-49.0960278	-20.0252778	WGS84
Anopheles arthuri A	-49.0960278	-20.0252778	WGS84
Anopheles arthuri A	-49.0960278	-20.0252778	WGS84
Anopheles arthuri A	-49.0960278	-20.0252778	WGS84
Anopheles arthuri A	-49.0960278	-20.0252778	WGS84
Anopheles arthuri A	-49.0960278	-20.0252778	WGS84
Anopheles arthuri A	-49.0960278	-20.0252778	WGS84
Anopheles arthuri A	-49.0960278	-20.0252778	WGS84
Anopheles arthuri A	-49.0960278	-19.9812778	WGS84
Anopheles arthuri A	-49.0960278	-19.9812778	WGS84
Anopheles arthuri A	-49.0960278	-19.9812778	WGS84
Anopheles arthuri A	-49.0960278	-19.9812778	WGS84
Anopheles arthuri A	-49.0960278	-19.9812778	WGS84
Anopheles arthuri A	-49.0765	-20.0252778	WGS84
Anopheles arthuri A	-43.3192667	-21.6358194	WGS84
Anopheles arthuri A	-43.3192667	-21.6358194	WGS84
Anopheles arthuri A	-51.355101	-18.765823	WGS84
Anopheles arthuri A	-51.355101	-18.765823	WGS84
Anopheles arthuri A	-51.355101	-18.765823	WGS84
Anopheles arthuri A	-51.355101	-18.765823	WGS84
Anopheles arthuri A	-51.355101	-18.765823	WGS84
Anopheles arthuri A	-51.355101	-18.765823	WGS84

Species	Longitude	Latitude	Datum
Anopheles arthuri A	-51.355101	-18.765823	WGS84
Anopheles arthuri A	-51.355101	-18.765823	WGS84
Anopheles arthuri A	-48.192263	-18.65366	WGS84
Anopheles arthuri A	-48.192263	-18.65366	WGS84
Anopheles arthuri A	-48.192263	-18.65366	WGS84
Anopheles arthuri A	-47.046312	-22.873985	WGS84
Anopheles arthuri A	-49.6883056	-10.5861667	WGS84
Anopheles arthuri A	-43.1826	-21.529742	WGS84
Anopheles arthuri A	-43.1826	-21.529742	WGS84
Anopheles arthuri A	-43.2008556	-21.538836	WGS84
Anopheles arthuri A	-43.2008556	-21.538836	WGS84
Anopheles arthuri A	-43.2008556	-21.538836	WGS84
Anopheles arthuri A	-43.2008556	-21.538836	WGS84
Anopheles arthuri A	-43.2008556	-21.538836	WGS84
Anopheles arthuri A	-43.2008556	-21.538836	WGS84
Anopheles arthuri A	-43.2008556	-21.538836	WGS84
Anopheles arthuri A	-44.9152778	-20.7463889	WGS84
Anopheles arthuri A	-44.9152778	-20.7463889	WGS84
Anopheles arthuri A	-44.9152778	-20.7463889	WGS84
Anopheles arthuri A	-44.916666	-20.7544444	WGS84
Anopheles arthuri A	-44.916666	-20.7544444	WGS84
Anopheles arthuri A	-44.916666	-20.7544444	WGS84

Species	Longitude	Latitude	Datum
Anopheles arthuri A	-44.916666	-20.7544444	WGS84
Anopheles arthuri A	-44.916666	-20.7544444	WGS84
Anopheles arthuri A	-44.916666	-20.7544444	WGS84
Anopheles arthuri A	-44.916666	-20.7544444	WGS84
Anopheles arthuri A	-44.916666	-20.7544444	WGS84
Anopheles arthuri A	-44.916666	-20.7544444	WGS84
Anopheles arthuri A	-44.8788889	-20.7675	WGS84
Anopheles arthuri A	-44.8788889	-20.7675	WGS84
Anopheles arthuri A	-44.8788889	-20.7675	WGS84
Anopheles arthuri A	-44.8788889	-20.7675	WGS84
Anopheles arthuri A	-44.8788889	-20.7675	WGS84
Anopheles arthuri A	-44.8788889	-20.7675	WGS84
Anopheles arthuri A	-44.8788889	-20.7675	WGS84
Anopheles arthuri A	-44.8788889	-20.7675	WGS84
Anopheles arthuri A	-44.8533333	-20.7547222	WGS84
Anopheles arthuri A	-44.8533333	-20.7547222	WGS84
Anopheles arthuri A	-44.8430556	-20.7388889	WGS84
Anopheles arthuri A	-44.8430556	-20.7388889	WGS84
Anopheles arthuri A	-44.8430556	-20.7388889	WGS84
Anopheles arthuri A	-44.8430556	-20.7388889	WGS84
Anopheles arthuri A	-44.8430556	-20.7388889	WGS84
Anopheles arthuri A	-44.8430556	-20.7388889	WGS84

Species	Longitude	Latitude	Datum
Anopheles arthuri A	-44.8430556	-20.7388889	WGS84
Anopheles arthuri A	-44.8430556	-20.7388889	WGS84
Anopheles arthuri A	-44.8430556	-20.7388889	WGS84
Anopheles arthuri A	-44.8430556	-20.7388889	WGS84
Anopheles arthuri A	-44.8430556	-20.7388889	WGS84
Anopheles arthuri A	-44.8430556	-20.7388889	WGS84
Anopheles arthuri A	-44.8430556	-20.7388889	WGS84
Anopheles arthuri A	-44.8833333	-20.4644444	WGS84
Anopheles arthuri A	-44.8833333	-20.4644444	WGS84
Anopheles arthuri A	-44.8833333	-20.4644444	WGS84
Anopheles arthuri A	-44.8833333	-20.4644444	WGS84
Anopheles arthuri A	-44.8833333	-20.4644444	WGS84
Anopheles arthuri A	-44.8833333	-20.4644444	WGS84
Anopheles arthuri A	-44.8836111	-20.4647222	WGS84
Anopheles arthuri A	-44.8836111	-20.4647222	WGS84
Anopheles arthuri A	-44.8836111	-20.4647222	WGS84
Anopheles arthuri A	-44.8836111	-20.4647222	WGS84
Anopheles arthuri A	-44.8836111	-20.4647222	WGS84
Anopheles arthuri A	-44.8836111	-20.4647222	WGS84
Anopheles arthuri A	-44.8836111	-20.4647222	WGS84
Anopheles arthuri A	-44.8836111	-20.4647222	WGS84
Anopheles arthuri A	-44.8836111	-20.4647222	WGS84

Species	**Longitude**	**Latitude**	**Datum**
Anopheles arthuri A	-45.2483	-20.88	WGS84
Anopheles arthuri A	-45.2483	-20.88	WGS84
Anopheles arthuri C	-63.5553889	-10.2686389	WGS84
Anopheles arthuri C	-63.5553889	-10.2686389	WGS84
Anopheles arthuri C	-63.5553889	-10.2686389	WGS84
Anopheles arthuri C	-63.5763611	-10.2909167	WGS84
Anopheles arthuri C	-63.5763611	-10.2909167	WGS84
Anopheles arthuri C	-63.5763611	-10.2909167	WGS84
Anopheles arthuri C	-63.5763611	-10.2909167	WGS84
Anopheles arthuri C	-65.4998333	-10.6376389	WGS84
Anopheles arthuri C	-65.4998333	-10.6376389	WGS84
Anopheles arthuri C	-65.4998333	-10.6376389	WGS84
Anopheles arthuri C	-65.4998333	-10.6376389	WGS84
Anopheles arthuri C	-65.4998333	-10.6376389	WGS84
Anopheles arthuri C	-65.4998333	-10.6376389	WGS84
Anopheles arthuri C	-65.4998333	-10.6376389	WGS84
Anopheles arthuri C	-65.4998333	-10.6376389	WGS84
Anopheles arthuri C	-65.4998333	-10.6376389	WGS84
Anopheles arthuri C	-65.4998333	-10.6376389	WGS84
Anopheles arthuri C	-65.4998333	-10.6376389	WGS84
Anopheles arthuri C	-65.4998333	-10.6376389	WGS84
Anopheles arthuri C	-65.4998333	-10.6376389	WGS84

Species	Longitude	Latitude	Datum
Anopheles arthuri D	-44.9152778	-20.7463889	WGS84
Anopheles arthuri D	-44.9166667	-20.7544444	WGS84
Anopheles arthuri D	-44.8791667	-20.7669444	WGS84
Anopheles arthuri D	-44.8430556	-20.7388889	WGS84
Anopheles arthuri D	-44.8377778	-20.7266667	WGS84
Anopheles arthuri D	-44.8836111	-20.4647222	WGS84
Anopheles arthuri D	-44.7188889	-20.4430556	WGS84
Anopheles arthuri D	-44.7188889	-20.4430556	WGS84
Anopheles CP Form	-39.8843889	-19.0834472	WGS84
Anopheles CP Form	-44.9152778	-20.7463889	WGS84
Anopheles CP Form	-44.8430556	-20.738888	WGS84

6. RESULTS

6.1 PHYLOGENETIC ANALYSIS

A total of 61 samples were used in this study. Sequences of *CAD* and *CAT* were obtained for the samples as seen in Table 5. After alignment, there were 55 sequences of the *white* gene each 606 bp long, 55 unique *COI* sequences of 658 bp long, 45 *CAD* sequences at 715 bp long and 55 *CAT* sequences at 720 bp long.

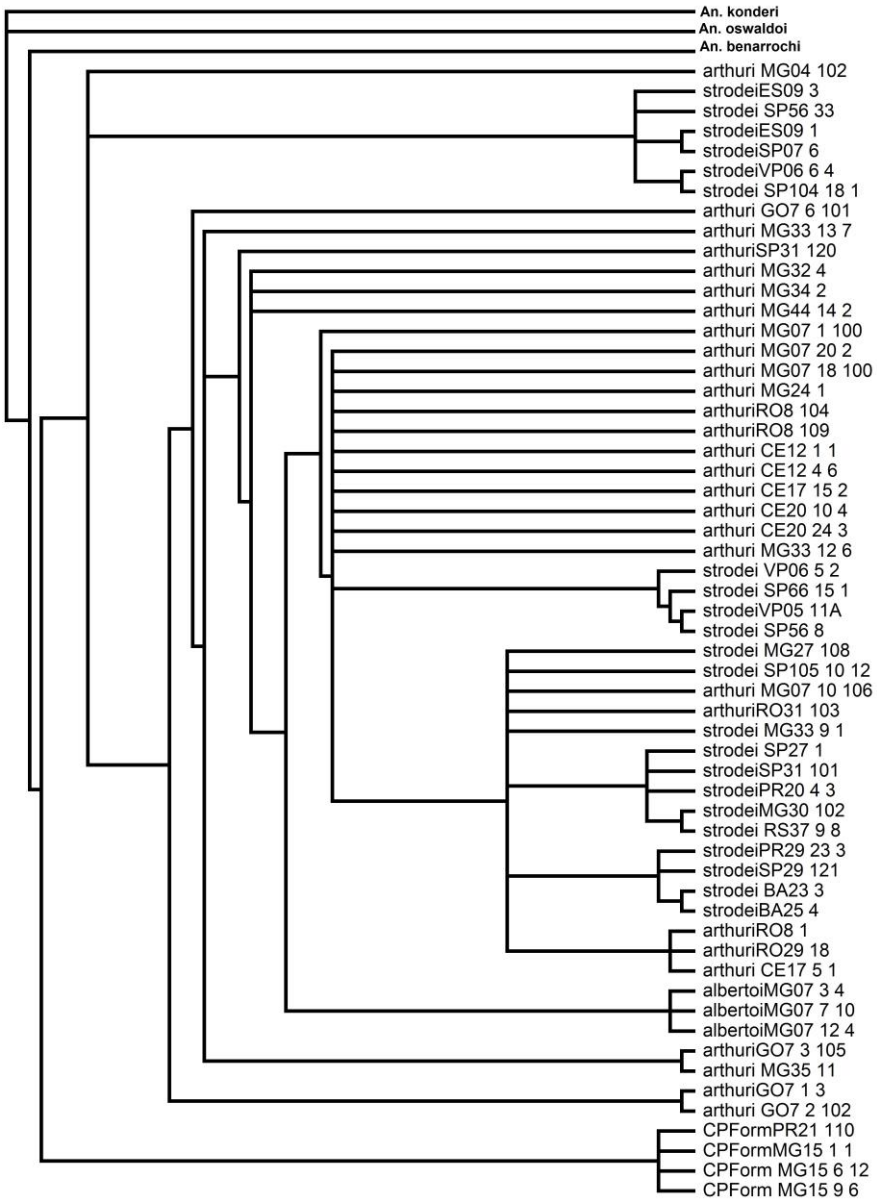
Table 5. Sample codes and information about the newly sequences generated for the *CAT* and *CAD* protein coding single copy nuclear genes, and GenBank accession numbers for the *COI* mitochondrial gene and *white* protein coding single copy nuclear gene

Sample codes	*CAT*	*CAD*	*COI*	*white*
BA233	Yes	No	KC330234	KC330308
BA254	Yes	Yes	KC330235	KC330309
CE1211	Yes	Yes	KC330250	KC330325
CE1246	Yes	Yes	KC330251	KC330326
CE17152	Yes	Yes	KC330253	KC330328
CE1751	Yes	Yes	KC330252	KC330327
CE20104	Yes	Yes	KC330254	KC330329
CE20243	Yes	Yes	KC330255	KC330330
ES091	Yes	Yes	GU226664	GU226730
ES093	Yes	Yes	GU226665	GU226731
G0713	Yes	Yes	KC330244	KC330319
G072102	Yes	Yes	KC330245	KC330320
G073105	Yes	Yes	KC330246	KC330321
G076101	Yes	No	KC330247	KC330322
MG03102	No	Yes	GU226679	GU226748
MG04102	Yes	Yes	GU226680	GU226751
MG0710106	Yes	Yes	GU226684	GU226756
MG071100	Yes	Yes	GU226683	GU226752
MG07124	Yes	Yes	GU226678	GU226678
MG0718100	Yes	Yes	GU226685	GU226753-GU226755
MG07202	Yes	Yes	GU226686	GU226750
MG0734	Yes	Yes	GU226676	GU226742
MG07710	Yes	Yes	GU226677	GU226743-GU226746
MG0763	No	Yes	GU226681	GU226749
MG1511	Yes	Yes	JN413711	KC330316
MG15612	Yes	No	JN413712	KC330317
MG1596	Yes	No	KC330243	KC330318
MG241	Yes	Yes	GU226682	GU226757
MG27108	Yes	Yes	GU226669	GU226735
MG30102	Yes	Yes	GU226670	GU226736
MG33112	No	Yes	KC330258	KC330340
MG324	Yes	Yes	KC330257	KC330332
MG33126	Yes	Yes	KC330259	KC330334
MG33137	Yes	No	KC330256	KC330331

Sample codes	*CAT*	*CAD*	*COI*	*white*
MG3391	Yes	Yes	KC330242	KC330333
MG342	Yes	Yes	KC330260	KC330335
MG349	No	No	KC330261	KC330336
MG3511	Yes	Yes	KC330262	KC330338
MG44142	Yes	Yes	KC330263	KC330339
PR2043	Yes	Yes	KC330233	KC330307
PR21110	Yes	Yes	GU226691	GU226762
PR29233	Yes	Yes	GU226671	GU226737
R02918	Yes	Yes	KC330248	KC330323
R081	No	Yes	GU226681	GU226759
R08104	Yes	Yes	GU226690	GU226760
R08109	Yes	Yes	GU226689	GU226761
RO31103	Yes	Yes	KC330249	KC330324
RS3798	Yes	No	KC330236	KC330310
SP076	Yes	Yes	GU226674	GU226740
SP10418	Yes	Yes	KC330240	KC330314
SP1051012	Yes	No	KC330241	KC330315
SP271	Yes	No	GU226672	GU226738
SP29121	Yes	Yes	GU226675	GU226741
SP31101	No	Yes	KC330232	KC330306
SP31120	Yes	Yes	GU226687	GU226758
SP5633	Yes	Yes	KC330238	KC330312
SP568	Yes	No	KC330237	KC330311
SP66151	Yes	No	KC330239	KC330313
VP0511A	Yes	Yes	GU226668	GU226734
VP0652	Yes	No	GU226667	GU226733
VP0664	Yes	Yes	GU226666	GU226732

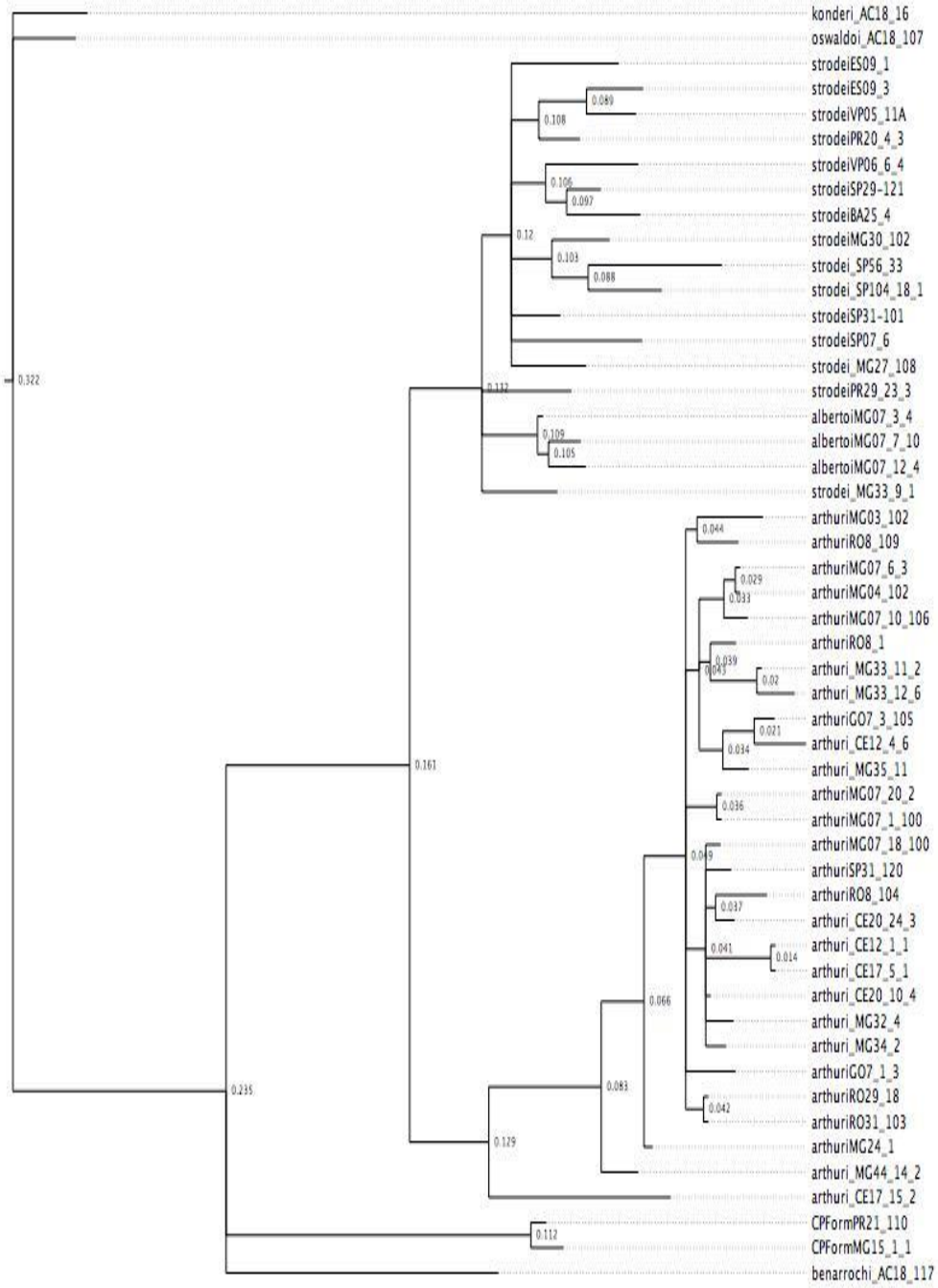
The Bayesian analysis on the *CAT* gene was performed using the TPM1uf + I + G model as selected using the AICc with JModeltest (pinvar = 0.6590). Results of the analysis performed using 55 *CAT* sequences was unable to resolve *An. strodei* from *An. arthuri*. There was no evidence of divergence in the *CAT* gene Bayesian tree except for *Anopheles* CP Form, which was completely resolved (Figure 1).

Figure 1. Bayesian tree generated using the *CAT* sequences from species of the *An. strodei* subgroup. *An. konderi*, *An. oswaldoi* and *An. benarrochi* were used as outgroup.



Source: Created by author using MrBayes

Figure 2. Bayesian tree generated using the *CAD* sequences from species of the *An. strodei* subgroup. The node labels indicate Bayesian Posterior Probabilities. *An. konderi*, *An. oswaldoi* and *An. benarrochi* were used as outgroup.

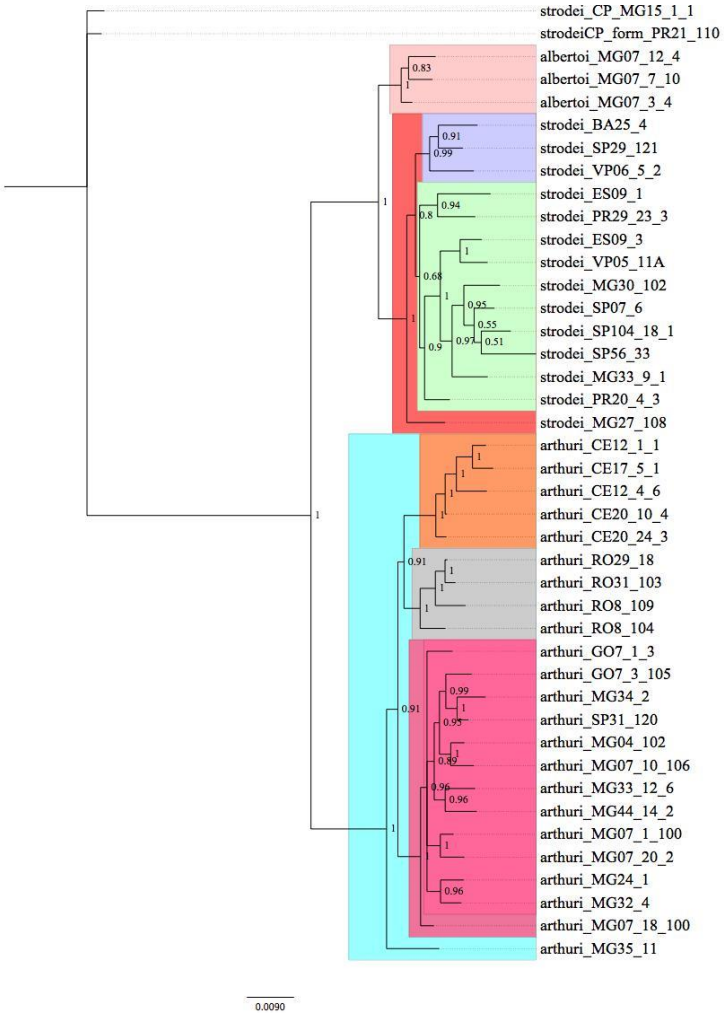


Source: Created by author using MrBayes

The Bayesian analysis using 45 unique *CAD* sequences was performed using the evolutionary model TVM + I + G (Pinvar = 0.5330) chosen by JModeltest 2.0. The analysis

resolved the complex into three not well supported clades (BPP: 0.112, 0.161 and 0.112). *Anopheles* CP Form was resolved into its own clade, *An. strodei* and *An. albertoi* were resolved into another clade and *An. arthuri* was found to be monophyletic (Figure 2). The Bayesian posterior probabilities show a low level of support for these clades (Figure 3).

Figure 3. Bayesian tree generated using concatenated DNA sequences of the nuclear *white*, *mitochondrial CO1* and nuclear *CAD* genes of species of the *An. strodei* subgroup. The node labels indicate Bayesian posterior probability greater than 0.5

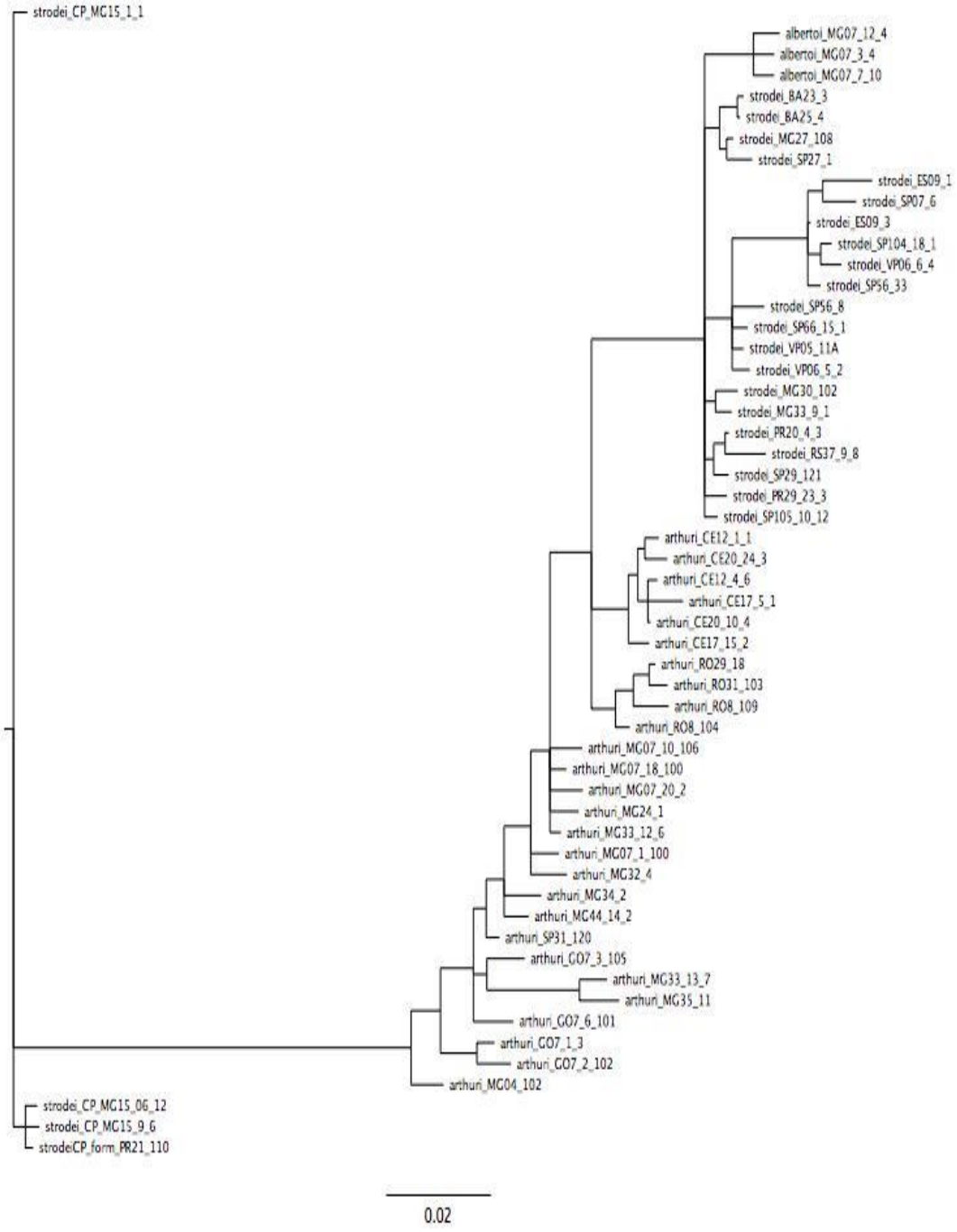


Source: Created by author using MrBayes

The gene tree formed from the Bayesian analysis of the three concatenated genes (*white, CAD, COI)* shows support for 8 clades. *An. arthuri* A-D was supported with high BPPs (0.91, 0.91,1 values) with the exception of the one lone individual of *An. arthuri* D who had a BPP <0.50 which is likely just a problem of having only one individual. The apparent polyphyly of *An. arthuri* makes sense in a geographical context, as well, since they were

collected in distinct areas. *An. albertoi* was resolved from *An. strodei* with a BPP of 1, indicating very strong support and *An. strodei* was resolved into two clades and a sister group. Despite the fact that there is evidence to suggest that *An. strodei* is in fact, polyphyletic, this is not supported in a geographical sense as in the situation with *An. arthuri*. *Anopheles strodei* A consists of three individuals collected from Lucelia, SP, Frutal, MG and Pindamoinangaba, SP. *Anopheles strodei* B consists of 10 individuals and *An. strodei* C, a sister group of *An. strodei* A and B, consists of one individual collected in Oliveira, MG (Figure 3). The Bayesian posterior probabilities shown are those that are higher than 0.5. The Bayesian analysis using the concatenated gene matrix of *CAT, COI* and *white* was uninformative and unable to resolve *An. arthuri* and *An. strodei* (Figure 3). *An.* CP Form was fully resolved.

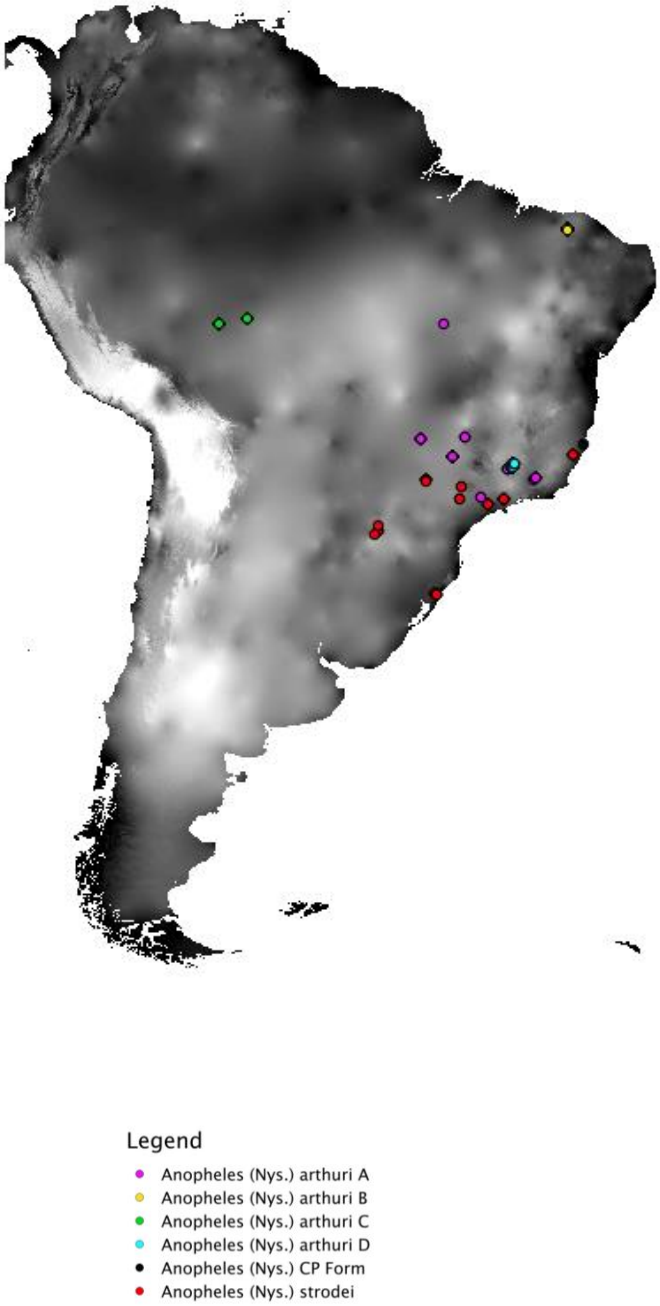
Figure 4. Bayesian topology generated for species of the *Anopheles strodei* group using the concatenated gene matrix using *CAD, CAT* and *white* sequences. Outgroup was excluded from the Bayesian analyses.



Source: Created by author using MrBayes and edited using TreeFig

6.2 NICHE MODELING USING THE MAXENT SOFTWARE

Figure 5. *Anopheles* collection localities used in the niche modeling analyses using Maxent software.



Source: Created by author using Maxent

The model using all 19 variables performed well. The training area under the curve (AUC) value for *An. arthuri* A was 0.987. The AUC for *An. arthuri* B was 0.996. The AUC value

for *An. arthuri* C was 0.905. The AUC value for *An. arthuri* D was 0.988. The AUC value *for An.* CP Form was 0.645 and the AUC value for *An. strodei* was 0.984.

In of the species included in this analysis the variable that provided the strongest contribution to the model for *An. arthuri* A, B and D and *An. strodei* was Bio18, precipitation in the warmest quarter. For *An. arthuri* C, the variable that provided the greatest contribution to the model was Bio4, Temperature Seasonality (standard deviation *100). For *Anopheles* CP Form, the only variable included in the analysis was Bio19, precipitation of the coldest quarter.

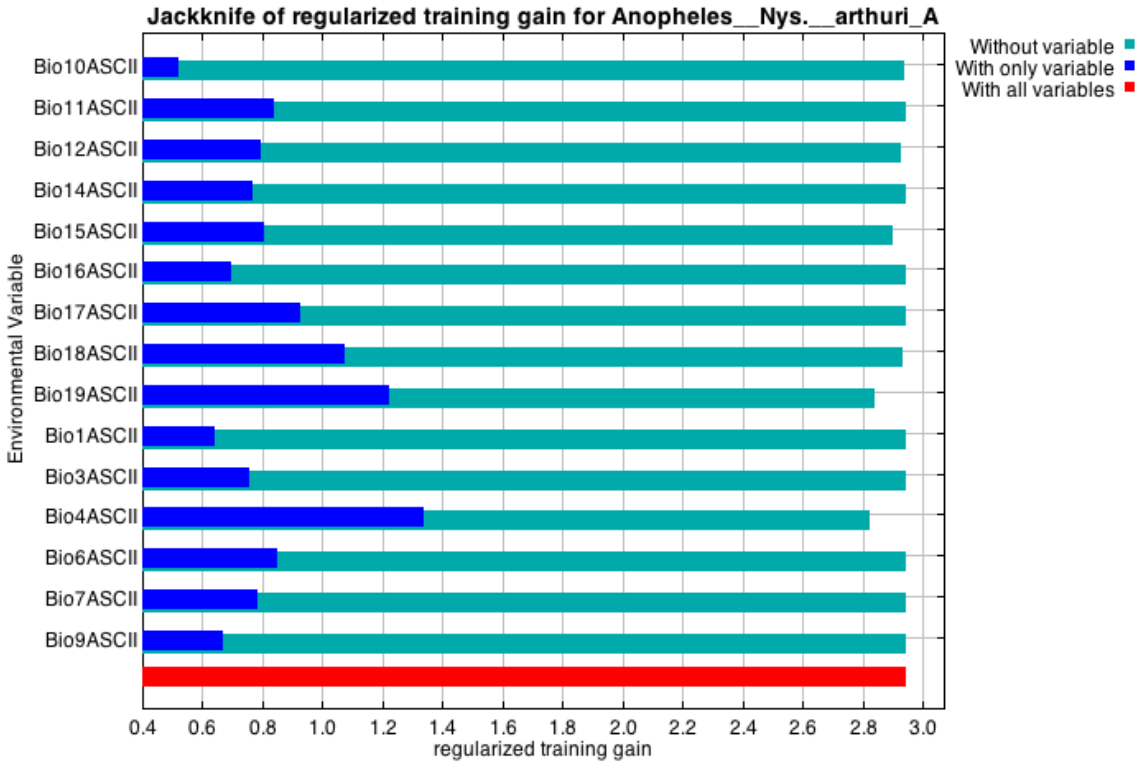
Table 6. Percent of contribution of variables to *Anopheles* niche modeling using the MAXENT software.

Variable	**% contribution to the model**	**Model AUC**
Model for *An. arthuri* A		
Bio18	36.1	0.987
Bio15	27.5	
Bio4	17.5	
Bio16	9.1	
Bio19	7.1	
Bio14	1.3	
Bio17	0.5	
Bio12	0.4	
Bio6	0.3	
Bio10	0.2	
Model for *An. arthuri* B		
Bio18	30.1	0.996
Bio15	28.5	
Bio19	19.9	
Bio4	12.0	
Bio17	5.3	
Bio3	2.8	

Variable	**% contribution to the model**	**Model AUC**
Bio7	1.4	
Model for *An. arthuri* C		
Bio4	60.5	0.905
Bio5	24.1	
Bio11	14	
Bio6	1.4	
Model for *An. arthuri* D		
Bio18	51.3	0.988
Bio19	32.6	
Bio17	16.1	
Model for *An. strodei*		
Bio18	40.8	0.984
Bio4	24.8	
Bio11	19.4	
Bio6	8.0	
Bio8	2.9	
Bio5	2.2	
Bio9	1.9	
Anopheles CP Form		
Bio19	100	0.645

When jackknife testing of variable importance was done, Bio4, temperature seasonality, provided the most training gain when used as the only variable for *An. arthuri* A. The variable that provided the least amount of training gain for *An. arthuri* A when used alone was Bio10, mean temperature of warmest quarter.

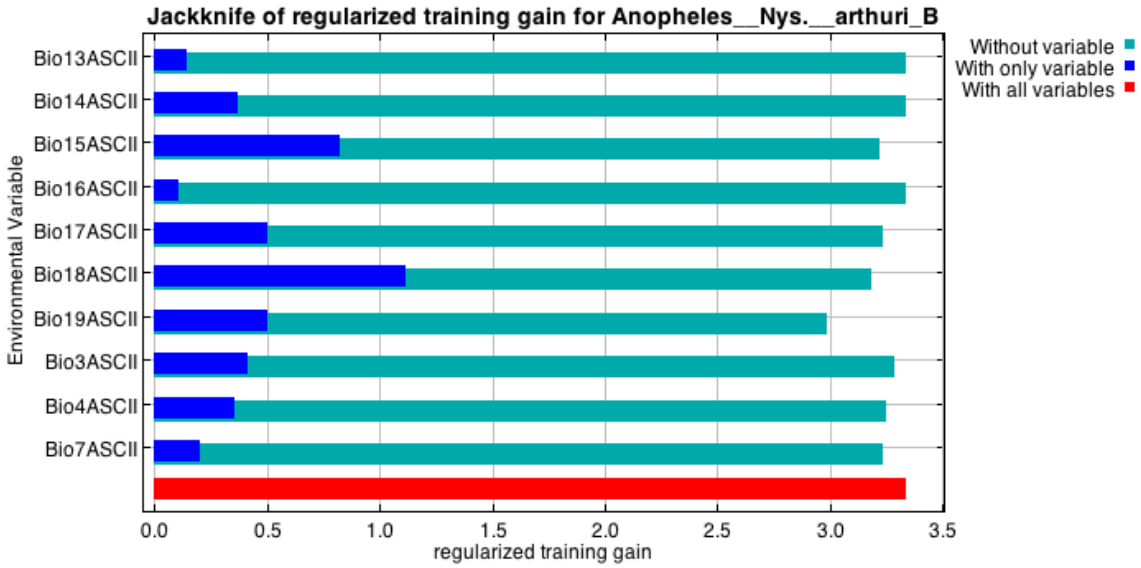
Figure 6. Jackknife test of variable importance for *An. arthuri* A.



Source: Created by author using Maxent

When jackknife testing of variable importance was done for *An. arthuri* B, the variable that provided the most regularized training gain when used alone was Bio18, precipitation of the warmest quarter, while Bio16 and Bio14, precipitation of warmest quarter and precipitation of driest month respectively, provided the least training gain when used alone.

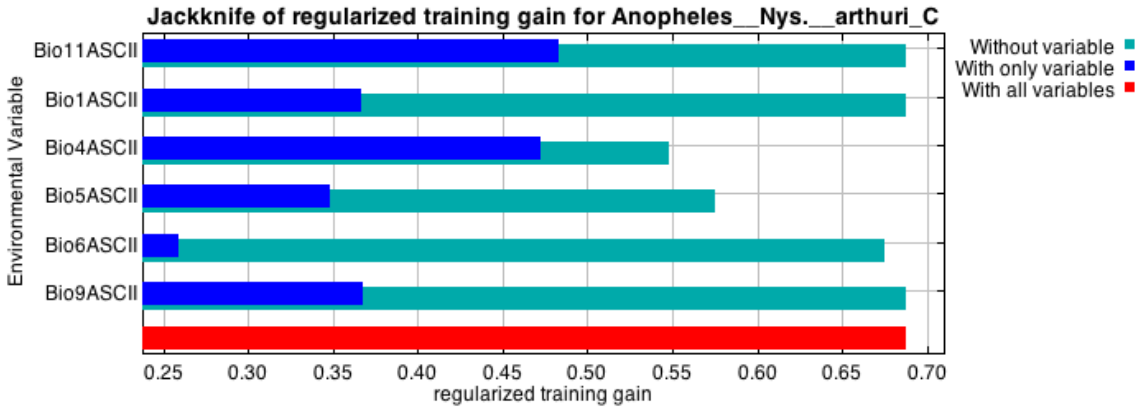
Figure 7. Jackknife test of variable importance for *An. arthuri* B. The light blue bar represents test gain without that particular variable, the dark blue bar represents gain with only that variable and the red bar represents all variables.



Source: Created by author using Maxent

When jackknife testing of variable importance was done for *An. arthuri* C, the variable that provided the most regularized training gain when used alone was Bio11, mean temperature of coldest month, while Bio6, minimum temperature of coldest month provided the least training gain when used alone.

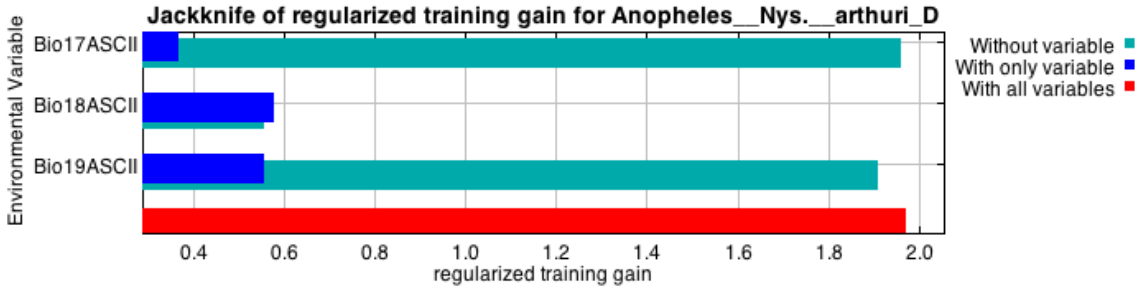
Figure 8. Jackknife test of variable importance for *An. arthuri* C. The light blue bar represents test gain without that particular variable included. The dark blue bar represents gain with only that variable included. The red bar represents all variables included.



Source: Created by author using Maxent

When jackknife testing of variable importance was done for *An. arthuri* D, the variable that provided the most regularized training gain when used alone was Bio18, Precipitation of Warmest Quarter, while Bio17, Precipitation of Driest Quarter provided the least training gain when used alone.

Figure 9. Jackknife testing of variable importance for *An. arthuri* D

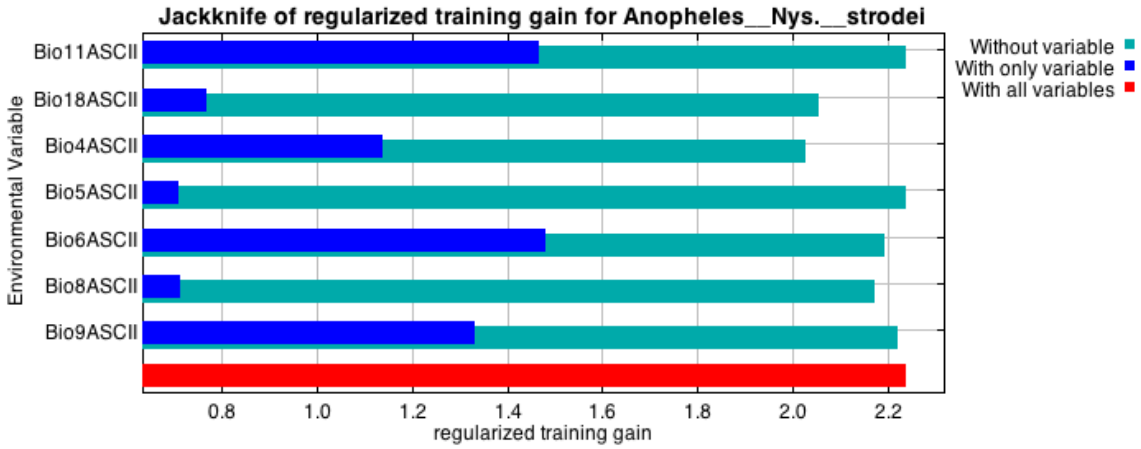


Source: Created by author using Maxent

Jackknife testing of variable importance was performed for *An. strodei* and both Bio 11, mean temperature of coldest quarter, and Bio6, min temperature of coldest month were found to provide the most unique information explaining the species distribution. Bio8,

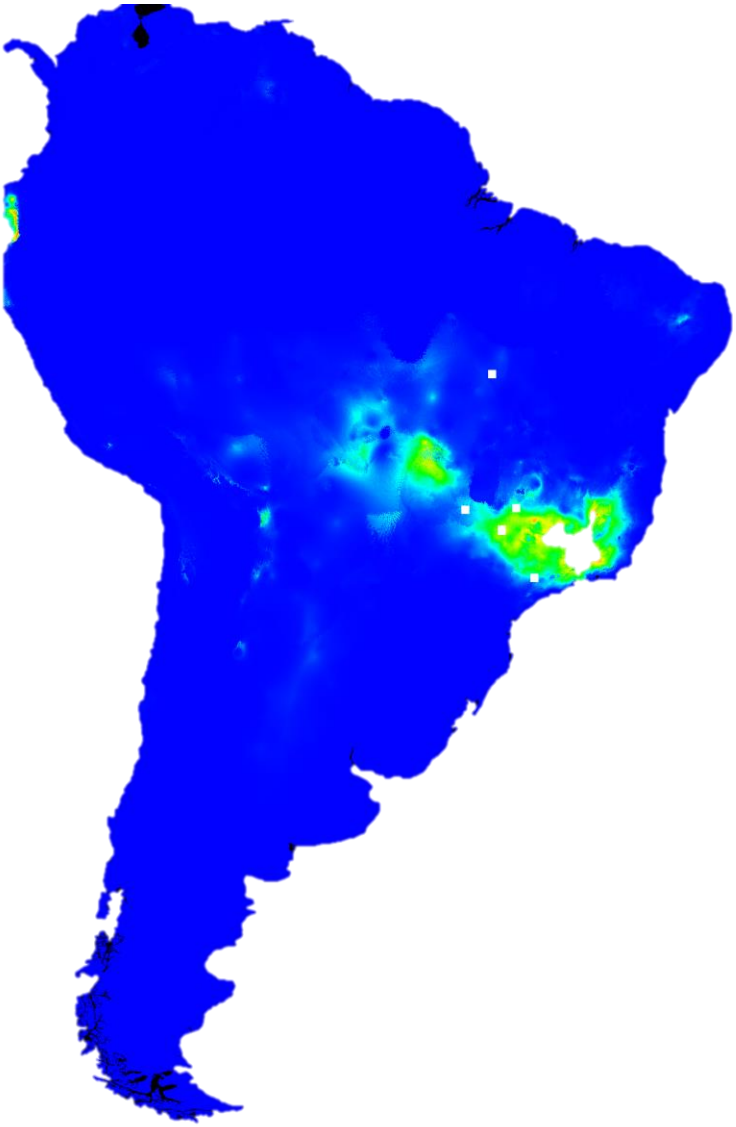
Mean Temperature of Wettest Quarter and Bio5, Max Temperature of Warmest Month, were found to provide the least amount of useful information for species distribution.

Figure 10. Jackknife testing of variable importance for *An. strodei*



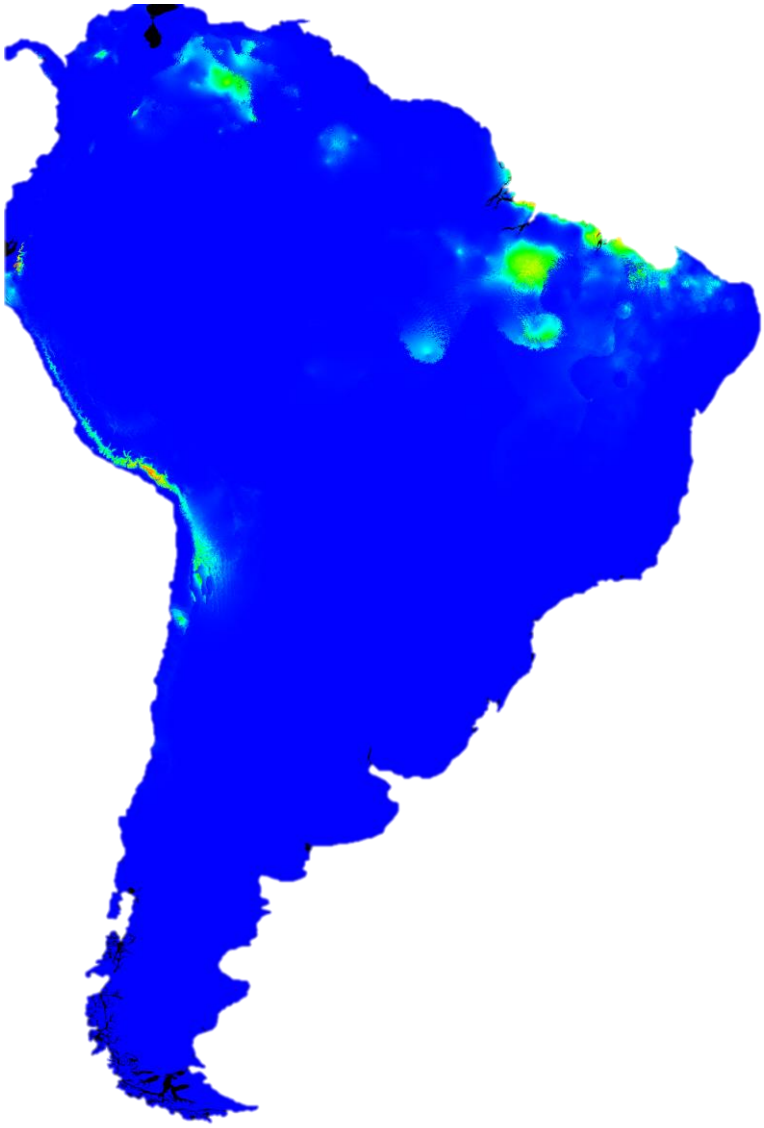
Source: Created by author using Maxent

Figure 11. Representation of the predicted *An. arthuri* A. White dots show presence locations used for model training. Warmer colors demonstrate areas with better-predicted conditions for the occurrence of the species.



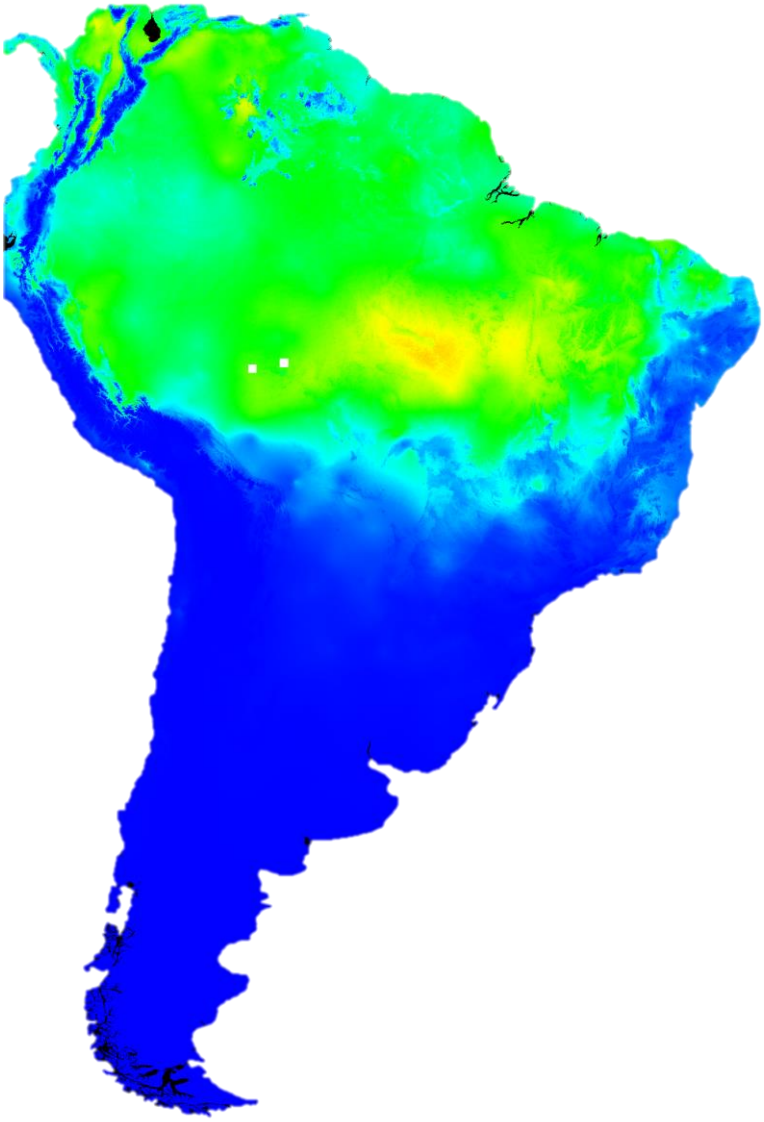
Source: Created by author using Maxent

Figure 12 Representation of the predicted *An. arthuri* B distribution. White dots show presence locations used for model training. Warmer colors demonstrate areas with better predicted conditions for the occurrence of the species.



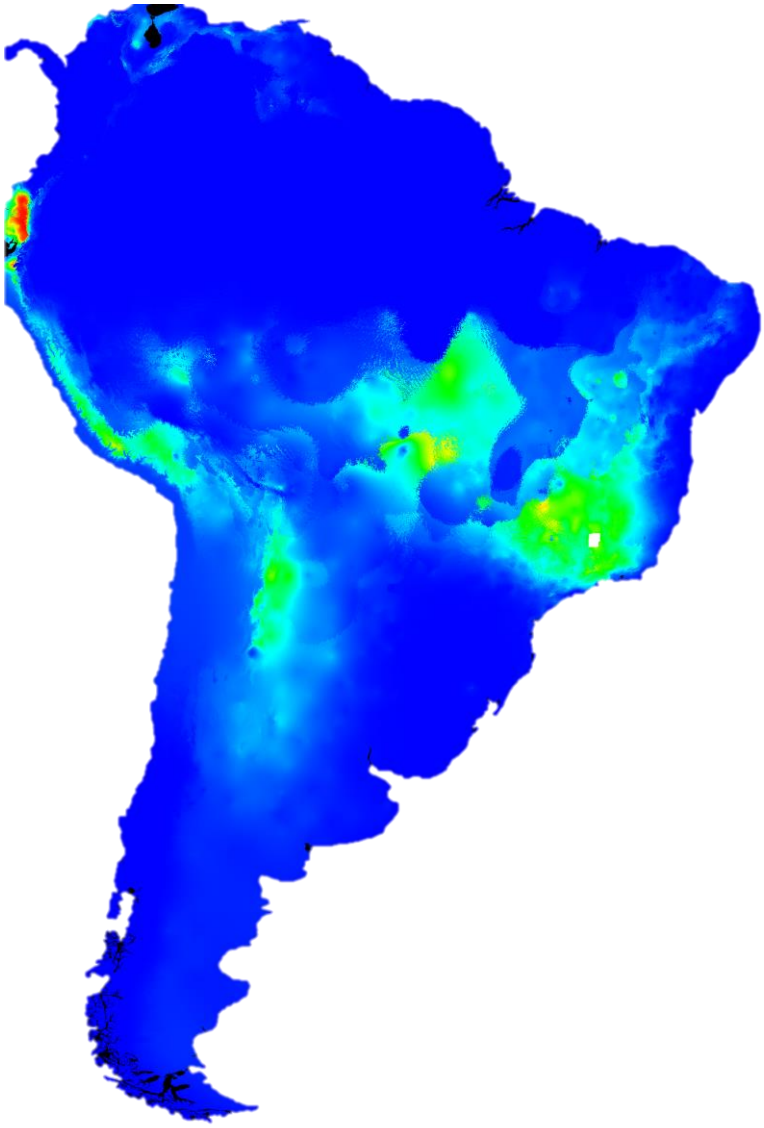
Source: Created by author using Maxent

Figure 13. Representation of the predicted *An. arthuri* C. White dots show presence locations used for model training. Warmer colors demonstrate areas with better predicted conditions for the occurrence of the species.



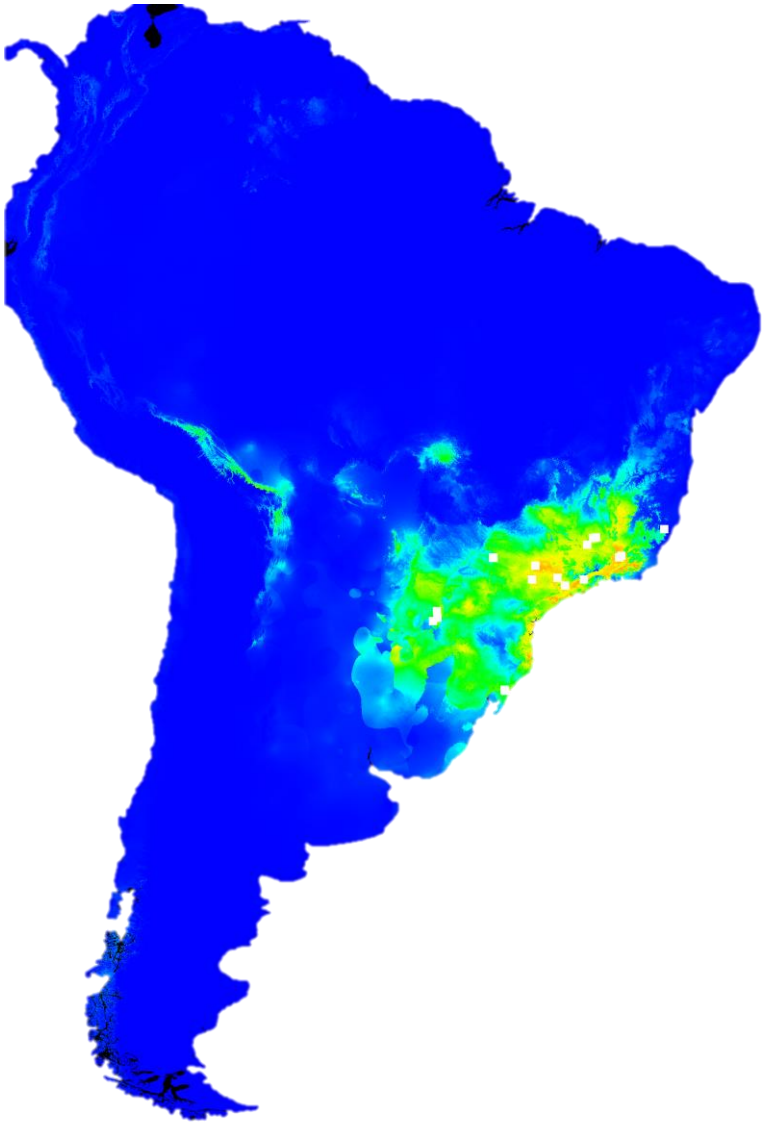
Source: Created by author using Maxent

Figure 14. Representation of the predicted *An. arthuri* D distribution. White dots show presence locations used for model training. Warmer colors demonstrate areas with better predicted conditions for the occurrence of the species.



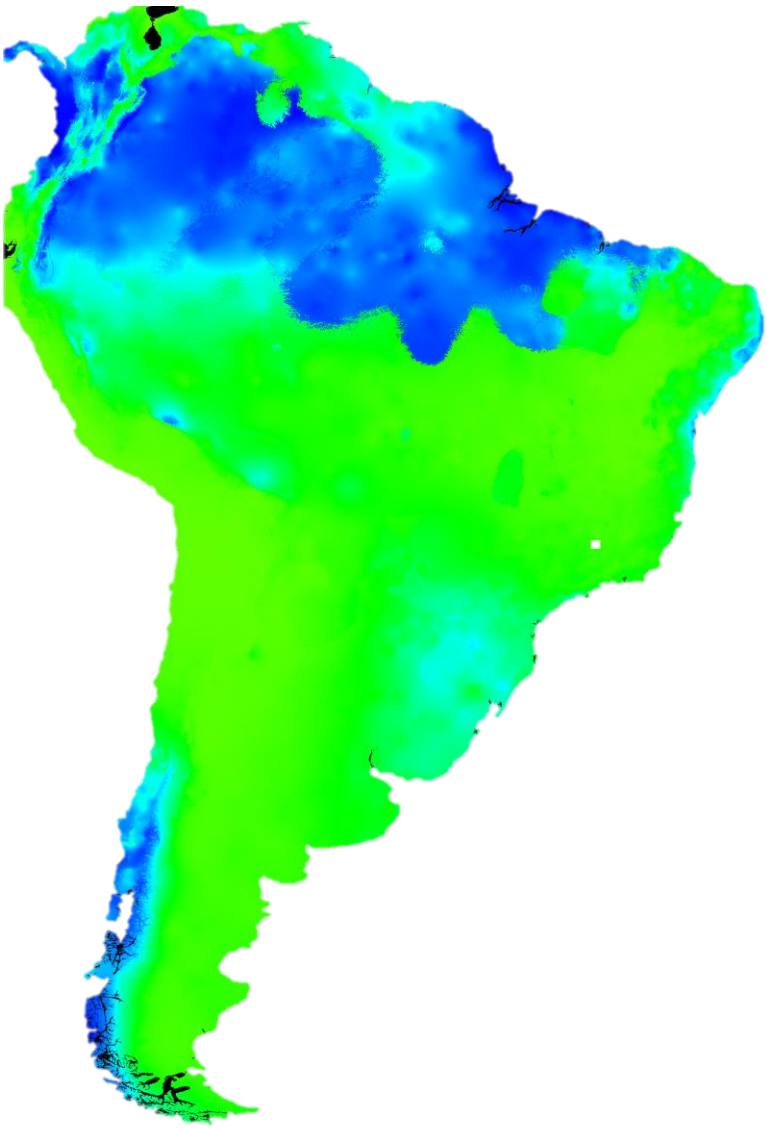
Source: Created by author using Maxent

Figure 15. Representation of the predicted *An. strodei*. White dots show presence locations used for model training. Warmer colors demonstrate areas with better predicted conditions for the occurrence of the species.



Source: Created by author using Maxent

Figure 16. Representation of the predicted *Anopheles* CP Form distribution. White dots show presence locations used for model training. Warmer colors demonstrate areas with better predicted conditions for the occurrence of the species.



Source: Created by author using Maxent

7. DISCUSSION

7.1 PHYLOGENETIC ANALYSIS

The *An. strodei* subgroup is an understudied subgroup of mosquitoes with unsure classifications. There are a number of proposed species, however, correct delineation is problematic as species of the subgenus *Nyssorhynchus* are difficult to identify based on morphological characters alone due to their physical similarities (Sallum, 2010). Some differences can be easily identified in their eggs, as in the case of *An. albertoi,* which lacks floaters on the eggs (Sallum, 2010).

For ease of delineation and accuracy, new molecular techniques have been employed in the study of species of the Strodei Subgroup. Sallum et al. (2010) used molecular markers coupled with classical taxonomy to investigate the *An. strodei* subgroup and added *An. albertoi, An.* CP Form and *An. arthuri* to the subgroup. Subsequently, Bourke et al. (2013) used three molecular markers to examine the relationships in the *An. strodei* subgroup and found support for the three species and proposed three new species designated as *An. arthuri* B-D. *Anopheles arthuri* A is the nominal species.

In the present study, the goal was to look at these seven species as well as *An. strodei* in the *An. strodei subgroup* and establish their relationships. To that end, Bayesian analyses were performed using four molecular markers: *white, CAD, COI,* and *CAT* nuclear genes.

The concatenated gene tree using *CAD, COI* and *white* nuclear genes (Figure 2) resolved the subgroup into 8 clades. This tree supported *An. arthuri, An. albertoi* and *An.* CP Form proposed by Sallum as well as the four proposed species, *An. arthuri* A-D, by Bourke et al. (2013) and resolved *An. strodei* into a strongly supported clade (BPP of 1) with three possible lineages within it. The first possible lineage in the *An. strodei* clade is comprised

of just three individuals, collected in three distinct localities, São José da Vitória, Bahia, in the northeast of Brazil, Lucelia, São Paulo and Pindamonhagaba, São Paulo. The BPP supporting this clade was less than 0.5.

The second possible lineage is comprised of 10 individual specimens collected in different localities throughout Brazil including localities in the states of São Paulo, Minas Gerais, Espírito Santo and Paraná. The BPP supporting this clade is 0.68. The lineage leading to the third lineage in the *An. strodei* clade, which would be considered a sister group, is comprised of just one lone individual, collected in Minas Gerais state, Brazil. The support seen for this clade was also less than 0.5. The low BPPs, and lack of geographical congruency in collection points suggest that while it is indeed possible that these are new species, unknown previously to science, it is more likely that these are simply distinct populations of *An. strodei* s.s.

The gene tree using only the *CAT* gene was not found to be informative and was unable to separate any of the species apart from *Anopheles* CP Form, which was also found in these other two studies cited above to be extremely different.

The gene tree using only CAD was able to separate *An. strodei* and *An. albertoi* into one clade but was unable to resolve them from each other. *An.* CP Form was fully resolved. *An. arthuri* was resolved into its own clade but was not resolved into the proposed four species.

7.2 ECOLOGICAL NICHE MODELING

The predicted South American map contained 224 presence points. Of these, there were 3 *Anopheles* CP Form, 8 presence points for *An. arthuri* D, 20 presence points for *An. arthuri* C, 96 presence points for *An. arthuri A*, 30 presence points for *An. arthuri* B and 67 presence points for *An. strodei* for a total of 224 total presence records.

Ecological niche models are certainly useful for investigating and predicting species distribution in the context of biogeographical analyses, however, they can also provide support for newly discovered species for which there is little geographical presence data, or behavior data, as in the case of *An. arthuri* in the present study (Raxworthy et al., 2009). One important consideration in judging closely related candidate species is whether they occupy their own niche and whether or not there is interbreeding (Wielstra et al., 2012). There appears to be a high level of genetic divergence in the *An. arthuri* clade, supporting the idea of evolutionary independence, however, spatial distribution evidence could help to make this a more compelling case. Based on the predicted species distributions seen in figures 10-13, there seems to be a possibility for overlap in the niches of *An. arthuri* A and D and to a lesser extent between *An. arthuri* B and C, however, their proposed distributions are all quite different from one another.

There is a paucity of studies looking at the applications of ecological niche modeling and how it can be applied to species delimitation. Raxworthy et al (2013) used ecological niche modeling to support their elevation of three subspecies of geckos to species status in Madagascar. They concluded that ENM could be a highly useful tool when looking at species for which little distribution data is known or for species with low vagality and a well-defined spatial distribution.

In this case, little is known of the distribution of *An. arthuri* and mosquitoes are known for having a low vagality with a flight range of typically less than one mile (Rios and Connelly, 2015).

8. CONCLUSIONS

This project reaffirmed the 4 proposed species *An. arthuri* A-D and corroborated *An.* CP Form. *An. strodei* s.l. was found to be a monophyletic clade.

Fourty-five unqiue *CAD* sequences and 55 unique *CAT* sequences were published for for the Strodei Subgroup.

Proposed distributions were created for 6 species in the Strodei Subgroup.

9. REFERENCES

Beard, Charles. B, Eisen, Rebecca J., Barker, Christopher M., Garofalo, Jada F., Hahn, Micah, Hayden, Mary, Monaghan, Andrew J., Ogden, Nicholas H., Schramm, Paul J., Vector borne diseases. 2016. Retrieved June 1, 2016 from https://health2016.globalchange.gov/vectorborne-diseases

Besansky, N.J., Fahey, G.T., Utility of the white Gene in Estimating Phylogenetic Relationships Among Mosquitoes (Diptera: Culicidae), Molecular Biology and Evolution, 1997, 14:4

Bickford, D. et al., Cryptic species as a window on diversity and conservation, Trends in Ecology and Evolution, 2006, 22:3.

Bourke, B.P, Oliveira, T. P., Suesdek, L., Bergo, E.S., Sallum, M.A.M, 2013, A multi-locus approach to barcoding in the *An. An. strodei subgroup* (Diptera: Culicidae), Parasites & Vectors 2013, 6:111.

Bôtto-Menezes, Camila, Azucena Bardají, Giselane dos Santos Campos, Silke Fernandes, and Kara Hanson. "Costs Associated with Malaria in Pregnancy in the Brazilian Amazon, a Low Endemic Area Where *Plasmodium vivax* Predominates." *PLOS Neglected Tropical Diseases*, 2016.

Bradshaw, W.E., Blood Feeding and Capacity for Increase in the Pitcher Plant Mosquito, *Wyeomia smithii,* Environmental Entomology, 1979, 9:1.

CDC, Dengue, April 2014, Retrieved May 11, 2014 from http://www.cdc.gov/Dengue/

CDC, Ross and the Discovery that Mosquitoes Transmit Malaria Parasites, November 2012, Retrieved March 6, 2014, from http://www.cdc.gov/malaria/about/history/ross.html

Corona, M, and G.E. Robinson. "Genes of the antioxidant system of the honey bee: annotation and phylogeny." *Insect molecular biology* 15, no. 5 (2006): 687-701.
Cruickshank, R.H., Molecular markers for the phylogenetics of mites and ticks
Systematic & Applied Acarology, 2002, 7, 3:14.

Danforth, Bryan N, Jennifer Fang, and Sedonia Sipes. "Analysis of family-level relationships in bees (Hymenoptera: Apiformes) using 28S and two previously unexplored nuclear genes: CAD and RNA polymerase II." *Molecular Phylogenetics and Evolution* 39 (2006): 358–372.

Danforth, Bryan. *Resources.* 2009.
http://www.danforthlab.entomology.cornell.edu/resources.html (accessed May 1, 2016).

Dayrat, B., Towards Integrative Taxonomy, Biological Journal of the Linnean Society, 2005, 85. London.

Duarte, Ana Maria, et al. "Natural infection in anopheline species and its implications for autochthonous malaria in the Atlantic forest in Brazil." *Parasites & Vectors* 6 (2013): 58.

Etymologia: *An.*. Emerg Infect Dis [Internet]. 2012 Sep [*date cited*]. http://dx.doi.org/10.3201/eid1809.ET1809.

Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., Li, J., G. Lohmann, L., A. Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. M. Overton, J., Townsend Peterson, A., J. Phillips, S., Richardson, K., Scachetti-Pereira, R., E. Schapire, R., Soberón, J., Williams, S., S. Wisz, M. and E. Zimmermann, N. (2006), Novel methods improve prediction of species' distributions from occurrence data. Ecography, 29: 129–151.

Faran, M.E., A Revision of the Albimanus Section of the subgenus Nyssorhynchus of An., Contributions of the American Entomological Institute, 1980, Vol 15.

Foster, P.G., Bergo, E.S., Bourke, B.P., Oliveira, T.M.P., Nagaki, S.S., Sant'Ana, D.C., Sallum, M.A.M., Phylogenetic Analysis and DNA-based Species Confirmation in *An.* (*Nyssorhynchus*), 2013, PLoS One, 8:2.

Folley, D.H., Klein, T.A., Kim, H.C., Brown, T., Wilkerson, R.C., Rueda, L.M., Validation of Ecological Niche Models for Potential Malaria Vectors in the Republic of Korea, 2010, Journal of the American Mosquito Control Association, 26:2.

Freitas LA, Russo CAM, Voloch CM, Mutaquiha OCF, Marques LP, et al., Diversification of the Genus *An.* and a Neotropical Clade from the Late Cretaceous, 2015, PLoS ONE 10(8): e0134462. doi: 10.1371/journal.pone.0134462.

Galvão, A.L.A. Chaves para a determinação das espécies do subgênero Nyssorhynchus do Brasil, Arquivos de Higiene Saúde Pública, 1943, 8:19.

Githeko, Andrew K., Lindsay, Steve W., Confalonieri, Ulisses E., & Patz, Jonathan A. (2000). Climate change and vector-borne diseases: a regional analysis. *Bulletin of the World Health Organization*, *78*(9), 1136-1147.

Griffits, T.H.D, Impounded Waters and Malaria, Southern Medical Journal 19 (1926): 367-70.

Gubler, D.J., The Global Threat of Emergent/Reemergent Vector Borne Diseases, Vector-Borne Diseases: Understanding the Environmental, Human Health, and Ecological Connections, Workshop Summary, 2008.

Gubler, D.J., Resurgent vector-borne diseases as a global health problem, Emerging Infectious Diseases, 1998, 4:3.

Harbach R.E., Subgenus Nyssohynchus, Mosquito Taxonomic Inventory, http://mosquito-taxonomic-inventory.info/subgenus-ltemgtnyssorhynchusltemgt

Hebert, P.D.N, Penton, E.H. et al., Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator, PNAS 2004, 101:41.

Hoef-Emden, Kerstin. http://www.uni-koeln.de/~aeb25/species.html.
Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones and A. Jarvis, 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25: 1965-1978

Jörger, K.M., Noremburg, J.L., Wilson, N.J., Schrödl, M., Barcoding against a paradox? Combined molecular species delineations reveal multiple cryptic lineages in elusive meiofaunal sea slugs, *BMC Evolutionary Biology* 2012, 12:245.

Keillen M Martins-Campos et al., Integrated vector management targeting *An. darlingi* populations decreases malaria incidence in an unstable transmission area, in the rural Brazilian Amazon. Malaria Journal 2012, 11:351.

Krzywinski, J., Wilkerson, R.C., Besansky, N.J., Toward Understanding Anophelinae (Diptera, Culicidae) Phylogeny: Insights from a Nuclear Single-copy genes and the Weight of Evidence. Systematics Biology 2001, 50:4.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ and Higgins DG. *Bioinformatics 2007 23(21): 2947-2948.*

Laporta, Gabriel, et al. "Plasmodium falciparum in the southeastern Atlantic forest: a challenge to the bromeliad-malaria paradigm?" *Malaria Journal* 14 (2015): 181.

Laporta, Gabriel, et al. Malaria vectors in South America: current and future scenarios. Parasites and Vectors. (2015) 8:426.

Linthicum, K.J., A revision of the Argyritarsis Section of the subgenus Nyssorhynchus of An. (Diptera: Culicidae), Mosquito Systematics, 1988, 20:1.

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

Maddison, W. P. and D.R. Maddison. 2015. Mesquite: a modular system for evolutionary analysis. Version 3.04 http://mesquiteproject.org

Mallet, Jim. *BIODIVERSITY AND SPECIES.* http://www.ucl.ac.uk/~ucbhdjm/courses/b242/Spp/SppPP.pdf (accessed May 1, 2016).

Portal Brasil. October 2009. http://www.brasil.gov.br/meio-ambiente/2009/10/biomas-brasileiros (accessed May 1, 2016).

Monaghan, M.T., Balke, M. et al., DNA-based species delineation in tropical beetles using mitochondrial and nuclear markers. Philos Trans R Soc Lond B Biol Sci., 2005, 360:1462.

Moroni, R.B., Maia, J.F., Tadei, W.P., Santos, J.M.M., Genetic variability among *An.* species belonging to the *Nyssorhynchus* and *An.* subgenera in the Amazon region, Revista da Sociedade Brasileira de Medicina Tropical, 2010, 43:4.

Natural History Museum of the University of Kansas, NHM of KU, Desktop GARP FAQ, Retrieved December 1, 2014 from http://www.nhm.ku.edu/desktopgarp/FAQ.html.

Neafsey, Daniel E. et al., Highly evolvable malaria vectors: The genomes of 16 *An.* mosquitoes, Science, 2015, 247:6217.

Ohio State University, OSU, Bayesian Estimation of Species Trees, Retrieved December 1, 2014 from https://www.stat.osu.edu/~dkp/BEST/introduction/.

Oliveira-Ferreira, J., Lourenco-de-Oliveira, R., Deane, L.M., Teva, A., Daniel-Ribeiro, C.T., Natural malaria infections in Anophelines in Rondonia State, Brazilian Amazon, The American Journal of Tropical Medicine and Hygiene, 1990. 43(1):6-10

Padial, J.M., Miralles, A. et al., The Integrative Future of Taxonomy, Frontiers in Zoology 2010, 7:16.

Pan-American Health Organization, Estado do Acre é homenageado como campeão na luta contra a malária nas Américas 2013, November 19, 2013 (accessed April 1, 2014 from www.paho.org/bra)

Peterson, A.T., Predicting species´ geographic distributions based on ecological niche modeling, The Cooper Ornithological Society, 2010, 103:599-605.

Posada D. In press. jModelTest: Phylogenetic Model Averaging. Molecular Biology and Evolution. Guindon S and Gascuel O (2003). A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood". *Systematic Biology* 52: 696-704.

QGIS Development Team, <2016>. QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.osgeo.org

Saez, A.G., Lozano, E., Body doubles. Nature 2005, 433:111.

Sallares, R. Malaria and Rome: A history of Malaria in ancient Italy, 53, Oxford University Press, 2002, Oxford.
Sallum, M.A.M., Peyton, E.L., Harrison, B.A., Wilkerson, R.C., Revision of the Leucosphyrus group of An. (Celia) (Diptera, Culicidae), Revista Brasileira de Entomologia, 2005, 49:Sup.

Sallum, M.A.M, Foster, P.G., Santos, C.L.S., Flores, D.C., Motoki, M.T., Bergo, E.S., Resurrection of Two Species From Synonymy of *An.* (*Nyssorhynchus*) *strodei* Root, and Characterization of a Distinct Morphological Form From the Strodei Complex (Diptera: Culicidae), Journal of Medical Entomology 2010, 47:4.

Service, M., Medical Entomology for Students: Third Edition, Cambridge University Press, 2004, Oxford.

Sequencher® version 5.4.1 sequence analysis software, Gene Codes Corporation, Ann Arbor, MI USA http://www.genecodes.com

Shepard, Donald , et al. "Economic and Disease Burden of Dengue Illness in India." *American Journal of Tropical Medicine and Hygiene*, 2014: 1235-1242.

Stevenson J, St. Laurent B, Lobo NF, Cooke MK, Kahindi SC, Oriango RM, et al. Novel vectors of malaria parasites in the western highlands of Kenya [letter]. Emerg Infect Dis.

Suaya, J A, et al. "Cost of Dengue Cases in Eight Countries in the Americas and Asia: A Prospective Study." *The American Journal of Tropical Medicine and Hygiene*, 2009: 846-855.

Surendran, S.N., Sarma, D.K., Jude, P.J., Kemppainen, P., Kanthakumaran, N., Gajapathy, K., Peiris, L.B.S., Ramasamy, R., Walton, C., Molecular characterization and identification of members of the *An. subpictus* complex in Sri Lanka, Malaria Journal, 2013, 12:304.

Sum, J.S., Lee, W.C., Amir, A. Braima, K.A., Jeffery, J., Abdul-Aziz, N.M., Fong, M.Y., Lau, Y.L., Phylogenetic study of six species of *An.* mosquitoes in Peninsular Malaysia based on inter-transcribed spacer region 2 (ITS2) of ribosomal DNA, Parasites & Vectors, 2014, 7:309.

Tadei, W.P., Thatcher, B.D., Malaria Vectors in the Brazilian Amazon: An. of the subgenus Nyssorhynchus, Revista do Instituto de Medicina Tropical de São Paulo, 2000, 42:2.

The Ideal Barcoding Gene
http://biotechlearn.org.nz/themes/barcoding_life/the_ideal_barcoding_gene (accessed May 1, 2016).

The University of California Museum of Paleontology, Berkeley, and the Regents of the University of California. *Biological Species Concept.*

http://www.evolution.berkeley.edu/evosite/evo101/VA1BioSpeciesConcept.shtml. (accessed Feb 1, 2016).

van Proosdij, A. S. J., Sosef, M. S. M., Wieringa, J. J. and Raes, N. (2015), Minimum required number of specimen records to develop accurate species distribution models. Ecography. doi: 10.1111/ecog.01509.

Wang, Yan, Jiannong Xu, and Yajun Ma. "Molecular Characterization of Cryptic Species of *Anopheles Barbirostris* van Der Wulp in China." *Parasites & Vectors*7 (2014): 592. *PMC*. Web. 8 July 2016.

Warren, D L, and S N Seifert. "Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria." *Ecological Applications* 21 (2011): 335-342.

Warrell, D.A., Gilles, H.M. eds. Essential Malariology, 4th ed., 59, Hodder Arnold, 2002, London.

Wilson E.O. Taxonomy as a fundamental discipline. Philosophical Transactions of the Royal Society of London B 2004, 359:739. London.

World Health Organization. "World Malaria Report 2015." *WHO*. January 1, 2015. http://apps.who.int/iris/bitstream/10665/200018/1/9789241565158_eng.pdf?ua=1 (accessed January 1, 2016).

Youfen, Li, Jeong-Soon Park, Jian-Hong Deng, and Yidong Bai. "Cytochrome c Oxidase Subunit IV is Essential for Assembly and Respiratory Function of the Enzyme Complex." *Journal of Bioenergetics and Biomembranes* 38 (2006): 283-291.

Zhang, N., Zeng, L., Shan, H., Ma, H., Highly conserved low-copy nuclear genes as effective markers for phylogenetic analyses in angiosperms, New Phytologist, 2012, 195.

Susan Elaine Greni

Endereço para acessar este CV: http://lattes.cnpq.br/8412653822920550

Última atualização do currículo em 07/05/2014

Possui graduação em Biologia pela University of Nebraska - Lincoln(2011). Tem experiência na área de Saúde Coletiva, com ênfase em Saúde Pública. **(Texto gerado automaticamente pela aplicação CVLattes)**

Identificação

Nome

Susan Elaine Greni

Nome em citações bibliográficas

GRENI, S. E.

Endereço

Endereço Profissional

Universidade de São Paulo.
Avenida Doutor Arnaldo - de 601/602 ao fim
Sumaré
01255000 - São Paulo, SP - Brasil
Telefone: (11) 30852329

Formação acadêmica/titulação

2014

Mestrado em andamento em Saúde Pública (Conceito CAPES 6).
Universidade de São Paulo, USP, Brasil. Orientador: .
Bolsista do(a): .

2007 - 2011

Graduação em Biologia.
University of Nebraska - Lincoln.
Orientador: Brigitte Tenhumberg.
Bolsista do(a): .

Maria Anice Mureb Sallum

Endereço para acessar este CV: http://lattes.cnpq.br/1200247772482237

Última atualização do currículo em 05/05/2016

Possui graduação em Ciências Biológicas pela Universidade de São Paulo (1976), mestrado (1991) e doutorado (1994) em Saúde Pública pela Universidade de São Paulo. Possui título de livre docente, pela Universidade de São Paulo, desde 2010. Atualmente é professor adjunto do Departamento de Epidemiologia da mesma universidade. Membro do Comitê Editorial das revistas: Infection, Genetics and Evolution, Memórias do Instituto Oswaldo Cruz e Revista Brasileira de Entomologia. Tem experiência na área de Parasitologia, com ênfase em Entomologia e Malacologia de Parasitos e Vetores, atuando principalmente nos seguintes temas: Entomologia Médica, Culicidae, Anopheles, Culex, taxonomia morfológica e molecular, ecologia, biodiversidade e evolução. **(Texto informado pelo autor)**

Identificação

Nome

Maria Anice Mureb Sallum

Nome em citações bibliográficas

SALLUM, M. A. M.;Sallum, Maria Anice M.;Sallum, Maria Anice Mureb;Mureb Sallum, Maria A.;Mureb Sallum, María A.;Mureb Sallum, Maria Anice;Sallum, Maria AM;Sallum, Maria Anice;Maria Anise Sallum;Sallum, Maria Anise;SALLUM, M.A.M.;SALLUM, MARIA

Endereço

Endereço Profissional

Universidade de São Paulo, Faculdade de Saúde Pública.
AV. DR. ARNALDO, 715 Sala 210
CERQUEIRA CESAR
01246904 - São Paulo, SP - Brasil
Telefone: (11) 30617731
Fax: (11) 32821898

Formação acadêmica/titulação

1991 - 1994

Doutorado em Saúde Pública (Conceito CAPES 6).
Universidade de São Paulo, USP, Brasil.
Título: REVISÃO DA SEÇÃO SPISSIPES DE CULEX (MELANOCONION)(DIPTERA:CULICIDAE), Ano de obtenção: 1994.
Orientador: OSWALDO PAULO FORATTINI.
Palavras-chave: Culex (Melanoconion); Culicidae; Sistemática.
Grande área: Ciências da Saúde

1989 - 1991

Mestrado em Saúde Pública (Conceito CAPES 6).
Universidade de São Paulo, USP, Brasil.
Título: ESTUDO TAXONOMICO DO CIBÁRIO NA SEÇÃO SPISSIPES DE CULEX (MELANOCONION) (DIPTERA: CULICIDAE),Ano de Obtenção: 1991.
Orientador: OSWALDO PAULO FORATTINI.
Palavras-chave: Culex (Melanoconion); Culicidae; Sistemática.
Grande área: Ciências da Saúde

1979 - 1979

Especialização em Entomologia Medica. (Carga Horária: 560h).
Universidade de São Paulo, USP, Brasil.

1973 - 1976

Graduação em Ciências Biológicas.
Universidade de São Paulo, USP, Brasil.