Ministério da Saúde
Fundação Oswaldo Cruz
Centro de Pesquisas René Rachou
Programa de Pós-graduação em Ciências da Saúde

Genomic study of *Anopheles (Nyssorhynchus) aquasalis* Curry, 1932. A Neotropical human malaria vector

por

Luis Eduardo Martínez Villegas

Belo Horizonte
2015

TESE DBCM-CPQRR L.E.M. VILLEGAS 2015

LUIS EDUARDO MARTÍNEZ VILLEGAS

Genomic study of *Anopheles (Nyssorhynchus) aquasalis* Curry, 1932. A Neotropical human malaria vector

Tese apresentada ao Programa de Pós-graduação em Ciências da Saúde do Centro de Pesquisas René Rachou, como requisito parcial para obtenção do título de Doutor em Ciências - área de concentração de Biologia Celular e Molecular.

Orientação: Paulo Filemón Paollucci Pimenta
Coorientação: Guilherme Oliveira.

2015

Catalogação-na-fonte
Rede de Bibliotecas da FIOCRUZ
Biblioteca do CPqRR
Segemar Oliveira Magalhães CRB/6 1975

V732g
2015

Villegas, Luis Eduardo Martínez.

Genomic study of Anopheles (Nyssorhynchus) aquasalis, Curry 1932. A Neotropical human malaria vector / Luis Eduardo Martínez Villegas. – Belo Horizonte, 2015.

xvii, 119 f: il.: 210 x 297 mm.

Bibliografia: 92 – 126

Tese (doutorado) – Tese para obtenção do título de Doutor(a) em Ciências pelo Programa de Pós-Graduação em Ciências da Saúde do Centro de Pesquisas René Rachou. Área de concentração: Biologia Celular e Molecular.

1. Malária/genética 2. Plasmodium/genética 3. Anopheles/genética 4. Microbiota/genética 5. Metagenômica/métodos 6. Genoma mitocondrial/genética I. Título. Pimenta, Paulo Filemon Paolucci (Orientação). III. Oliveira, Guilherme Corrêa (Coorientação).

CDD – 22. ed. – 616.936 2

LUIS EDUARDO MARTÍNEZ VILLEGAS

Genomic study of *Anopheles (Nyssorhynchus) aquasalis* Curry, 1932. A Neotropical human malaria vector

Tese apresentada ao Programa de Pós-graduação em Ciências da Saúde do Centro de Pesquisas René Rachou, como requisito parcial à obtenção do título de Doutor em Ciências - área de concentração de Biologia Celular e Molecular.

Banca examinadora:

Prof. Dr. Paulo Filemón Paolucci Pimenta (CPqRR/FIOCRUZ) Presidente
Dr. Victor Pylro (CPqRR/FIOCRUZ) Titular
Dr. Luiz Fabiano Borges Oliveira (CPqRR/FIOCRUZ) Titular
Prof. Dr. Aristóteles Góes Neto (UEFS) Titular
Dr. Héctor Díaz-Albiter (IOC/FIOCRUZ) Titular
Prof. Dr. Nágila Francinete Costa Secundino (CPqRR/FIOCRUZ) Suplente

Tese defendida e aprovada em Belo Horizonte, 10/04/2015

Exploration is wired into our brains.

If we can see the horizon, we want to know what's beyond.

Buzz Aldrin.

ACKNOWLEDGEMENTS

¨Ad astra per aspera.¨

I would like to express my gratitude to the ¨Programa de Estudantes-Convênio de Pós-Graduação - PEC-PG, CAPES for the financial support provided to embark on this endeavor. To the ¨Centro de Pesquisas Renê Rachou – Fundação Oswaldo Cruz (CPqRR-FIOCRUZ MINAS)¨, the Board of Graduate Studies and its administrative staff, for their diligent support and assistance each time I needed it, which were many.

To attain a PhD degree implies that a fellow and senior researcher believes in your potential to fulfill the requirements of the program. Thus I would like to express my profound gratitude to Dr. Paulo F.P. Pimenta for accepting me in his laboratory without hesitation, for his patience with the time it took me to adapt to a new country and the PhD life, for allowing me to think freely and be curious, for the financial support, and the scientific / personal advice he has kindly provided through the years. Indeed, becoming a PhD requires much more than scientific skills, it demands character and social skills, all of which I realized during my time with a person with such a long trajectory. I thank Dr. Nagila Secundino, vice chief at the Medical Entomology Laboratory for her support, advice, and encouragement in key moments during the PhD experience.

I would also like to express my gratitude to Dr. Guilherme Oliveira and Dr. Angela Volpini at the Bioinformatics core facility (CEBio). The chance to interact with all the personnel at CEBio, to participate in their seminars and meetings made me realize how much I enjoy the challenge of learning new skills such as using bioinformatics' tools. The time there allowed me to learn how to have a productive conversation with a bioinformatician / expert user of ¨omic¨ tools. To both of them, thanks for opening the doors of CEBio, for their financial support and help whenever I needed it.

Special thanks to all my dear friends and colleagues, both past and current members of the Medical Entomology Laboratory and CEBio. The journey has been more interesting, productive and joyful because of the chance I got to meet and learn from excellent individuals.

Many people collaborated with me to get to this point in my career. But the PhD really began in 2009 with a trip to visit Carolina Barillas at the NIH. Thank you Carolina for receiving me without actually knowing me. I have so much respect and admiration for you. I would like to thank you for believing I could go abroad and become a curious and insightful researcher. I constantly try to remember your challenging approach to things: think big, aim

high but keep it simple. That visit to NIH was indeed a breaking point, incited by my dear friend and colleague Rodrigo, to whom I would like to extend a sincere acknowledgement for his exemplary commintment to science and friendship.

In science, you don´t always follow a scheme. I had the opportunity to brainstorm a lot during my brief and unplanned visit to the University of California at Davis with the group of Dr. Gregory Lanzaro. Not only did I generate data there, the ¨State of California¨ and the Vector Genetics Lab people gave me many ideas and positive input on how to re-formulate my research questions. For two months I experienced 24/7 academia in sunny California.

I have always been curious and passionate about science. I also try to put myself at the service of others whenever it is possible. I am positively sure I inherited that trait from my loving and awesome parents. Maria del Carmen and José Roberto (a.k.a. Mom and Dad), thanks to you both for overflowing me with love, education and examples of patience, hard work and courage. I am just doing what you raised me for.

To my two older brothers, Roberto and Juan Pablo, thanks for having my back. I admire your many stories and examples of struggle, success and family ties. To both of you (and my superb sisters-in-law) I owe the privilege of having amazing nephews and nieces. To each of them: Carlo, Mercy, and Mariana; Alex, Tato and Bailee: you are the force that drives me to be a good and honest person. Never quit on any goal you have; be smart, and be safe.

When I was 18 I had cancer. It changed my life. It made me realize that being kind, respectful, curious and brave are traits that should accompany everyday of your life. As you progress in your professional life you realize that any achievement, big or small, tastes better if you fought hard around the clock to get it and, if along the way, you made the greatest discovery: in science and in life in general, you are never alone…so what I treasure more than knowledge is the friends I made in these 4 years.

To Yesid, Eugenia, Jose Manuel, Vicky, Jorge and Marinely: ¨Gracias, son mi família hispanoamericana¨. Thank you for your help, support, kindness and most of all for the many, many laughs we shared. Having friends spread across America is indeed the most cherished treasure I carry from this PhD endeavor.

Thus, I finish saying to my family, friends, and my home country Guatemala: you were always in my mind propelling me to do my best. ¨Obrigado, Thank you, Gracias¨.

Suporte Financeiro

O Programa de Estudantes-Convênio de Pós-Graduação - PEC-PG, CAPES.
Processo: 571610-1.

CONTENTS

			Page
		LIST OF FIGURES	xi
		LIST OF TABLES	xii
		LIST OF ABBREVIATIONS	xiii
		Resumo	xvi
		Abstract	xvii
1		**Introduction**	
	1.1	The anopheline mosquitoes	18
	1.2	Arthropods acting as natural pathogen vectors	21
	1.3	Human malaria and Neotropical anopheline vectors	21
	1.4	*Anopheles aquasalis* bionomics	25
	1.5	Malaria transmission laboratory models	25
	1.6	On metagenomics and the *Anopheles* holobiont	26
	1.7	Consideration of *Anopheles* genomes and those of New World vectors	30
		Objectives	
2	2.1	**General Objective**	34
	2.2	**Specific Objectives**	
3		**Chapter 1: Exploring biological patterns in *Anopheles* associated microbiota: Is there a core bacterial assembly shared between African and American anophelines.**	35
		Abstract	36
	3.1	**Introduction**	37
	3.2	**Methods**	
		3.2.1 Mosquito rearing	39
		3.2.2 Sample Collection	39
		3.2.3 DNA extraction	39
		3.2.4 Ion Torrent PGM Library preparation and sequencing	40
		3.2.5 Shotgun metagenomic analysis	40
		3.2.6 Compilation of *Anopheles*-associated microbiota: searching for a geographically determined *Anopheles* core microbiota	41
		3.2.7 Ecological diversity analysis	42
	3.3	**Results**	
		3.3.1 Data summary	44
		3.3.2 Bacterial community composition	44
		3.3.3 Compilation of *Anopheles*-associated microbiota: searching for the *Anopheles* core microbiota	44
		3.3.4 β-diversity exploration: comparing bacterial profiles between aquatic stages of *A.aquasalis* and *A. gambiae*	49
	3.4	**Discussion**	
		3.4.1 *Anopheles aquasalis* associated microbiota	52
		3.4.2 The *Anopheles* core microbiota	56
		3.4.3 Ecological multivariate analysis between	59
	3.5	**Conclusion**	61

Page

4 **Chapter 2: Preliminary Characterization of the Mitochondrial Genome of *Anopheles (Nyssorhynchus) aquasalis* Curry, 1932. A Neotropical model of human malaria transmission** 63

Abstract 64

4.1 **Introduction**

4.1.1 Insect mitochondrial genomes 65

4.1.2 Anopheline phylogenetics and divergence times 66

4.2 **Methods**

4.2.1 *Anopheles aquasalis* colony 68

4.2.2 Single mosquito DNA extraction 68

4.2.3 Whole genome shotgun sequencing 68

4.2.4 Mitochondrial genome assembly 69

4.2.5 Sequence analysis: composition and genomic features of the *A. aquasalis* mitogenome 69

4.2.6 Comparative analyses between anophelines from different geographic regions 70

4.2.7 Phylogenetic analysis and molecular dating 70

4.3 **Results**

4.3.1 Composition and genomic features 72

4.3.2 Comparative composition and identity analyses between selected anophelines 78

4.3.3 Phylogeny and molecular dating analyses 79

4.4 **Discussion** 83

4.5 **Conclusion** 90

Bibliography 92

Appendices

Appendix 1: Additional files; supplementary figures and tables 127

Appendix 2: Doctoral Thesis related publications 135

LIST OF FIGURES

No.	Title:	Page
Figure 1	Life cycle of the anopheline mosquitoes.	19
Figure 2	Global distributions of the dominant *Anopheles* vectors of human malaria.	22
Figure 3	The malaria parasite cycle within the mosquito vector and the vertebrate host.	23
Figure 4	Human malaria global transmission map and mortality according to Murray *et al*. (2012).	24
Figure 5	Relative abundance of the bacterial community (f-OTUs) associated to *A. aquasalis* pupae.	45
Figure 6	The *A. aquasalis* associated bacterial community.	46
Figure 7	Number of bacterial genera per anopheline host.	47
Figure 8	Shared and exclusive bacterial genera identified in anopheline human malaria vectors grouped by *Anopheles* continental distribution.	48
Figure 9	Relative abundance of the African-American *Anopheles* shared bacterial family-rank OTUs associated to five aquatic stages of anopheline human malaria vectors.	49
Figure 10	Combined ordination and clustering analyses comparing the family-rank OTUs abundance profiles of five anopheline aquatic life stages (African-American shared f-OTUs matrix).	51
Figure 11	The complete mitogenome of *A. aquasalis*.	75
Figure 12	*A. aquasalis* mtDNA predicted tRNA structures.	77
Figure 13	Compositional patterns of *A. aquasalis* and other anopheline mitochondrial genomes.	80
Figure 14	Radar depiction of the nucleotide identity comparison between mitochondrial PCG's and rRNA genes from *A. aquasalis* and other anophelines.	81
Figure 15	Support of the *Anopheles* phylogeny using the concatenated DNA sequences of all the mitochondrial protein coding genes.	86
Figure 16	Phylogenetic tree of *Anopheles* using the concatenated DNA sequences of all the mitochondrial protein coding genes.	87
Figure S1	Combined ordination and clustering analyses comparing the bacterial family-rank OTUs abundance profiles of five anopheline aquatic life stages (full dataset).	93
Figure S2	*Anopheles*-associated microbiota distributed by continent.	94

LIST OF TABLES

No.	Title:	Page
Table 1	List of the insect species used in this study with their corresponding GenBank number. For anopheline species, the continent in which they exert their malaria vectorial activity is shown. The sequence length reflects the number of base pairs assembled not considering N's.	73
Table 2	Organization and gene features of the *A. aquasalis* mitochondrial genome.	74
Table 3	Base composition analysis of the *A. aquasalis* whole mitochondrion DNA sequence, and other annotated features such as individual PCG's, and concatenated rRNA and tRNA genes.	76
Table 4	Codon usage in the *A. aquasalis* mtDNA. Codon frequency and Relative synonymous Codon Usage (RSCU) are shown for each codon.	78
Table 5	Top 10 codons (classified by count) identified in the protein coding genes of the *A. aquasalis* mitogenome.	78
Table 6	Protein coding genes nucleotide sequence identity comparison between *A. aquasalis* and other selected anophelines.	79
Table 7	Mean divergence times in MYA (million of years ago), and 95% credibility intervals for selected nodes.	82
Table S1	Bacterial community composition at family taxonomic rank associated to *A. aquasalis* laboratory reared pupae.	95
Table S2	Index of bacterial genera showing their shared or exclusive distribution among anopheline human malaria vectors classified according to the continent in which they exert vectorial activity.	96
Table S3	Transformed (Arcsin $\sqrt{x}$) African-American anopheline core f-OTUs abundance matrix.	97
Table S4	Comparative strand asymmetry analysis based on AT and GC skews between the *A. aquasalis* mtDNA and selected anopheline species.	98
Table S5	Comparative compositional analysis represented in AT% between the *A. aquasalis* mtDNA and selected anopheline species.	99

LIST OF ABBREVIATIONS

˚C: degrees centigrade

A: Adenine

AGC: *Anopheles* Genomes Cluster

AMPs: Antimicrobial peptides

***ATP8*:** mitochondrially encoded ATP synthase 8 gene

***ATPase6*:** mitochondrial adenosine triphosphatase 6 (ATPase 6) gene

BLAST: Basic Local Alignment Search Tool

bp: base pairs

BRIG: Blast Ring Image Generator

C: Citocine

CDS: Coding sequence

***COI-III*:** Citochrome C oxidase genes

***CYTB*:** Cytochrome B gene

DHU-arm: dihydrouracil-arm

DVS: Dominant vector species/species complexes

EDTA: Ethylenediaminetetraacetic acid

f-OTUs: Family rank operational taxonomic units

G: Guanidine

g-OTUs: Genus rank operational taxonomic units

IGNs: intergenic nucleotides

iTOL: Interactive Tree Of Life

kb: kilo base

lrRNA: long ribosomal RNA gene

M5RNA: Non-redundant multisource ribosomal RNA annotation

MG-RAST QC: Metagenomics Rapid Annotation using Subsystem Technology quality control

MG-RAST: Metagenomics Rapid Annotation using Subsystem Technology

mL: milliliter

mM: millimolar

MRCA: Most recent common ancestor

mtDNA: Mitochondrial Deoxyribonucleic acid

MYA: Million years ago

N/A: Not Applicable

NCBI: [illegible]

***ND*#:** NADH dehydrogenase genes

NGS: Next Generation Sequencing

NMDS: Non-metric MultiDimensional Scaling

NP-MANOVA: Non-parametric multivariate analysis of variance

N's: Nucleotides

ORF: Open reading frame

OTUs: Operational taxonomic units

PBS: Phosphate Buffered Saline

PCG's: Protein coding genes

PCR: Polymerase Chain Reaction

***Pfs47*: *Plasmodium falciparum* surface protein 47 gene

PGM: Personal genome machine

pH: negative logarithm of the activity of the (solvated) hydronium ion / *P*otenz power + *H* (symbol for hydrogen)

RNA: Ribonucleic acid

RNAse: Ribonuclease

rRNA: Ribosomal ribonucleic acid

RSCU: Relative synonymous codon usage

SDS: Sodium dodecyl sulfate

srRNA: short ribosomal RNA gene

T: Thymine

TE: Tris EDTA buffer

Tris-HCl: Tris hydrochloric acid

tRNA: transfer ribonucleic acid

v/v: volume / volume

WHO: World Health Organization

µl: microlitre

Estudo genômico de *Anopheles* (*Nyssorhynchus*) *aquasalis* Curry, 1932. Um modelo neotropical da transmissão da malaria humana.

Resumo

A malária humana é uma doença provocada por parasitas do gênero *Plasmodium*, os quais na natureza requerem de um mosquito anofelíno para completar o seu ciclo de vida e serem transmitidos a um hospedeiro humano. Nas Américas, o Brasil tem a maior incidência de malária, sendo responsável por 41% dos casos. Com o aparecimento do sequenciamento de nova geração e das ferramentas bioinformática relacionados, grandes avanços foram alcançados em relação à montagem de genomas e transcriptomas de anofelinos, assim como na exploração de estratégias de paratransgenesis para interromper a transmissão da malária. No entanto, os vetores neotropicais da malária encontram-se longe dos vetores da África e Ásia no que refere a estes conhecimentos. Este estudo é parte de um esforço contínuo para montar o genoma do *Anopheles aquasalis*, um vetor neotropical da malária humana, que atualmente posiciona-se como um excelente modelo de transmissão da malária no Brasil.

Em paralelo ao sequenciamento do genoma, e para maximizar os dados gerados, optamos por focar em duas tarefas pontuais e viáveis: explorar a diversidade e composição do consórcio bacteriano associado ao anofelino; assim como montar e caracterizar o genoma mitocondrial desta espécie.

O sequenciamento metagenômico ¨shotgun¨ e o programa MG-RAST foram utilizados para fazer um ¨screening¨ das bactérias associadas à pupas de *A .aquasalis* criadas em laboratório. O consórcio bacteriano predito é composto por 74 gêneros contendo bactérias marinhas e bioluminescentes. No nível taxonômico de família bacteriana, identificamos 14 OTUs compartilhadas entre anofelinos americanos e africanos. Além disso, foram comparadas cinco comunidades bacterianas associadas a duas espécies de anofelinos: *A. aquasalis* e *Anopheles gambiae*. Foi identificada uma associação significativa (NPMANOVA $p < 0{,}05$) entre a composição da comunidade bacteriana e o ambiente aquático (laboratório ou condições semi-naturais) nas quais cada hospedeiro anofelino foi criado.

Atualmente, o entendimento da filogenia do gênero *Anopheles* é limitado e as informações sobre o tempo de divergência dentro da linhagem de mosquitos é escassa. Apresentamos a sequencia de 15,393 pb correspondente ao genoma mitocondrial de *A. aquasalis*. Quando comparado com outros mitogenomas anofelinos relevantes, observou-se alta similaridade na composição dos genomas assim como características estruturais conservadas. Através de análises Bayesianas, reconstruímos as relações filogenéticas e estimamos a data de divergência entre 22 anofelinos e outras espécies de dípteros. Descobrimos que o mais recente ancestral entre as subfamílias *Nyssorhynchus* e *Anopheles* + *Cellia* existiu ~ 83 milhões anos atrás (MYA). Estimou-se que *A. aquasalis* divergiu do complexo do *Anopheles albitarisis* faz ~ 28 MYA, e faz ~ 38 MYA do *Anopheles darlingi*.

A distribuição estreita e o peculiar nicho ecológico do *A. aquasalis*, além de considerar a sua adaptação a ambientes larvários com água salobra fizeram nos perguntar se a sua história evolutiva deixou uma marca na arquitetura do seu genoma, assim como sobre a estrutura da comunidade bacteriana associada a este anofelino.

Genomic study of *Anopheles (Nyssorhynchus) aquasalis* Curry, 1932. A Neotropical model of human malaria transmission.

Abstract

Human malaria is a malady caused by *Plasmodium* parasites, which in nature, require an anopheline mosquito to complete their life cycle and be transmitted to a human host. In the Americas, Brazil has the largest incidence of malaria, accounting for 41% of the cases. With the advent of Next Generation Sequencing and related bioinformatics' tools, great leaps forward were attained regarding the assembly of anopheline genomes, transcriptomes; in addition to the exploration of paratransgenesis as means to interrupt malaria transmission. Nonetheless, Neotropical malaria vectors still lag behind those from Africa and Asia on such matters. This study is part of an ongoing effort to assemble the genome of *Anopheles aquasalis*, a Neotropical human malaria vector currently positioned as a key malaria transmission model in Brazil.

In parallel to the genome sequencing study, and to maximize the NGS sequencing data generated, we opted to focus in two punctual and feasible tasks: exploring the diversity and composition of this anopheline's associated bacterial consortium; plus, assemblying and characterizing, the mitochondrial genome of this species.

Shotgun metagenomic sequencing and the MG-RAST suite were used to survey the bacteria associated to laboratory reared *A. aquasalis* pupae. The predicted bacterial consortium is composed of 74 genera and contains marine and bioluminescent bacteria. At the bacterial family rank, we identified 14 OTUs shared between African and American anophelines. In addition, we compared five *Anopheles* associated bacterial communities from two species: *A. aquasalis* and *Anopheles gambiae*. We found a significant association (NPMANOVA $p < 0.05$) between the bacterial community composition and the aquatic environment (laboratory or semi-natural conditions) in which each *Anopheles* host was reared.

The current understanding of the *Anopheles* phylogeny is limited and information regarding the time of deep lineage divergences within mosquitoes is scarce. Here we also present the assembled 15,393 bp mitochondrial genome of *A. aquasalis*. When compared with other relevant anopheline mitogenomes, high composition similarity and conserved features were observed. Through Bayesian analyses, we reconstructed the phylogenetic relationships and estimated the date of divergence between 22 anopheline and other dipteran species. We found that the most recent ancestor between *Nyssorhynchus* and *Anopheles* + *Cellia* subfamilies was extant ~83 million years ago. It was estimated that *A. aquasalis* diverged from the *Anopheles albitarisis* complex ~28 MYA and ~38 MYA from *Anopheles darlingi*.

The narrow distribution and peculiar niche of *A. aquasalis*, plus considering its adaptation to brackish-water larval environments makes us wonder if its evolutionary history left a mark upon its genome architecture, and also on the bacterial community structure associated to it.

1 Introduction

The Arthropoda phylum represents approximately 80% of the extant animal species, being insects the most abundant class. Arthropods are components of multiple food chains; they consume and recycle debris, bacteria, fungus and algae, within the ecosystems they inhabit. When considering this vast taxonomic and ecologic diversity of food sources, it is not surprising that they evolved the ability to exploit the nutrients present in vertebrate blood. The occurrence of hematophagy (the habit of feeding upon blood) has emerged and evolved independently at least 21 times in different arthropod taxa. Such a strategy was certainly exploited promptly by viruses, protozoa, and helminths as means to increase their ¨mobility¨ and to find novel vertebrate hosts (Marquardt, 2005, Becker *et al*., 2010)

1.1 The anopheline mosquitoes

Mosquitoes (Diptera: Culicidae) are an ancient and monophyletic group composed of at least 4,500 species. The origin of this insect class predates the Jurassic period (Marquardt, 2005). The first basal split of the ancestral mosquito lineage gave rise to two deeply diverged subfamilies, *Culicinae* (containing *Aedes aegypti* and *Culex pipiens quinquefasciatus*) and *Anophelinae* (containing *Anopheles gambiae*), an estimated 145-200 million years ago (MYA) (Krzywinski *et al*., 2006). Although subfamily *Culicinae* contains important vectors of arboviruses and filarial worms, only subfamily *Anophelinae* contains vectors of human malaria (Besansky, 2008).

The *Anophelinae* subfamily (~500 species) probably arose in the neotropics (Krzywinski *et al*., 2001). It is divided into three genera, of which the largest is by far the *Anopheles* genus (484 species), which contains the major human malaria vectors. In turn, *Anopheles* consists of six subgenera whose relationships are largely uncertain due to their relatively ancient and rapid divergence (Krzywinski *et al*., 2001).

The divergence relationships and radiation of anophelines throughout the known world probably started with geologic events associated with the breakup of Pangaea, about 150 MYA. The most studied subgenera are *Anopheles* (189 species) and Old World *Cellia* (239 species), sister taxa that together contain most of the important malaria vectors. Interestingly, though malaria exerts a considerable epidemiologic burden upon Latin American countries (WHO, 2014), Neotropical vectors from the *Nyssorhynchus* subgenus are lagging behind sibling anophelines when regarding biologic and genomic knowledge.

Anophelines are holometabolous as all the other insects from the endopterygote superorder. This means they undergo complete metamorphosis, transitioning from aquatic to terrestrial environments (Becker *et al*., 2010). They have four developmental stages in their lifespan: egg, larvae, pupae and adult (**Figure 1**). Briefly, the larval stage is a period of active feeding and growth; the pupal stage is a period of reconstruction in which larval tissues are histolyzed and rebuilt according to the adult body plan; the adult stage is a period of dispersal and reproduction (Marquardt, 2005, Becker *et al*., 2010).

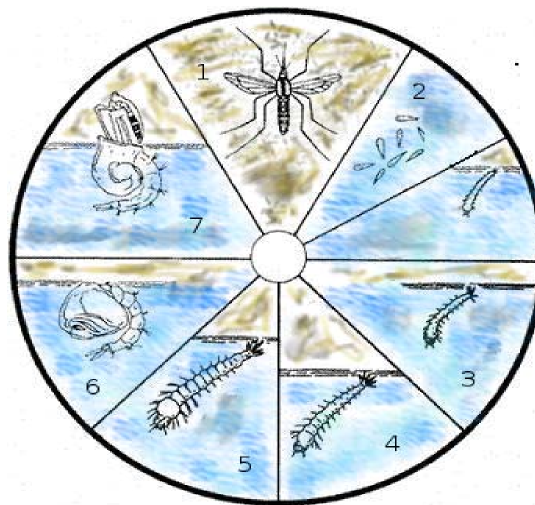


Figure 1 Life cycle of the anopheline mosquitoes. Throughout their holometabolous lifespan anophelines undergo the following stages: adult (1), egg (2), larval (3-5) and pupae (6). Modified from:
http://dvm5.blogspot.com.br/2010/10/family-culicidaemosquitoestheoryentomol.htm -May, 2013.

At night, females lay their eggs in batches of 70 to 100 on the surface of water. The type of water where the eggs are laid (larval ecological niche) can be considered an indicative trait of the mosquito species. Such selection may depend on cues such as bacteria-derived chemicals (Lindh *et al*., 2008).

The anopheline larvae lie just below the surface of the water and generally feed on bacteria and algae, turning into pupae after 7 to14 days during a five-minute process (Moll *et al*., 2001, Gonzalez-Ceron *et al*., 2003, Becker *et al*., 2010). Pupae are comma-shaped, and are the least active stage of the *Anopheles* lifecycle. After two to four days the pupa metamorphoses into an adult mosquito. The adults emerge during late evening and are able to fly within minutes (Marquardt, 2005, Becker *et al*., 2010).

In majority of anopheline species, adults are active during the night, starting at dusk. They usually mate during flight. The male is attracted to the female by the tone of her wing beat, and has antennae that act as sound receptors. Once mated, the female searches out a blood meal, following sensory cues such as host odor, carbon dioxide and convection currents. She then seeks out a resting place, which may be indoors or outdoors depending on the species (Marquardt, 2005, Becker *et al*., 2010).

1.2 Arthropods acting as natural pathogen vectors

Mosquitoes have been described as the most dangerous animal in the world. Many pathogens perpetuate in nature through transmission cycles that involve vertebrate hosts and hematophagous arthropod vectors. The vector-borne disease cycle comprises a dynamic interaction between the members of the cycle and the environment where transmission occurs. Pathogens involved in this cycles need to infect and replicate (and / or develop) in both vertebrate and invertebrate hosts. When feeding on blood to support vitellogenesis, or to supply nutrients to metabolic pathways, the vector may ingest and be infected by pathogens in the vertebrate blood or skin. In subsequent blood meals, the vector may transmit the pathogen to other susceptible vertebrates. Commonly, the pathogen exerts little or no deleterious effects upon its compatible arthropod host, while infection of vertebrate hosts (especially tangential hosts) may result in significant morbidity and / or mortality. At first glance, vector-borne cycles appear to be an inefficient mode of transmission and maintenance for pathogens; however, the multiple and diverse pathogens that rely on such strategy suggests it is in the long run, and evolutionary success (Marquardt, 2005, Becker *et al*., 2010).

Vectorial transmission of pathogens involves, in most cases, hematophagous female arthropods like mosquitoes and sand flies. Nevertheless it may involve male insects such as ticks. The mechanism through which arthropods transmit pathogens can be biological or mechanical. Biological transmission involves the reproduction or development of the infectious agent within the insect, before it is transmitted to the next vertebrate host. In contrast, for mechanical transmission the pathogen is merely inoculated by the arthropod in a physical contact manner (most often involving mouthparts of the insect) while attempting to feed. Protozoans, like the malaria parasite *Plasmodium*, are transmitted in a cyclopropagative mode involving the transition from gametes to infective sporozoites as explained further below. This strategy implies that, in the next blood meal, the vector will transmit more pathogenic units than the amount it originally ingested (Marquardt, 2005). Nevertheless, in the case of malaria parasites and their anopheline vectors, this cyclopropagative transmission mode may not be that efficient, or successful, since *Plasmodium* parasite numbers undergo a

severe bottleneck during its development in the mosquito gut (Wang & Jacobs-Lorena, 2013).

The success of the malaria cyclopropagative cycle, and the ulterior transmission to other vertebrate host depends on the compatibility of the *Anopheles-Plasmodium* duo (Molina-Cruz & Barillas-Mury, 2014). This compatibility may be determined by multiple factors that range from genetics and insect innate immunity (Dong *et al*., 2006, Jaramillo-Gutierrez *et al*., 2009), to the effect exerted by the microbial symbionts within the mosquito (Minard *et al*., 2013)

From the 467 formally recognized species approximately 70 have the capacity to transmit human malaria parasites (Service & Townson, 2002) being 41 considered as dominant vector species (DVS) capable of transmitting malaria at a level of major concern to public health (Hay *et al*., 2010). The distribution of this anophelines species across the world can be seen in **Figure 2**.

1.3 Human malaria and Neotropical anopheline vectors

Human malaria is a malady that threatens an estimated of 3.2 billion people in 97 countries with ongoing transmission (WHO, 2014). It is caused by *Plasmodium* parasites, which in nature need an anopheline mosquito to complete their life cycle and be transmitted to a human host (**Figure 3**). Even though its mortality burden worldwide may be underestimated as reported Murray *et al*. (2012) (refer to **Figure 4**), the current World Malaria Report (WHO, 2014) registered 198 million reported cases and 584,000 deaths in 2013. In the Americas, 119 million people are at risk in 21 affected countries, with an estimate of 700,000 cases and 800 deaths in 2013 (WHO 2014). Brazil has the largest incidence of malaria in the Americas, with 41% of the cases (WHO 2012). Out of the 267,000 cases reported in Brazil in 2011, 84% were caused by *Plasmodium vivax* (Ministerio da saúde SVS, 2013).

The etiological agents of human malaria are apicomplexans of the Plasmodiidae family. All share numerous characteristics, including asexual reproduction in a vertebrate host and sexual reproduction in a definitive host (a mosquito, in the case of the *Plasmodium* species that infect all mammals, including humans). The family contains about twelve genera, of which one is *Plasmodium*, which itself is divided into numerous subgenera, and into hundreds of different species, of which five infect humans (*Plasmodium falciparum*, *P. vivax*, *Plasmodium ovale*, *Plasmodium malariae and Plasmodium knowlesi*) (Consoli & Oliveira, 1994, WHO, 2014).

While feeding upon an infected vertebrate host, one mosquito will ingest, on average, 10^3 gametocytes in an infected blood meal. Within minutes after the infective blood meal, and inside the lumen of the midgut, these gametocytes will go through maturation, generating

micro-, and macrogametocytes that will be fertilized producing a diploid zygote (Sinden 1999). The mature zygote then differentiates into the mobile form of the parasite known as the ookinete, in a process which can take up to 16-24 hours, depending on the *Plasmodium* species (Ghosh *et al*., 2000, Dinglasan *et al*., 2009).

This above mentioned process involves the exflagellation of the gametocytes in the mosquito's midgut. Exflagellation occurs mainly due to the effect of fluctuations in temperature, pH, and the production of xanturenic acid by the mosquito (Billker *et al*. 1998). The zygote will eventually differentiate into an ookinete (Abraham & Jacobs-Lorena, 2004). This mobile form of the parasite will move and penetrate the peritrophic matrix and pass over the intestinal epithelium before transforming into the oocyst. About ~14 days later, the oocyst ruptures, releasing thousands of sporozoites into the hemocoel. The sporozoites are only capable of invading the salivary gland (Abraham & Jacobs-Lorena, 2004).

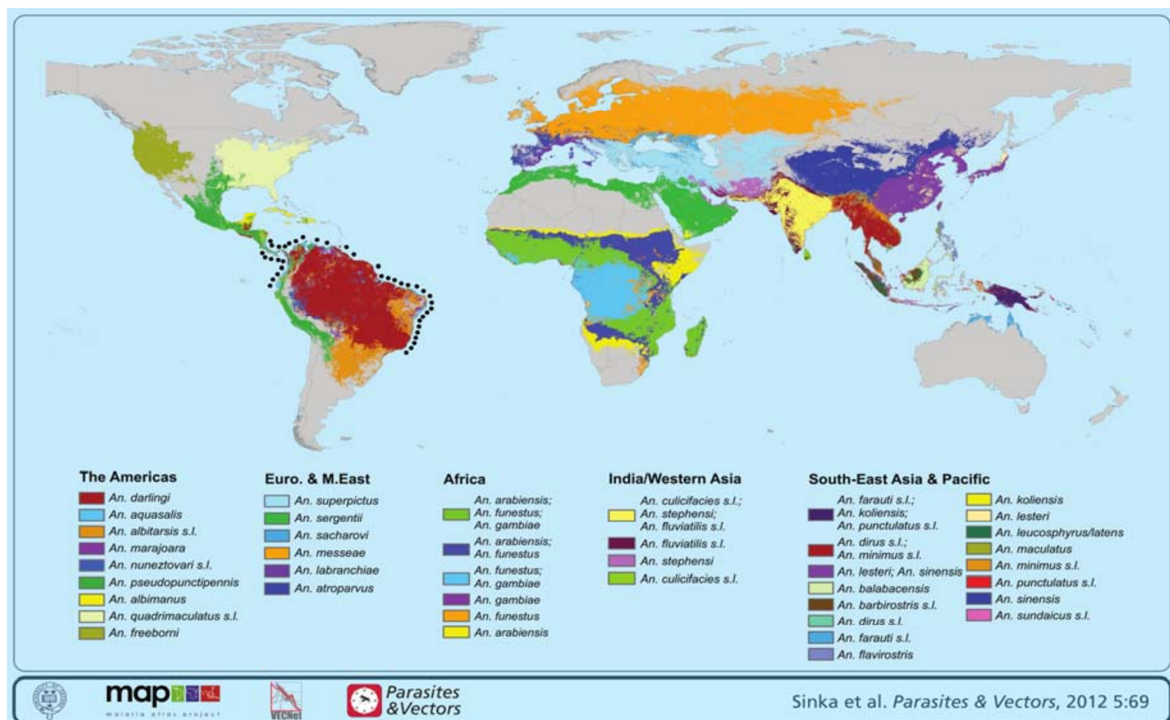


Figure 2 Global distributions of the dominant *Anopheles* vectors of human malaria. Figure illustrating (with black dots) the coastal distribution of *A. aquasalis*. The color legend shows the geographical distributions of the dominant human malaria vectors per continental blocks. Figure modified from: http://www.map.ox.ac.uk/browse-resources/ - April, 2013).

The invasion of the salivary gland by sporozoites is very inefficient; with usually less than 20% of the total number of parasites produced being capable to invade the organ (Hillyer *et al*. 2007). By means of a specific recognition receptor present in the salivary gland of the

anopheline vector, these parasites are able to adhere and penetrate the basal lamina of the gland (Sinden & Matuschewski, 2005). After invasion, the sporozoite will remain stored in this cavity until a new blood meal is ingested, when a small number of sporozoites can enter the secretory duct and can be released into the vertebrate host along with the saliva (Pimenta *et al*. 1994). When the mosquito bites another vertebrate host, transmission is completed by release of sporozoites from the salivary glands (Wang & Jacobs-Lorena, 2013).

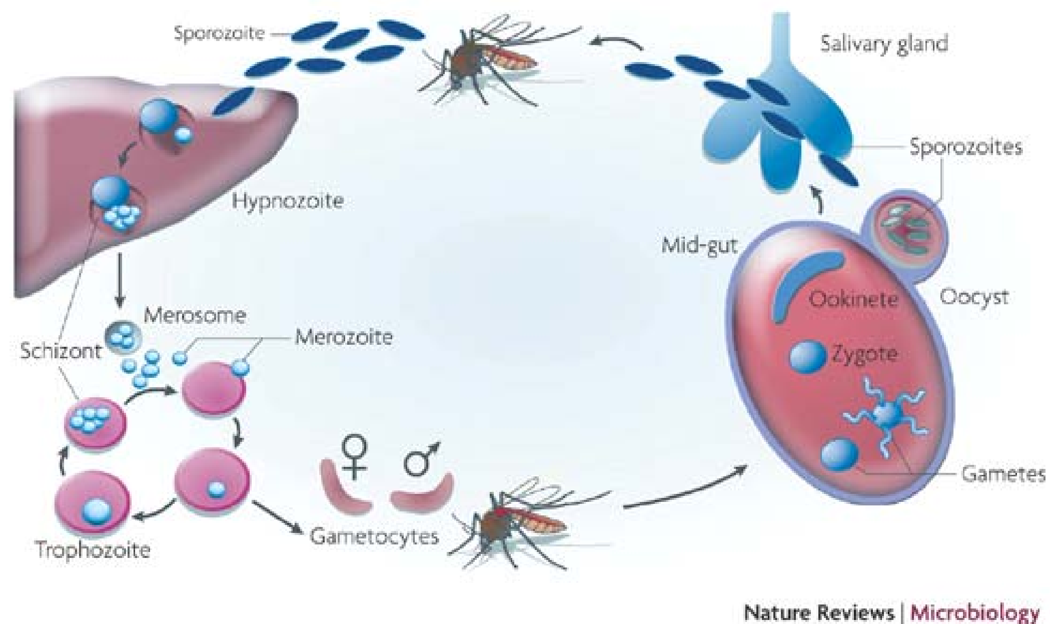


Figure 3 The malaria parasite cycle within the mosquito vector and the vertebrate host. During the blood meal, the mosquito bite releases hundreds of sporozoites into the vertebrate host's bloodstream, which then migrate to the liver and form parasitophorous vacuoles in hepatocytes. Depending on the *Plasmodium* species, the sporozoites can either remain dormant as hypnozoites or develop into thousands of merozoites. Budding of parasite vesicles, called merosomes, releases merozoites, which invade erythrocytes and then replicate. Some parasites differentiate into male or female gametocytes, which are taken up by mosquitoes through a blood meal. Once inside the mosquito, gametocytes undergo rapid transition into activated gametes, which migrate out of the mid-gut wall, thereby forming the oocyst. Meiosis leads to the formation of sporozoites which, following rupture of the oocyst, migrate into the salivary gland, ready to begin the cycle again (Pain & Hertz-Fowler, 2009).

The predominant form of malaria in Africa has *P. falciparum* as etiological agent, whereas in Asia and the Americas it is more common to find *P. vivax* malaria, which produces fewer fatalities but can be severe. It is interesting to notice that *P. vivax* and *P. ovale* parasites,

once inside their vertebrate host, may remain dormant in the liver as the hypnozoite form for months or years, producing no outward manifestations of disease but resulting in relapses months or years after an individual has left a malaria endemic region. This has been observed and proposed as an adaptation to a temperate or subtropical climate where mosquitoes may not be present throughout the year (Pain & Hertz-Fowler, 2009).

When studying malaria transmission it is relevant to consider that not all *Anopheles* species are competent vectors of *Plasmodium* pathogens (Service & Townson, 2002). For example, amongst the 33 *Anopheles* species described in the Brazilian Amazon region, the main malaria vectors are *Anopheles darlingi, Anopheles albitarsis and Anopheles aquasalis* (WHO, 2014). In general terms, the success or failure of vector control policies can be intimately linked with the comprehension of *Anopheles* bionomics in the local malaria transmission context (The Malaria Atlas Project: http://www.map.ox.ac.uk/explore/mosquito-malaria-vectors/ - April 2013). Below, we describe the bionomics of an anopheline malaria vector currently acting as a laboratory model to study malaria transmission in Latin America, *A. aquasalis*.

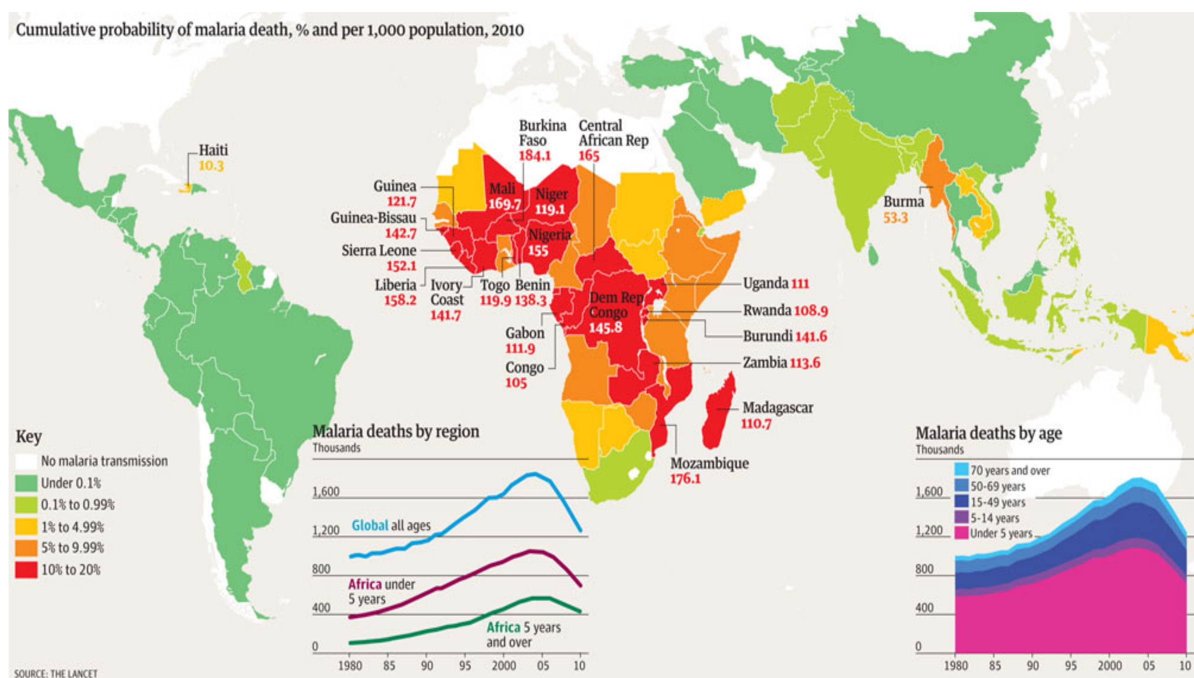


Figure 4 Human malaria global transmission map and mortality, according to Murray *et al.* (2012). World map illustrating malaria transmission and the amount of deaths per region, age and 1000 inhabitants.

1.4 *Anopheles aquasalis* bionomics

A. aquasalis is an opportunistic anopheline species that thrives mostly in sunny habitats with emerging vegetation. It adapted to both brackish and fresh water environments, though it is believed that *A. aquasalis* prefers clean water such as natural pools, streams, mangroves, ponds, and ditches (Manguin *et al*., 1993, Grillet, 2000). The demarcation of the *A. aquasalis* territory to coastal regions and its tolerance to salt water could be an evolutionary adaptation, consequence of a behavioral trait involving avoiding competition for food with other *Anopheles* during the larval phases. This weakness, pushed *A. aquasalis* towards coastal ecological niches inserting the mosquito in the large and varied marine trophic chain (Sinka *et al*., 2010).

As it is shown in **Figure 2**, *A. aquasalis* is a saltwater tolerant species distributed along the American Atlantic coastline from Nicaragua to São Paulo, Brazil; and at the Pacific coast from Costa Rica to Ecuador (Sinka *et al*., 2010, Zimmerman, 1992). It has also been reported in Venezuela, Guyana, the Antilles and Trinidad and Tobago (Berti *et al*., 1993, Laubach *et al*., 2001, Faran 1981, Chadee *et al*., 1992).

This species can penetrate 8 to 10 miles from the coast lines, since they have a flight capacity of up to 8 kilometers (Sinka *et al*., 2010). *A. aquasalis* is exophilic, zoophilic and crepuscular, but in the drier areas, it can be endophilic, biting human hosts. The females are opportunist, feeding both intra- and peri-domiciliary in animals or humans. They begin to bite during sunset, reaching maximum activity in the early evening, and then subsequently decreasing (Flores-Mendoza *et al*., 1996). Usually before and after the blood meal, the mosquitoes rest in peri-domestic habitats. Under this conditions and characteristics *A. aquasalis* is considered an important *P. vivax* vector in the Americas. In situations where the mosquito density becomes high, females can be vectors of human malaria, especially in the absence of domestic animals, which are its usual food source (Giglioli, 1963). This mosquito species has been associated with several outbreaks of malaria in multiple Latin countries (Berti *et al*., 1993, Deane, 1986, Laubach *et al*., 2001, Mouchet *et al*., 2008). As mentioned before, human malaria is still a public health problem and a pending endeavor to tackle by research groups throughout malaria burdened countries, relying on local transmission models and knowledge pertaining to Neotropical vectors.

1.5 Malaria transmission laboratory models

Model anopheline vectors reared under laboratory conditions enable studies on their biology and behavior aiming to characterize details of their susceptibility to the *Plasmodium* species, thus providing a greater understanding of malaria disease dynamics. To the best of

our knowledge and according to reviewed literature related to *Anopheles* species, currently there are only two colonized American human malaria vector species routinely maintained under laboratory colonies for experimental transmission studies: *A. aquasalis* and *Anopheles albimanus* (Pimenta *et al*., 2015). In the case of *A. aquasalis*, several studies have been performed an published by our group regarding the susceptibility this species presents towards *P. vivax* infections and the immune response elicited (Bahia *et al*., 2010, Bahia *et al*., 2011, Bahia *et al*., 2013, Rios-Velasquez *et al*., 2013). However, the recent report of adaptation of *A. darlingi* to laboratory conditions (Villareal *et al*., 2013) could be a major breakthrough for the understanding of the biology of *Plasmodium* transmission by mosquito vectors in Brazil and Latin America (Molina-Cruz & Barillas-Mury, 2014). One of the trending and promising approaches to improve malaria control strategies, and further understand vector-parasite dynamics, is the comprehension and intentional modification of the *Anopheles* microbiota (Minard *et al*., 2013). Nonetheless, not many studies have been produced on this topic referring Neotropical anophelines (Villegas &Pimenta, 2014).

1.6 On metagenomics and the *Anopheles* holobiont

The term ¨hologenome¨ is defined as the sum of the genetic information of the host and the symbiotic microorganisms it harbours (Rosenberg and Zilber-Rosenberg, 2011). The term ¨holobiont¨ (Greek, from holos, whole; bios, life; -ont, to be; whole unit of life) is applied when describing a long-term physical association between different living organisms (Margulis, 1993). Theoretically, this definition encompasses all symbiotic associations (along the mutualism–parasitism continuum) spanning all taxa. However, in most cases, the term holobiont is restricted to the host and its associated mutualistic symbionts. The hologenome theory of evolution considers that the holobiont is the unit under natural selection in evolution (Brucker & Bordenstein, 2013). Recently, Dheilly (2014) elegantly argued that new perspectives on the study of host–parasite interactions, such as *Plasmodium-Anopheles*-Humans, opened with evidence suggesting that all of the diverse microorganisms associated with the hosts and parasite play a part in the co-evolutionary history of diseases.

Amongst the metazoans, insects are by far the most diverse and abundant clade (Basset *et al*., 2012). Their success can be explained in part by the relationships they have established with beneficial members of their associated microbiome. The term microbiota defines the microbial communities that stably or transiently colonize insect epithelia as well as intracellular compartments and target organs. They may vary from bacteria to viruses, yeasts and protists. The bacterial component of this ecosystem is to date the most studied and characterized (Ng *et al*., 2011a, Gendrin & Christophides 2013, Minard *et al*., 2013). These

symbiotic microbiomes or consortia are beneficial to their insect hosts in many ways (Dillon & Dillon 2004; Azambuja *et al*., 2005; Thomas *et al*., 2012; Engel & Moran 2013), including: dietary supplementation, enhancement of digestive mechanisms, tolerance to environmental perturbations, protection from parasites (Degnan & Moran 2008) and pathogens (Nartey *et al*., 2013) maintenance and/or enhancement of host immune system homeostasis. Furthermore, the absence or elimination of the microbial fauna, and even the modification of its composition can reduce the fitness of the harboring insect (Thomas *et al*., 2012). This observed influence of the microbiome on its host has been referred to as the extended phenotype and can range from mutualism to parasitism as well.

Recently, the study of microorganisms living in insect guts has increased considerably. The last decade has seen the publication of multiple relevant studies ranging from diversity screening metagenomic surveys (Baumann 2005, Lindh *et al*., 2005, Carpi *et al*., 2011, Lindh & Lehane 2011, Ng *et al*., 2011a, Ng *et al*., 2011b, Chavshin *et al*., 2012, Djadid *et al*., 2011) to molecular studies on how gut bacteria interact with the host's immune system and response towards infection (Azambuja *et al*., 2005, Chouaia *et al*., 2010, Boissiere *et al*., 2012).

It is not within the scope of this thesis to review and provide an exhaustive analysis on metagenomics (as a computational tool), nor the architecture and dynamics of this micro-ecosystem within cullicine vectors. Recent revisions cover these topics substantially and creatively (Dillon & Dillon 2004, Engel & Moran 2013, Gendrin & Christophides 2013, Minard *et al*., 2013). Our aim is to briefly describe some of the recent advancements that malaria vector control has generated regarding microbiota and its association with vector competence traits. Many of them have been greatly enhanced by the use of metagenomic tools which have allowed us to discover and explore how microbial species could be used in paratrangenesis and malaria transmission blocking strategies.

Metagenomics emerged as a derivation of classic microbial genomics with the key difference being that it bypasses the requirement for obtaining pure cultures for sequencing. (Glass *et al*., 2010, Kim *et al*., 2013) We now have the ability to obtain genomic information directly from microbial communities in their natural habitats and study them in a concerted manner describing their species composition and even predict the potential genomic functions and metabolic capabilities they possess (Wooley *et al*., 2010, Williamson & Yooseph, 2012).

As Next Generation Sequencing (NGS) has skyrocketed, our potential to generate genomic data (Ansorge, 2009) benchmarking has gained relevance providing guidance to experimental biologists that encounter themselves with the myriad of bioinformatic tools available (Delcher *et al*., 2007, Huson *et al*., 2007, Meyer *et al*., 2008, Angly *et al*., 2009, Clemente *et al*., 2010, Glass *et al*., 2010, Gerlach & Stoye 2011, Jiang *et al*., 2012). As users

of such technology, we would like to stress that when designing experiments that encompass metagenomic data generation, it is imperative to consider points such as: sampling techniques, DNA/RNA extraction protocols, sequencing platforms, assembly, taxonomic binning, gene annotation tools, statistical analysis and data/meta-data sharing formats (Wommack *et al*., 2008, Tanenbaum *et al*., 2010, Wooley *et al*., 2010, Thomas *et al*., 2012). The availability of standardized procedures (Field *et al*., 2008, Tanenbaum *et al*., 2010) and platforms for data storage and sharing are becoming increasingly important to ensure that the output of individual projects can be assessed and compared (Thomas *et al*., 2012).

Metagenomic screening assays are now being used to determine the diversity of microorganisms and viruses residing in arthropod vectors of medical importance. Such assays allow human health agencies and research groups to monitor endemic infections, perform real-time surveillance of newly emerging zoonotic pathogens, discover etiological agents and unravel how they associate with and within their host (Bishop-Lilly *et al*., 2010, Carpi *et al*., 2011, Ng *et al*., 2011a, Ng *et al*., 2011b, Mokili *et al*., 2012).

Due to their importance as vectors of malaria, anopheline mosquitos have been the target of multiple efforts to profile their microbiota (Gendrin & Christophides 2013). Behind this efforts lies the knowledge that bacterial living in the midgut have been found to: modulate the response of the mosquitos towards *Plasmodium* infection (Pumpuni *et al*., 1993, Dong *et al*., 2009, Boissiere *et al*., 2012, Eappen *et al*., 2013); and have the potential to block infections and can be used as genetic transformation vehicles (Pumpuni *et al*., 1993, Dong *et al*., 2009, Weiss & Aksoy 2011, Boissiere *et al*., 2012, Eappen *et al*., 2013, Ricci *et al*., 2012). Below, we summarize some of the key findings regarding the impact of microbiota on the *Plasmodium*-Cullicidae interaction model.

Both, laboratory and field mosquito strains have been found to be associated with microbial organisms that colonize the gut in particular. They primarily consist of Gram-negative bacteria of the *Enterobacteriaceae* family. For example, field populations of *Anopheles gambiae* and *Anopheles funestus* were found to contain 16 bacteria species spanning14 genus (Lindh *et al*., 2005). Laboratory populations of *A. gambiae* and *Anopheles stephensi* presented as well a wide variety of bacteria, especially of the genus *Asaia*, *Enterobacter*, *Mycobacterium, Sphingomonas, Serratia* and *Chryseobacterium* (Favia *et al*., 2007, Dong *et al*., 2009). Bacteria of the *Asaia* genus were found also in *Aedes aegypti* mosquitoes (Pidiyar *et al*., 2004, Rani *et al*., 2009, Gaio *et al*., 2011). In addition, beyond the digestive tract, studies have shown that species of this genus are also able to colonize the salivary gland and ovaries of mosquitoes and are usually acquired through vertical transmission (Favia *et al*., 2007).

It has been shown how gut bacteria may have an impact on vectorial competence by inhibiting the sporogonic development of malaria parasites within the mosquito vector (Pumpuni *et al*., 1993, Pumpuni *et al*., 1996, Gonzalez-Ceron *et al*., 2003, Dong *et al*., 2009, Cirimotich *et al*., 2011). Pumpuni *et al*. (1993 & 1996) also showed, whilst manipulating the bacterial content, that Gram-negative bacteria inhibit the oocysts formation in whole or in part, and that the same action was not observed with Gram-positive bacteria.

Evidence of this influence of the intestinal microbiota in the life cycle of parasites has been demonstrated for other insects such as sand flies and tsetse flies (Schlein *et al*., 1985, Welburn & Maudlin 1999).

Recent studies suggest that the presence of Enterobacter species in the gut of *Anopheles arabiensis* originating from Zambia, act directly on *P. falciparum,* blocking the development of the parasite, making this population refractory to infection. This refractoriness was associated with the generation of reactive oxygen species (ROS) that interfere with the development of the parasite and kills it before its invasion in the intestinal epithelium (Cirimotich *et al*., 2011).

Previous studies suggest that bacteria in the gut lumen modify the intestinal environment and inhibit the development of parasites by action of the immune system by overexpression of immunity genes, culminating in an increased rate of production of antimicrobial peptides – AMPs- (Ratcliffe & Whitten 2004, Michel & Kafatos 2005). Such peptides are likely to have a key role not only in the control of pathogenic or symbiotic bacteria but also in the development of infections by parasites (Beard *et al*., 2001; Boulanger *et al*., 2004). Interestingly, the mosquito's immune system acts against bacterial growth and also eliminates a large number of parasites modulating the intensity of infection when mosquitoes infected with *Plasmodium berghei* or *P. falciparum* (Meister *et al*., 2009).

It has also been proposed that certain bacteria taxa can induce a reductive environment within the mosquito midgut, thus aiding in the detoxification of reactive oxygen and nitrogen species, a fact that would allow for an aggressive immune response of the mosquito when infected by the parasite (Wang *et al*., 2011).

As concluding remarks to this section, we would like to highlight the fact that vector biology has made great advancements the past few years and many results have been attained by "synergic" approaches with computational science as a key element. Many interesting theories are now being discussed and explored regarding the hologenomic basis of speciation (Nikoh *et al.*, 2008, Rosenberg & Zilber-Rosenberg 2011, Ni *et al*., 2012, Brucker & Bordenstein, 2013) and how bacteria and viruses may be shaping the genomes and phenotypes of harboring organisms (Gorski *et al*., 2003, Crochu *et al*., 2004, Degnan &

Moran 2008, Keeling & Palmer 2008, Klasson *et al*., 2009, Rohwer *et al*., 2009, Holmes 2011, Rosario & Breitbart 2011, Reyes *et al*., 2012, Stern *et al*., 2012, Horie *et al*., 2013, Husnik *et al*., 2013, Ioannidis *et al*., 2013, Seed *et al*., 2013).

When we take into consideration that there are 1 million bacteria and 10 million viral particles per milliliter of surface seawater (Suttle 2005, Ng *et al*., 2011b, Rosario & Breitbart 2011), maximizing the NGS sequencing data generated by an ongoing *A. aquasalis* genome project becomes an opportunity to explore many of these new avenues. This vast surrounding and potentially associated microcosms may have left its mark upon the coevolving larval stages of this species whilst developing in brackish waters.

1.7 Consideration of *Anopheles* genomes and those of New World vectors

The publication of the *A. gambiae sensu strict* (Holt et al. 2002) and the *P. falciparum* (Gardner et al. 2002) genomes, both in 2002, marked a breaking point in the field of malaria vector biology research. The *Anopheles* project wrapped together decades of classic genetics knowledge allowing us to better understand issues such as chromosome and gene architecture. It also allowed vector biologists to plunge into the area of comparative genomics through which the first comparisons made (Christophides *et al*., 2002, Zdobnov *et al*., 2002) addressed matters like the composition of the immunity-related gene repertoire. In the post-genome era, several genetic engineering tools and strategies for vector control have arisen, been implemented and assessed (Alphey *et al*., 2002, Lycett & Kafatos 2002, Takken & Knols 2009, Isaacs *et al*., 2011, Sumitani *et al*., 2013). Nevertheless, the high diversity and plasticity that *Plasmodium* parasites have shown in vertebrate and invertebrate hosts has led to the assumption that the parasites evolve faster, and adapt rapidly, more so than human and anopheline hosts (Carius *et al*., 2001, Cohuet *et al*., 2010). As a consequence of this phenomenon and with the experiences accumulated since, the vector biology community understood that sequencing the genomes of multiple mosquito and parasite species would be imperative to understand and manipulate the vector-parasite interactions.

For this purpose, efforts were joined and channeled via the *Anopheles* Genomes Cluster (AGC), which, in 2008, set the basis of what would become the first anopheline comparative genomics consortium (Besansky, 2008). The committee identified and selected 16 mosquito species whose genomes were recently published (Neafsey *et al*., 2015) and made available through the Vector Base (Megy *et al*., 2012). Unfortunately, *A. albimanus* is the only American vector listed in the project, with no attention paid to the Amazon mosquitoes that are vectors of the majority of the human cases in the continent.

The evolutionary vector-parasite dynamics, vectorial competence traits and mosquito

behavior could have been shaped by multiple factors such as specific genotype combinations. Experimental evidence and theories explaining how the genomic composition of a mosquito species determines whether it is refractory or susceptible towards infection by a species (strains) of *Plasmodium* parasite have been published (Billingsley & Sinden 1997, Norris *et al*., 2001, Osta *et al*., 2004, Lambrechts *et al*., 2005, Riehle *et al*., 2007, Harris *et al*., 2010).

There is also a great body of literature connecting vector biology with non-genetic components like ecological factors (Schmid-Hempel & Ebert 2003, Lambrechts *et al*., 2005, Tripet *et al*., 2008, Tripet 2009, Wolinska & King 2009).

As stated by the AGC (Besansky 2008, Moreno et al. 2010), sequencing the genome of mosquito species that capture and represent the evolutionary and phenotypic divergence within the anopheline vectors distributed throughout the world is critical. It is the consensus among the community that, to envision a eukaryote genome project, requires looking at it as continuum process of innovation, re-sequencing and annotation (Li *et al*., 2006, Sharakhova *et al*., 2007, Li *et al*., 2010a, Moreno *et al*., 2010). Together with the *An. gambiae* s.s. genome, other annotated anopheline assemblies will provide a platform for gaining genome-wide evolutionary and population genetic insights into the mechanisms of speciation, and the biological processes that influence the ability of mosquitoes to transmit malaria parasites to humans.

It has also been brought to the attention of the vector community that the genomic aspects of vectorial capacity and competence have not been uniformly studied (Cohuet *et al*., 2006, Cohuet *et al*., 2010) and some have been largely overlooked both in terms of the species analyzed and the gene families addressed by experimental biology. For example, rapid progression has been made regarding mosquito immunity, insecticide resistance and olfaction genetics. However, genetic determinants of parasite virulence, mosquito adaptation to human environments and the evolutionary forces exerted on vector by the parasite and the microbiome associated with them are still lagging behind. The area of comparative genomics is rapidly evolving and developing tools. Therefore, the number of questions that vector biology can answer through sequenced and published genomes has expanded (Zdobnov *et al*., 2002, Reddy *et al*., 2012).

Major analysis themes now include topics such as molecular evolution and speciation, chemoreception, circadian rhythm, development, repetitive and transposable elements, reproduction, secretomes, rearrangements of chromosomal architectures, neuropeptides and behavior, blood/sugar metabolism, etc.

The Neotropical vectors represent an interesting target to understand how competent malaria transmission evolved in a different ecological setting and following different human

settling conditions as well (Fagundes *et al*., 2008, Hubbe *et al*., 2010, O'Rourke & Raff 2010, Yalcindag *et al*., 2012) It is believed that the interactions between the actors of the malaria transmission triad, humans, Neotropical vectors and *Plasmodium* parasites, are relatively recent in the American continent. For example, the main Neotropical malaria vector, *A. (Nyssorhynchus) darlingi*, which diverged from *A. (Cellia) gambiae* approximately 100 MYA, could have evolved in a human- and parasite-free environment for several million years (Moreno *et al*., 2010).

When we add up all of the biological evidence and take into account that Malaria is a malady that still imposes a high burden upon the people that live in the Amazon basin - >500 thousand cases are reported on a yearly basis- (Molina-Cruz & Barillas-Mury, 2014), sequencing the genome of a Neotropical vector seems important. Thus, in 2013, this became a reality with the publication and upload onto Vector Base of the *A. darlingi* genome (Marinotti *et al*., 2013). This project was performed due to the interest of the Brazilian National Council for Research, and set a cornerstone for future basic and applied comparative genomics studies. Such research endeavors will be able to start answering long sought-after answers regarding the biology of malaria in an American context and will focus on generating genetic and chemical tools (e.g. insecticides, bacterial larvicides, paratrangenesis strategies) for vector control that better adjust the ecological and public health conditions in Latin America.

Nonetheless, currently, there is a practical void between the *A. darlingi* genome and its use in experimental biology research. The absence of laboratory colonies of this species in Brazil, and the highly heterozygous nature of its genome assembly may hinder its potential as a research model. *A. aquasalis* is a species that has viable and operating colonies throughout Brazil and it has been used in experimental infections and transmission assays with multiple *Plasmodium* species (Pimenta *et al*., 2015). Therefore, this species positions itself as a top malaria model for the understanding of malaria transmission within the Brazilian context.

Next Generation Sequencing technology (NGS) has evolved into an impressive tool that ranges from genome assembly to microbiome screening (Mardis, 2011). As mentioned before, the peculiar bionomics of the *A. aquasalis* mosquito (Sinka *et al*., 2010) has prompted us to expand and explore other ‘omic’ areas. In particular, the reconstruction of the associated consortium of bacteria and viruses that could be predicted from the massive parallel sequencing process is of interest.

As a final thought, we believe that tailored measures of vector control that respond to local conditions and transmission patterns are direly needed in our region. Targeted interventions based on the growing existence of (phylo)-genomic /immune response / microbiota data, pertaining to Neotropical vectors and indigenous *Plasmodium* parasites,

could prove to enhance the already existing control strategies. Building the capacity to generate and use comparative genomics data from local anopheline species is thus, a justified effort.

2 Objectives

2.1 General Objective

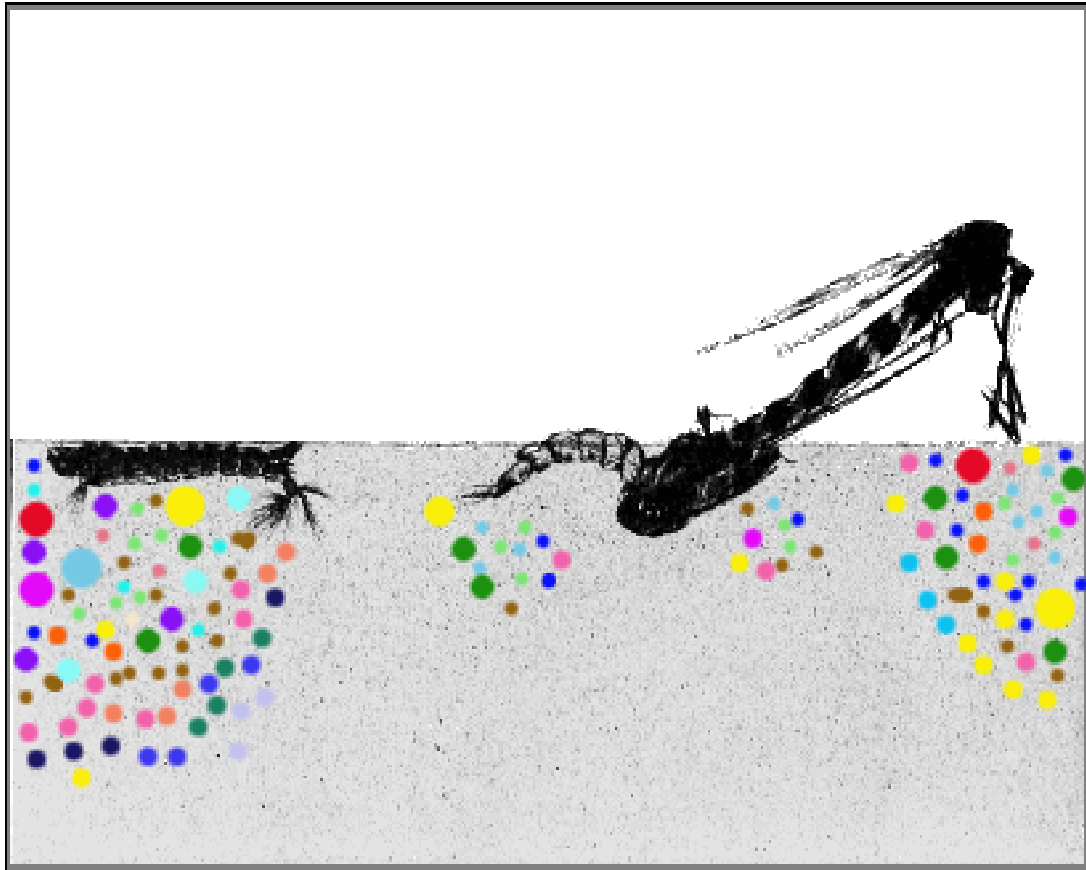
Perform a massive parallel NGS genomic characterization of the *A. aquasalis* holobiont, mainly focusing on its mitochondrial genome and bacterial associates.

2.2 Specific Objectives

a) To maximize the use of the massive parallel sequencing data by identifying the bacterial consortium associated to this anopheline.
b) To explore the (β-) diversity and composition of the *A. aquasalis* (laboratory reared pupae) associated bacterial community.
c) To seek the existence of the *Anopheles* genus ¨core microbiota¨.
d) To explore the potential use of β-diversity profile patterns to describe *Anopheles* biology.
e) To generate and assemble genomic sequences pertaining to the mitochondrial genome of *A. aquasalis*.
f) To perform a preliminary characterization (composition and phylogeny) of the mitochondrial genome of *A. aquasalis*.

Research questions

1) Is there an *Anopheles* ¨core microbiota¨?
2) Could the *A. aquasalis*-associated microbiota diversity reflect the evolutionary history of its host?
3) Can bacterial abundance profiles reveal ecological patterns representative of biological characteristics from their anopheline hosts?
4) When did *A. aquasalis* diverge from its Neotropical sibling species? Particularly, from the other brackish-water adapted species within the *Anopheles albitarsis* complex.



(Modified from Underwood, 1903)

3 Chapter 1:

Exploring biological patterns in *Anopheles* associated microbiota: Is there a core bacterial assembly shared between African and American anophelines?

Exploring biological patterns in *Anopheles* associated microbiota: Is there a core bacterial assembly shared between African and American anophelines?

Abstract

Nowadays, anophelines are being studied from a microbial ecology perspective to understand the impact of their associated microbiota on vectorial competence. Here we present the predicted bacterial community associated to *Anopheles aquasalis*, a saltwater tolerant malaria vector. We searched for bacterial taxa shared between members of the *Anopheles* genus, and explored biological patterns based on bacterial profiles.

Shotgun metagenomic sequencing and the MG-RAST suite were used to survey the bacteria associated to laboratory reared *A. aquasalis* pupae. The identified operational taxonomic units (OTUs) were added to a previously published index of *Anopheles*-associated bacteria. The updated index was then filtered by a geographic criterion to identify the shared genera between African and American anophelines. In addition, the *A. aquasalis* bacterial profile was used to perform explorative ß-diversity comparisons against published data from aquatic stages of *A. gambiae* (semi-natural and laboratory reared samples). Two relative abundance data matrices were built. One including all the family rank OTUs represented in the compared groups, and a reduced set encompassing data from the f-OTUs shared between African and American anophelines. Multivariate ordination and non-parametric statistical analyses were then applied to the matrices.

The bacterial consortium is composed of 74 genera and contains marine and bioluminescent bacteria. From the index, 14 f-OTUs were identified as shared between African and American anophelines. Within the composition of both abundance matrices, we found a significant association (NPMANOVA $p < 0.05$) between the bacterial community composition and the aquatic environment in which each compared anopheline group was reared. The presence of brackish water-related bacteria, not described previously in African or Asian anophelines, supports the idea that the larval niche influences the diversity of the microbial community. The meta-analysis performed on current data regarding *Anopheles*-associated bacteria suggests that a core group of bacterial families could exist when considering the continental distribution of the anophelines. Nevertheless, more extensive and homogeneous profiling studies are required to identify this "core" group. Bacterial abundance profiles may reflect the effects that biotic and abiotic factors have upon the bacterial community composition. Understanding the interacting mosquito-microbiota duo could further open new avenues to develop novel malaria control strategies.

3.1 Introduction

Human malaria is a malady that threatens an estimated of 3.2 billion people in 97 countries with ongoing transmission (WHO 2014). It is caused by *Plasmodium* parasites, which in nature need an anopheline mosquito to complete their life cycle and be transmitted to a human host. However, not all *Anopheles* species are competent vectors of *Plasmodium* pathogens (Service & Townson, 2002). Amongst the 33 *Anopheles* species described in the Brazilian Amazon region, the main malaria vectors are *A. darlingi, A. albitarsis and A. aquasalis* (WHO, 2014). *A. aquasalis* is a saltwater tolerant species distributed along the American Atlantic coastline from Nicaragua to São Paulo, Brazil; and at the Pacific coast from Costa Rica to Ecuador (Sinka *et al*., 2010, Zimmerman, 1992). The species evolutionary adaptation to saltwater breeding sites could have originated as a mechanism to avoid competition for food with other anophelines during the larval phases (Sinka *et al*., 2010). Thus, while evolving in a coastal ecological niche, this Neotropical anopheline may have interacted closely with the vast marine trophic chain.

Ecological, physiological and genetic determinants with influence over vectorial competence have been described and include species-specific evolutionary histories, mosquito immune system structure, behavior and life history (Sinden *et al*., 2004, Mitri & Vernick 2013, Lefèvre *et al*., 2013). Additionally, non-genetic factors may as well affect vectorial competence by modulating mosquito immunity and inhibiting *Plasmodium* sporogonic development (Lefèvre *et al*., 2013). Bacterial members of the anopheline inner-ecosystem have been found to act as effectors upon the aforementioned biological processes (Dimopoulos *et al*., 1997, Dong *et al*., 2009, Eappen *et al*., 2013, Pumpuni *et al*., 1993 & 1996; Gonzalez-Ceron *et al*., 2003, Cirimotich *et al*., 2011).

Anophelines exhibit a niche-specific global distribution (Sinka *et al*., 2012). As a genus, they were driven to adapt through geographic and evolutive time to their particular niche. Throughout that process, preferential larval site selection by ovipositing females probably exposed their progeny to particular microbial communities and perhaps even different immune pressures (Mitri & Vernick 2012, Lefèvre *et al*., 2013). Studies have shown that the aquatic environment in which the immature mosquito life stages develop partly shapes the adult associated bacterial community (Duguma *et al*., 2013, Dada *et al*., 2014, Wang *et al*., 2011, Boissière *et al*., 2012). Bacterial community diversity and abundance throughout the mosquito's life are dynamic and respond to multiple factors such as feeding, blood digestion and molting (Minard *et al*., 2013, Wang *et al*., 2011, Dharne *et al*., 2006). The

transition from pupae to adult represents a turning point regarding bacterial abundance and diversity (Moll *et al*., 2001, Wang *et al*., 2011), since the mosquito midgut is partially sterilized during metamorphosis (Moll *et al*., 2001). Nevertheless, this incomplete sterilization process would allow some direct transmission of bacterial OTUs from pupae to adults and thus bacterial taxa acquired from the breeding site would be transstadially transmitted (Lindh *et al*., 2008, Moll *et al*., 2001). Moreover, shortly after emerging, adults ingest water from their breeding site re-acquiring some of the previously egested bacteria (Lindh *et al*., 2008). The dynamics of this interacting holobiont through the mosquito's life stages going from an aquatic to a terrestrial environment require further studies.

Though the bacterial community structure shifts in composition and diversity throughout the host's life history, stable associations between bacterial symbionts and their anopheline host have been found and hinted as a potential ¨core microbial consortium¨ (Wang *et al*., 2011, Boissière *et al*., 2012). Identifying steadily *Anopheles*-associated abundant bacterial OTUs would enhance paratransgenesis based control strategies (Coutinho-Abreu *et al*., 2010, Wang & Jacobs Lorena, 2013). Candidate taxa such as *Asaia*, *Pantoea* and *Serratia* have generated interesting results on this topic (Capone *et al*., 2013, Damiani *et al*., 2010, Favia *et al*., 2007, Djadid *et al*., 2011, Bisi and Lampe, 2011, Wang et al., 2012, Wang & Jacobs Lorena, 2013, Bando *et al*., 2013).

We recently compiled data regarding *Anopheles*-associated microbiota (Villegas & Pimenta, 2014). The publication had two main goals. First, it reviews current knowledge regarding paratrangenesis strategies to control malaria transmission. Secondly, the meta-analysis performed, showed how our current views on the *Anopheles* holobiont mostly represents African and Asian species whereas knowledge regarding the microbiota associated to Neotropical anophelines is scarce. No bacterial taxonomic profile had been reported for a Neotropical human malaria vector using Next Generation Sequencing (NGS). [Refer to **Appendix 2** to read the publication.]

Here we present the predicted bacterial community associated to *A. aquasalis* pupae. With the generated data we updated the previous meta-analysis, searched for *A. aquasalis* exclusive symbionts, and identified a potential "core" group of bacterial OTUs shared among the *Anopheles* genus, according to the continent in which they are distributed. Additionally, with the predicted *A. aquasalis* microbial abundance profile and published data from *A. gambiae* aquatic stages (Wang *et al*., 2011) we performed explorative β-diversity analyses.

Our comprehension of the inner ecosystem of anophelines is still limited. As more data is available from other anopheline species our knowledge regarding the biological role of the bacterial community associated to the *Anopheles* genus will expand.

3.2 Methods

3.2.1 Mosquito rearing

A. aquasalis were obtained from a colony established at the Medical Entomology Laboratory at Centro de Pesquisas René Rachou-FIOCRUZ (Fiocruz, Minas Gerais). The mosquitoes came from a colony established in 1995 in Rio de Janeiro (Arruda *et al*., 1982, Silva *et al*., 2006, Carvalho *et al*., 2002), and are kept at laboratory conditions as described recently by Rios-Velásquez *et al*. (2013) and Costa-da-Silva *et al*. (2014).

3.2.2 Sample collection

Pupae (130 pupae ≈ 12-15 hrs old) were collected and washed, as follows, in order to detach any bacteria and biofilm from the exoskeleton surface and crevices: they were transferred to a small container with sterile molecular grade water for 15 minutes and then sequential 2 min rinses with chloride (10% v/v) and phosphate buffered saline (PBS) were performed. Finally they were washed with 70% ethanol and allowed to air dry on top of a sterile mesh and kimwipe paper before DNA extraction.

3.2.3 DNA extraction

Genomic DNA was isolated from 130 *A. aquasalis* pupae using the method of Bradfield and Wyatt (1984) (Bradfield & Wyatt, 1983) with slight modifications. Briefly, surface cleaned pupae were transferred onto a frozen mortar (frozen by pouring liquid nitrogen into it and allowing it to evaporate). The frozen pupae were then triturated to a fine powder and dissolved in 15 ml of 0.5% SDS, 100 mM EDTA pH 8.0, 200 μg/ml proteinase K (Sigma-Aldrich, St. Louis, MO, USA) and incubated at 55 ^{o}C overnight. The sample was extracted twice with equal volumes (15ml) of phenol + chloroform / isoamyl alcohol 24:1 and the supernatant removed with a blunt end pipette tip and transferred to a 50 mL tube. Two volumes of ice-cold 95% ethanol were added and the sample centrifugated at 4700 rpm for 5 min at 15 ˚C to precipitate the sample. After decantation of the supernantant, the DNA pellet was rinsed briefly with 75% ethanol and resuspended in 15 ml of 10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 20 mM NaCl by heating to 65 ˚C. DNAse-free RNAse (Roche, Pleasanton, CA, USA) was added and incubated according to manufacturer’s instructions. Afterwards, SDS to a final concentration of 0.5% and 100 μg/ml proteinase K were added, followed by incubation at 55 ˚C for 1 hour. The sample was extracted twice as described above, ethanol precipitated

and washed with 75% ethanol. The DNA pellet was resuspended in 500 µl of TE buffer.

3.2.4 Ion Torrent PGM Library preparation and sequencing

The genomic DNA sample was subjected to Next Generation Sequencing (NGS) based on ion semiconductor technology using the Ion Torrent PGM system as per the manufacturer's instructions. Two Ion Torrent adapter-ligated libraries were built according to the Ion Xpress Plus Fragment Library Kit protocol (Life Technologies, Carlsbad, CA, USA). Briefly, 500 ng of genomic DNA were enzimatically fragmented, end-repaired and ligated to the Ion Torrent library adapters. Following AMPure bead purification (Beckman Coulter, Brea, CA, USA), the adapter-ligated products were nick-translated and PCR-amplified for a total of 10 cycles. The resulting libraries were purified using AMPure beads and subjected to emulsion PCR using the Ion Xpress Template 200 Kit (Life Technologies, CA, Carlsbad, USA) according to the manufacturer's instructions. Next, ISP's were recovered and template-positive ISP's enriched using Dynabeads MyOne Streptavidin C1 beads (Life Technologies, Carlsbad, CA, USA). ISP enrichment was confirmed using the Qubit 2.0 fluorometer (Life Technologies, Carlsbad, CA, USA). The sample was prepared for sequencing using the Ion Sequencing Kit protocol (Life Technologies, Carlsbad, CA, USA). The complete sample was loaded onto an Ion 316 chip and sequenced on the PGM system for 110 cycles.

After sequencing, the individual sequence reads from each library were filtered by the PGM proprietary software to remove low quality and polyclonal sequences. Sequence reads matching the PGM 3′ adaptor were trimmed automatically. The quality-approved, trimmed and filtered sequences from each of the libraries were then uploaded to the MG-RAST v3.3.6 open-source server (Rapid Annotation using Subsystems Technology for Metagenomes, http://metagenomics.nmpdr.org) (Meyer *et al*., 2008) for downstream analyses regarding taxonomic profiling of the microbial consortium associated to *A. aquasalis* laboratory reared pupae.

3.2.5 Shotgun metagenomic analysis.

The uploaded Ion Torrent reads were further filtered using the MG-RAST QC pipelines (Meyer *et al*., 2008). The quality filtering was performed using the default parameters as recommended in the MG-RAST v3.3.6 manual (ftp://ftp.metagenomics.anl.gov/data/manual/mg-rast-manual.pdf) and it involved the removal of replicated reads, filtering out reads with ambiguous bases and screening against a model host organism genome.

The taxonomic profiling of the *A. aquasalis* holobiont was conducted following the

recommendations to perform shotgun metagenomic analyses with the MG-RAST suite described by Glass *et al*. (2010). Briefly, for taxonomic analyses of the small subunit (SSU) and large-subunit (LSU) rRNA reads, the metagenomic sequences were compared against the M5RNA database available through MG-RAST using a maximum e-value of 1E-5, a minimum identity of 97% and a minimum alignment length of 50 bp. The relative abundance of the predicted bacterial taxa (at the family rank) was computed and visualized on a histogram using Excel. A phylogenetic tree of the predicted bacterial genera associated to the sample was generated using iTOL online tool v2.1: http://itol.embl.de/ (Letunic & Bork, 2011). All the analyzed unassembled sequences are available (upon request) at the MG-RAST suite under the ID numbers 4523102.3 and 4526005.3.

3.2.6 Compilation of *Anopheles*-associated microbiota: searching for a geographically determined *Anopheles* core microbiota.

Recently, we conducted a thorough revision and enrichment of the microbiota index published by Minard *et al*. (2013) (Villegas & Pimenta, 2014). The compiled data represents the bacterial operational taxonomic units (OTUs) reported as associated with multiple anophelines. The predicted bacterial genera associated to *A. aquasalis* pupae were included in the *Anopheles* Microbiota Index (**Additional file 1**).

Briefly, the genus rank OTUs (g-OTUs) included in the index were selected based on the criterion that they accounted exclusively for bacteria identified as associated with anopheline mosquitoes collected on the field or laboratory reared in the continent in which they act as malaria vectors. We then grouped the Index data according to the continent in which each *Anopheles* species is distributed. This “by-continent” classified index represents the microbiota reported for anophelines from Africa, Asia, Europe and America (**Additional file 1**, spreadsheet: Microbiota Index classified by Continent).

To identify and quantify the shared and exclusive bacterial genera between each continent group, a Venn diagram was generated with the online tool Venny: http://bioinfogp.cnb.csic.es/tools/venny/ (Oliveros, 2007). The observed distribution of the *Anopheles*-associated bacterial genera per continent group allowed us to identify the g-OTUs shared between African and American anophelines, which we later used for ecological analyses.

Complementarily, to visualize the correlation between continent groups and bacterial g-OTUs, the online tool Circos: http://mkweb.bcgsc.ca/tableviewer/visualize/ (Krzywinski *et al*., 2009) was used to plot a subset of the index representing the bacterial genera shared at

least by two continent groups (**Additional file 3**).

3.2.7 Ecological diversity analysis

Community diversity analyses were performed to explore for potential biological patterns explained by bacterial abundance signature profiles when comparing samples representing American (*A. aquasalis*) and African (*A. gambiae*) moquitoes.

A recent publication by Wang *et al.* (2011) regarding the dynamic gut microbiota across life stages of *A. gambiae* provided us the relative abundance of bacterial taxa associated to aquatic stages of this anopheline species. We retrieved the abundance of family rank OTUs (f-OTUs) found in *A. gambiae* pupae and larvae reared both in laboratory conditions and in a semi-natural microcosmos (tagged as "Kenya").

To perform ecological community comparisons based on β-diversity measures, two matrices were built encompassing abundance data from 5 groups: *A. aquasalis* laboratory reared pupae (from the present study), *A. gambiae* laboratory reared pupae, *A. gambiae* Kenya pupae, *A. gambiae* laboratory larvae and *A. gambiae* Kenya larvae (data obtained from table S9 in Wang *et al.* (2011)). The first matrix contains the relative abundance of all the bacterial families identified in each of the five groups (**Additional file 2**). With the Venn diagram analysis we identified the bacterial genera shared between the African and American anophelines included in the index. Thus, for the second matrix, we used a reduced data set with the relative abundance of these "core" bacterial taxa at the family rank (**Table S3**).

The meta-analysis performed to build the bacterial index relied on multiple taxonomic surveys generated with multiple detection techniques. It is accepted that methods of detection and taxonomic classification strategies pose challenges and induce biases that may impact their detection efficiency, sensitivity and profiling accuracy; particularly at the species and genus ranks (Liu *et al*., 2011, Prakash & Taylor, 2012, Sharma *et al*., 2012, Sharpton, 2014). Thus we opted to work at the taxonomic family rank, encompassing diversity and composition data from a more conservative approach.

To determine if an underlying biological pattern could be observed within the full and reduced data sets, the variation in community structure was explored by multivariate measures based on pairwise distances between the five defined sample units. A variance stabilizing transformation was applied to both relative abundance matrices to normalize the data sets (Ramette 2007, Mason *et al*., 2013). The relative abundances of the f-OTUs in each compared group were expressed as the arcsin ($\sqrt{x}$)**.**

Exploratory multivariate analyses were performed on both of the transformed data sets of bacterial abundance profiles following suggested parameters and settings for microbial

ecology analyses (Anderson *et al*., 2011; Ramette, 2007, Legendre & De Caceres, 2013). To visualize the sample dissimilarities, hierarchical clustering (with their corresponding cophenetic coefficient) and Non-metric MultiDimensional Scaling (NMDS) analyses of the Bray-Curtis Similarity Index were generated using the Rstudio programming environment version 0.98.507 equipped with the R statistical software version 3.0.2 (R Core Team, 2013) using the stats package (part of R) and the vegan package version 2.0-10 (Oksanen *et al*., 2013). To test the strength and significance of the NMDS, a Non-parametric multivariate analysis -NP-MANOVA- (Anderson, 2001) was applied to both distance matrices employing the Adonis function implemented in the MASAME applications available at: https://sites.google.com/site/mb3gustame/hypothesis-tests/manova/npmanova (Buttigieg & Ramette, 2014). For both matrices the tests were performed with 5000 permutations. For this purpose, an additional explanatory grouping matrix was generated with dummy variables indicating the aquatic environment (laboratory or semi natural microcosmos) and life stage (pupae or larvae) of the tested groups. The adonis function computes an effect size value R^2 representing the percentage of variation within a distance matrix explained by the tested dummy variable.

3.3 Results

3.3.1 Data summary

Ion Torrent PGM sequencing of the two built libraries generated 4,606,079 sequence reads. After processing the reads with the MG-RAST QC pipeline, a total of 231,780 sequences were found to contain ribosomal RNA (rRNA) genes. To further explore, the metagenomic analysis statistics are available on request at the MG-RAST site under the ID numbers 4523102.3 and 4526005.3.

3.3.2 Bacterial community composition

The taxonomic profiling criteria applied to survey the pupae-associated microbiome identified 1,028 16S rRNA gene sequences. Bacteria derived sequences were assigned to 34 unique f-OTUs, with 1 to 392 sequences per f-OTU and 107 bacterial sequences unclassified at that taxonomic rank. Detailed community composition and relative abundance at the family level are presented in **Figure 5** and **Table S1**. The most abundant f-OTUs were *Flavobacteriaceae* (38.13%), *Aeromonadaceae* (28.02%) and *Enterobacteriaceae* (9.14%), which represent 75.29% of the bacterial community. A total of 10.41% of the Bacteria derived reads remained unclassified at the family rank. At the genus rank a total of 74 g-OTUs were identified and are shown in **Figure 6**.

Within the predicted bacterial consortium, marine, halotolerant and bioluminescent g-OTUs, such as *Shewanella*, *Pseudoalteromonas*, *Photobacterium*, *Alivibrio*, *Robiginitalea*, *Oceanimonas, Psychrobacter,* and *Photorhabdus* were identified. Known insect symbionts *Asaia* and *Burkholderia* were not identified within the surveyed community.

3.3.3 Compilation of *Anopheles*-associated microbiota: searching for the *Anopheles* core microbiota.

A previously published compilation of *Anopheles*-associated microbiota (Villegas & Pimenta, 2014) was updated by adding recent studies and the *A. aquasalis* predicted bacterial consortium we present in this paper (index in the **Additional file 1**). The new version of the index includes data from 21 published papers in which 272 unique bacterial genera have been reported as associated to 16 *Anopheles* species (**Figure 7**). The first published index lacked data regarding Neotropical anophelines and was enriched towards *A. gambiae* and *A. stephensi*. That bias led us to look for a broad scope approach to categorize the data in a

manner that allowed the search of a core bacterial assembly across the available screened anophelines.

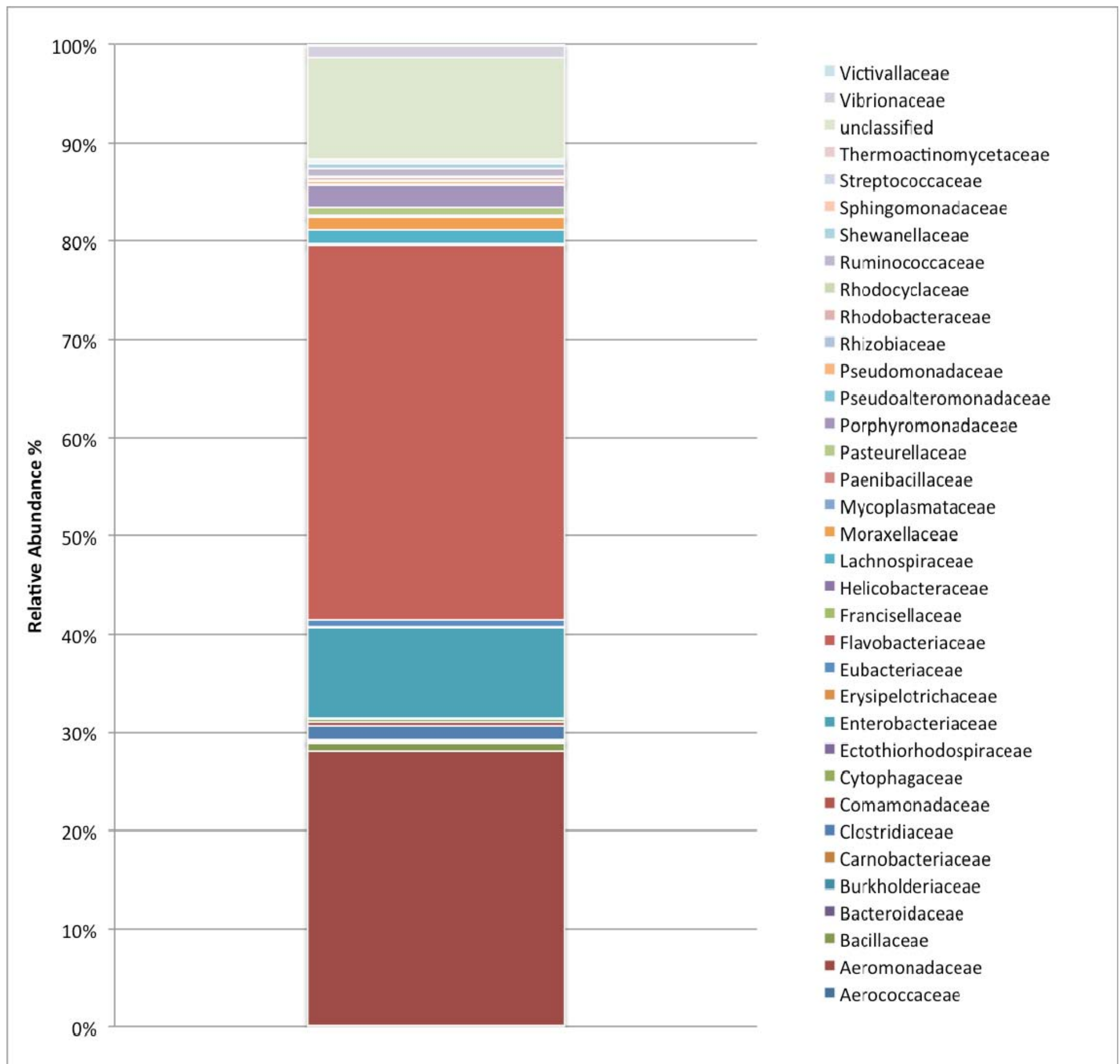


Figure 5 Relative abundance of the bacterial community (f-OTUs) associated to *A. aquasalis* pupae. The graph shows the relative abundance of the microbiota associated to laboratory reared *A. aquasalis* pupae at the family taxonomic rank - f-OTUs - (see relative abundance in table S1). As suggested by Glass *et al*. (2010) shotgun taxonomic profiling was performed with MG-RAST comparing the data to the M5RNA database using a maximum e-value of 1e-5, a minimum identity of 97% and a minimum alignment length of 50 bp. The "Unclassified" group encompasses reads that could not be assigned to the family level but were recognized as bacteria.

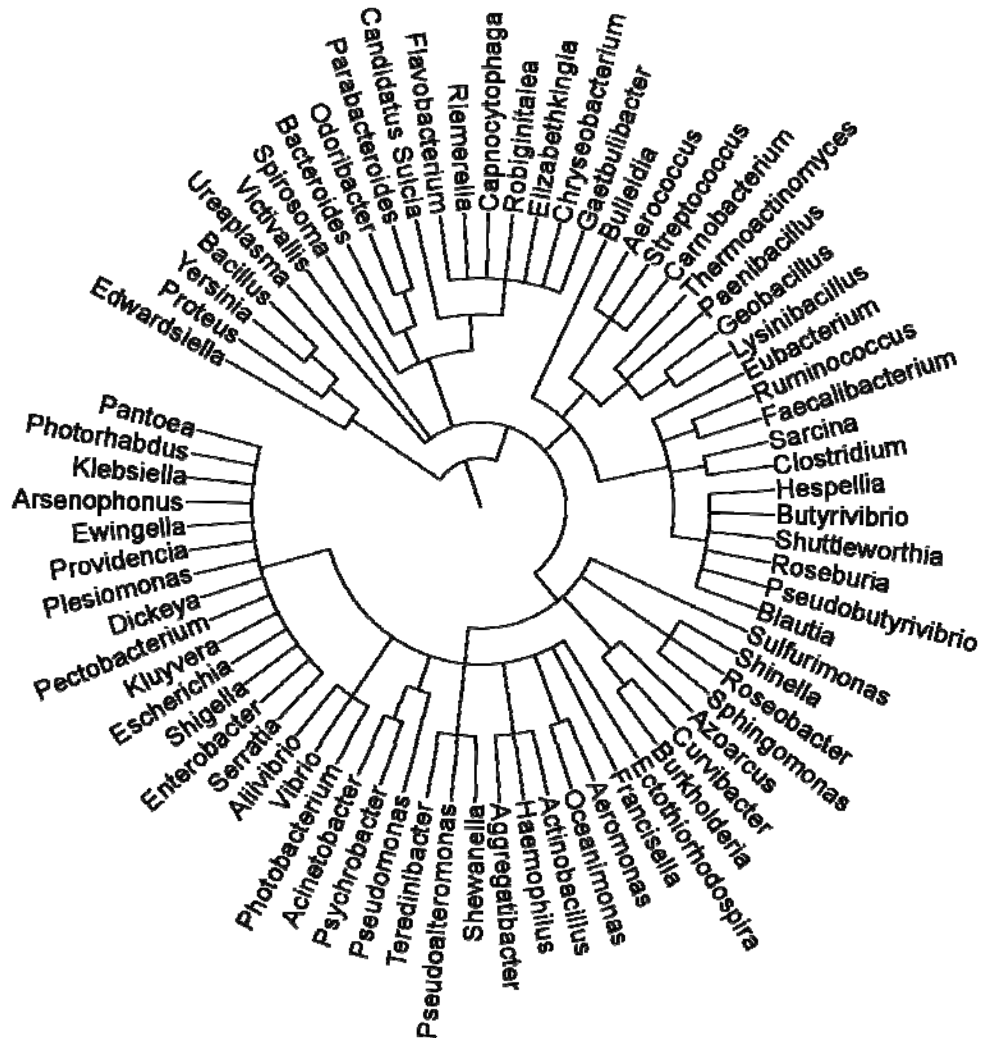


Figure 6 The *A. aquasalis* associated bacterial community. Bacterial genus rank OTUs (g-OTUs) predicted to be associated to laboratory reared *A. aquasalis* pupae. The phylogenetic tree was generated using the iTOL online tool v2.1: http://itol.embl.de/ (Letunic & Bork, 2011).

When grouped by the geographic (by-continent) distribution of the anopheline hosts, the 272 unique bacterial genera distributed as follows: Africa (83), America (79), Asia (189) and Europe (3) as shown in **Figure 8**. The shared and exclusive genera between the continent categories represent the following percentages. Exclusive: 11.03% to Africa, 17.28% to America, 48.90% to Asia. Shared: 5.88% between Africa, America and Asia; 2.21% between Africa and America; 10.29% between Africa and Asia; 3.31% between America and Asia; and 0.73% between Africa, Asia and Europe. A single genus, *Serratia*, is shared by the four continents representing 0.37%. (**Figure 8** and **Table S2**). The 23 bacterial genera shared between African and American anophelines (Figure 4) belong to the following 14 f-OTUs: *Aeromonadaceae, Bacillaceae, Bacteroidaceae, Burkholderiaceae, Clostridiaceae, Enterobacteriaceae, Flavobacteriaceae, Moraxellaceae, Paenibacillaceae, Pasteurellaceae, Pseudomonadaceae, Sphingomonadaceae, Streptococcaceae*, and *Vibrionaceae*. An additional visual aid to explore the correlation between the by-continent categories and bacterial genera is shown in **Figure S2**.

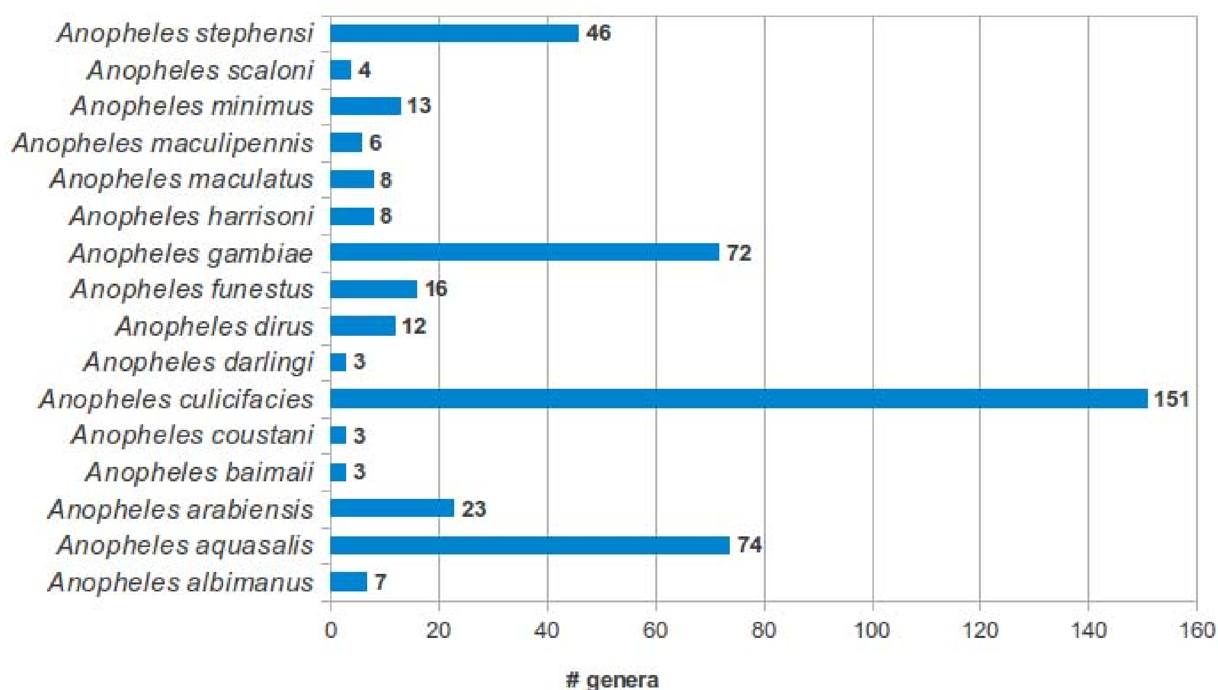


Figure 7 Number of bacterial genera per anopheline host. Current count of bacterial genera reported as associated to *Anopheles* host species. In detail: 272 unique bacterial genera were found associated to 16 *Anopheles* species according to 21 published papers. Refer to Additional file 1 to search through the compiled data.

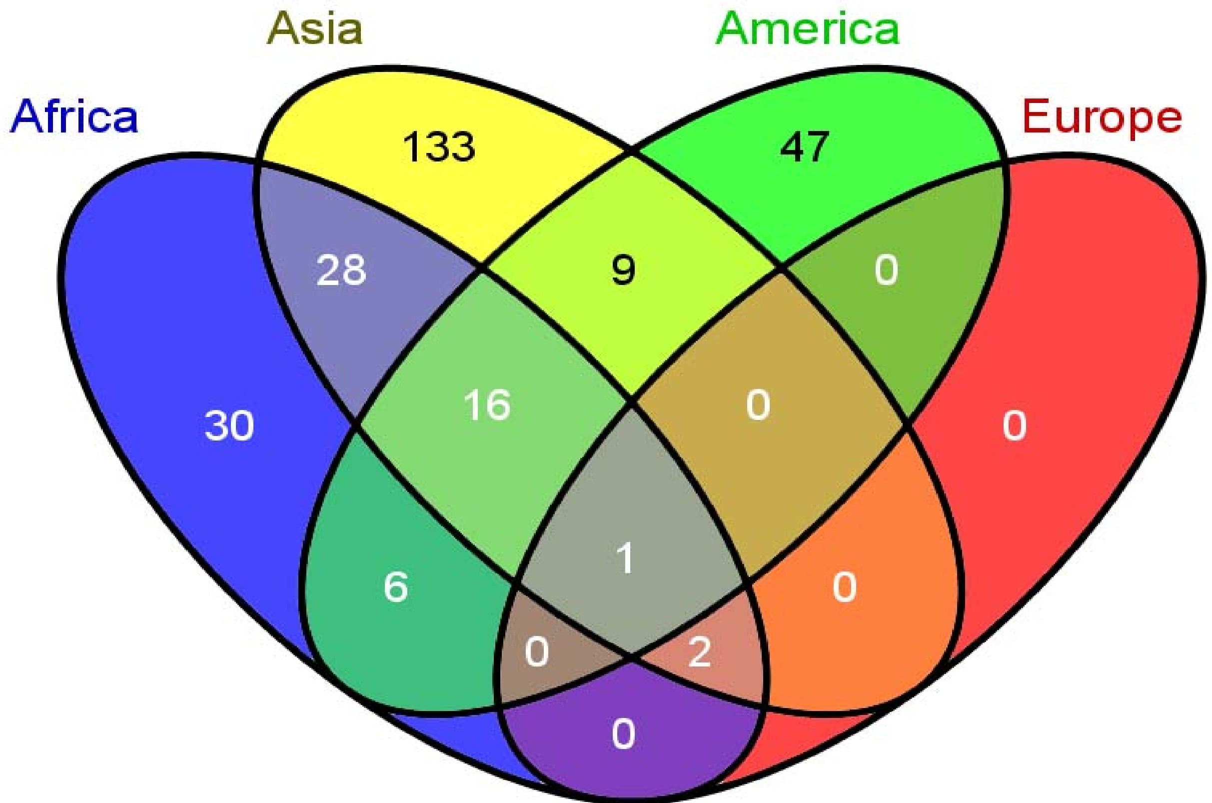


Figure 8 Shared and exclusive bacterial genera identified in anopheline human malaria vectors grouped by *Anopheles* continental distribution. Using a continent-oriented filtering criterion (refer to Methodology) the 272 unique bacterial genera reported to date as associated to human malaria vectors, distribute as follows: Africa (83), America (79), Asia (189) and Europe (3). The shared and exclusive genera distribute among the continent categories as follows: 11.03% exclusive to Africa; 17.28% exclusive to America; 48.90% exclusive to Asia; 5.88% shared by Africa, America and Asia; 2.21% shared by Africa and America; 10.29% shared by Africa and Asia; 3.31% shared by America and Asia; 0.73% shared by Africa, Asia and Europe. A single genus (*Serratia*) is shared by the four continents (0.37%). Notice that 23 genera are shared between Africa and America. A "by-continent" bacterial genera index is available in the Additional file 1, spreadsheet 2, while the list of the shared and exclusive bacterial genera is shown in the table S2.

3.3.4 *β*-diversity exploration: comparing bacterial profiles between aquatic stages of *A. aquasalis* and *A. gambiae*.

Explorative community diversity analyses were performed to determine if a biological or ecological pattern could be revealed when comparing the five tested groups based on their bacterial abundance profiles. In particular we tested if the bacterial profile of the identified African-American core f-OTUs (**Figure 9**) was as informative as the full dataset (**Additional file 2**) when revealing a potential pattern.

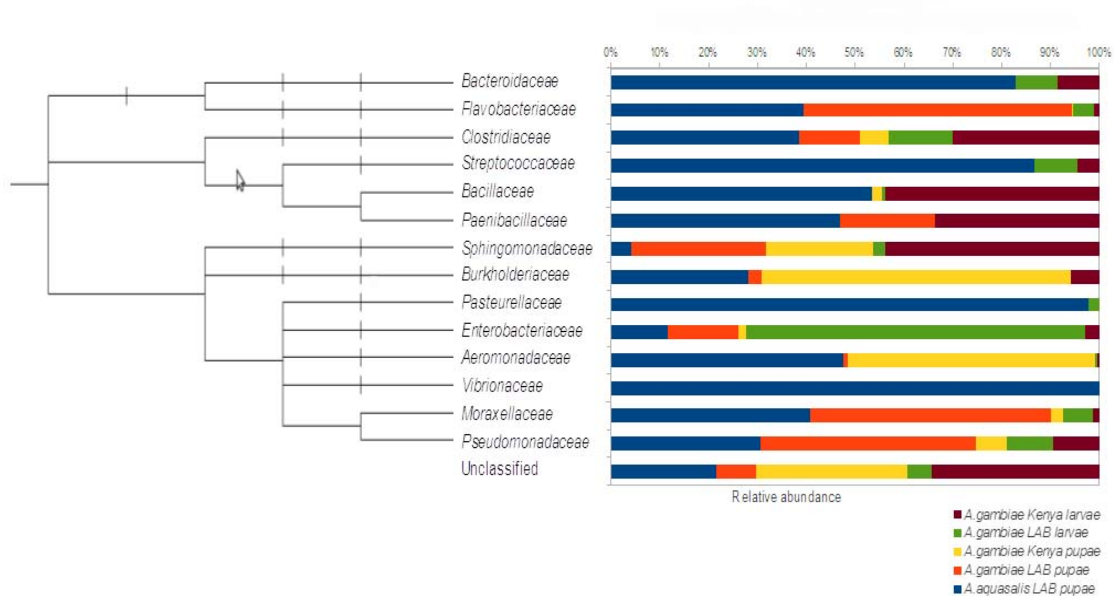


Figure 9 Relative abundance of the African-American *Anopheles* shared bacterial family-rank OTUs associated to five aquatic stages of anopheline human malaria vectors. The bars represent the relative abundance of the 14 African-American *Anopheles* core bacterial f-OTUs, within five aquatic stages corresponding to *A. gambiae* (4) and *A. aquasalis* (1). Three of the groups were reared in laboratory conditions and two in semi-natural conditions (Microcosmos setup Wang *et al*., 2011; tagged in the key as ¨Kenya¨). The phylogenetic tree of the f-OTUs was generated using iTOL (Letunic & Bork, 2011). The Unclassified bar represents bacterial tags that could not be assigned to a family taxon by MG-RAST. They may represent not yet identified bacterial richness associated to mosquitoes or their surroundings.

The Bray-Curtis distance metrics were used to compare the abundance profiles of the tested groups, using hierarchical clustering and NMDS analyses to identify the structure within the datasets. As shown in **Figure S1B** a distinct clustering pattern was revealed when examining the transformed microbial abundance matrix encompassing the full dataset consisting of 126 f-OTUs. The potential main structure within the dataset resolved along dimension 1 and suggested that, regardless of the *Anopheles* species, the laboratory reared groups from both *A. gambiae* and *A. aquasalis* samples have a more similar bacterial profile clustering apart from the semi-naturally reared groups. Additionally, a life stage dependent gradient may be grouping pupae apart from larvae along dimension 2. The low NMDS stress value (stress ≈ 0) supported the presence of a potential ecological gradient driven by the type of aquatic environment (or breeding site). Thus, to test the strength and significance of the NMDS clusters, a NP-MANOVA was applied using an additional explanatory matrix indicating the aquatic origin (laboratory or semi natural microcosmos) and life stage (pupae or larvae) of the tested groups. The applied adonis function showed that 49% of the variation within the matrix, explained in dimension 1 of the NMDS plot, was statistically significant ($R^2 = 0.49$, $p < 0.05$) and responds to the influence of the aquatic environment or breeding site. The variation potentially explained by the anopheline life stage (22%) was not statistically significant ($p > 0.05$).

When examining the abundance matrix encompassing the bacterial profiles of the 14 African-American core f-OTUs the same clustering pattern was revealed (**Figure 10C**). As described above, the potential structure within the dataset resolved along dimension 1 and suggested that, regardless of the *Anopheles* species, the laboratory reared anophelines had a more similar bacterial profile clustering apart from those reared in a semi-natural environment. Again, a potential life stage dependent gradient could be grouping pupae apart from larvae along dimension 2. The low NMDS stress value (stress ≈ 6.12e-05) supported the presence of a potential ecological gradient driven by the type of aquatic breeding site. The strength and significance of the NMDS clusters were tested by applying a NP-MANOVA using an explanatory dummy matrix as explained above. The adonis test showed that 42% of the variation within the matrix, explained across dimension 1 of the NMDS plot, was statistically significant ($R^2 = 0.42$, $p < 0.05$) and responds to the aquatic environment or breeding site. The variation potentially explained by the anopheline life stage (35%) was not significant ($p > 0.05$).

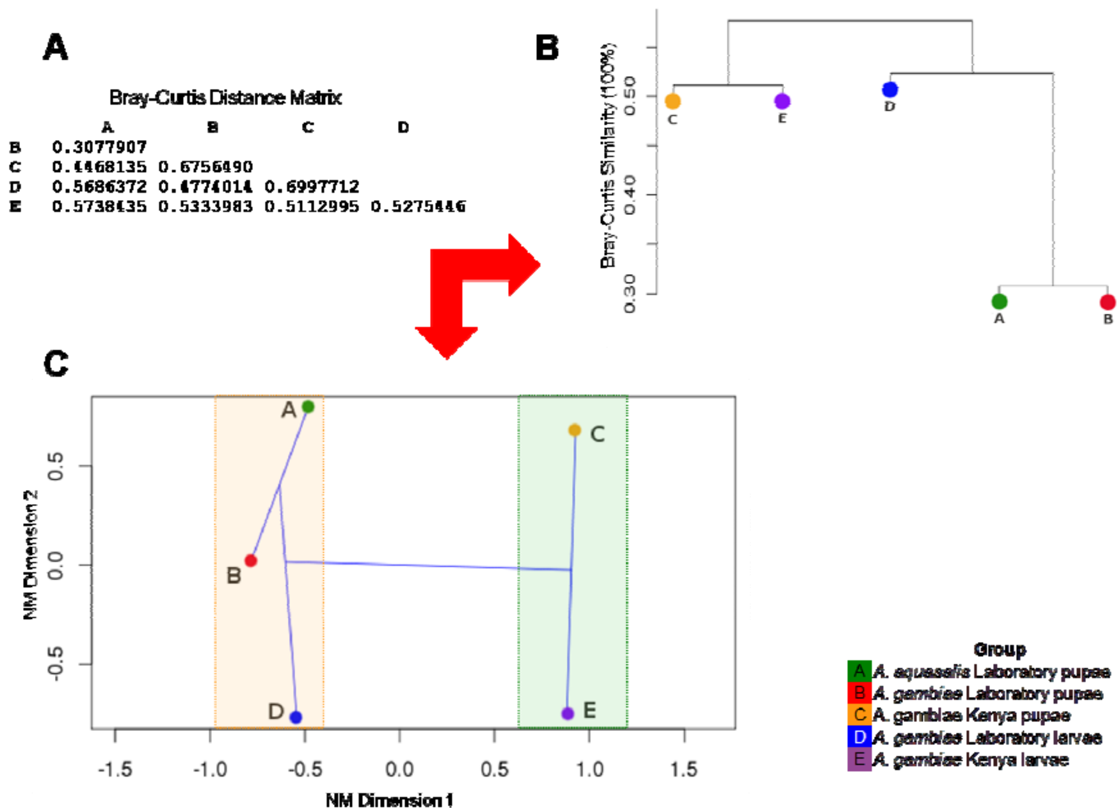


Figure 10 Combined ordination and clustering analyses comparing the family-rank OTUs abundance profiles of five anopheline aquatic life stages (African-American shared f-OTUs matrix). To reduce the impact of data dispersion, the relative abundance matrix (Table S3) was transformed (Arcsin $\sqrt{x}$) to create a normal distribution of the data. The abundance data corresponds to five aquatic stage anopheline groups described in the color/letter key. Panel "A" shows the Bray-Curtis pairwise distance matrix between the anopheline groups tested. Panel B shows the hierarchical clustering dendrogram of the five compared anopheline aquatic stages built upon the Bray-Curtis dissimilarities; the dendrogram was constructed using the "average" method (cophenetic index = 0.74). Panel C shows the NMDS ordination of the five anopheline groups based on the distance rank analysis of their "core" bacterial f-OTUs profiles. The cluster analysis was superimposed to depict the primary connections onto the ordination space. The potential main structure within the dataset is highlighted as rectangular shaded areas resolved along Dimension 1. Regardless of the *Anopheles* species the laboratory reared groups mapped onto the ordination space shaded in orange, whereas the micro-cosmos reared groups (Kenya) mapped onto the ordination space shaded in green. Thus, it appears as if the nurturing aquatic environment acts as an abiotic factor driving the ordination of the groups. An additional gradient based on the life stage may be present along Dimension 2 separating pupae from larvae (NMDS stress value = 6.12e-05; iterations = 31).

3.4 Discussion

Advances in NGS sequencing and biocomputing have allowed whole shotgun metagenomic surveys to reveal the composition and biological functions of complex microbial communities associated to particular biomes and even host organisms (Segata *et al*., 2013, Sharpton, 2014, and references therein). The Ion Torrent PGM device is a cost effective and scalable microbial ecology sequencing platform (Whitheley *et al*., 2012). It has been used recently in multiple microbiome taxonomic surveys ranging from the human fecal microbiome (Milani *et al*., 2013) to buffalo rumen (Singh et al., 2014), reef waters & coral surfaces (Lim *et al*., 2014), and permafrost-affected soils (Frank-Fahle *et al*., 2014).

Microbial community surveys performed with 16S rRNA gene sequences derived both from shotgun metagenomic sets, as well as PCR targeted amplicons, are similar at broad taxonomic rank predictions (Kalyushnaya *et al*., 2008, Poretsky *et al*., 2014). Both methods are capable of yielding nearly identical estimates of diffrences between bacterial comunities in terms of comunity diversity and composition (showing a significant correlation in detection capabilities) (Fierer *et al*., 2012, Sharpton *et al*., 2011). It has been found that for 16S rRNA gene sequencing surveys the number of reads per sample is not paramount to reveal and detect the biological patterns that define the ecosystem structure [~from 100 to 1,000 reads] (Gilbert *et al*., 2010, Kuczynski *et al*., 2010; Sanschagrin & Yergeau, 2014). Thus to explore an detect relevant biological patterns within microbial communites, the ecological diversity analyses (ordination methods) applied become the keystone to interrogate the abundance profiles (Anderson *et al*., 2011, Ramette, 2007, Kuczynski *et al*., 2010)

This study is part of an ongoing project aiming to sequence the genome of *A. aquasalis* to further support its use as an experimental malaria transmission model in Brazil. To maximize the use of whole genomic DNA extracted from laboratory-reared pupae, we performed a whole-metagenome shotgun sequencing to survey, primarily, the composition of the bacterial community associated to this Neotropical species.

3.4.1 *Anopheles aquasalis* associated microbiota

It is thought that insect-associated microbiota is composed of a few dominant bacterial taxa, particularly when compared to mammals (Engel&Moran, 2013, Dong *et al*., 2009, Dillon&Dillon 2004, Minard *et al*., 2013). A reduced bacterial diversity is characteristic of laboratory-reared *Anopheles* mosquitoes, when compared to their wild counterparts (Wang *et*

al., 2011; Boissière *et al*., 2012). For example, the bacterial community of laboratory-reared *A. gambiae* pupae contained 38 f-OTUs whereas 80 f-OTUs were identified in pupae reared in semi-natural conditions (Wang *et al*., 2011). We found a similar number of bacterial f-OTUs predicted to be associated to laboratory reared *A. aquasalis* pupae (34 f-OTUs). Though these recounts were obtained by different metagenomic approaches, it has been observed that 16S rRNA datasets, both amplicon and shotgun derived, yield concordant diversity and composition estimates which reflect, overall, the bacterial community structure of a same sample (Fierer *et al*., 2012).

The Flavobacteria *Elizabethkingia* spp. is known to thrive when associated to laboratory mosquitoes, in which, low bacterial richness and diversity seems to reflect mosquito food source (Avgustin *el al*., 1997) and aquatic environment (Wang *et al*., 2011; Boissière *et al*., 2012; Gendrin & Christophidis, 2013). Members of the *Flavobacteriaceae* family were identified as the predominant OTUs within the predicted *A. aquasalis*-associated bacterial community. This finding supports previous reports of *Flavobacteriaceae* being the predominant bacterial family associated to insectary-reared *A. gambiae* pupae (Wang *et al*., 2011) and the existence of a significant correlation (evaluated with redundancy analysis) between this bacterial family and adult laboratory mosquitoes (Boissière *et al*., 2012).

Dipteran aquatic stages, larvae in particular, are rich in members of *Enterobacteriaceae* as observed in *Culex tarsalis* (Duguma *et al*., 2013), *Drosophila* spp. (Chandler *et al*., 2011) and *Musca domestica* (Martínez-Falcón *et al*., 2001). Similar abundances of this predominant family are found in both *A. gambiae* (Wang *et al*., 2011) and *A. aquasalis* laboratory-reared pupae (11.31% and 9.14% respectively).

Taxa within the Cyanobacteria phyla have been described in semi-naturally reared *A. gambiae* aquatic stages (Wang *et al*., 2011), accounting for up to ~40% of the community. These photosynthetic bacteria may act as a food source for mosquito larvae in natural habitats (Vasquez-Martínez *et al*., 2002, Thiery *et al*., 1991). Nevertheless, in laboratory-reared *A. gambiae* larva and pupae (Wang *et al*., 2011) and *A. aquasalis* pupae they are almost absent (relative abundance < 0.1%).

Breeding site water samples possess a higher OTU diversity and abundance than mosquito larvae and pupae (Duguma *et al*., 2013, Dada *et al*., 2014, Wang *et al*., 2011). This suggests that only a subset of bacteria from the aquatic environment can effectively establish a stable community within the mosquito host. The midgut in particular would act as selective micro-environment throughout each life stage (Coon *et al*., 2014, Gimonneau *et al*., 2014). Larval instars feed upon bacteria (Thiery *et al*., 1991, Merrit *et al*., 1992) on which they rely not only as a food source, but also to complete their life cycle (Coon *et al*., 2014). Thus, it is

possible that bacterial OTUs identified in meta-taxogenonomic studies of aquatic life stages represent ingested transitory members of the community (allochthonous members). Moreover, the low bacterial OTU abundance detected in laboratory-reared mosquitoes may be due to other microbes, such as algae and fungi, acting as the main food source of larvae, as observed in *A. gambiae* (Gimnig *et al*., 2002, Kaufman *et al*., 2006).

Many reads in metagenomic datasets are left unclassified. Currently, up to ~90% of the sequences may remain unidentified due to the lagging availability of curated sequences in reference databases (Huson *et al*., 2009). Within the *A. aquasalis* predicted microbiota, 10.41% of the OTU's were tagged as unclassified (derived from Bacteria) when assessed with the M5RNA database in MG-RAST. The abundance of unclassified sequences in laboratory and semi-natural *A. gambiae* pupae was reported to be 3.89 % and 14.99 % respectively (Wang *et al*., 2011). Insect gut communities are known to be complex (Engel & Moran, 2013). One factor influencing the bacterial community structure is the food source of each insect. In termites for example, hundreds of bacterial species have been identified as novel. This suggests that many of the unidentified bacterial organisms are termite gut-adapted specialists and not environmental bacteria ingested with food (Hongoh *et al*., 2006, Ohkuma & Brune, 2010). We consider that the percentage of Bacteria-derived unclassified sequences in laboratory-reared *A. aquasalis* pupae could reflect unexplored bacterial diversity present in Neotropical anophelines. Furthermore, they could represent brackish water specialists that co-evolved with this anopheline. Nevertheless, to date the microbiota associated to other salt-water tolerant anophelines has not been explored in depth to draw comparisons.

Anophelines possess distinct bionomic traits (Sinka, 2013, Tripet *et al*., 2008). Most breed in fresh waters though several species adapted to saltern larval niches and are implicated in coastal malaria transmission (Sinka, 2013). For example, *Anopheles melas* and *Anopheles merus* from Africa, *Anopheles farauti* s.s. and *Anopheles irenicus* in Australasia, *Anopheles sundaicus* in Southeast Asia, and *A. Aquasalis* in America exhibit this bionomic trait (Rejmánková *et al*., 2013 and references therein). Evolving in brackish water environments might have established a stable interaction between the mosquitoes and microbial partners thriving in saltern waters. It has been suggested that the most dominant taxa remain associated to its anopheline host even after laboratory colonization (Wang *et al*., 2011; Gendrin & Christophides, 2013), though the diversity and relative abundance of the community members may change. Interestingly, marine, halotolerant and bioluminescent g-OTUs such as *Photobacterium* (Vezzi *et al*., 2005), *Aliivibrio* (Urbanczyk *et al*., 2007), *Robiginitalea* (Cho & Giovannoni, 2004), *Oceanomonas* (Brown *et al*., 2001), *Photorhabdus* (Duchaud *et al*., 2003), *Psychrobacter* (Bakermans *et al*., 2006) *Shewanella* (Chikuma *et al*.,

2004), were indentified in *A. aquasalis* reared under insectary conditions. Furthermore, these genera have not been reported in African anophelines (**table S2**) and they may reflect this anopheline's ecological niche influence throughout its evolutionary history.

In fact, the evolutionary origin of bioluminescence in marine bacteria has been linked to the capacity of these OTUs to reduce reactive oxygen species allowing them to adapt to their harsh environment (Widder, 2010). Wang *et al*. (2011) performed a prospective comparative genomics study predicting the bacterial redox capabilities of members of the *A. gambiae* microbiota. The authors suggest that gut bacterial redox capacity would be connected to the invertebrate's immune response by allowing redox homeostasis to sustain the anti-malarial nitrogen oxides production. Profiling the microbial consortium of other salt-tolerant anophelines may provide further evidence on whether marine bacteria co-evolved and participated in the adaptation of these mosquitoes to coastal niches. It is argued that, within a holobiont, most bacterial taxa form a strong metabolic or functional association with adjacent taxa (host and other symbionts) inhabiting the same physical niche. This co-occurrence would allow the microbial consortium to specialize and adapt to highly localized niches (Caporaso *et al*., 2011, Faust *et al*., 2012). The stability and high abundance of bacterial core members (and their housekeeping functions), has been found to contrast with the greater phylotype variability and lower abundance of niche-specific OTUs, which although rare, are consistently present to provide key metabolic pathways / metabolites to their host and other bacterial partners (Human Microbiome Project Consortium, 2012).

Amongst the predicted bacterial community in *A. aquasalis* pupae, two absences drew our attention. First, no OTUs from the *Asaia* genus were identified. *Asaia* has been identified in anopheline and culicine mosquitoes, reported to have positive effects on mosquito fitness, and shown to be horizontally and vertically transmitted under laboratory conditions (Favia *et al*. 2007, Chouaia *et al*. 2010, Damiani et al., 2008). Nevertheless, recent studies (Djadid *et al*., 2011, Chavshin *et al*., 2012, Rani *et al*., 2009, Sharma *et al*., 2014) were not able to identify *Asaia* in field-collected and/or laboratory-reared Asian malaria vectors from Iran and India. The positive impact it has upon mosquito fitness and its apparent absence in Neotropical mosquitoes could imply that other *Acetobacteriaceae* may be performing its functions within the Neotropical anopheline holobiont. Being a promising candidate for paratransgenic malaria control (Damiani *et al*., 2010, Minard *et al.,* 2013), the fact that several authors have not found it in field collected and laboratory reared *A. stephensi*, could hinder its potential as a candidate for malaria transmission control in Asia. Notwithstanding, the absence of this genus in the screened mosquitoes could be due to the detection strategies used (Djadid *et al*., 2011; Rani *et al*., 2009) or geographic and ecological niche related factors (Sharma *et*

al., 2014). The fact that this genus has been described in field collected South-east Asian (none *A. stephensi*) and African anophelines (Manguin *et al*., 2013, Favia *et al*., 2007, Gimonneau *et al*., 2014, Boissière *et al*., 2012) stresses the need to perform more thorough profiling efforts encompassing representative anophelines from malaria burdened regions in which paratransgenesis strategies are being considered as a transmission-control strategy.

Secondly, the members of the *Burkholderia* genus, other known insect symbionts, were not detected in colonized *A. aquasalis* pupae. This genus was also absent in adult laboratory-reared *A. gambiae* mosquitoes (Bossierè *et al*., 2012) and throughout the life stages of semi-naturally reared *A. gambiae* (Wang *et al*., 2011). This insect symbiont has been associated to plant feeding insects as a positive fitness effector that provides protection against entomopathogenic fungi (Santos *et al*., 2004). Thus, in anophelines, it may not be essential for growth and reproduction since this genus is missing in colonized mosquitoes (Bossierè *et al*., 2012).

The previous observations bring forth the potential existence of negative relationships or associations between bacterial taxa (Faust *et al*., 2012). In Asian anophelines, it was proposed that the co-occurrence of *Pseudomonas fluorescens* and *Serratia* might inhibit the colonization of the anopheline midgut by *Acinetobacter*. The absence of enterobacteria was also noted in anopheline samples whenever *Sphingomonas* was detected (Manguin *et al*., 2013). Since *Asaia* has been so commonly reported in other anophelines we cannot rule out that there may be negative associations excluding it from the microbiota of Neotropical anophelines.

3.4.2 The *Anopheles* core microbiota

The *Anopheles* holobiont and the identity of the core bacterial consortium (or bacteria mediated metabolic functions), if any, is a trending topic particularly when linked with its potential implication with vectorial capacity and as the basis for malaria transmission control strategies (Minard *et al*., 2013, Wang & Jacobs-Lorena, 2013, Wang *et al*., 2011, Bahia *et al*., 2014).

The insect midgut harbours most of the microbiota though other organs and specialized cells do as well (Engel&Moran, 2013). Mirroring the concept of the bacterial human enterotype (Evans *et al*., 2013, Turnbaugh *et al*., 2007), the *Anopheles* bacterial partners would act as metabolic exchangers/enhancers (Russell *et al*., 2013) and as key elements of the invertebrate immune response (Nyholm & Graf, 2012) influencing the *Anopheles*-*Plasmodium* interactions (Dong *et al*., 2009, Bian *et al*., 2013, Molina-Cruz *et al*.,

2008).

Current knowledge on *Anopheles* associated microbiota was reviewed by Gendrin and Christophides (2013). Their data compilation found 41 bacterial g-OTUs reported in more than one *Anopheles* species according to 23 studies. Nevertheless, they found no single bacterial genus associated to all of the surveyed anophelines. Therefore, the authors suggest there is no presumable or proven obligate symbiont species associated to the *Anopheles* genus.

Nonetheless, the identity of an *Anopheles* core microbiota has been hinted from a broader taxonomic perspective in recent publications (Boissière *et al*., 2012, Gendrin & Christophides 2013). Wang *et al*. (2011) observed that, at a family taxonomic level, *Enterobacteriaceae*, *Flavobacteriaceae* and *Pseudomonadaceae* encompassed 80.9% and 94.3% of the microbiota associated to laboratory-reared and field collected adult *A. gambiae* mosquitoes respectively. Shared phyla and genus level OTUs were also reported between two life stages of *A. gambiae* and *Anopheles coluzzi* (Gimonneau *et al*., 2014).

Associations between anopheline mosquitoes and core bacterial taxa have been observed, mostly, in African models. Neotropical malaria vectors lag behind those from Africa and Asia in terms of the wealth of information regarding mosquito-parasite-microbiome interactions (Molina-Cruz & Barillas-Mury, 2014, Villegas & Pimenta, 2014). On this matter, Molina-Cruz & Barillas-Mury (2014) present a hypothesis in which geographic and evolutionary distance would act as a key molding factor of the *Plasmodium*-*Anopheles* interaction and its establishment in the Americas. Such connection would be based on the compatibility between the parasites' protein *Pfs47* and a yet to be discovered mosquito protein. This protein-protein interaction is thought to be involved in the parasites' ability to adapt to new vectors evading the immune response (Molina-Cruz *et al*., 2013). Interestingly, they propose that the geographic genetic structure of *Pfs47* would be largely determined by selective pressures imposed by the immune system of anopheline mosquitoes which may differ according to geographic regions or ecological niches. Thus, if a potential geography-determined genetic structure played a role in shaping the compatibility between the *Plasmodium* parasites and its anopheline vectors, then the fact that anophelines adapted to specific ecological niches across earth (Sinka *et al*., 2012) gains relevance from a microbial ecology perspective.

The microbial components of the particular niches in which the anopheline species evolved plausibly acted as underlying forces shaping the mosquito immune system via feeding or immune challenges (Minard *et al*., 2013, Tripet, 2009). Furthermore, it is known that the dynamic microbial consortium associated to the midgut of anophelines (Wang *et al*.,

2011, Gimonneau *et al*., 2014) is acquired primarily from the specific aquatic environments in which each species develops (Lindh *et al*., 2008; Wang *et al*., 2011; Boissière *et al*., 2012, Gimonneau *et al*., 2014, Coon *et al*., 2014). Thus, the *Anopheles*-associated microbiota, established throughout evolutionary and geographic time, might have factored in the process of compatibility development between vector species and *Plasmodium* species (or strains), particularly acting as a modulator of the anopheline innate immune system (reviewed in Cirimotich *et al*., 2010; Gendrin & Christophides 2013; Minard *et al*., 2013 and Clayton *et al*., 2014).

Therefore, categorizing current data into broader groups based on a geographic *Anopheles* distribution criterion (**Additional file 1**) allowed us to explore the existence of an *Anopheles* genus core microbiota. The performed meta-analysis encompassed 21 publications, based on multiple detection techniques and taxonomic profiling methods, and applied to different anopheline life stages. We consider this filtering criterion as a valid, initial, approach to glimpse into this trending topic linking it with a current hypothesis on the origins of the *Anopheles*-*Plasmodium* compatibility and particularly from our perspective, its radiation from Africa to America.

From the filtered bacterial index *Serratia* emerged as the only genus shared amongst the anopheline groups generated (**Figures 8** and **S2**). Interestingly it is one of the main candidates for paratransgenic malaria transmission control (Wang & Jacobs-Lorena, 2013). Other candidates for such strategies like *Asaia* and *Pantoea* (Coutinho-Abreu *et al*., 2010, Villegas & Pimenta *et al.,* 2014) were found shared by at least three continental groups (**Table S2** and **Figure S2**) highlighting the absence of *Asaia* in American anophelines.

We identified 23 bacterial genera shared between African and American anophelines (**Figure 8** and **Table S2**), which at the bacterial family rank, correspond to 14 f-OTUs (**Figure 9**). Whether the association between African and American anophelines, with key members of these f-OTUs remains stable throughout life stages or across anopheline species is unknown. Nevertheless, amongst the 14 identified core OTUs, the three dominant taxa described previously as adult ¨*Anopheles* core families¨ (*Enterobacteriaceae*, *Flavobacteriaceae* and *Pseudomonadaceae*) are present (Wang *et al*., 2011, Boissière *et al*., 2012, Gendrin & Christophides 2013).

It is important to consider that the taxonomic level at which the identity of each bacterial partners is reported, may not be as relevant as the particular metabolic functions they provide to their host throughout its life history. It is possible that different OTUs within a bacterial family perform the same functions in different anopheline life stages or species. It has been observed, in mammals and other ecosystems that, bacterial members of a

phylogenetic lineage can perform the same metabolic functions when established in similar ecological conditions even in different physical or geographic sites (Ley *et al*., 2008, Fierer 2012). Therefore, future studies should consider exploring the metabolic potential of the bacterial partners of the *Anopheles* genus.

As more metagenomic data becomes available, signature bacterial profiles, based on the abundance of the *Anopheles* genus core microbiota, may become available. Relevant ecological and biological shifts in community abundance most likely occur among key core members that, by definition, are well established in the ecosystem (Magurran & Henderson, 2003). Signature f-OTU or g-OTU abundance profiles could describe anopheline species at specific life stages and, furthermore, they may reflect bionomic traits, environmental conditions or particular phenotypes (infection, insecticide resistance, etc). This extra layer of knowledge would enhance vector biology comprehension and enable induced dysbiosis and paratransgenesis strategies. In this study we identified 14 African-American core f-OTUs. We performed an explorative multivariate analysis based on our *A. aquasalis* findings and published data (Wang *et al*., 2011) regarding the bacterial profiles of *A. gambiae* aquatic stages to test if the profiles (full dataset and core f-OTUs) could reflect differences between African and American anopheline samples or reveal underlying biological gradients.

3.4.3 Ecological multivariate analysis between anopheline aquatic stages

The community composition of the microbiota within its host may represent an ancestral footprint of evolution (Brucker and Bordenstein, 2013). As observed in plant and animal communities, shifts in diversity and relative abundance of key bacterial taxa could be due to broad-scale gradients driven by biotic and abiotic factors within their ecological niche (Fierer *et al*., 2012).

The dynamic nature of bacterial communities has been described in African anophelines throughout their holometabolous life history (Wang *et al*., 2011, Gimonneau *et al*., 2014). Shifts in diversity and abundance profiles would reflect the transition from aquatic to terrestrial environments, different nutrient sources (e.g. bacteria, nectar, blood), responses to host genetics, intra- and inter- bacterial community interactions, or environmental factors (Minard *et al*., 2013, Engel & Moran, 2013, Gendrin & Christophides, 2013).

We used hierarchical clustering and NMDS analyses of the Bray-Curtis distance metrics to compare the abundance profiles of the five anopheline groups included in the two built matrices. Whether the clustering patterns would respond to the anopheline species, nurturing aquatic habitats (semi-natural vs. laboratory) or life stages (pupae vs. larvae) was unknown. Interestingly, both the full dataset and core matrices identified the same significant

environmental pattern.

Regardless of *Anopheles* species and life stage, an environmental factor (nurturing aquatic environment) significantly determined the clustering pattern of the compared groups based on their f-OTU abundance profiles. Laboratory reared pupae and larvae were significantly more similar in terms of their bacterial community composition (**Figures 10** and **S1**). Paucity in bacterial diversity associated to adult laboratory reared mosquitoes has been described (Boissière *et al*., 2012, Gendrin & Christophides 2013). It seems that laboratory rearing also modulates the bacterial abundance profiles of *Anopheles* aquatic stages which appear to be more similar amongst them than with those reared in semi-natural conditions. Though expected, this observation attained by comparing different anopheline species and life stages further highlights the importance of cautiously interpreting malaria transmission and paratransgenesis assays based on laboratory models (Aguilar *et al*., 2005, Wang & Jacobs-Lorena, 2013, Coutinho-Abreu *et al*., 2010). Particularly, when bacterial challenges or abundance measurements are fundamental to the findings.

Multivariate ecological analyses rely on bacterial taxa abundances acting as real-time interactive biomarkers whose shifts reflect host phenotypes or responses towards ecological/biological gradients (Ley *et al*., 2008, Fierer *et al*., 2012, Kuczynski *et al*., 2010) in a sort of host discriminating microbial signature (Mason *et al*., 2014; Knights *et al*., 2011). Reducing the dimensionality of the datasets used to indentify signature profiles is critical to improve multivariate analyses or machine learning methods (Knights *et al*., 2011).

The reduced data set of 14 core f-OTUs (based on the geographic filtering criterion) was capable of identifying the same main ecological pattern as the full data set (126 families as shown in **Additional file 2**). Thus, the variation contained in the African-American core f-OUT distance matrix (**Table S3**) seems to be representative of the variation within the whole bacterial communities tested. This suggests that, at the bacterial family rank, it may be possible to discriminate between anopheline samples based on their bacterial signature.

In addition to our study, other mosquito associated microbial assemblies have been explored from this perspective (Boissière *et al*., 2012, Tchioffo *et al*., 2013, Dada *et al*., 2014, Gimonneau *et al*., 2014, Duguma *et al*., 2013, Osei-Poku *et al*., 2012). When evaluating the correlation between the geographic origin of adult anophelines (larval niche) and the composition of their associated microbiota, apparently contradictory results have been reported (Osei-Poku *et al*., 2012; Boissière *et al*., 2012, Tchioffo *et al*., 2013). Gendrin & Christophides (2013) discuss the basis of these opposing observations. They highlight that differences on the experimental design and the anopheline molecular forms analyzed might have influenced the results. To these arguments we would add the effect of the similarity

measuring methods used to detect biological or ecological patterns between the bacterial community structures (Kuczynski *et al*., 2010, Anderson *et al*., 2011, Legendre & De Cáceres, 2013, Lozupone *et al*., 2007). The UniFrac pairwise distance method used by Osei-Poku *et al*. (2012) relies on phylogenetic divergence information to compare community structures. Such an approach requires substantial additional decisions about the bacterial phylogenetic tree and the rate of environment switching (Kuczynski *et al*., 2010, Lozupone *et al*., 2005). On such terms, bacterial communities from different geographic points, in the same country, may not differ significantly regarding the phylogenetic origin of the bacterial taxa. The developers of this community similarity measure (Lozupone *et al*., 2005) found that the phylogenetic origin of the bacterial members of a community play a minor role upon shaping the composition of the community when compared to the effect of the ecological niche since bacterial types are dispersed widely in similar habitat types. Thus it is the environmental gradients that induce most of the modulating effects on bacterial OTU abundance profiles. From this perspective, when focusing on the relative abundance of bacterial OTUs and their response to biological or ecological effectors, the Bray-Curtis distance method may be more informative to identify patterns when the question posed to the dataset does not seek to give more weight to bacterial phylotypes.

3.5 Conclusion

No NGS taxonomic profiling had been performed on Neotropical anophelines. The presence of brackish water-related bacteria, not described previously in African or Asian anophelines, supports the idea that the aquatic larval niche influences the diversity and composition of the microbial community within the mosquito.

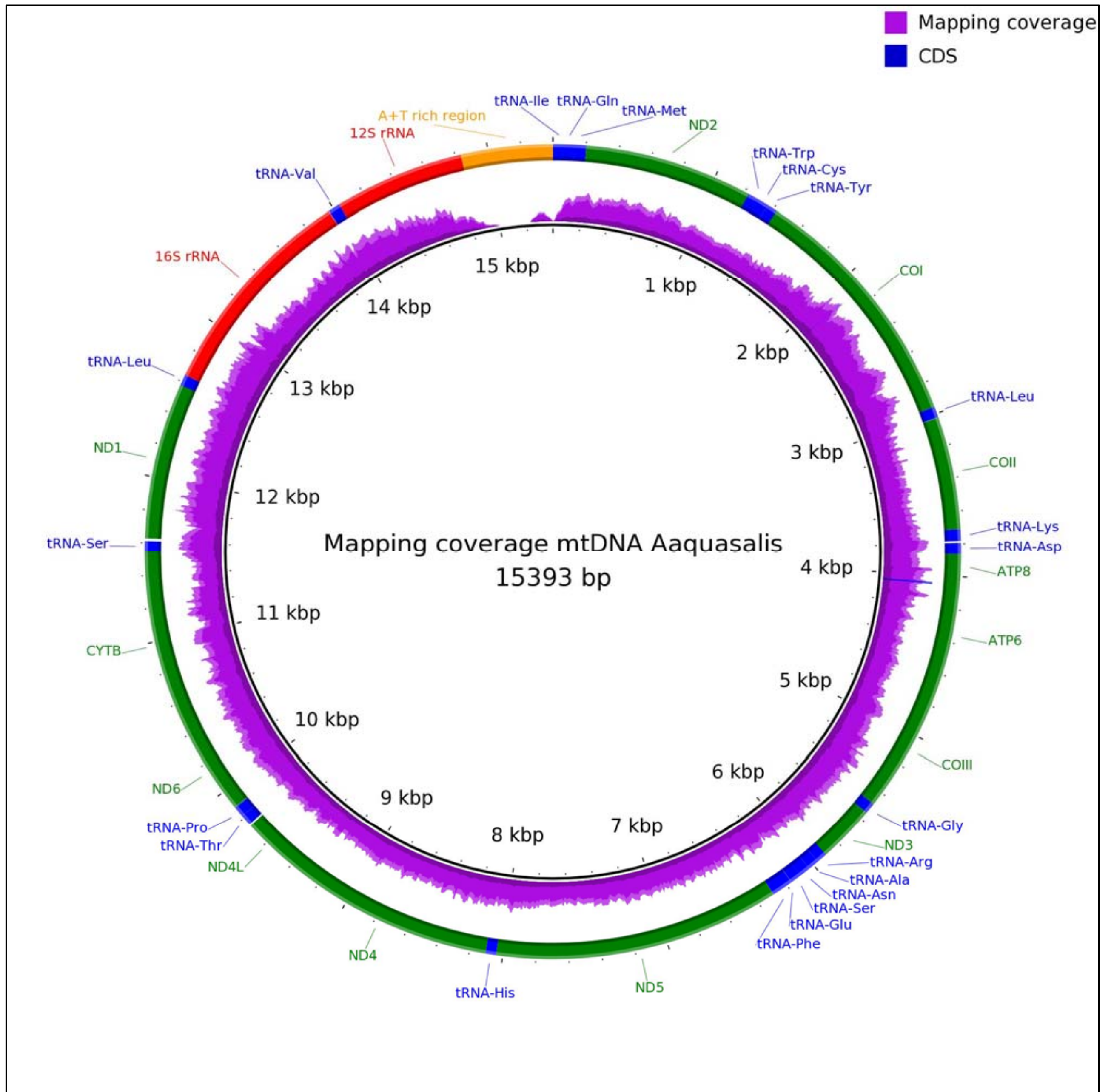
More metagenomic data from multiple species, life stages, ecological niches and phenotypes (blood fed, malaria infected, fed with nectar, insecticide resistant) will certainly broaden our knowledge on the microbial consortium associated to the *Anopheles* genus and the potential functions it plays within the mosquito host. Defining a core assembly of *Anopheles*-associated bacteria is a task that requires further meta-analysis to explore the results obtained with different filtering or grouping questions such as core per: anopheline species; life stage across anopheline species; aquatic ecotypes, etc.

This explorative meta-analysis is a reflection of current and available data, though we are aware it may be missing reports. Based on a hypothesis linking the geographic distribution of anophelines, and their compatibility with *Plasmodium* parasites, a core group of bacterial

families was identified amongst the *Anopheles* genus, when comparing between African and American anophelines. Nevertheless, more inclusive and homogeneous profiling studies (NGS and bioinformatics homogeneous methodologies) are required to better characterize this "core group" and the underlying factors that define it.

Bacterial abundance profiles (as both full and reduced data sets from this study did) may reflect the effects that biotic and abiotic factors have upon the bacterial community composition. Effectors such as nurturing larval habitats (Boissière *et al*., 2012, Tchioffo *et al*., 2013), laboratory colonization, and *Plasmodium* infection (Boissière *et al*., 2012) could act as "fine tuning" modulators of key OTU abundances within the vector-associated microbial constortium. Thus, signature profiles would reflect particular host ¨biological states¨. Understanding the dynamics of the interacting mosquito-parasite-microbiota triad may open new avenues to develop novel malaria control strategies.

Global climate change may already be altering the transmission dynamics of mosquito vector-borne diseases such as malaria, lymphatic filariasis, and dengue in many parts of the world (Afrane *et al*., 2012, Ermert *et al*., 2012, Sternberg & Thomas, 2014). One of the driving factors of such phenomenon would be the expansion of saline and brackish water bodies due to a rise in sea levels in coastal zones. As a consequence the densities of salinity-tolerant mosquitoes like *A. sundaicus*, *A. aquasalis* and *Culex sitiens* could increase. In addition, fresh water mosquito vectors like *A. culicifacies*, *A. stephensi*, *A. aegypti*, and *Aedes albopictus* could adapt to salinity (Ramasamy & Surendran, 2012). Future efforts to better understand the microbiome's connection with ecological niche driven factors that shaped vector competence, behavior, and distribution of anophelines across the world will be of use to better prepare for future malaria transmission scenarios.



(Martínez, 2015; unpublised)

4 Chapter 2:

Preliminary Characterization of the Mitochondrial Genome of *Anopheles (Nyssorhynchus) aquasalis* Curry, 1932. A Neotropical model of human malaria transmission.

Preliminary Characterization of the Mitochondrial Genome of *Anopheles (Nyssorhynchus) aquasalis* Curry, 1932. A Neotropical model of human malaria transmission.

Abstract

Whole mitogenome sequences (mtDNA) have been widely exploited for insect ecology studies, in particular using them as molecular markers to reconstruct phylogenies, or to infer phylogeographic relationships and dispersion routes.

Sequencing whole mitogenomes (amplification-free) is now a feasible endeavor performed by integrating Next Generation shotgun sequencing platforms with fast bionformatics' pipelines. Gene sequences, especially those of mitochondrial protein coding genes, have been frequently utilized as a tool to estimate time of divergence between different taxa. Nevertheless, the current understanding of the *Anopheles* phylogeny is limited and information regarding the time of deep lineage divergences within mosquitoes (Culicidae) is scarce, even in the postgenomic era.

Here we report the complete 15,393 bp mtDNA sequence of *Anopheles aquasalis*, a Neotropical human malaria vector and laboratory malaria transmission model in Brazil. When comparing its structure and base composition with other relevant and available anopheline mtDNA sequences high similarity and conserved features were observed. The mtDNA sequences of 22 anophelines and other dipteran siblings were analyzed using a Bayesian approach to reconstruct phylogenetic relationships and to estimate the date of divergence of the tested species. Bayesian analysis supports the conclusion that the most recent ancestor of *Nyssorhynchus* and *Anopheles* + *Cellia* was extant ~83 million years ago. It was estimated that *A. aquasalis* diverged from the *Anopheles albitarisis* complex ~28 MYA and ~38 MYA from *Anopheles darlingi*. Additional sampling and publication of African, Asian and North American anophelines will improve the resolution of the *Anopheles* phylogeny and their early dispersal routes.

4.1 Introduction

4.1.1 Insect mitochondrial genomes

The mitochondrial genome of arthropods has been studied extensively, with over 600 complete genomes currently available. This wealth of mtDNA-related information sets arthropods only second to vertebrates when considering the most studied metazoan phyla. Amongst the arthropod mitogenomes sequenced to date, insects represent approximately 80% (Cameron, 2014).

The insect mitogenome is a small double-stranded circular molecule of 14-20 kb in length containing 37 genes including: 13 protein-coding genes (PCG), 22 transfer RNA genes (tRNA) and two ribosomal RNA genes (small and large ribosomal subunits -rRNA-). The 13 PCG code for oxidative phosphorylation proteins classified as follows: 7 from complex I (ND1, ND2, ND3, ND4, ND4L, ND5, and ND6), one from complex III (cytochrome B), three from complex IV (COI, COII, and COIII), and two from complex V (ATPase6 and ATPase8). It additionally contains an A+T rich region that acts as a control region involved in the initiation of transcription and replication (Wolstenholme, 1992, Boore, 1999). The order of the genes within the genome is also highly conserved and it can be traced to the ancestral gene arrangement from the Bilateria (Mwinyi *et al*., 2009), which differs only slightly from ancestral ecdysozoan and arthropod mt genomes (Braband *et al*., 2010).

Mitochondrial genome sequences provide molecular resources to perform population genetic studies. They are useful as well to infer phylogenetic and phylogeographic relationships (Masta *et al*., 2009; Hickerson *et al*., 2010; Santini *et al*., 2013, Zhao *et al*., 2013, Perseke *et al*., 2013). Within these mtDNA sequences, informative patterns arise when the relative gene order, codon usage bias, gene expressivity, and RNA secondary structures are studied (Panday *et al*., 2010, Behura & Severson, 2012, Plazzi *et al*., 2013). Some of these traits have been used as additional sources of phylogenetic signal (Perseke *et al*., 2013, Simon & Hadrys, 2013, Wang *et al*., 2013b). Thus, inherent characteristics of the mitogenome such as, high copy number per cell, low recombination rate, high mutation rate, and dominantly maternal inheritance turn it into a powerful tool to perform evolutionary studies (Wang *et al*., 2013b, Silverio *et al*., 2014).

To enhance the results generated by insect phylogenetic studies performed with nuclear markers or mitogenomes, the relationship analyses and the inferences drawn from them, must take into account knowledge about the ecosystem harboring the subject of study

and the role the insect has within natural cycles. In this line of thought, arthropods are known participants of natural cycles including plant pollination, decomposition, and vectorial pathogen transmission (Yu *et al*., 2012).

4.1.2 Anopheline phylogenetics and divergence times

A total of 537 species of *Anopheles* are currently known, if the unnamed members of species complexes are taken into account (Harbach, 2011). From these species 87% have been formally named. Nevertheless, and until recently, little work had been done to understand the evolution, divergence time, and phylogenetic relationships between representative species within this genus despite its medical importance worldwide (Logue *et al*., 2013, Harbach, 2013).

Due to the existence of species complexes and incipient species, studies of this nature are challenging when regarding the *Anopheles* genus, moreover when considering the paucity of genetic data for most species besides *A. gambiae* (Krzywinski *et al*., 2001a, della Torre *et al*., 2001, Logue *et al*., 2013, Harbach *et al*., 2013). Nevertheless, the recent publication by the *Anopheles* Genomes Cluster regarding the sequencing of 16 anopheline nuclear genomes (Neafsey *et al*., 2015) may provide, in the near future, new full mitogenomes as parallel assemblies of the genomic data produced. This would enable more accurate phylogenetic reconstructions and enhance the estimates of the divergence times between the members of this genus.

The current hypothesis regarding *Anopheles* evolution is mostly based on the extant geographic distribution of *Anopheles* mosquitoes (Krzywinski *et al*., 2006). It proposes that major mosquito lineages like *Anopheles* originated in western Gondwana during the Early Cretaceous or late in the Jurassic period approximately 100 to 145 MYA (Krzywinski *et al*., 2006, Reidenbach *et al*., 2009). *Anopheles* probably emerged in what now is South America and, following rapid diversification, colonized most of the Earth's habitable environments migrating across the world aided by land bridges (Krzywinski & Besansky 2003, Krzywinski *et al*., 2006).

It is hypothesized that the *Anopheles* subgenera *Nyssorhynchus* and *Anopheles* + *Cellia* diverged ~ 79-100 MYA (Moreno *et al*., 2010, Marinotti *et al*., 2013, Logue *et al*., 2013), suggesting that their most recent common ancestor might have lived before the geological split of western Gondwana ~95 MYA (Marinotti *et al*., 2013).

The most ancient human colonization of the American continent is still a matter of discussion and is estimated to have occurred 15–20 millennia ago (Stringer, 2011). Thus, ancestral Neotropical anophelines evolved in an environment devoid of humans or human

ancestral species for several million years (Marinotti *et al*., 2013).

The human malaria cycle probably evolved in Africa, where malaria parasites along with their anopheline hosts and other hominid primates interacted (Harbach, 2013). As malaria-infected humans migrated out of Africa, they carried *P. falciparum* with them, leaving their African mosquito vectors behind. This migration gave rise to a journey that led the adaptation of the *Plasmodium* parasite to at least 34 different anopheline mosquito species worldwide (Sinka *et al*., 2012). According to historical, archaeological and genetic evidence, human malaria was introduced into the Americas by Europeans who transferred both, *P. falciparum* and *P. vivax* (the most prevalent malaria parasite species), to the indigenous population of the American continent (Castro & Singer, 2005, Culleton *et al*., 2011, Yalcidang *et al*., 2012). Therefore, interactions between Neotropical malaria vectors, humans, and malaria parasites, are relatively recent (Marinotti *et al*., 2013). The *Plasmodium*-compatible anophelines were surely already adapted to feeding on primates, including the ancestors of *Homo sapiens* (Harbach, 2013).

Efforts to interrupt malaria transmission and comprehend anopheline bionomics would benefit from the enhancement of the currently known phylogeny of *Anopheles* mosquitoes. This positive outcome would be based on the elucidation of descent relationships of genes associated with innate immune responses, refractoriness, insecticide resistance, and other genetically determined ecological and behavioral traits key to human malaria transmission (Foley *et al*., 1998). Furthermore, a clearer panorama of the evolutionary relationships among *Anopheles* species could help answer whether traits required for the transmission of human blood-borne pathogens evolved only once in an ancestral population or, if such traits were acquired independently by different species throughout their evolutionary history (Logue *et al*., 2013).

Here we present a preliminary characterization of the mitogenome of *A. aquasalis* assembled by Next Generation shotgun sequencing. We compared the mtDNA sequence, and some of its genome features, with other selected anopheline mitogenomes. Additionally, we applied Bayesian analysis to reconstruct phylogenetic relationships and estimate divergence times amongst anopheline human malaria vectors. The implications of this study's findings were briefly discussed with regard to the evolutionary history of anophelines in general and would certainly benefit from the existence of more mitogenomes representative of anophelines from Africa, Asia and North America to better describe the *Anopheles* dispersal routes throughout evolutionary and geological times.

4.2 Methods

4.2.1 *Anopheles aquasalis* colony

A. aquasalis were obtained from a colony established at the Medical Entomology Laboratory at Centro de Pesquisas René Rachou-FIOCRUZ (Fiocruz, Minas Gerais). The mosquitoes came from a colony established in 1995 in Rio de Janeiro (Arruda et al., 1982; Silva *et al*., 2006, Carvalho *et al*., 2002), and are kept at laboratory conditions as described recently by Rios-Velásquez *et al*. (2013) and Costa-da-Silva *et al*. (2014).

4.2.2 Single mosquito DNA extraction

Genomic DNA from a single *A. aquasalis* mosquito was extracted using the Qiagen DNeasy® blood and tissue kit (Qiagen, Hilden,Germany) according to the protocol for purification of insect DNA with a minor modification: EB rather than AE buffer was used to avoid possible interference of EDTA with the Nextera enzyme. The purified *Anopheles* genomic DNA was then quantified with the Qubit HS (Invitrogen, USA) system to proceed and build the genomic library.

4.2.3 Whole genome shotgun sequencing

The genomic DNA was processed using the Nextera-based DNA sample preparation kit (Epicentre Biotechnologies, Madison, WI). Thirty nanograms of sample DNA were fragmented utilizing 5 µl of Tagment DNA Enzyme with 25 µl of Tagment Buffer. Tagmentation reactions (DNA fragmentation and addition of the 19-bp inverted repeat to the fragment ends) were performed by incubating the sample for 5 min at 55°C followed by purification of the tagmented DNA using the Qiagen MinElute protocol (QIAGEN, Germany). Purified DNA was eluted from the column with 11 µl of nuclease-free water. Purified DNA (5 µl) was used as the template in a 20-µl volume for limited-cycle PCR (5 cycles) and processed as outlined in the Nextera protocol. Amplified DNA was purified using the AMPure Bead cleanup (Beckman Coulter, USA) according to the manufacturer's protocol. The fragment size distribution of the tagmented DNA was analyzed utilizing a 2100 Bioanalyzer with a 7500 DNA assay kit (Agilent Technologies, Santa Clara, CA). Fragments of ~600 bp long were carried out for sequencing. The library was sequenced on one lane of an Illumina HiSeq2000 instrument to generate paired-end reads. Sequencing performed by the

DNA sequencing facility at the University of California at Berkeley.

4.2.4 Mitochondrial genome assembly

Sequences were *de novo* assembled using Velvet v1.2.10 (Zerbino & Birney, 2008) with a k-mer size of 41, according the scripts and parameters suggested by the Velvet Manual (http://www.ebi.ac.uk/~zerbino/velvet/Manual) and suggestions from the personnel at the UC Davis Vector Genetics Laboratory (http://popi.ucdavis.edu/vgl/).

The assembled contigs were aligned to the mtDNA sequence of *An. gambiae* (GenBank No. L20934.1) using the MUMmer v3.0 software (Kurtz *et al.*, 2004) to identify and confirm the *de novo* assembly of the *A. aquasalis* mitogenome.

4.2.5 Sequence analysis: composition and genomic features of the *A. aquasalis* mitogenome

The assembled mitogenome was manually inspected for repeats at the beginning and end of the assembly to infer circularity. Automatic annotation of the mitogenome fasta file was performed with MITOS (Bernt *et al.*, 2012), followed by a more elaborate and manual curation based on the GenBank file format. Manual inspection of the predicted Protein Coding Genes (PCG's), ribosomal RNA (rRNA) genes, transfer RNA (tRNA) genes, and the AT rich region was performed with Artemis -release 16- (Rutherford *et al.*, 2000). The nucleotide sequences of PCG's were translated based on the invertebrate mtDNA genetic code.

The manual curation of coding regions and rRNA genes was mainly carried by sequence comparison with published insect mitogenome sequences such as *A. gambiae* (GenBank No. L20934.1) and *A. darlingi* (GenBank No.GQ918273.1) amongst others. Careful attention was given to PCG's comparing the predicted Open reading frames (ORF's) to the Uniprot database (http://www.uniprot.org/) giving more weight to similarity hits that had experimental validation.

The MITOS annotated tRNA's were verified and their secondary structures predicted with the tRNAscan-SE search server v1.21 (Lowe & Eddy, 1997) with default settings: the invertebrate mitochondrial codon predictors and a cove score cut off of 5. Tool available at: http://lowelab.ucsc.edu/tRNAscan-SE/. Some tRNA genes could not be detected by tRNAscan-SE. They were identified by direct comparison and sequence similarity to tRNA's of other dipterans or anophelines. They were modeled with RNAStructure ® (Bellaousov *et al.*, 2013). To maintain a uniform format, all of the 22 figures were generated with

RNAStructure ®.

To visualize the annotated mitogenome, a circular representation of it was generated with Blast Ring Image Generator (BRIG) (Alikhan *et al*., 2011). For such purpose a GenBank formatted file and the *.fasta file were employed according the Brig 0.95 Manual. Available at: http://ufpr.dl.sourceforge.net/project/brig/BRIGMANUAL.pdf.

Nucleotide composition analyses, expressed as AT%, were performed for individual PCG's, full mtDNA, concatenated PCG's, concatenated tRNA's, lrRNA (*16S*), srRNA (*12S*) and concatenated rRNA's. For the aforementioned targets, composition bias based on strand asymmetry values were estimated using the following formulae for skews: AT skew = [A-T]/[A+T] and GC skew = [G-C]/[G+C] as proposed by Perna & Kocher (1995) on an Excel ® spreadsheet. Codon bias was assessed estimating the relative synonymous codon usage (RSCU). All the above compositional analyses were performed using MEGA v6.0 (Tamura *et al*., 2013). The phylogenetic and comparative analyses performed henceforth relied upon available sequences and published literature regarding anopheline and culicine species detailed in **Table 1**.

4.2.6 Comparative analyses between anophelines from different geographic regions

Comparative analyses regarding nucleotide composition and strand asymmetry were performed between *A. aquasalis* and four other anophelines representative of different geographic regions: *Anopheles punctulatus* (GenBank No. JX219738.1) from South East Asia, *A. gambiae* (GenBank No. L20934.1) from Africa, *A. darlingi* North (GenBank No. GQ918272.1) from Central America, and *A. darlingi* South (GenBank No. GQ918273.1) from South America. For each mtDNA genome, base composition (expressed as AT%) and strand asymmetry (AT and GC skew) were calculated as explained above.

Additionally, a CDS nucleotide similarity comparison between *A. aquasalis*, the above cited anophelines, plus *Anopheles albitarisis* (GenBank No. HQ335349.1) --another South American brackish-water tolerant species-- was performed using the Blast2sequence online tool (Zhang *et al*., 2000). Available at: http://blast.ncbi.nlm.nih.gov/Blast.cgi#**.**

4.2.7 Phylogenetic analysis and molecular dating

To further our knowledge on the *Anopheles* genus phylogeny, as well as to estimate the divergence time or split between *A. aquasalis* and other Neotropical anophelines, we

reproduced the phylogeny and molecular dating analyses performed by Logue *et al*. (2013) using the tools and parameters suggested by the authors adding the assembled *A. aquasalis* mitogenome.

Briefly, 21 insect mitogenomes were selected from the ones used by Logue *et al*. (2013) and their sequences retrieved from the NCBI databank (**Table 1**). Then, for each of the 13 PCG's, the following actions were performed: the DNA sequences were translated into amino acid sequences using the invertebrate mt genetic code, then they were aligned to each other with the Muscle alignment engine, and the aligned amino acid sequences were reverse-translated back into nucleotide sequences. All the above steps were performed with the online tool Translator X (Abascal *et al*., 2010) using default parameters. Online tool available at: http://translatorx.co.uk/. Afterwards, the aligned sequences from all 13 mt genes were concatenated using FASconCAT (Kück & Meusemann, 2010). The concatenated PCG sequences from the 22 mitogenomes were analyzed with jModeltest v0.1.1 (Posada, 2008) to determine the best nucleotide substitution model according to the Akaike Information Criterion.

Bayesian phylogenies were reconstructed using BEAST v1.7.5 (Drummond & Rambaut, 2007) with the following program parameters: an uncorrelated lognormal relaxed clock model allowing for rate heterogeneity among species, the GTR + G substitution model, the SRD06 model of partitioning, which allows estimation of nucleotide substitution parameters separately for the 1st + 2nd and 3rd codon positions (this apparently provides a better fit for protein-coding nucleotide data), and a Yule model for tree reconstruction. With the above parameters, three independent runs of 20 million generations were preformed, saving the generated trees every 1,000 generations. All runs were then combined after a burn-in of 10% using LogCombiner v1.7.2 and afterwards Tracer v1.5 was used to verify the mixing of the Markov chains (both tools are part of the BEAST pipeline). The maximum credibility tree was determined using TreeAnnotator v1.7.2 and visualized with FigTree v1.4 available at: http://tree.bio.ed.ac.uk/software/figtree/ .

Divergence times were estimated using BEAST v1.7.5 following the instructions provided by the developers at: http://beast.bio.ed.ac.uk/. The *Drosophila*-*Anopheles* divergence time was set as the calibration point using a prior distribution normally distributed around a mean of 260 million years ago (MYA) ranging from 243 to 276 MYA as suggested by Gaunt & Miles (2002).

4.3 Results

4.3.1 Composition and genomic features

The complete mitogenome of *A. aquasalis* was assembled into a single contig of 15,393bp. The expected 37 genes in animal mtDNA, comprising 13 protein coding genes, two rRNA genes (*12S* and *16S*), 22 tRNA genes and a control region were identified (shown in **Table 2** and **Figure 11)**. Though not yet deposited, refer to the **Additional file 4**, Aaquasalis_mt.gb, to inspect the annotated features in the GenBank format. A short region of 229 bp, located within the A-T rich region, failed to be assembled due to low coverage.

The annotated genes are encoded on both the heavy (22 genes) and light (15 genes) strands with some ORF's overlapping adjacent genes. In total there are 38 overlapping nucleotides between 14 neighboring genes with the junctions spanning from 1 to 7 bp. Excluding the control region, there were 44 intergenic nucleotides (IGN's) at 7 locations with their lengths ranging from 1 to 17 bp. As in other dipterans and metazoans (Boore, 1999; Zhao *et al*., 2013, Plazzi *et al*., 2013) the most common start codon was ATG (6 PCG's). Incomplete or truncated termination codons were annotated in the following PCG's: *COI*, *COII*, *COIII*, *ND5*, and *ND4* (**Table 2**).

The mtDNA of *A. aquasalis*, as in other anophelines or insects (Moreno *et al*., 2010, Li *et al*., 2013), includes 22 tRNA genes with anticodons representing 20 different amino acids, with a length ranging from 64 -72 bp, and a total length of 1477 bp when concatenated. The lengths of the *12S* and *16S* rRNA genes are 793 and 1324 bp respectively, both being encoded on the light strand (L). The ends of both rRNA genes were assumed to extend to the boundaries of flanking genes as it is suggested for metazoans (Boore *et al*., 2005). As in other dipterans species, the *16S* rRNA gene is flanked by the $tRNA^{Leu}$ and $tRNA^{Val}$ genes while the *12S* rRNA gene is between $tRNA^{Val}$ and the control region. Their A-T contents were 82.5% and 79.9% respectively, and they both resemble the composition of other dipterans compared by Zhao *et al*. (2013).

The 22 predicted secondary structures of individual tRNA's are shown in **Figure 12**. All folded into the classic cloverleaf secondary structure except *tRNA-Ser* (GCT) in which the dihidrouridine stem “simply” formed a loop.

Table 1 List of the insect species used in this study with their corresponding GenBank number. For anopheline species, the continent in which they exert their malaria vectorial activity is shown. The sequence length reflects the number of base pairs assembled not considering N's.

Species	Family	Length (bp)	GenBank No.	Vector of malaria in (continent)	Reference
Anopheles aquasalis	*Culicidae*	15393		South America	This study
Anopheles punctulatus (isolate ITN_PNG-18)	*Culicidae*	15412	JX219738.1	Oceania	Logue *et al*., 2013
Anopheles farauti 4 (isolate 7_10-11)	*Culicidae*	15412	JX219735.1	Oceania	Logue *et al*., 2013
Anopheles farauti 4 (isolate 8_11-12)	*Culicidae*	15412	JX219736.1	Oceania	Logue *et al*., 2013
Anopheles hinesorum	*Culicidae*	15336	JX219734.1	Oceania	Logue *et al*., 2013
Anopheles koliensis (isolate ESP001B)	*Culicidae*	15412	JX219743.1	Oceania	Logue *et al*., 2013
Anopheles dirus A (isolate A1)	*Culicidae*	15404	JX219731.1	Southeast Asia	Logue *et al*., 2013
Anopheles dirus A (isolate A2)	*Culicidae*	15404	JX219732.1	Southeast Asia	Logue et al., 2013
Anopheles cracens (isolate B1)	*Culicidae*	15412	JX219733.1	Southeast Asia	Logue et al., 2013
Anopheles albitarsis F	*Culicidae*	15418	HQ335349.1	South America	Krzywinski *et al*., 2011
Anopheles albitarsis G	*Culicidae*	15474	HQ335346.1	South America	Krzywinski *et al*., 2011
Anopheles deaneorum	*Culicidae*	15424	HQ335347.1	South America	Krzywinski *et al*., 2011
Anopheles janconnae	*Culicidae*	15425	HQ335348.1	South America	Krzywinski *et al*., 2011
Anopheles oryzalimnetes	*Culicidae*	15422	HQ335345.1	South America	Krzywinski *et al*., 2011
Anopheles darlingi North	*Culicidae*	15386	GQ918272.1	Central America	Moreno *et al*., 2010
Anopheles darlingi South	*Culicidae*	15385	GQ918273.1	South America	Moreno et al., 2010
Anopheles quadrimaculatus (A strain Orlando)	*Culicidae*	15455	L04272.1	North America	Mitchell *et al*., 1993
Anopheles gambiae	*Culicidae*	15363	L20934.1	Africa	Beard *et al*., 1993
Culex pipiens	*Culicidae*	14856	HQ724614.1	NA	Atyame *et al*., 2011
Aedes aegypti	*Culicidae*	16655	EU352212.1	NA	Unpublished
Aedes albopictus	*Culicidae*	16665	NC_006817.1	NA	Unpublished
Drosophila melanogaster	*Drosophillidae*	19517	U37541.1	NA	Lewis *et al*., 1995

The nucleotide composition of the *A. aquasalis* mtDNA is biased towards a high A+T content. The overall AT% content in the whole mitochondrion sequence was 77.16%, and concordantly it had a 22.84% GC. The AT% for individual PCG's, long and short rRNA genes, and concatenated rRNA's and tRNA's are summarized in **Table 3**.

Table 2 Organization and gene features of the *A. aquasalis* mitochondrial genome.

Gene	Strand	Location	Length (bp)	IGN	Codon		AT %
					Start / Anti	Stop	
$tRNA^{Ile}$	H	1-68	68	-3	GAT		
$tRNA^{Gln}$	L	66-134	69	-1	TTG		
$tRNA^{Met}$	H	134-202	69	0	CAT		
ND2	H	203-1228	1026	-2	ATT	TAA	81.97
$tRNA^{Trp}$	H	1127-1295	69	-1	TCA		
$tRNA^{Cys}$	L	1295-1358	64	0	GCA		
$tRNA^{Tyr}$	L	1359-1423	65	-2	GTA		
COI	H	1422-2958	1537	0	TCG	T[a]	69.14
$tRNA^{Leu}$	H	2959-3024	66	2	TAA		
COII	H	3027-3711	685	0	ATG	T[a]	73.98
$tRNA^{Lys}$	H	3712-3782	71	11	CTT		
$tRNA^{Asp}$	H	3794-3861	68	0	GTC		
ATP8	H	3862-4023	162	-7	ATT	TAA	82.10
ATPase6	H	4017-4697	681	-1	ATG	TAA	74.01
COIII	H	4697-5483	787	0	ATG	T[a]	70.74
$tRNA^{Gly}$	H	5484-5550	67	0	TCC		
ND3	H	5551-5904	354	-2	ATA	TAA	76.55
$tRNA^{Arg}$	H	5903-5966	64	0	TCG		
$tRNA^{Ala}$	H	5967-6032	66	0	TGC		
$tRNA^{Asn}$	H	6033-6100	68	-1	GTT		
$tRNA^{Ser}$	L	6100-6166	67	1	GCT		
$tRNA^{Glu}$	H	6168-6233	66	-2	TTC		
$tRNA^{Phe}$	L	6232-6298	67	0	GAA		
ND5	L	6299-8039	1741	0	GTG	T[a]	78.22
$tRNA^{His}$	L	8040-8104	65	0	GTG		
ND4	L	8105-9446	1342	-7	ATG	T[a]	76.81
ND4L	L	9440-9739	300	10	ATG	TAA	83.00
$tRNA^{Thr}$	H	9750-9817	68	0	TGT		
$tRNA^{Pro}$	L	9818-9883	66	2	TGG		
ND6	H	9886-10410	525	-1	ATT	TAA	84.76
CYTB	H	10410-11546	1137	-2	ATG	TAA	72.56
$tRNA^{Ser}$	H	11545-11610	66	17	TGA		
ND1	L	11628-12584	957	-6	ATA	TAA	76.70
$tRNA^{Leu}$	L	12579-12644	66	0	TAG		
16S rRNA	L	12645-13968	1324	2			
$tRNA^{Val}$	L	13971-14042	72	0	TAC		
12S rRNA	L	14043-14835	793	0			
Control region		14836-15393	558	0			

[a] Termination codons completed via polyadenylation

IGN: Intergenic nucleotide; minus indicates overlapping between genes.

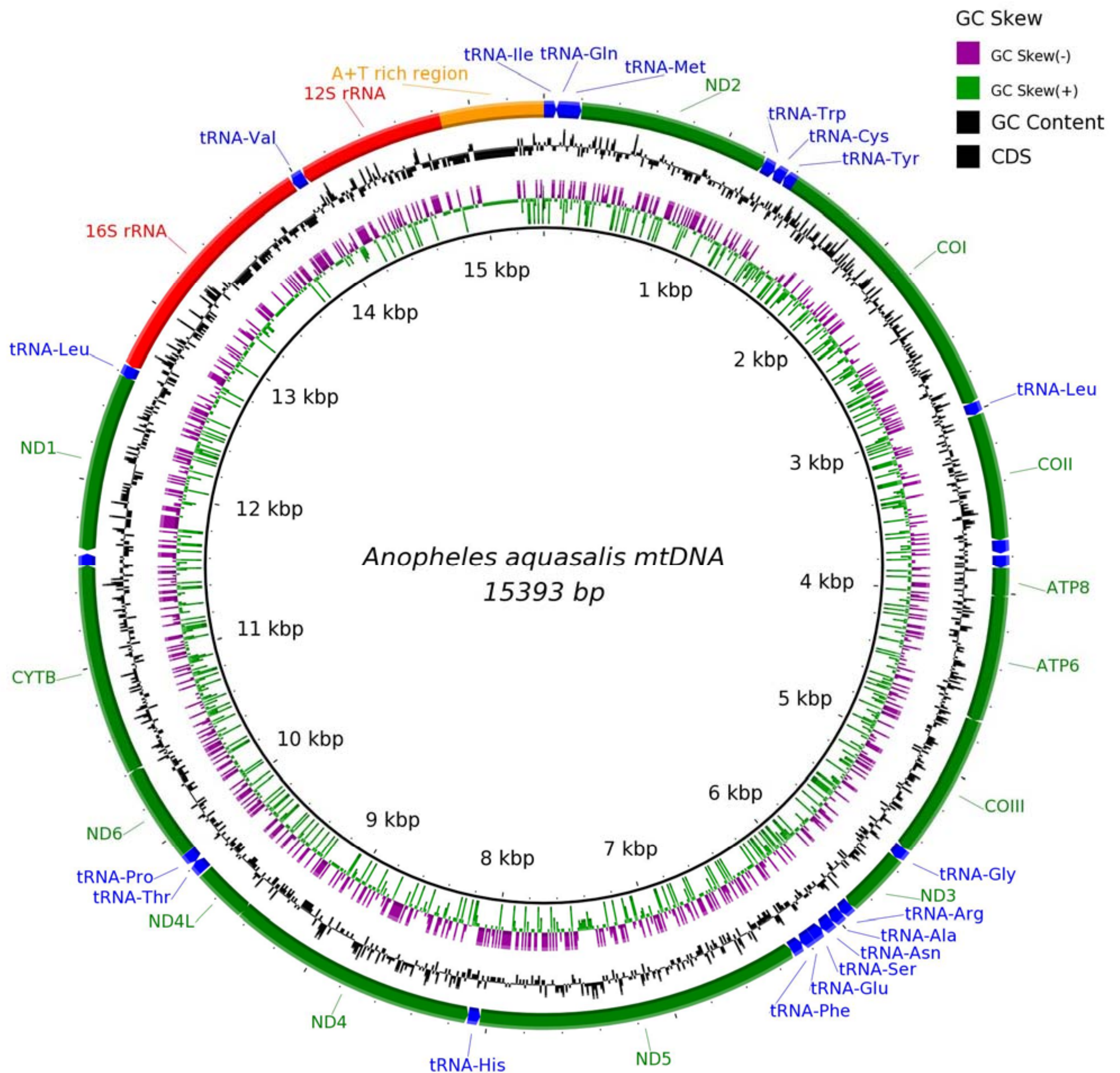


Figure 11 The complete mitogenome of *A. aquasalis*. BRIG visualization showing the protein coding genes, rRNA's and tRNA's in the mtDNA genome of *A. aquasalis*. The black inner ring shows the GC content on the outer surface, whereas AT content is shown on the inner surface. Strand asymmetry of the mitogenome is shown by the GC (+) and (-) skews according to the color key shown in the legend.

Base composition, measured by strand asymmetry analyses (AT and GC skews), were also computed and are shown in **Table 3**. Whole anopheline mitogenomes have AT- and GC-skews similar to the usual strand compositional biases reported for other metazoan mtDNA (Moreno *et al*., 2010, Wang *et al*., 2013b). The complete *A. aquasalis* mitogenome has a positive AT skew and a negative GC skew for the majority strand (also known as light -L-), which means that this strand is richer in A and C.

The *A. aquasalis* mitogenome contains 3743 codifying codons whose usage is shown in **Table 4**. Almost all of the codons were present except AGG (S). There are 29 codons (out of 32) of the NNU and NNA types with a relative codon usage bias (RSCU) greater than 1. A strong bias towards A+T rich codons was observed, being the five most prevalent codons (in descending order): UUA (L), UUU (F), AUU (I), AUA (M), and AAU (N) as shown in **Table 5**. The most used codon is UUA (L) and the less used codon in the genome is CGC(R). The mitogenome is rich in leucine while cysteine is the rarest amino acid.

Table 3 Base composition analysis of the *A. aquasalis* whole mitochondrion DNA sequence, and other annotated features such as individual PCG's, and concatenated rRNA and tRNA genes.

Feature	AT%	AT skew	GC skew
ND2	81.97	-0.0868	-0.1459
COI	69.16	-0.1214	-0.0127
COII	74.01	-0.0533	-0.0562
ATP8	82.10	-0.1429	-0.3793
ATP6	74.01	-0.1111	-0.1638
COIII	70.78	-0.0736	-0.0783
ND3	76.55	-0.0775	-0.1325
ND5	78.23	-0.1733	0.2454
ND4	76.83	-0.2454	0.2476
ND4L	83.00	-0.2851	0.2549
ND6	84.76	-0.0562	-0.3250
CYTB	72.56	-0.0836	-0.0641
ND1	76.70	-0.2834	0.3184
PCG	75.90	-0.1441	0.0354
Genome	77.16	0.0373	-0.1684
lrRNA	82.48	-0.0714	0.3362
srRNA	79.95	-0.0347	0.2704
rRNA_concat	81.53	-0.0579	0.3095
tRNA_concat	78.54	0.0103	0.1293

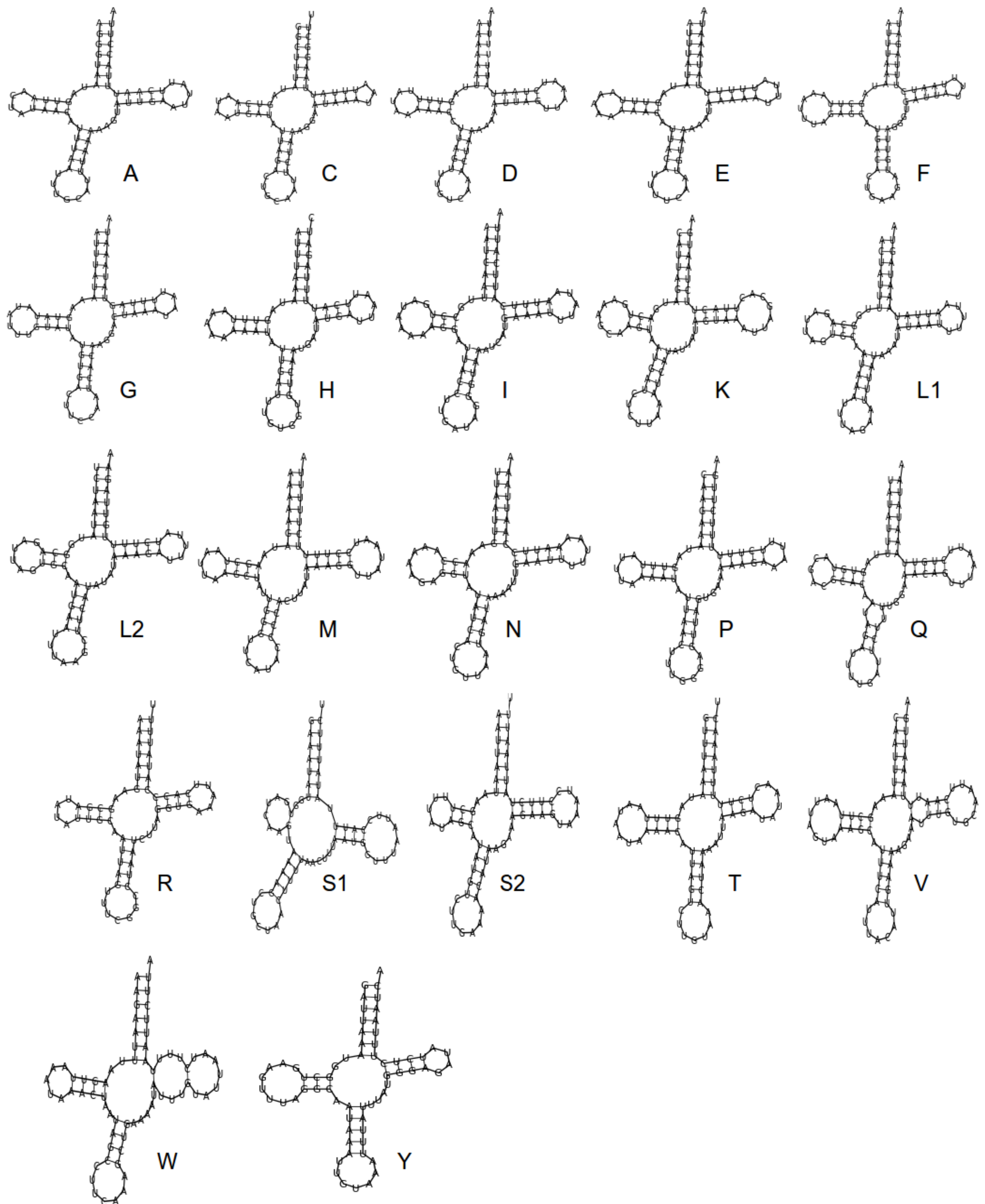


Figure 12 *A. aquasalis* mtDNA predicted tRNA structures. 22 tRNA's were identified in the mitogenome of *A. aquasalis* and their cloverleaf secondary structures predicted with RNAStructure ®. Using **Dayhoff's single letter amino acid code**.

Table 4 Codon usage in the *A. aquasalis* mtDNA. Codon frequency and Relative synonymous Codon Usage (RSCU) are shown for each codon.

Codon	Count	RSCU	Codon	Count	RSCU	Codon	Count	RSCU	Codon	Count	RSCU
UUU(F)	328	1.78	UCU(S)	91	2.54	UAU(Y)	159	1.72	UGU(C)	36	1.80
UUC(F)	40	0.22	UCC(S)	13	0.36	UAC(Y)	26	0.28	UGC(C)	4	0.20
UUA(L)	479	4.53	UCA(S)	81	2.26	UAA(*)	43	1.26	UGA(W)	74	1.85
UUG(L)	49	0.46	UCG(S)	6	0.17	UAG(*)	25	0.74	UGG(W)	6	0.15
CUU(L)	44	0.42	CCU(P)	65	2.24	CAU(H)	70	1.69	CGU(R)	14	1.10
CUC(L)	11	0.10	CCC(P)	8	0.28	CAC(H)	13	0.31	CGC(R)	2	0.16
CUA(L)	41	0.39	CCA(P)	39	1.34	CAA(Q)	75	1.47	CGA(R)	31	2.43
CUG(L)	10	0.09	CCG(P)	4	0.14	CAG(Q)	27	0.53	CGG(R)	4	0.31
AUU(I)	311	1.89	ACU(T)	93	2.14	AAU(N)	170	1.81	AGU(S)	44	1.23
AUC(I)	18	0.11	ACC(T)	9	0.21	AAC(N)	18	0.19	AGC(S)	8	0.22
AUA(M)	204	1.77	ACA(T)	68	1.56	AAA(K)	77	1.40	AGA(S)	44	1.23
AUG(M)	27	0.23	ACG(T)	4	0.09	AAG(K)	33	0.60	AGG(S)	0	0.00
GUU(V)	67	1.54	GCU(A)	82	2.23	GAU(D)	70	1.47	GGU(G)	30	0.63
GUC(V)	7	0.16	GCC(A)	14	0.38	GAC(D)	25	0.53	GGC(G)	6	0.13
GUA(V)	93	2.14	GCA(A)	46	1.25	GAA(E)	78	1.71	GGA(G)	124	2.61
GUG(V)	7	0.16	GCG(A)	5	0.14	GAG(E)	13	0.29	GGG(G)	30	0.63

Table 5 Top 10 codons (classified by count) identified in the protein coding genes of the *A. aquasalis* mitogenome.

Aminoacid	Codon	Count	RSCU
Leucine	UUA(L)	479	4.53
Phenylalanine	UUU(F)	328	1.78
Isoleucine	AUU(I)	311	1.89
Methionine	AUA(M)	204	1.77
Asparagine	AAU(N)	170	1.81
Tyrosine	UAU(Y)	159	1.72
Glycine	GGA(G)	124	2.61
Valine	GUA(V)	93	2.14
Threonine	ACU(T)	93	2.14
Serine	UCU(S)	91	2.54

4.3.2 Comparative composition and identity analyses between selected anophelines

Compositional patterns based upon AT% and strand asymmetry were sought by comparing the mtDNA sequences from *A. aquasalis* and other human malaria vectors selected for their representative geographic distribution. The computed data and comparative approach rendered the plots shown in **Figure 13**. Overall, the plots represent the similar trends of AT% and skew patterns between the compared features, amongst the evaluated anophelines. Matching profiles between invertebrtates (mollusks) had been reported before by Plazzi *et al.*

(2013) via this type of plots. The observed trend had a few exceptions, like the estimated GC skews from *COI* (from *A. darlingi* North and *A. darlingi* South) and *COII* (from *A. darlingi* South).

Table 6 Protein coding genes nucleotide sequence identity comparison between *A. aquasalis* and other selected anophelines.

Gene (CDS)	Identity				
	vs *A. albitarsis* F	vs *A. darlingi* N	vs *A. darlingi* S	vs *A. gambiae*	vs *A. punctulatus*
ND2	93.55	92.98	92.98	89.06	86.42
COI	93.29	91.72	91.66	89.31	89.27
COII	92.84	91.53	91.09	89.03	88.75
ATP8	94.44	98.11	96.85	91.82	90.74
ATP6	92.95	90.74	91.33	90.01	87.51
COIII	91.74	91.74	91.48	89.45	88.19
ND3	92.04	92.93	92.93	87.85	84.74
ND5	92.70	92.87	92.70	87.77	86.90
ND4	91.74	90.98	91.20	87.58	86.89
ND4L	93.33	95.33	95.66	91.00	90.00
ND6	94.47	92.00	92.19	87.33	86.83
CYTB	96.65	92.34	92.17	89.97	89.79
ND1	93.75	93.52	93.62	90.13	89.73
16s RNA	98.03	97.50	97.73	94.31	94.15
12s RNA	98.86	98.36	97.85	95.11	93.86
	AMERICA			AFRICA	SE ASIA / OCEANIA

When comparing *A. aquasalis* coding nucleotide sequences and rRNA genes with their corresponding counterparts from the selected anopheline species, higher identity values were obtained with the American anophelines than with those from Africa and Asia. This was the general trend across all the compared PCG's and rRNA genes (**Table 6** and **Figure 14**).

4.3.3 Phylogeny and molecular dating analyses

The protein coding DNA sequences of the *Anopheles* mtDNA genomes included in this part of the study were analyzed considering, in addition, the following outgroups: *D. melanogaster*, *Cx. Pipiens*, *Ae. albopictus*, and *Ae. Aegypti* (GenBank numbers shown in **Table 1**). The aligned and concatenated sequences from the 22 mitogenomes resulted in a block of 11,514 nucleotides. According to the Akaike Information Criterion, the best nucleotide substitution model for this data set was the General Time Reversible with gamma distribution (GTR + G) model.

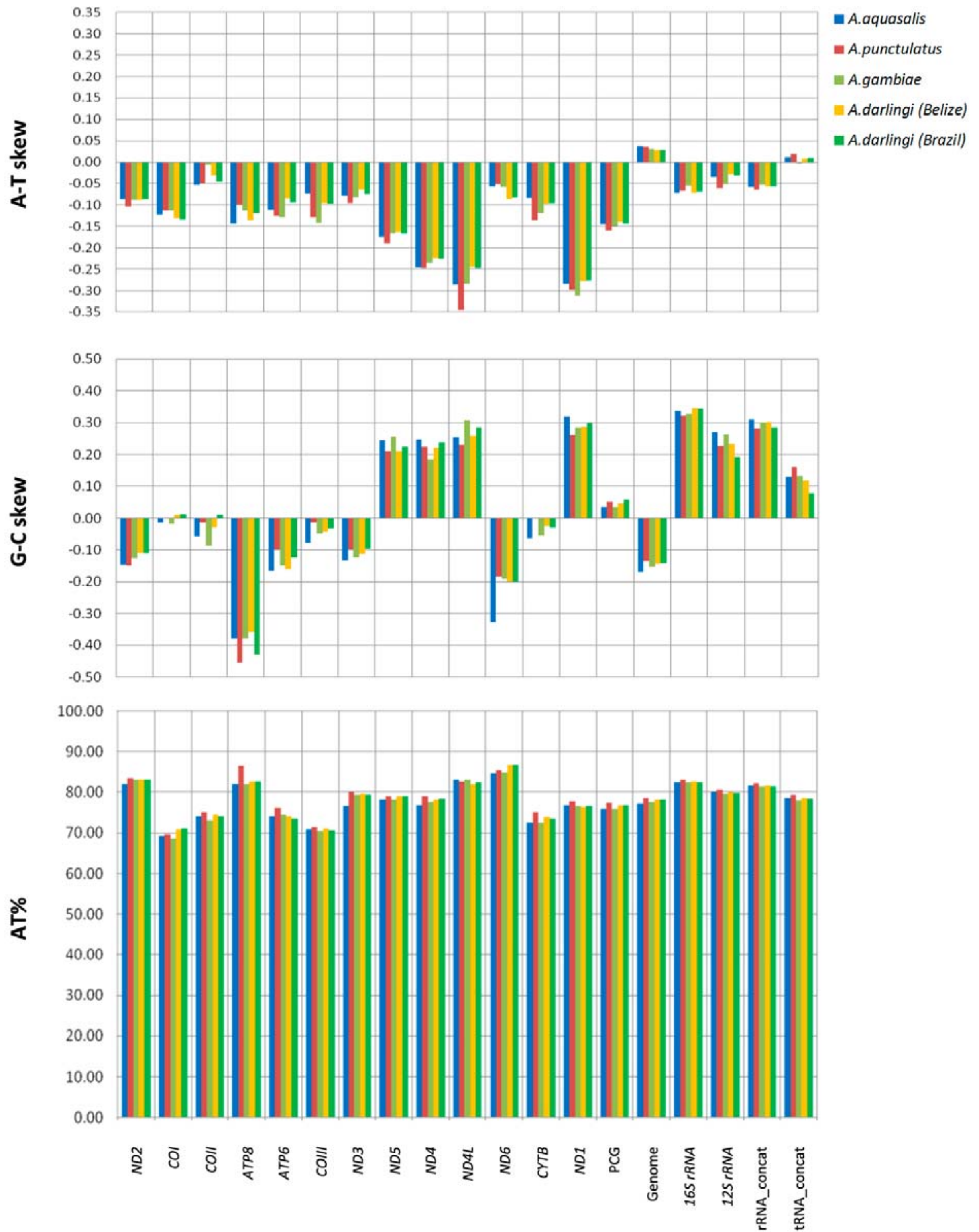


Figure 13 Compositional patterns of *A. aquasalis* and other anopheline mitochondrial genomes. A-T content expressed as AT%, AT skew, and GC skew were estimated and plotted for each single PCG and for other genomic regions according to the legend below the chart. Refer to the Supplementary Tables 4 and 5 for specific values.

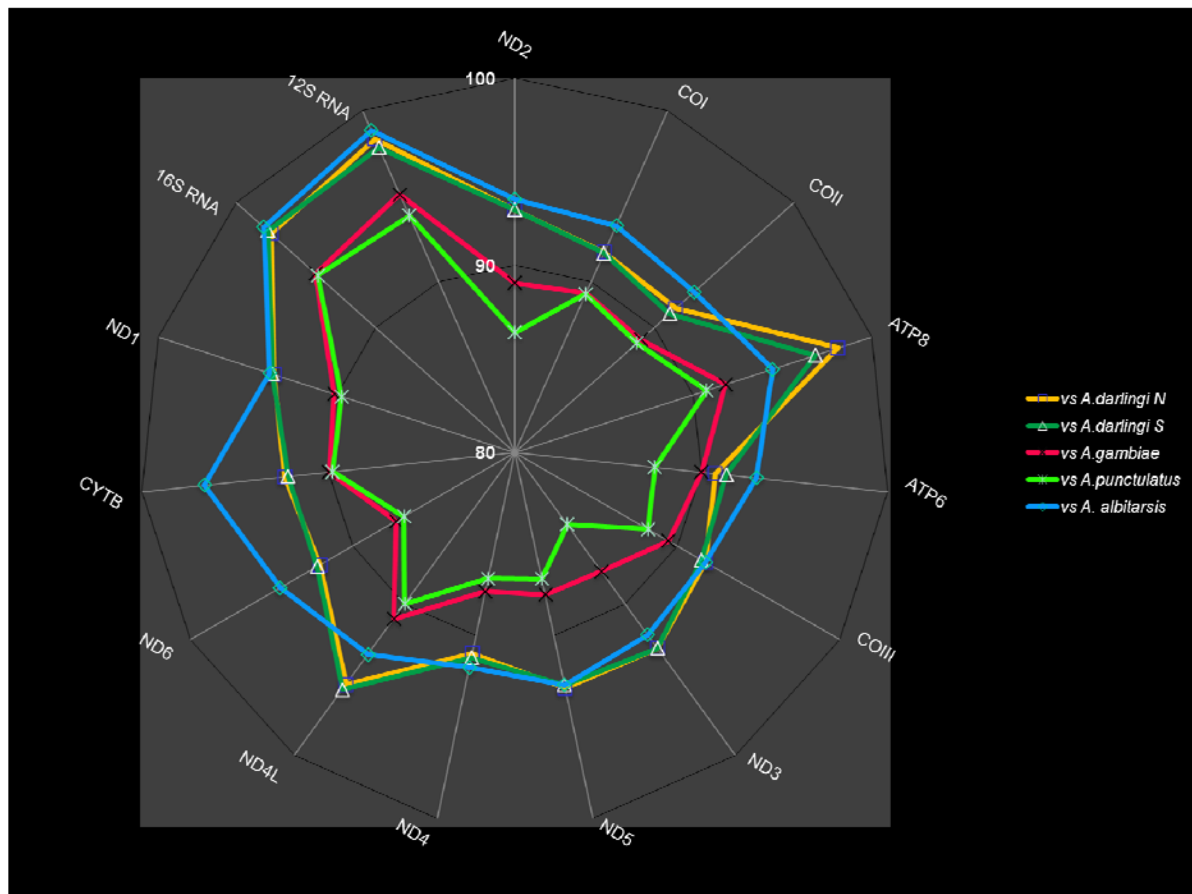


Figure 14 Radar depiction of the nucleotide identity comparison between mitochondrial PCG's and rRNA genes from *A. aquasalis* and other anophelines. Nucleotide identity (%) between *A. aquasalis* and selected *Anopheles* species ranging from 80 to 100%. The scale is shown along the main vertical axis of the radar plot.

A phylogenetic tree was reconstructed using a Bayesian analysis with BEAST v1.7.5 (Drummond & Rambaut, 2007). All the phylogenetic relationships were supported with robust posterior probabilities greater than 90%, with exception of the position of *A. gambiae* (42%) and the internal nodes amongst the South East Asia and Oceania anophelines included (ranging from 85 to 88%). The reconstructed *Anopheles* phylogenetic tree is shown in **Figure 15**. From examining the tree topology we observed a deep divergence between two *Anopheles* geographical lineages. The results, when considering the continental distribution of the anopheline species studied, clearly separate Central and South American anophelines. The other lineage seems to have radiated first, to generate North American anophelines, followed

by those in Africa, and from there the anophelines present in South East Asia and Oceania. This radiation pattern resembles the one published by Logue *et al*. (2013). The monophyletic clades corresponding to the *A. punctulatus*, *A. dirus* and *A. albitarsis* groups are also clear.

The currently available fossil record for mosquitoes is poor (Poinar *et al*., 2000, Harbach, 2013). Thus, we estimated the divergence times amongst anopheline species using the *Drosophila-Anopheles* divergence time (set at 260 MYA) as the only calibration point time (Gaunt & Miles, 2002). We dated the most recent common ancestor (MRCA) of all *Anopheles* to 83.23 MYA with a 95% credibility interval ranging from 54.33 to 115.88 MYA as shown in **Table 7**. Amongst Neotropical anophelines, the MRCA within the *A. albitarsis* complex and *A. aquasalis* dates to 28.56 MYA with a 95% credibility interval ranging from 17.10 to 42.12 MYA. This MRCA is younger than the one shared between *A. darlingi* and the *A. albitarsis* complex, which dates back to 38.98 MYA as it can be observed in **Table 7** and **Figure 16**.

Table 7 Mean divergence times in MYA (million of years ago), and 95% credibility intervals for selected nodes.

MRCA	Mean (mya)	95% Credibility (mya)
Drosophila / *Anopheles* (Calibration point ~260mya)	258.85	[239.70-278.86]
Anophelinae / *Culicinae*	136.63	[90.13-189.15]
Anopheles genus	83.23	[54.33-115.88]
Anopheles gambiae / SE Asia - Oceania anophelines	65.50	[43.54-89.80]
Anopheles darlingi / *Anopheles albitarsis* complex	38.98	[23.54-58.32]
Anopheles aquasalis / *Anopheles albitarsis* complex	28.56	[17.10-42.12]

4.4 Discussion

Greater integration of nuclear and mitochondrial genomic studies is necessary to further our understanding of the *Anopheles* genomic evolution, the phylogenetic relationships of vectorial competence traits, and the co-evolutionary history of this genus with human malaria parasites. Thus, in the current "omics" era, extending the wealth of representative anopheline mitogenomes availables is a necessary and feasible task.

Like in most metazoans, the assembled mitochondrial genome of *A. aquasalis* is a small, double-stranded circular DNA molecule. Encompassed in 15,393 base pairs, we annotated a densely packaged set of 13 protein-coding genes, 22 tRNA genes, and two rRNA genes as shown in **Figure 11**. Highly conserved features regarding genome structure, composition and organization have been reported among insects (Cameron, 2014) such as anophelines (Moreno *et al*., 2010).

When regarding genome architecture, the occurrence of overlapping open reading frames implies a compact genome structure. This feature can be observed in compactly packaged mtDNA sequences of culicines, anophelines, and other insect mitogenomes (Stewart & Bechenbach, 2005, Behura *et al*., 2011, Li *et al*., 2013). The reason for such characteristic can be (structurally) attributed to the small intergenic distances between consecutive genes (intergenic distance < x bp): < 23 bp in *A. aegypti*; < 24 bp in *C. quinquefasciatus*; < 21 bp in *A. albopictus*; < 16 bp in *A. gambiae*; < 18 bp in *Anopheles quadrimaculatus*; < 30 bp in *Drosophila yakuba*, and < 30 bp in *D. melanogaster* (Behura *et al*., 2011) and in the case of *A. aquasalis* < 17 bp. Once a free-living prokaryote, mitochondria underwent genome reduction as it became an obligate endosymbiont tightly harbouring a remnant of the eubacterial genome of its ancestor (Timmis *et al*., 2004).

The intergenic distances do not account for the control region, which in the *A. aquasalis* mitogenome spans 558bp. This major non-coding region is also known as the A+T-rich region, and it plays a role in initiation of transcription and replication (Wolstenholme, 1992). The length of this region is highly variable among different insects due to its high rate of nucleotide substitution, insertions/deletions, and the presence of a variable number of tandem repeats (Fauron & Wolstenholme, 1980, Inhorita *et al*., 1997).

Almost all the tRNA sequences annotated could be folded into typical cloverleaf secondary structures exposing the adequate anticodon triplet, except for the DHU-arm of $tRNA^{Ser}$, which is entirely absent as it has been observed in other insects (Li *et al*., 2013,

Zhao *et al*., 2013 and references therein). Deficiencies of tRNA genes are often observed in protozoans, fungi, algae, plants and low metazoans (Scheneider, 2000). Aberrant loops, non-Watson-Crick matches or short arms may induce aberrant tRNA's to lose their function, but a "corrective" post-transcriptional RNA editing mechanism has been proposed (Tomita & Weiner, 2001, Li & Guan, 2010b).

The tRNA genes are embedded in variable regions within the mitogenome (Cameron, 2013). Throughout evolution, these regions underwent rearrangements more often than protein coding regions. Therefore, tRNA order is nowadays explored as an additional tool for comparative phylogenetic analysis between species (Silverio *et al*., 2014).

The nucleotide composition of the complete *A. aquasalis* mtDNA sequence is clearly biased towards a high A+T content (77.16%) as it can be seen in **Table 3**. This result was similar to the base composition described for other anophelines with which a comparison (considering strand asymmetry also) was drawn and is discussed further below.

We report here that the majority strand (L) of the *A. aqualsalis* mtDNA has a compositional bias (positive AT skew and a negative GC skew). This means that this strand is richer in A and C. Though this is the trend in insects, exceptions exist in arthropod mitogenomes in which strand asymmetry is reversed towards a composition with less A than T and less C than G on the majority strand (Wei *et al*., 2010). The underlying mechanism that leads to the strand bias has been generally related to the asymmetric replication, and transcription processes. During both, one strand is transiently in a single-stranded state and thereby exposed to more DNA damage. This phenomenon has been widely considered to bias the occurrence of mutations between the two complementary mtDNA strands (Hassanin *et al*., 2005).

Most of the start codons of the annotated PCG's followed the ATN rule described previously in other anopheline mitogenomes (Beard *et al*., 1993, Moreno *et al*., 2010). This was not the case for *COI* and *ND5*; genes that have TCG and GTG start respectively (**Table 2**). The latter start codon characteristic has been described in other anophelines (Mitchell *et al*., 1993; Beard *et al*., 1993; Moreno *et al*., 2010). As observed in *A. darlingi* (Moreno *et al*., 2010), eight genes use the complete stop codon TAA except for *COI*, *COII*, *COIII*, *ND5*, and *ND4*, which terminate with a single T (**Table 2**). The incomplete stop codon is a reported phenomenon in mitogenomes of insects (Moreno *et al*., 2010, Behura *et al*., 2011, Wang *et al*., 2013b, Zhao *et al*., 2013) and it has been proposed that the complete termination codon (TAA) is synthesized by a post-transcriptional polyadenilation reaction (Wei *et al*., 2009).

The total number of non stop codons within the predicted PCG's (3743) was similar to that reported for *A. darlingi* and *A. gambiae* (3733), differing slightly to that of *A. quadrimaculatus* (3715) (Moreno *et al*., 2010). The nucleotide bias was also reflected in the codon usage within the annotated PCG's. A strong bias towards A+T rich codons like TTA (L), TTT (F), ATT (I), ATA (M), and AAT (N) is a compositional trend that follows the pattern reported for other Culicidae species (Moreno *et al*., 2010, Behura *et al*., 2011). When compared to other insects, such as aphids, the same five codons were the most prevalent, with F and I exchanging positions (Wang *et al*., 2013b).

Among the 32 most frequently used codons (RSCU > 1) 29 were of the NNU and NNA type. This indicates that the third position of the codons are mostly U/A which differs from the expected for dipterans in which, frequently used codons have G or C in the third position (Behura & Severson, 2012).

The mitochondrial genome of insects displays unique characteristics such as high codon bias, low GC_3 content, and highly conserved gene arrangement (Cameron, 2013). These characteristics are associated with well-conserved patterns of Shine-Dalgarno sequences in the transcripts of mitochondrial functional proteins. In sum, the aforementioned characteristics promote the high expression level of the mitochondrial genes (Duret & Mouchiroud, 1999, Hiraoka *et al*., 2009). In insects and nematodes, this high gene expression would be linked with maintaining mitochondrial physiology under extreme conditions, potentially leading to increased stress resistance (Panday *et al*., 2010). In the case of anophelines, it is interesting to consider linking this increased cellular stress resistance hypothesis with the detoxification mechanisms (Jaramillo-Guitierrez *et al*., 2010) towards the oxidative stress triggered as a response to the immune challenge posed by the *Plasmodium* infection (Bahia *et al*., 2013; Shrinet *et al*., 2014).

The sequence identity comparison and compositional patterns analyses we performed showed how the nucleotide sequence of protein-coding genes are a conserved feature of anopheline mitogenomes, in which most differences have been reported within the control region (Oliveira *et al*., 2007, Moreno *et al*., 2010). In terms of sequence identity (all values > 86%), a trend was distinguished (**Figure 14**). In general, the *A. aquasalis* PCG's are more similar to those of the other Neotropical anophelines tested, when compared to those from *A. gambiae* and *A. punctulatus* (African and Asian malaria vectors). Higher sequence similarity could be obeying the phylogeographic and divergence history of anophelines. Notice that, in average, the highest similarity values were obtained with *A. albitarsis*, another brackish-water

Neotropical anopheline, from which *A. aquasalis* diverged ~28 MYA (**Figure 16**). The identity between rRNA sequences ranged from 93% to 98% (**Table 6**). Similar identity values described in aphids prompted Wang *et al*. (2013b) to suggest that potentially conserved (predicted) rRNA secondary structures could enhance phylogenetic analyses between insect species.

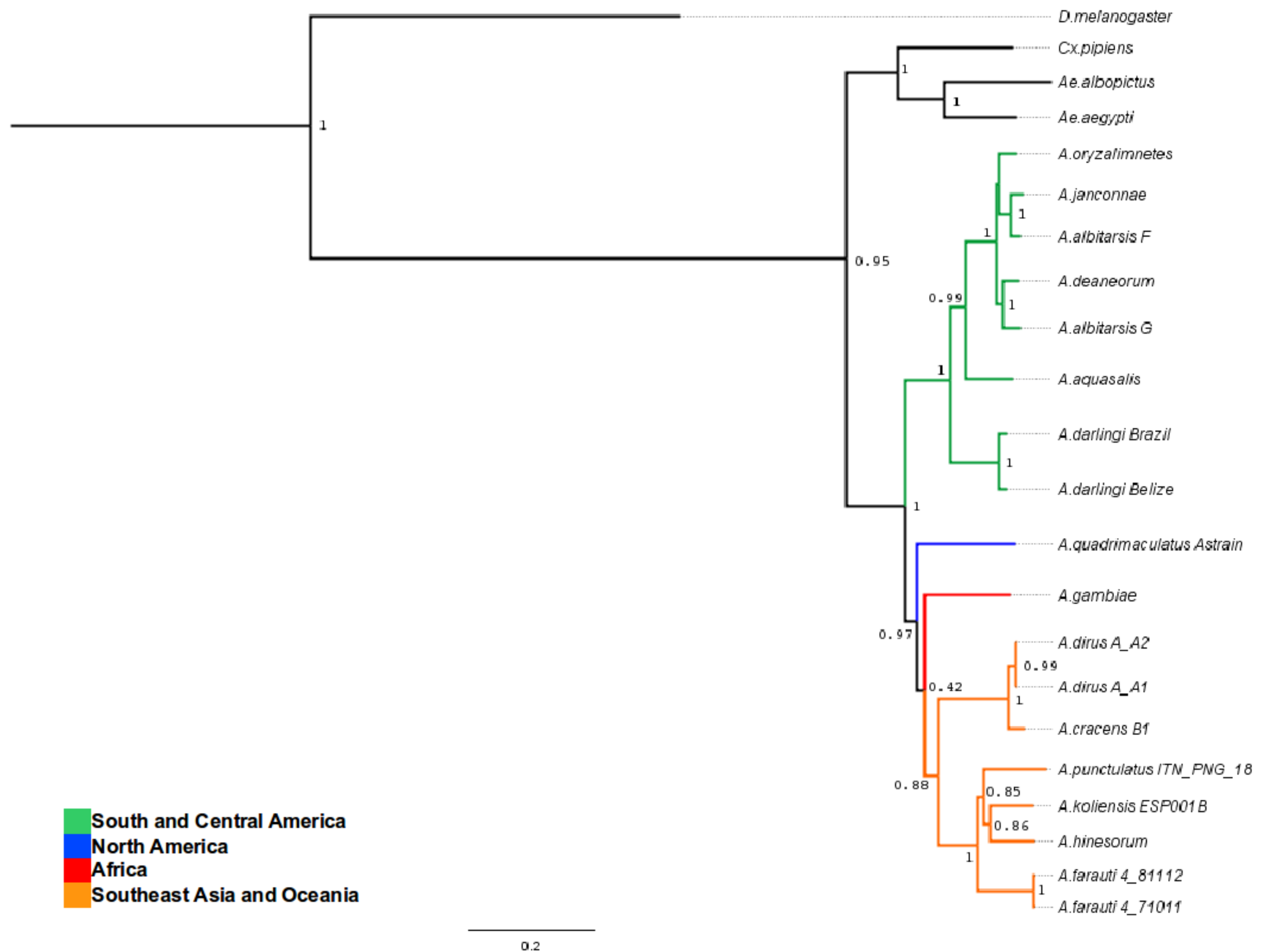


Figure 15 Support of the *Anopheles* phylogeny using the concatenated DNA sequences of all the mitochondrial protein coding genes. The values on the nodes correspond to the posterior probabilities supporting the tree topology.

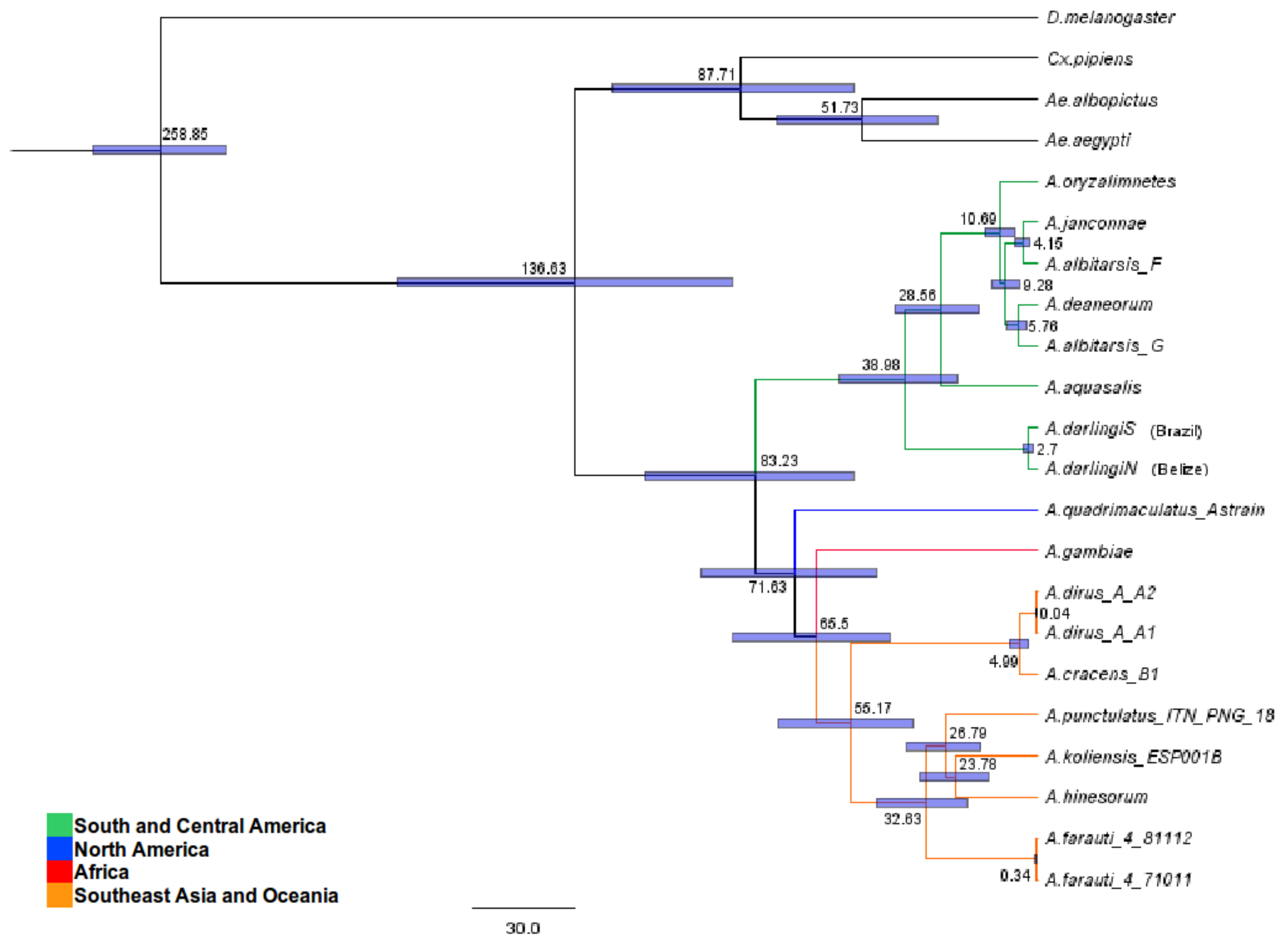


Figure 16 Phylogenetic tree of *Anopheles* using the concatenated DNA sequences of all the mitochondrial protein coding genes. The values on the tree nodes correspond to the mean divergence time (MYA) estimated for each event. The bars illustrate the 95% credibility intervals of the divergence times. Refer to Table 7 for selected key divergence events.

Human malaria vectors from different species complexes are not closely related. This implies that at least some of the "competent vector" traits arose independently, multiple times in different anopheline lineages. Furthermore, the presence of both vector and non-vector species in the same species complex implies either rapid loss or rapid gain of vector traits.

The evolutionary history of genes associated with vectorial capacity might have been driven by a rapid evolution process; such process got triggered when (compatible) mosquitoes began interacting with humans and *Plasmodium* parasites. This would mean that these genes

are not highly conserved genomic sequences originated from a single ancestor (Besansky, 2008). These, amongst other observations, were key elements of the rationale behind the effort to sequence the genome of multiple malaria vectors by The *Anopheles* Genomes Cluster.

The recent publication of 16 anopheles genomes by the above mentioned scientific consortium (Neafsey *et al*., 2015) might have generated concomitantly, mitogenomes that will broaden our perspective on the phylogenetic and divergence relationships between the members of the *Anopheles* genus. This would be particularly important since both our study, and Logue *et al*. (2013), showed the paradoxical situation regarding the scarcity of mitogenomes from dominant malaria vectors from Africa (other than *A. gambiae*) and the Middle East, areas burdened by ongoing human malaria (WHO, 2014). For instance, we opted to leave out the mitogenome of *Anopheles funestus* due the amount of gaps in the mtDNA PCG's sequences (GenBank accession number NC008070).

The reconstructed *Anopheles* phylogeny showed a deep divergence between two main *Anopheles* lineages (**Figure 15**) seemingly driven by their phylogeographic relations and the earth's geologic eras. The tree topology we obtained is consistent with the current hypothesis regarding the origin of *Anopheles* mosquitoes in the Gondwana supercontinent during the Cretaceous period (Krzywinski & Besansky, 2003, Krzywinski *et al*., 2006). *Anopheles* would have radiated from what is now South America, migrated into Africa and then colonized Europe and North America (with aid of land bridges), migrating also through Asia, and into the Pacific (Logue *et al*., 2013).

However, as discussed by Logue *et al*. (2013), the paucity regarding mitogenomes from African and Asian anophelines, plus the lack of mtDNA sequences from European *Anopheles*, precludes the chance of determining accurately if African anophelines are ancestral to those in Europe and North-America, or if North-American anophelines radiated from South America. This particular issue is bring forth since both our results, and those of Logue *et al*. (2013), represent a phylogeny in which the position of *Anopheles quadrimaculatus* relative to African and non-American anophelines remains unclear.

Recently, Harbach (2013) stated that interpreting the current distributions of anophelines in an evolutionary context is problematic. This author reviewed existing alternative hypotheses that, based on recent studies, suggest a different scenario for the evolution of the extant groups of the *Anophelinae* subfamily. The alternative scenario closely reflects the ideas presented by Christophers (1933), in which the ancestral lineage of

Anopheles existed before the breakup of Pangaea and subsequently diversified into the modern subgenera and species after the separation of the continents. Nonetheless, it is not the scope of this study to dig deeper into this matter.

Regarding the estimated divergence times, we dated the *Anopheles* MRCA to ~ 83 MYA. The node age (**Figure 16**) differs from the ~ 79 MYA reported by Moreno *et al*. (2010) and the ~ 93 MYA estimated by Logue *et al*. (2013). These molecular dating estimates, though different, are in agreement with (they all predate) the breakup of western Gondwana and the loss of land connections between South America and Africa (~ 95 MYA). This geologic event might have prompted the divergence of *Anopheles* + *Cellia* from *Nyssorhynchus*, and further on, the divergence between *Cellia* and *Anopheles* ~ 71 MYA, (or ~ 81 MYA according to Logue et al. (2013)). The geographic sorting of lineages thus, would coincide with the loss of land bridges between Africa-Europe and Europe-North America, potentially explaining the absence of *Cellia* in the New World and *Nyssorhynchus* in the Afro-Eurasian continents (Moreno *et al*., 2010; Krzywinski *et al*., 2001).

This study suggests that *A. aquasalis* diverged from the *A. albitarsis* complex ~ 28 MYA. This node age implies that their MRCA would be younger than the one shared between *A. darlingi* and the *A. albitarsis* complex which dates back to ~ 38 MYA (**Figure 16**). If we consider the bionomics of Neotropical anophelines, then *A. aquasalis* would have diverged from the *A. albitarisis* complex, adapting to its narrow coastal ecological niche, outcompeted by inland (*A. darlingi*) and brackish-water adapted (*A. albitarsis*) sibling species (Sinka *et al*., 2010).

Differences observed regarding radiation and divergence node ages have been reported by other authors (Krzywinski *et al*., 2006, Moreno *et al*., 2010, Harbach, 2013). Particularly, the choices regarding data inclusion have a direct effect on phylogenomic studies. Therefore, the use of full mitogenomes, individual or concatenated genes (PCG's, rRNA's, tRNA's) or nuclear genes, generate different molecular evolutionary histories (Moreno *et al*., 2010; Cameron, 2014). Additionally, observed discrepancies may be due to: incomplete species sampling representative of the geographical regions and /or the subgenera studied; the use of mutation rates; and the use of more than one calibration point (Krzywinski *et al*., 2006; Moreno *et al*., 2010, Logue *et al*., 2013). For example, Moreno *et al*. (2010) used as calibration point, the Anophelinae and Culicinae divergence age estimated to be 120 MYA (Rai & Black, 1999). Nevertheless, the only available, and rigorously calculated date for the

most recent split involving a mosquito lineage and a sister taxon, is the divergence time of 259.9 MYA between *Drosophila* and *Anopheles* estimated by Gaunt & Miles (2002) (Krzywinski *et al*., 2006). The pitfalls inherent in using molecular clocks to date divergence events are well known and estimated ages should be interpreted with caution (Krzywinski *et al*., 2006).

Molecular dating of the divergence events reported in this study occurred long before humans arrived to America 15-20 millenia ago (Stringer, 2011). Evidence suggests that *Plasmodium* parasites were introduced into the Americas via African slave trade routes during the European invasion (Yalcindag *et al*., 2012). In order for the transmission cycle to establish, the parasite had to adapt to local mosquitoes. Then, in accordance with the present geographic distribution of anophelines (Sinka *et al*., 2012), and the known slave trade disembarking ports, some of the first indigenous vectors that *P. falciparum*-infected humans encountered would have been: *A. albimanus*, *A. aquasalis*, *A. darlingi*, *A. albitarsis,* and *A. quadrimaculatus* (Molina-Cruz & Barillas-Mury, 2014).

4.5 Conclusion

Considering the evolutionary distance between Old and New World anophelines, it is not surprising that there are marked genetic, ecological and behavioral differences between them (Sinka *et al*., 2012). This probably resulted in a more stringent selection of *Plasmodium* as it adapted to these new vectors (Molina-Cruz & Barillas-Mury, 2014). Multiple Neotropical anophelines became human malaria vectors independently from each other, and as previously observed for South East Asian and Oceania anophelines (Logue *et al*., 2013), this suggests that co-occurrences of malaria transmission-related traits of anophelines are the result of convergent evolution (White *et al*., 2011).

Co-evolutionary studies of parasites and vectors require phylogenies for the mosquitoes to gain insights into their interspecific and co-evolutionary relationships (Harbach, 2013). Recently, Molina-Cruz and Barillas-Mury (2014), proposed a hypothesis linking the *Plasmodium* protein PFS47 and a still unknown interacting protein from *Anopheles*, as a key determinant of mosquito-parasite compatibility and vectorial competence “emergence” in American indigenous anophelines (Molina-Cruz *et al*., 2013, Molina-Cruz & Barillas-Mury, 2014).

Therefore, we finish this study strongly supporting the need of a better and more accurate *Anopheles* phylogeny. This, in turn, could aid reconstructing the history of malaria transmission in the New World and the inheritance of vectorial competence traits amongst anopheline sibling species, from an American context point of view. In addition, transgenic approaches to manipulate vector receptiveness to infection are, currently, a potential/enticing malaria control strategy. A natural classification of *Anopheles* predictive of biological and ecological traits could facilitate the manipulation of vector genomes by informing the dynamics that introduced genes would follow (Harbach, 2013).

Bibliography

Abascal F, Posada D, Knight RD, Zardoya R. Parallel evolution of the genetic code in arthropod mitochondrial genomes. PLoS Biol. 2006;4(5):e127.

Abascal F, Zardoya R, Telford MJ. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res. 2010;38(Web Server issue):W7-13.

Abraham EG, Jacobs-Lorena M. Mosquito midgut barriers to malaria parasite development. Insect Biochem Mol Biol. 2004; 34: 667-671. PM:15242707

Afrane YA, Githeko AK, Yan G. The ecology of *Anopheles* mosquitoes under climate change: case studies from the effects of deforestation in East African highlands. Annals of the New York Academy of Sciences. 2012;1249(1):204-10.

Aguilar R, Dong Y, Warr E, Dimopoulos G. *Anopheles* infection responses; laboratory models versus field malaria transmission systems. Acta tropica. 2005;95(3):285-91.

Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics. 2011;12:402.

Alphey L, Beard CB, Billingsley P, Coetzee M, Crisanti A, Curtis C, Eggleston P, Godfray C, Hemingway J, Jacobs-Lorena M. Malaria control with genetically manipulated insect vectors. Science. 2002; 298: 119-121. PM:12364786

Anderson MJ. A new method for non-parametric multivariate analysis of variance. Austral ecology. 2001;26(1):32-46.

Anderson MJ, Crist TO, Chase JM, Vellend M, Inouye BD, Freestone AL, et al. Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist. Ecol Lett. 2011;14(1):19-28.

Angellotti MC, Bhuiyan SB, Chen G, Wan XF. CodonO: codon usage bias analysis within and across genomes. Nucleic Acids Res. 2007;35(Web Server issue):W132-6.

Ansorge WJ. Next-generation DNA sequencing techniques. N Biotechnol. 2009; 25: 195-203. PM:19429539

Arruda ME, Rios RI, Arantes PC, Oliveira AC, Nascimento LP. Laboratory maintence of *Anopheles albitarsis* and *Anopheles aquasalis* by induced copulation. Memórias do Instituto Oswaldo Cruz. 1982;77(1):89-91.

Atyame CM, Delsuc F, Pasteur N, Weill M, Duron O. Diversification of *Wolbachia* endosymbiont in the *Culex pipiens* mosquito. Mol Biol Evol. 2011;28(10):2761-72.

Avguštin G, WALLACE RJ, Flint HJ. Phenotypic diversity among ruminal isolates of *Prevotella ruminicola*: proposal of *Prevotella brevis* sp. nov., *Prevotella bryantii* sp. nov., and *Prevotella albensis* sp. nov. and redefinition of *Prevotella ruminicola*. International journal of systematic bacteriology. 1997;47(2):284-8.

Azambuja P, Garcia ES, Ratcliffe NA. Gut microbiota and parasite transmission by insect vectors. Trends Parasitol. 2005; 21: 568-572. PM:16226491

Bahia AC, Dong Y, Blumberg BJ, Mlambo G, Tripathi A, BenMarzouk-Hidalgo OJ, et al. Exploring *Anopheles* gut bacteria for *Plasmodium* blocking activity. Environmental microbiology. 2014;16(9):2980-94.

Bahia AC, Kubota MS, Tempone AJ, Araújo HR, Guedes BA, Orfanó AS, et al. The JAK-STAT pathway controls *Plasmodium vivax* load in early stages of *Anopheles aquasalis* infection. PLoS Negl Trop Dis. 2011;5(11):e1317.

Bahia AC, Kubota MS, Tempone AJ, Pinheiro WD, Tadei WP, Secundino NF, et al. *Anopheles aquasalis* Infected by *Plasmodium vivax* displays unique gene expression profiles when compared to other malaria vectors and plasmodia. PLoS One. 2010;5(3):e9795.

Bahia AC, Oliveira JHM, Kubota MS, Araújo HR, Lima JB, Ríos-Velásquez CM, et al. The role of reactive oxygen species in *Anopheles aquasalis* response to *Plasmodium vivax* infection. PloS one. 2013;8(2):e57014.

Bakermans C, Ayala-del-Río HL, Ponder MA, Vishnivetskaya T, Gilichinsky D, Thomashow MF, et al. *Psychrobacter cryohalolentis* sp. nov. and *Psychrobacter arcticus* sp. nov., isolated from Siberian permafrost. International journal of systematic and evolutionary microbiology. 2006;56(6):1285-91.

Bando H, Okado K, Guelbeogo WM, Badolo A, Aonuma H, Nelson B, et al. Intra-specific diversity of *Serratia marcescens* in *Anopheles* mosquito midgut defines *Plasmodium* transmission capacity. Scientific reports. 2013;3.

Basset Y, Cizek L, Cuenoud P, Didham RK, Guilhaumon F, Missa O, Novotny V, Odegaard F, Roslin T, Schmidl J. Arthropod diversity in a tropical forest. Science. 2012; 338: 1481-1484. PM:23239740

Baumann P. Biology bacteriocyte-associated endosymbionts of plant sap-sucking insects. Annu Rev Microbiol. 2005; 59: 155-189. PM:16153167

Beard CB, Dotson EM, Pennington PM, Eichler S, Cordon-Rosales C, Durvasula RV 2001. Bacterial symbiosis and paratransgenic control of vector-borne Chagas disease. Int J Parasitol. 2001; 31: 621-627. PM:11334952

Beard CB, Hamm DM, Collins FH. The mitochondrial genome of the mosquito *Anopheles gambiae*: DNA sequence, genome organization, and comparisons with mitochondrial sequences of other insects. Insect Mol Biol. 1993;2(2):103-24.

Becker N. Petric D, Zgomba M, Boase C, Dahl C, Madon M, Kaiser A. 2010. Mosquitoes and their control. 2nd Edition. Springer. 577p.

Behura SK, Lobo NF, Haas B, deBruyn B, Lovin DD, Shumway MF, et al. Complete sequences of mitochondria genomes of *Aedes aegypti* and *Culex quinquefasciatus* and comparative analysis of mitochondrial DNA fragments inserted in the nuclear genomes. Insect Biochem Mol Biol. 2011;41(10):770-7.

Behura SK, Severson DW. Comparative analysis of codon usage bias and codon context patterns between dipteran and hymenopteran sequenced genomes. PLoS One. 2012;7(8):e43111.

Bellaousov S, Reuter JS, Seetin MG, Mathews DH. RNAstructure: Web servers for RNA secondary structure prediction and analysis. Nucleic Acids Res. 2013;41(Web Server issue):W471-4.

Bernt M, Braband A, Middendorf M, Misof B, Rota-Stabelli O, Stadler PF. Bioinformatics methods for the comparative analysis of metazoan mitochondrial genome sequences. Mol Phylogenet Evol. 2013;69(2):320-7.

Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, et al. MITOS: improved de novo metazoan mitochondrial genome annotation. Mol Phylogenet Evol. 2013;69(2):313-9.

Besansky NJ. Genome Analysis of Vectorial Capacity in Major Anopheles Vectors Of Malaria Parasites. The *Anopheles* Genomes Cluster Committee. 2008. Accessed on line at: https://www.vectorbase.org/sites/default/files/ftp/documents/Anopheles_Genomes_4Aug08.pdf. On March 10, 2011.

Bian G, Joshi D, Dong Y, Lu P, Zhou G, Pan X, et al. *Wolbachia* invades *Anopheles stephensi* populations and induces refractoriness to *Plasmodium* infection. Science. 2013;340(6133):748-51.

Billingsley PF, Sinden RE. Determinants of malaria-mosquito specificity. Parasitol Today. 1997; 13: 297-301. PM:15275055

Billker O, Shaw MK, Margos G, Sinden RE. The roles of temperature, pH and mosquito factors as triggers of male and female gametogenesis of *Plasmodium berghei* in vitro. Parasitology. 1997; 115 (Pt 1): 1-7. PM:9280891

Bishop-Lilly KA, Turell MJ, Willner KM, Butani A, Nolan NM, Lentz SM, Akmal A, Mateczun A, Brahmbhatt TN, Sozhamannan S. Arbovirus detection in insect vectors by rapid, high-throughput pyrosequencing. PLoS Negl Trop Dis. 2010; 4: e878. PM:21085471

Bisi DC, Lampe DJ. Secretion of anti-*Plasmodium* effector proteins from a natural *Pantoea* agglomerans isolate by using PelB and HlyA secretion signals. Applied and environmental microbiology. 2011;77(13):4669-75.

Boissière A, Tchioffo MT, Bachar D, Abate L, Marie A, Nsango SE, et al. Midgut microbiota of the malaria mosquito vector *Anopheles gambiae* and interactions with *Plasmodium falciparum* infection. PLoS pathogens. 2012;8(5):e1002742.

Boore JL, Macey JR, Medina M. Sequencing and comparing whole mitochondrial genomes of animals. Methods Enzymol. 2005;395:311-48.

Boore JL. Animal mitochondrial genomes. Nucleic Acids Res. 1999;27(8):1767-80.

Boulanger N, Lowenberger C, Volf P, Ursic R, Sigutova L, Sabatier L, Svobodova M, Beverley SM, Spath G, Brun R. Characterization of a defensin from the sand fly *Phlebotomus duboscqi* induced by challenge with bacteria or the protozoan parasite *Leishmania major*. Infect Immun. 2004; 72: 7140-7146. PM:15557638

Braband A, Cameron SL, Podsiadlowski L, Daniels SR, Mayer G. The mitochondrial genome of the onychophoran *Opisthopatus cinctipes* (Peripatopsidae) reflects the ancestral mitochondrial gene arrangement of Panarthropoda and Ecdysozoa. Mol Phylogenet Evol. 2010;57(1):285-92.

Bradfield J, Wyatt G. X-linkage of a vitellogenin gene in *Locusta migratoria*. Chromosoma. 1983;88(3):190-3.

Brown GR, Sutcliffe IC, Cummings SP. Reclassification of [Pseudomonas] doudoroffii (Baumann et al. 1983) into the genus *Oceanomonas* gen. nov. as *Oceanomonas doudoroffii* comb. nov., and description of a phenol-degrading bacterium from estuarine water as *Oceanomonas baumannii* sp. nov. International journal of systematic and evolutionary microbiology. 2001;51(1):67-72.

Brucker RM, Bordenstein SR. The capacious hologenome. Zoology. 2013;116(5):260-1.

Buttigieg PL, Ramette A. A guide to statistical analysis in microbial ecology: a community-focused, living review of multivariate data analyses. FEMS Microbiology Ecology. 2014;90(3):543-50.

Cameron SL. Insect mitochondrial genomics: implications for evolution and phylogeny. Annu Rev Entomol. 2014;59:95-117.

Capone A, Ricci I, Damiani C, Mosca M, Rossi P, Scuppa P, et al. Interactions between *Asaia*, *Plasmodium* and *Anopheles*: new insights into mosquito symbiosis and implications in malaria symbiotic control. Parasit Vectors. 2013;6(1):182.

Caporaso JG, Lauber CL, Costello EK, Berg-Lyons D, Gonzalez A, Stombaugh J, et al. Moving pictures of the human microbiome. Genome Biol. 2011;12(5):R50.

Carius HJ, Little TJ, Ebert D. Genetic variation in a host-parasite association: potential for coevolution and frequency-dependent selection. Evolution. 2001; 55: 1136-1145. PM:11475049

Carpi G, Cagnacci F, Wittekindt NE, Zhao F, Qi J, Tomsho LP, Drautz DI, Rizzoli A, Schuster SC. Metagenomic profile of the bacterial communities associated with *Ixodes ricinus* ticks. PLoS One. 2011; 6: e25604. PM:22022422

Carvalho SCGd, Martins Junior AdJ, Lima JBP, Valle D. Temperature influence on embryonic development of *Anopheles albitarsis* and *Anopheles aquasalis*. Memórias do Instituto Oswaldo Cruz. 2002;97(8):1117-20.

Chadee DD, Kitron U. Spatial and temporal patterns of imported malaria cases and local transmission in Trinidad. The American journal of tropical medicine and hygiene. 1999;61(4):513-7.

Chan YC, Roos C, Inoue-Murayama M, Inoue E, Shih CC, Pei KJ, et al. Mitochondrial genome sequences effectively reveal the phylogeny of Hylobates gibbons. PLoS One. 2010;5(12):e14419.

Chandler JA, Lang JM, Bhatnagar S, Eisen JA, Kopp A. Bacterial communities of diverse *Drosophila* species: ecological context of a host–microbe model system. PLoS Genetics. 2011;7(9):e1002272.

Chavshin AR, Oshaghi MA, Vatandoost H, Pourmand MR, Raeisi A, Enayati AA, et al. Identification of bacterial microflora in the midgut of the larvae and adult of wild caught *Anopheles stephensi*: a step toward finding suitable paratransgenesis candidates. Acta tropica. 2012;121(2):129-34.

Chen WJ, Koch M, Mallatt JM, Luan YX. Comparative analysis of mitochondrial genomes in diplura (hexapoda, arthropoda): taxon sampling is crucial for phylogenetic inferences. Genome Biol Evol. 2014;6(1):105-20.

Chikuma S, Kasahara R, Kato C, Tamegai H. Bacterial adaptation to high pressure: a respiratory system in the deep sea bacterium *Shewanella violacea* DSS12. FEMS microbiology letters. 2007;267(1):108-12.

Cho J-C, Giovannoni SJ. *Robiginitalea biformata* gen. nov., sp. nov., a novel marine bacterium in the family *Flavobacteriaceae* with a higher G+ C content. International journal of systematic and evolutionary microbiology. 2004;54(4):1101-6.

Chouaia B, Rossi P, Montagna M, Ricci I, Crotti E, Damiani C, et al. Molecular evidence for multiple infections as revealed by typing of *Asaia* bacterial symbionts of four mosquito species. Applied and environmental microbiology. 2010;76(22):7444-50.

Christophers SR. The fauna of British India, Including Ceylon and Burma. Diptera. Vol. IV. Family Culicidae. Tribe Anophelini. London, Taylor and Francis; 1933.

Christophides GK, Zdobnov E, Barillas-Mury C, Birney E, Blandin S, Blass C, Brey PT, Collins FH, Danielli A, Dimopoulos G. Immunity-related genes and gene families in *Anopheles gambiae*. Science. 2002; 298: 159-165. PM:12364793

Cirimotich CM, Dong Y, Clayton AM, Sandiford SL, Souza-Neto JA, Mulenga M, et al. Natural microbe-mediated refractoriness to *Plasmodium* infection in *Anopheles gambiae*. Science. 2011;332(6031):855-8.

Cirimotich CM, Dong Y, Garver LS, Sim S, Dimopoulos G. Mosquito immune defenses against *Plasmodium* infection. Developmental & Comparative Immunology. 2010;34(4):387-95.

Clayton AM, Dong Y, Dimopoulos G. The *Anopheles* innate immune system in the defense against malaria infection. Journal of innate immunity. 2013;6(2):169-81.

Clemente JC, Jansson J, Valiente G. Accurate taxonomic assignment of short pyrosequencing reads. Pac Symp Biocomput. 2010; 3-9. PM:19908352

Cohuet A, Harris C, Robert V, Fontenille D. Evolutionary forces on *Anopheles*: what makes a malaria vector? Trends Parasitol. 2010; 26: 130-136. PM:20056485

Cohuet A, Osta MA, Morlais I, Awono-Ambene PH, Michel K, Simard F, Christophides GK, Fontenille D, Kafatos FC. *Anopheles* and *Plasmodium*: from laboratory models to natural systems in the field. EMBO Rep. 2006; 7: 1285-1289. PM:17099691

Consoli RAGB, Lourenço-de-Oliveira R. 1994 Principais mosquitos de importancia sanitária no Brasil. (PDF) Editora Fundação Instituto Oswaldo, Rio de Janeiro, Brasil.

Consortium HMP. Structure, function and diversity of the healthy human microbiome. Nature. 2012;486(7402):207-14.

Coon KL, Vogel KJ, Brown MR, Strand MR. Mosquitoes rely on their gut microbiota for development. Molecular ecology. 2014;23(11):2727-39.

Costa-da-Silva AL, Marinotti O, Ribeiro JM, Silva MC, Lopes AR, Barros MS, et al. Transcriptome Sequencing and Developmental Regulation of Gene Expression in *Anopheles aquasalis.* PLoS neglected tropical diseases. 2014;8(7):e3005.

Coutinho-Abreu IV, Zhu KY, Ramalho-Ortigao M. Transgenesis and paratransgenesis to control insect-borne diseases: current status and future challenges. Parasitology international. 2010;59(1):1-8.

Crochu S, Cook S, Attoui H, Charrel RN, De Chesse R, Belhouchet M, Lemasson JJ, de Micco P, de L, X. Sequences of flavivirus-related RNA viruses persist in DNA form integrated in the genome of *Aedes* spp. mosquitoes. J Gen Virol. 2004 85: 1971-1980. PM:15218182

Culleton R, Coban C, Zeyrek FY, Cravo P, Kaneko A, Randrianarivelojosia M, et al. The origins of African *Plasmodium vivax*; insights from mitochondrial genome sequencing. PLoS One. 2011;6(12):e29137.

Da Silva AN, Dos Santos CC, Lacerda RN, Santa Rosa EP, De Souza RT, Galiza D, et al. Laboratory colonization of *Anopheles aquasalis* (Diptera: Culicidae) in Belém, Pará, Brazil. Journal of Medical Entomology. 2006;43(1):107-9.

Dada N, Jumas-Bilak E, Manguin S, Seidu R, Stenström T-A, Overgaard HJ. Comparative assessment of the bacterial communities associated with *Aedes aegypti* larvae and water from domestic water storage containers. Parasites & vectors. 2014;7(1):391.

Damiani C, Ricci I, Crotti E, Rossi P, Rizzi A, Scuppa P, et al. Mosquito-bacteria symbiosis: the case of *Anopheles gambiae* and *Asaia*. Microbial ecology. 2010;60(3):644-54.

Damiani C, Ricci I, Crotti E, Rossi P, Rizzi A, Scuppa P, et al. Paternal transmission of symbiotic bacteria in malaria vectors. Current Biology. 2008;18(23):R1087-R8.

de Castro MC, Singer BH. Was malaria present in the Amazon before the European conquest? Available evidence and future research agenda. J Archaeol Sci. 2005;32:337–340.

Deane LM. Malaria vectors in Brazil. Mem. Inst. Oswaldo Cruz. 1986; 81 (supl. II): 5-14.

Degnan PH, Moran NA. Diverse phage-encoded toxins in a protective insect endosymbiont. Appl Environ Microbiol. 2008; 74: 6782-6791. PM:18791000

Delcher AL, Bratke KA, Powers EC, Salzberg SL. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics. 2007; 23: 673-679. PM:17237039

della Torre A, Fanello C, Akogbeto M, Dossou-yovo J, Favia G, Petrarca V, et al. Molecular evidence of incipient speciation within *Anopheles gambiae* s.s. in West Africa. Insect Mol Biol. 2001;10(1):9-18.

Dharne M, Patole M, Shouche YS. Microbiology of the insect gut: tales from mosquitoes and bees. Journal of biosciences. 2006;31(3):293-5.

Dheilly NM. Holobiont–Holobiont Interactions: Redefining Host–Parasite Interactions. PLoS Pathog. 2014; 10(7): e1004093

Dillon R, Dillon V. The gut bacteria of insects: nonpathogenic interactions. Annual Reviews in Entomology. 2004;49(1):71-92.

Dimopoulos G, Richman A, Müller H-M, Kafatos FC. Molecular immune responses of the mosquito *Anopheles gambiae* to bacteria and malaria parasites. Proceedings of the National Academy of Sciences. 1997;94(21):11508-13.

Dinglasan RR, Devenport M, Florens L, Johnson JR, McHugh CA, Donnelly-Doman M, Carucci DJ, Yates JR, III, Jacobs-Lorena M. The *Anopheles gambiae* adult midgut peritrophic matrix proteome. Insect Biochem Mol Biol. 2009; 39: 125-134. PM:19038338

Djadid ND, Jazayeri H, Raz A, Favia G, Ricci I, Zakeri S. Identification of the midgut microbiota of *An. stephensi* and *An. maculipennis* for their application as a paratransgenic tool against malaria. PloS one. 2011;6(12):e28484.

Dong Y, Aguilar R, Xi Z,Warr E, Mongin E, Dimopoulos G. *Anopheles gambiae* immune responses to human and rodent *Plasmodium* parasite species. PloS Pathogens. 2006; 2(6): e52.

Dong Y, Manfredini F, Dimopoulos G. Implication of the mosquito midgut microbiota in the defense against malaria parasites. PLoS pathogens. 2009;5(5):e1000423.

Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 2007;7:214.

Duchaud E, Rusniok C, Frangeul L, Buchrieser C, Givaudan A, Taourit S, et al. The genome sequence of the entomopathogenic bacterium *Photorhabdus luminescens*. Nature biotechnology. 2003;21(11):1307-13.

Duguma D, Rugman-Jones P, Kaufman MG, Hall MW, Neufeld JD, Stouthamer R, et al. Bacterial communities associated with *Culex* mosquito larvae and two emergent aquatic plants of bioremediation importance. PloS one. 2013;8(8):e72522.

Duret L, Mouchiroud D. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. Proc Natl Acad Sci U S A. 1999;96(8):4482-7.

Eappen AG, Smith RC, Jacobs-Lorena M 2013. Enterobacter-activated mosquito immune responses to *Plasmodium* involve activation of SRPN6 in *Anopheles stephensi*. PLoS One. 2013; 8: e62937. PM:23658788

Eappen AG, Smith RC, Jacobs-Lorena M. Enterobacter-activated mosquito immune responses to *Plasmodium* involve activation of SRPN6 in *Anopheles stephensi*. PloS one. 2013;8(5):e62937.

Engel P, Moran NA. The gut microbiota of insects–diversity in structure and function. FEMS microbiology reviews. 2013;37(5):699-735.

Ermert V, Fink AH, Morse AP, Paeth H. The impact of regional climate change on malaria risk due to greenhouse forcing and land-use changes in tropical Africa. Environmental health perspectives. 2012;120(1):77.

Evans JM, Morris LS, Marchesi JR. The gut microbiome: the role of a virtual organ in the endocrinology of the host. Journal of Endocrinology. 2013;218(3):R37-R47.

Fagundes NJ, Kanitz R, Eckert R, Valls AC, Bogo MR, Salzano FM, Smith DG, Silva WA, Jr., Zago MA, Ribeiro-dos-Santos AK. Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. Am J Hum Genet. 2008; 82: 583-592. PM:18313026

Faran ME, Linthicum KJ. A handbook of the Amazonian species of *Anopheles* (*Nyssorhynchus*)(Diptera: Culcidae): publisher not identified; 1981.

Fauron CM, Wolstenholme DR. Extensive diversity among *Drosophila* species with respect to nucleotide sequences within the adenine + thymine-rich region of mitochondrial DNA molecules. Nucleic Acids Res. 1980;8(11):2439-52.

Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J, et al. Microbial co-occurrence relationships in the human microbiome. PLoS computational biology. 2012;8(7):e1002606.

Favia G, Ricci I, Damiani C, Raddadi N, Crotti E, Marzorati M, et al. Bacteria of the genus *Asaia* stably associate with *Anopheles stephensi*, an Asian malarial mosquito vector. Proceedings of the National Academy of Sciences. 2007;104(21):9047-51.

Field D, Garrity GM, Sansone SA, Sterk P, Gray T, Kyrpides N, Hirschman L, Glockner FO, Kottmann R, Angiuoli S. Meeting report: the fifth Genomic Standards Consortium (GSC) workshop. OMICS. 2008; 12: 109-113. PM:18564915

Fierer N, Leff JW, Adams BJ, Nielsen UN, Bates ST, Lauber CL, et al. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. Proceedings of the National Academy of Sciences. 2012;109(52):21390-5.

Flores-Mendoza C, Cunha R, Rocha D, Lourenço-de-Oliveira R. Determinação das fontes alimentares de *Anopheles aquasalis* (Diptera: Culicidae) no Estado de Rio de Janeiro, Brasil, pelo teste de precipitina.Revista Saúde Pública. 1996; 30(2).

Foley DH, Bryan JH, Yeates D, Saul A. Evolution and systematics of *Anopheles*: insights from a molecular phylogeny of Australasian mosquitoes. Mol Phylogenet Evol. 1998;9(2):262-75.

Frank-Fahle BA, Yergeau É, Greer CW, Lantuit H, Wagner D. Microbial functional potential and community composition in permafrost-affected soils of the NW Canadian Arctic. PloS one. 2014;9(1):e84761.

Gaio AO, Gusmao DS, Santos AV, Berbert-Molina MA, Pimenta PF, Lemos FJ. Contribution of midgut bacteria to blood digestion and egg production in *Aedes aegypti* (diptera: culicidae) (L.). Parasit Vectors. 2011; 4: 105. PM:21672186

Gaunt MW, Miles MA. An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks. Mol Biol Evol. 2002;19(5):748-61.

Gendrin M, Christophides GK. The *Anopheles* mosquito microbiota and their impact on pathogen transmission. *Anopheles* mosquitoes-New insights into malaria vectors. 2013.

Gerlach W, Stoye J. Taxonomic classification of metagenomic shotgun sequences with CARMA3. Nucleic Acids Res. 2011; 39: e91. PM:21586583

Ghosh A, Edwards MJ, Jacobs-Lorena M. The journey of the malaria parasite in the mosquito: hopes for the new century. Parasitol Today. 2000; 16: 196-201. PM:10782078

Giglioli G. Ecological change as a factor in renewed Malaria Transmission in an eradicated area. Bulletin of the World Health Organization. 1963; 29: 131-145.

Gimnig JE, Ombok M, Otieno S, Kaufman MG, Vulule JM, Walker ED. Density-dependent development of *Anopheles gambiae* (Diptera: Culicidae) larvae in artificial habitats. Journal of Medical Entomology. 2002;39(1):162-72.

Gilbert JA, Field D, Swift P, Thomas S, Cummings D, et al. The Taxonomic and Functional Diversity of Microbes at a Temperate Coastal Site: A ‘Multi-Omic’ Study of Seasonal and Diel Temporal Variation. PLoS ONE. 2010; 5(11): e15545.

Gimonneau G, Tchioffo MT, Abate L, Boissière A, Awono-Ambéné PH, Nsango SE, et al. Composition of *Anopheles coluzzii* and *Anopheles gambiae* microbiota from larval to adult stages. Infection, Genetics and Evolution. 2014;28:715-24.

Glass EM, Wilkening J, Wilke A, Antonopoulos D, Meyer F. Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. Cold Spring Harbor Protocols. 2010(2010):pdb. prot5368.

Gonzalez-Ceron L, Santillan F, Rodriguez MH, Mendez D, Hernandez-Avila JE. Bacteria in midguts of field-collected *Anopheles albimanus* block *Plasmodium vivax* sporogonic development. Journal of Medical Entomology. 2003;40(3):371-4.

Gorski A, Dabrowska K, Switala-Jelen K, Nowaczyk M, Weber-Dabrowska B, Boratynski J, Wietrzyk J, Opolski A. New insights into the possible role of bacteriophages in host defense and disease. Med Immunol. 2003; 2: 2. PM:12625836

Grillet ME. Factors associated with distribution of *Anopheles aquasalis* and *Anopheles oswaldoi* (Diptera: Culicidae) in a malarious area, northeastern Venezuela. J Med Entomol. 2000; 37: 231-238. PM:10730493

Hahn C, Bachmann L, Chevreux B. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads--a baiting and iterative mapping approach. Nucleic Acids Res. 2013;41(13):e129.

Harbach RE 2013. The Phylogeny and Classification of Anopheles, Anopheles mosquitoes - New insights into malaria vectors, Prof. Sylvie Manguin (Ed.), ISBN: 978-953-51-1188-7, InTech, DOI: 10.5772/54695. Available from: http://www.intechopen.com/books/anopheles-mosquitoes-new-insights-into-malaria-vectors/the-phylogeny-and-classification-of-anopheles

Harbach RE. Genus *ANOPHELES* Meigen, 1818. *Mosquito Taxonomic Inventory*. Book Genus ANOPHELES Meigen, 1818 Mosquito Taxonomic Inventory City. 2011.

Harris C, Lambrechts L, Rousset F, Abate L, Nsango SE, Fontenille D, Morlais I, Cohuet A. Polymorphisms in *Anopheles gambiae* immune genes associated with natural resistance to *Plasmodium falciparum*. PLoS Pathog. 2010; 6: e1001112. PM:20862317

Hassanin A, Léger N, Deutsch J. Evidence for multiple reversals of asymmetric mutational constraints during the evolution of the mitochondrial genome of metazoa, and consequences for phylogenetic inferences. Syst Biol. 2005;54(2):277-98.

Hay SI, Sinka ME, Okara RM, Kabaria CW, Mbithi PM, Tago CC, et al. Developing global maps of the dominant anopheles vectors of human malaria. PLoS Med. 2010;7(2):e1000209.

Henning W. Insect Phylogeny Wiley, Chichester, UK. 1981.

Hickerson MJ, Carstens BC, Cavender-Bares J, Crandall KA, Graham CH, Johnson JB, et al. Phylogeography's past, present, and future: 10 years after Avise, 2000. Mol Phylogenet Evol. 2010;54(1):291-301.

Hillyer JF, Barreau C, Vernick KD. Efficiency of salivary gland invasion by malaria sporozoites is controlled by rapid sporozoite destruction in the mosquito haemocoel. Int J Parasitol. 2007; 37: 673-681. PM:17275826

Hiraoka Y, Kawamata K, Haraguchi T, Chikashige Y. Codon usage bias is correlated with gene expression levels in the fission yeast *Schizosaccharomyces pombe*. Genes Cells. 2009;14(4):499-509.

Holmes EC. The evolution of endogenous viral elements. Cell Host Microbe. 2011; 10: 368-377. PM:22018237

Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, Wincker P, Clark AG, Ribeiro JM, Wides R. The genome sequence of the malaria mosquito *Anopheles gambiae*. Science. 2002; 298: 129-149. PM:12364791

Hongoh Y, Deevong P, Inoue T, Moriya S, Trakulnaleamsai S, Ohkuma M, et al. Intra-and interspecific comparisons of bacterial diversity and community structure support coevolution of gut microbiota and termite host. Applied and environmental microbiology. 2005;71(11):6590-9.

Horie M, Kobayashi Y, Suzuki Y, Tomonaga K. Comprehensive analysis of endogenous bornavirus-like elements in eukaryote genomes. Philos Trans R Soc Lond B Biol Sci. 2013; 368: 20120499. PM:23938751

Hubbe M, Neves WA, Harvati K. Testing evolutionary and dispersion scenarios for the settlement of the new world. PLoS One. 2010; 5: e11105. PM:20559441

Husnik F, Nikoh N, Koga R, Ross L, Duncan RP, Fujie M, Tanaka M, Satoh N, Bachtrog D, Wilson AC. Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. Cell. 2013; 153: 1567-1578. PM:23791183

Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic data. Genome research. 2007;17(3):377-86.

Inohira K, Hara T, Matsuura ET. Nucleotide sequence divergence in the A+T-rich region of mitochondrial DNA in *Drosophila simulans* and *Drosophila mauritiana*. Mol Biol Evol. 1997;14(8):814-22.

Ioannidis P, Johnston KL, Riley DR, Kumar N, White JR, Olarte KT, Ott S, Tallon LJ, Foster JM, Taylor MJ. Extensively duplicated and transcriptionally active recent lateral gene transfer from a bacterial *Wolbachia* endosymbiont to its host filarial nematode *Brugia malayi*. BMC Genomics. 2013; 14: 639. PM:24053607

Isaacs AT, Li F, Jasinskiene N, Chen X, Nirmala X, Marinotti O, Vinetz JM, James AA. Engineered resistance to *Plasmodium falciparum* development in transgenic *Anopheles stephensi*. PLoS Pathog. 2011; 7: e1002017. PM:21533066

Jaramillo-Gutierrez G, Molina-Cruz A, Kumar S, Barillas-Mury C. The *Anopheles gambiae* oxidation resistance 1 (OXR1) gene regulates expression of enzymes that detoxify reactive oxygen species. PLoS One. 2010;5(6):e11168.

Jaramillo-Gutierrez G, Rodrigues J, Ndikuyeze G, Povelones M, Molina-Cruz A, Barillas-Mury C. Mosquito immune responses and compatibility between *Plasmodium* parasites and anopheline mosquitoes. BMC Microbiol. 2009; 9: 154. PM:19643026

Jiang H, An L, Lin SM, Feng G, Qiu Y. A statistical framework for accurate taxonomic assignment of metagenomic sequencing reads. PLoS One. 2012; 7: e46450. PM:23049702

Kalyuzhnaya MG, Lapidus A, Ivanova N, Copeland AC, McHardy AC, et al. High-resolution metagenomics targets specific functional types in complex microbial communities. Nat Biotechnol. 2008; 26: 1029–1034.

Kaufman MG, Wanja E, Maknojia S, Bayoh MN, Vulule JM, Walker ED. Importance of algal biomass to growth and development of *Anopheles gambiae* larvae. Journal of Medical Entomology. 2006;43(4):669-76.

Keeling PJ, Palmer JD. Horizontal gene transfer in eukaryotic evolution. Nat Rev Genet. 2008; 9: 605-618. PM:18591983

Kim M, Lee KH, Yoon SW, Kim BS, Chun J, Yi H. Analytical tools and databases for metagenomics in the next-generation sequencing era. Genomics Inform. 2013; 11: 102-113. PM:24124405

Klasson L, Kambris Z, Cook PE, Walker T, Sinkins SP. Horizontal gene transfer between *Wolbachia* and the mosquito *Aedes aegypti*. BMC Genomics. 2009; 10: 33. PM:19154594

Knights D, Parfrey LW, Zaneveld J, Lozupone C, Knight R. Human-associated microbial signatures: examining their predictive value. Cell host & microbe. 2011;10(4):292-6.

Krzywinski J, Besansky NJ. Molecular systematics of *Anopheles*: from subgenera to subpopulations. Annu Rev Entomol. 2003;48:111-39.

Krzywinski J, Grushko OG, Besansky NJ. Analysis of the complete mitochondrial DNA from *Anopheles funestus*: an improved dipteran mitochondrial genome annotation and a temporal dimension of mosquito evolution. Mol Phylogenet Evol. 2006;39(2):417-23.

Krzywinski J, Li C, Morris M, Conn JE, Lima JB, Povoa MM, et al. Analysis of the evolutionary forces shaping mitochondrial genomes of a Neotropical malaria vector complex. Mol Phylogenet Evol. 2011;58(3):469-77.

Krzywinski J, Wilkerson RC, Besansky NJ. Evolution of mitochondrial and ribosomal gene sequences in anophelinae (Diptera: Culicidae): implications for phylogeny reconstruction. Mol Phylogenet Evol. 2001;18(3):479-87.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. Genome research. 2009;19(9):1639-45.

Kück P, Meusemann K. FASconCAT: Convenient handling of data matrices. Mol Phylogenet Evol. 2010;56(3):1115-8.

Kuczynski J, Liu Z, Lozupone C, McDonald D, Fierer N, Knight R. Microbial community resemblance methods differ in their ability to detect biologically relevant patterns. Nature methods. 2010;7(10):813-9.

Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. Genome Biol. 2004;5(2):R12.

Lambrechts L, Halbert J, Durand P, Gouagna LC, Koella JC. Host genotype by parasite genotype interactions underlying the resistance of anopheline mosquitoes to *Plasmodium falciparum*. Malar J. 2005; 4: 3. PM:15644136

Lefevre T, Vantaux A, Dabire KR, Mouline K, Cohuet A. Non-genetic determinants of mosquito competence for malaria parasites. PLoS pathogens. 2013;9(6):e1003365.

Legendre P, Cáceres M. Beta diversity as the variance of community data: dissimilarity coefficients and partitioning. Ecology letters. 2013;16(8):951-63.

Letunic I, Bork P. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. Nucleic acids research. 2011:gkr201.

Lewis DL, Farr CL, Kaguni LS. Drosophila melanogaster mitochondrial DNA: completion of the nucleotide sequence and evolutionary comparisons. Insect Mol Biol. 1995;4(4):263-78.

Ley RE, Lozupone CA, Hamady M, Knight R, Gordon JI. Worlds within worlds: evolution of the vertebrate gut microbiota. Nature Reviews Microbiology. 2008;6(10):776-88.

Li H, Liu H, Shi A, Stys P, Zhou X, Cai W. The complete mitochondrial genome and novel gene arrangement of the unique-headed bug *Stenopirates* sp. (Hemiptera: Enicocephalidae). PLoS One. 2012;7(1):e29419.

Li J, Riehle MM, Zhang Y, Xu J, Oduol F, Gomez SM, Eiglmeier K, Ueberheide BM, Shabanowitz J, Hunt DF. *Anopheles gambiae* genome reannotation through synthesis of ab initio and comparative gene prediction algorithms. Genome Biol. 2006; 7: R24. PM:16569258

Li R, Fan W, Tian G, Zhu H, He L, Cai J, Huang Q, Cai Q, Li B, Bai Y 2010. The sequence and de novo assembly of the giant panda genome. Nature. 2010a; 463: 311-317. PM:20010809

Li R, Guan MX. Human mitochondrial leucyl-tRNA synthetase corrects mitochondrial dysfunctions due to the tRNALeu(UUR) A3243G mutation, associated with mitochondrial encephalomyopathy, lactic acidosis, and stroke-like symptoms and diabetes. Mol Cell Biol. 2010b; 30(9):2147-54.

Li T, Gao C, Cui Y, Xie Q, Bu W. The complete mitochondrial genome of the stalk-eyed bug *Chauliops fallax* Scott, and the monophyly of Malcidae (Hemiptera: Heteroptera). PLoS One. 2013;8(2):e55381.

Lim YW, Cuevas DA, Silva GGZ, Aguinaldo K, Dinsdale EA, Haas AF, et al. Sequencing at sea: challenges and experiences in Ion Torrent PGM sequencing during the 2013 Southern Line Islands Research Expedition. PeerJ. 2014;2:e520.

Lindh J, Borg-Karlson A-K, Faye I. Transstadial and horizontal transfer of bacteria within a colony of *Anopheles gambiae* (Diptera: Culicidae) and oviposition response to bacteria-containing water. Acta tropica. 2008;107(3):242-50.

Lindh JM, Lehane MJ. The tsetse fly *Glossina fuscipes fuscipes* (Diptera: Glossina) harbours a surprising diversity of bacteria other than symbionts. Antonie Van Leeuwenhoek. 2011; 99: 711-720. PM:21203841

Lindh JM, Terenius O, Faye I. 16S rRNA gene-based identification of midgut bacteria from field-caught *Anopheles gambiae* sensu lato and *A. funestus* mosquitoes reveals new species related to known insect symbionts. Appl Environ Microbiol. 2005; 71: 7217-7223. PM:16269761

Littlewood DT, Lockyer AE, Webster BL, Johnston DA, Le TH. The complete mitochondrial genomes of *Schistosoma haematobium* and *Schistosoma spindale* and the evolutionary history of mitochondrial genome changes among parasitic flatworms. Mol Phylogenet Evol. 2006;39(2):452-67.

Liu B, Gibbons T, Ghodsi M, Treangen T, Pop M. Accurate and fast estimation of taxonomic profiles from metagenomic shotgun sequences. BMC genomics. 2011;12(Suppl 2):S4.

Logue K, Chan ER, Phipps T, Small ST, Reimer L, Henry-Halldin C, et al. Mitochondrial genome sequences reveal deep divergences among *Anopheles punctulatus* sibling species in Papua New Guinea. Malar J. 2013;12:64.

Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25(5):955-64.

Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. Applied and environmental microbiology. 2005;71(12):8228-35.

Lozupone CA, Hamady M, Kelley ST, Knight R. Quantitative and qualitative *β* diversity measures lead to different insights into factors that structure microbial communities. Applied and environmental microbiology. 2007;73(5):1576-85.

Lycett GJ, Kafatos FC. Anti-malarial mosquitoes? Nature. 2002; 417: 387-388. PM:12024196
Magurran AE, Henderson PA. Explaining the excess of rare species in natural species abundance distributions. Nature. 2003;422(6933):714-6.

Manguin S, Ngo CT, Tainchum K, Juntarajumnong W, Chareonviriyaphap T, Michon A-L, et al. Bacterial biodiversity in midguts of *Anopheles* mosquitoes, malaria vectors in Southeast Asia. *Anopheles* mosquitoes-new insights into malaria vectors Available from: intechopen com/books/anopheles-mosquitoes-new-insights-into-malaria-vectors/bacterial-biodiversity-in-midguts-of-anopheles-mosquitoes-malaria-vectors-in-southeast-asia[Links]. 2013.

Manguin S, Peyton EL, James AC, Roberts DR. Apparent changes in the abundance and distribution of *Anopheles* species on Grenada island. J Am Mosq Control Assoc. 1993; 9: 403-407. PM:8126473

Mardis ER. A decade's perspective on DNA sequencing technology. Nature. 2011; 470: 198-203. PM:21307932

Margulis L. Origins of species: acquired genomes and individuality. Biosystems. 1993;31(2-3):121-5.

Marquardt WC, Kondratieff BC, Black IV WC, Moore CG, James AA, Freier JE, Higgs S, Hagedorn HH. 2005. Biology of Disease Vectors. China. Elsevier Academic Press. 2A edição. 785p.

Martínez-Falcón AP, Durbán A, Latorre A, Antón J, de los Ángeles Marcos-García M. Bacteria associated with *Copestylum* (Diptera, Syrphidae) larvae and their cactus host *Isolatocereus dumortieri*. PloS one. 2011;6(11):e27443.

Masta SE, Longhorn SJ, Boore JL. Arachnid relationships based on mitochondrial genomes: asymmetric nucleotide and amino acid bias affects phylogenetic analyses. Mol Phylogenet Evol. 2009;50(1):117-28.

Megy K, Emrich SJ, Lawson D, Campbell D, Dialynas E, Hughes DS, Koscielny G, Louis C, Maccallum RM, Redmond SN. VectorBase: improvements to a bioinformatics resource for invertebrate vector genomics. Nucleic Acids Res. 2012; 40: D729-D734. PM:22135296

Meister S, Agianian B, Turlure F, Relogio A, Morlais I, Kafatos FC, Christophides GK. *Anopheles gambiae* PGRPLC-mediated defense against bacteria modulates infections with malaria parasites. PLoS Pathog. 2009; 5: e1000542. PM:19662170

Merritt R, Dadd R, Walker E. Feeding behavior, natural food, and nutritional relationships of larval mosquitoes. Annual review of entomology. 1992;37(1):349-74.

Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, et al. The metagenomics RAST server–a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC bioinformatics. 2008;9(1):386.

Michel K, Kafatos FC. Mosquito immunity against Plasmodium. Insect Biochem Mol Biol. 2005; 35: 677-689. PM:15894185

Milani C, Hevia A, Foroni E, Duranti S, Turroni F, Lugli GA, et al. Assessing the fecal microbiota: an optimized ion torrent 16S rRNA gene-based analysis protocol. PloS one. 2013;8(7):e68739.

Minard G, Mavingui P, Moro CV. Diversity and function of bacterial microbiota in the mosquito holobiont. Parasit Vectors. 2013;6(1):146.

Mitchell SE, Cockburn AF, Seawright JA. The mitochondrial genome of *Anopheles quadrimaculatus* species A: complete nucleotide sequence and gene organization. Genome. 1993;36(6):1058-73.

Mitri C, Vernick KD. *Anopheles gambiae* pathogen susceptibility: the intersection of genetics, immunity and ecology. Current opinion in microbiology. 2012;15(3):285-91.

Mokili JL, Rohwer F, Dutilh BE. Metagenomics and future perspectives in virus discovery. Curr Opin Virol. 2012; 2: 63-77. PM:22440968

Molina-Cruz A, Barillas-Mury C. The remarkable journey of adaptation of the *Plasmodium falciparum* malaria parasite to New World anopheline mosquitoes. Mem Inst Oswaldo Cruz. 2014;109(5):662-7.

Molina-Cruz A, DeJong RJ, Charles B, Gupta L, Kumar S, Jaramillo-Gutierrez G, et al. Reactive oxygen species modulate *Anopheles gambiae* immunity against bacteria and *Plasmodium*. Journal of biological chemistry. 2008;283(6):3217-23.

Molina-Cruz A, Garver LS, Alabaster A, Bangiolo L, Haile A, Winikor J, et al. The human malaria parasite Pfs47 gene mediates evasion of the mosquito immune system. Science. 2013;340(6135):984-7.

Moll RM, Romoser WS, Modrakowski MC, Moncayo AC, Lerdthusnee K. Meconial peritrophic membranes and the fate of midgut bacteria during mosquito (Diptera: Culicidae) metamorphosis. Journal of Medical Entomology. 2001;38(1):29-32.

Moreno M, Marinotti O, Krzywinski J, Tadei WP, James AA, Achee NL, et al. Complete mtDNA genomes of *Anopheles darlingi* and an approach to anopheline divergence time. Malar J. 2010;9:127.

Mouchet J, Carnevale P, Manguin S. Biodiversity of Malaria in the World. John Libbey Eurotext, Paris. 2008; 427p.

Murray, C. J. L., Rosenfeld, L. C., Lim, S. S., Andrews, K. G., Foreman, K. J., Haring, D., et al. Global malaria mortality between 1980 and 2010: a systematic analysis. Lancet. 2012; 379, 413–431. doi: 10.1016/S0140-6736(12)60034-8

Mwinyi A, Meyer A, Bleidorn C, Lieb B, Bartolomaeus T, Podsiadlowski L. Mitochondrial genome sequence and gene order of *Sipunculus nudus* give additional support for an inclusion of Sipuncula into Annelida. BMC Genomics. 2009;10:27.

Nartey R, Owusu-Dabo E, Kruppa T, Baffour-Awuah S, Annan A, Oppong S, Becker N, Obiri-Danso K. Use of *Bacillus thuringiensis* var israelensis as a viable option in an Integrated Malaria Vector Control Programme in the Kumasi Metropolis, Ghana. Parasit Vectors. 2013; 6: 116. PM:23607376

Neafsey DE, Christophides GK, Collins FH, Emrich SJ, Fontaine MC, Gelbart W, Hahn MW, Howell PI, Kafatos FC, Lawson D. The evolution of the *Anopheles* 16 genomes project. G3 (Bethesda). 2013; 3: 1191-1194. PM:23708298

Neafsey DE, Waterhouse RM, Abai MR, Aganezov SS, Alekseyev MA, Allen JE, et al. Mosquito genomics. Highly evolvable malaria vectors: the genomes of 16 *Anopheles* mosquitoes. Science. 2015; 347(6217):1258522.

Ng TF, Duffy S, Polston JE, Bixby E, Vallad GE, Breitbart M. Exploring the diversity of plant DNA viruses and their satellites using vector-enabled metagenomics on whiteflies. PLoS One. 2011a; 6: e19050. PM:21544196

Ng TF, Willner DL, Lim YW, Schmieder R, Chau B, Nilsson C, Anthony S, Ruan Y, Rohwer F, Breitbart M. Broad surveys of DNA viral diversity obtained through viral metagenomics of mosquitoes. PLoS One 6. 2011b; e20579. PM:21674005

Ni T, Yue J, Sun G, Zou Y, Wen J, Huang J. Ancient gene transfer from algae to animals: mechanisms and evolutionary significance. BMC Evol Biol. 2012; 12: 83. PM:22690978

Nikoh N, Tanaka K, Shibata F, Kondo N, Hizume M, Shimada M, Fukatsu T. *Wolbachia* genome integrated in an insect chromosome: evolution and fate of laterally transferred endosymbiont genes. Genome Res. 2008; 18: 272-280. PM:18073380

Norris DE, Shurtleff AC, Toure YT, Lanzaro GC. Microsatellite DNA polymorphism and heterozygosity among field and laboratory populations of *Anopheles gambiae* ss (Diptera: Culicidae). J Med Entomol. 2001; 38: 336-340. PM:11296845

Nyholm SV, Graf J. Knowing your friends: invertebrate innate immunity fosters beneficial bacterial symbioses. Nature Reviews Microbiology. 2012;10(12):815-27.

Ohkuma M, Brune A. Diversity, structure, and evolution of the termite gut microbial community. Biology of termites: a modern synthesis: Springer; 2011. p. 413-38.

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara R, et al. Package ‘vegan’. Community ecology package, version. 2013;2(9).

Oliveira MT, Azeredo-Espin AM, Lessinger AC. The mitochondrial DNA control region of Muscidae flies: evolution and structural conservation in a dipteran context. J Mol Evol. 2007;64(5):519-27.

Oliveros JC. VENNY. An interactive tool for comparing lists with Venn Diagrams. 2007.

O'Rourke DH, Raff JA. The human genetic history of the Americas: the final frontier. Curr Biol. 2010; 20: R202-R207. PM:20178768

Osei-Poku J, Mbogo C, Palmer W, Jiggins F. Deep sequencing reveals extensive variation in the gut microbiota of wild mosquitoes from Kenya. Molecular ecology. 2012;21(20):5138-50.

Osta MA, Christophides GK, Kafatos FC. Effects of mosquito genes on *Plasmodium* development. Science. 2004; 303: 2030-2032. PM:15044804

Pain A, Hertz-Fowler C. *Plasmodium* genomics: latest milestone. Nature Reviews Microbiology. 2009; 7: 180-181.

Pandey A, Suman S, Chandna S. Predictive role of mitochondrial genome in the stress resistance of insects and nematodes. Bioinformation. 2010;5(1):21-7.

Perna NT, Kocher TD. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J Mol Evol. 1995;41(3):353-8.

Perseke M, Golombek A, Schlegel M, Struck TH. The impact of mitochondrial genome analyses on the understanding of deuterostome phylogeny. Mol Phylogenet Evol. 2013;66(3):898-905.

Pidiyar VJ, Jangid K, Patole MS, Shouche YS. Studies on cultured and uncultured microbiota of wild *Culex quinquefasciatus* mosquito midgut based on 16s ribosomal RNA gene analysis. Am J Trop Med Hyg. 2004; 70: 597-603. PM:15210998

Pimenta PF, Touray M, Miller L. The journey of malaria sporozoites in the mosquito salivary gland. J Eukaryot Microbiol. 1994; 41: 608-624. PM:7866385

Pimenta PF, Orfano AS, Bahia AC, Duarte AP, Ríos-Velásquez CM, Melo FF, et al. An overview of malaria transmission from the perspective of Amazon *Anopheles* vectors. Mem Inst Oswaldo Cruz. 2015;110(1):23-47.

Plazzi F, Ribani A, Passamonti M. The complete mitochondrial genome of *Solemya velum* (Mollusca: Bivalvia) and its relationships with conchifera. BMC Genomics. 2013;14:409.
Posada D. jModelTest: phylogenetic model averaging. Mol Biol Evol. 2008;25(7):1253-6.

Poretsky R, Rodriguez-R LM, Luo C, Tsementzi D, Konstantinidis KT. Strengths and Limitations of 16S rRNA Gene Amplicon Sequencing in Revealing Temporal Microbial Community Dynamics. PLoS ONE. 2014; 9(4): e93827.

Prakash T, Taylor TD. Functional assignment of metagenomic data: challenges and applications. Briefings in bioinformatics. 2012;13(6):711-27.

Pumpuni C, Beier M, Nataro J, Guers LD, Davis J. *Plasmodium falciparum*: inhibition of sporogonic development in *Anopheles stephensi* by gram-negative bacteria. Experimental parasitology. 1993;77(2):195-9.

Pumpuni CB, Demaio J, Kent M, Davis JR, Beier JC. Bacterial population dynamics in three anopheline species: the impact on *Plasmodium* sporogonic development. The American journal of tropical medicine and hygiene. 1996;54(2):214-8.

Rai KS, Black WC. Mosquito genomes: structure, organization, and evolution. Adv Genet. 1999;41:1-33.

Ramasamy R, Surendran SN. Global climate change and its potential impact on disease transmission by salinity-tolerant mosquito vectors in coastal zones. Frontiers in physiology. 2012;3.

Ramette A. Multivariate analyses in microbial ecology. FEMS Microbiology Ecology. 2007;62(2):142-60.

Rani A, Sharma A, Rajagopal R, Adak T, Bhatnagar RK. Bacterial diversity analysis of larvae and adult midgut microflora using culture-dependent and culture-independent methods in lab-reared and field-collected *Anopheles stephensi*-an Asian malarial vector. BMC microbiology. 2009;9(1):96.

Ratcliffe NA, Whitten MMA. Vector immunity. 2004; 199-262

Reddy BP, Labbe P, Corbel V. *Culex* genome is not just another genome for comparative genomics. Parasit Vectors. 2012; 5: 63. PM:22463777

Reidenbach KR, Cook S, Bertone MA, Harbach RE, Wiegmann BM, Besansky NJ. Phylogenetic analysis and temporal diversification of mosquitoes (Diptera: Culicidae) based on nuclear genes and morphology. BMC Evol Biol. 2009;9:298.

Rejmánková E, Grieco J, Achee N, Roberts DR. Ecology of larval habitats. *Anopheles* Mosquitoes—New Insights into Malaria Vectors InTech ISBN. 2013:978-53.

Reyes A, Semenkovich NP, Whiteson K, Rohwer F, Gordon JI. Going viral: next-generation sequencing applied to phage populations in the human gut. Nat Rev Microbiol. 2012; 10: 607-617. PM:22864264

Ricci I, Damiani C, Capone A, DeFreece C, Rossi P, Favia G. Mosquito/microbiota interactions: from complex relationships to biotechnological perspectives. Curr Opin Microbiol. 2012; 15: 278-284. PM:22465193

Richards SL, Anderson SL, Yost SA. Effects of blood meal source on the reproduction of *Culex pipiens quinquefasciatus* (Diptera: Culicidae). Journal of Vector Ecology. 2012;37(1):1-7.

Riehle MA, Srinivasan P, Moreira CK, Jacobs-Lorena M. Towards genetic manipulation of wild mosquito populations to combat malaria: advances and challenges. J Exp Biol. 2003; 206: 3809-3816. PM:14506216

Rios-Velásquez CM, Martins-Campos KM, Simões RC, Izzo T, Dos Santos EV, Pessoa F, et al. Experimental *Plasmodium vivax* infection of key *Anopheles* species from the Brazilian Amazon. Malar J. 2013;12:460.

Rocha DdS, Fonseca AHd, Costa FA, Jurberg J, Galvão C. Desenvolvimento de *Rhodnius pictipes* Stal, 1872 alimentado através de membrana de silicone e em camundongos (Hemiptera, Reduviidae, Triatominae). Memórias do Instituto Oswaldo Cruz. 1997;92:553-8.

Rohwer F, Prangishvili D, Lindell D. Roles of viruses in the environment. Environ Microbiol. 2009; 11: 2771-2774. PM:19878268

Rosario K, Breitbart M. Exploring the viral world through metagenomics. Curr Opin Virol. 2011; 1: 289-297. PM:22440785

Rosenberg E, Zilber-Rosenberg I. Symbiosis and development: the hologenome concept. Birth Defects Res C Embryo Today. 2011; 93: 56-66. PM:21425442

Rubio-Palis Y, Zimmerman RH. Ecoregional classification of malaria vectors in the neotropics. Journal of Medical Entomology. 1997;34(5):499-510.

Russell CW, Bouvaine S, Newell PD, Douglas AE. Shared metabolic pathways in a coevolved insect-bacterial symbiosis. Applied and environmental microbiology. 2013;79(19):6117-23.
Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, et al. Artemis:

sequence visualization and annotation. Bioinformatics. 2000;16(10):944-5.

Sanschagrin S, Yergeau E. Next-generation sequencing of 16S ribosomal RNA gene amplicons. J Vis Exp. 2014; (90).

Santini F, Kong X, Sorenson L, Carnevale G, Mehta RS, Alfaro ME. A multi-locus molecular timescale for the origin and diversification of eels (Order: Anguilliformes). Mol Phylogenet Evol. 2013;69(3):884-94.

Santos AV, Dillon RJ, Dillon VM, Reynolds SE, Samuels RI. Ocurrence of the antibiotic producing bacterium *Burkholderia* sp. in colonies of the leaf-cutting ant *Atta sexdens rubropilosa*. FEMS microbiology letters. 2004;239(2):319-23.

Schlein Y, Polacheck I, Yuval B. Mycoses, bacterial infections and antibacterial activity in sandflies (Psychodidae) and their possible role in the transmission of leishmaniasis. Parasitology. 1985; 90 (Pt 1): 57-66. PM:3982854

Schmid-Hempel P, Ebert D. On the evolutionary ecology of specific immune defence. Trends Ecol Evol. 2003; 18: 27-32

Schneider A, Maréchal-Drouard L. Mitochondrial tRNA import: are there distinct mechanisms? Trends Cell Biol. 2000;10(12):509-13.

Seed KD, Lazinski DW, Calderwood SB, Camilli A. A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. Nature. 2013; 494: 489-491. PM:23446421

Segata N, Boernigen D, Tickle TL, Morgan XC, Garrett WS, Huttenhower C. Computational meta'omics for microbial community studies. Molecular systems biology. 2013;9(1).

Service MW, Townson H. The *Anopheles* vector. In: Gilles HM, Warrell DA, editors. Essential Malariology. Fourth edition ed. London: Arnold; 2002. p. 59–84.

Sharakhova MV, Hammond MP, Lobo NF, Krzywinski J, Unger MF, Hillenmeyer ME,

Bruggner RV, Birney E, Collins FH. Update of the *Anopheles gambiae* PEST genome assembly. Genome Biol. 2007; 8: R5. PM:17210077

Sharma P, Sharma S, Maurya RK, De TD, Thomas T, Lata S, et al. Salivary glands harbor more diverse microbial communities than gut in *Anopheles culicifacies*. Parasites & vectors. 2014;7(1):235.

Sharma VK, Kumar N, Prakash T, Taylor TD. Fast and accurate taxonomic assignments of metagenomic sequences using MetaBin. PloS one. 2012;7(4):e34030.

Sharpton TJ. An introduction to the analysis of shotgun metagenomic data. Frontiers in plant science. 2014; 5:209.

Shrinet J, Nandal UK, Adak T, Bhatnagar RK, Sunil S. Inference of the oxidative stress network in *Anopheles stephensi* upon *Plasmodium* infection. PLoS One. 2014;9(12):e114461.

Silverio MS, Rodovalho VeR, Bonetti AM, de Oliveira GC, Cuadros-Orellana S, Ueira-Vieira C, et al. Preliminary characterization of mitochondrial genome of *Melipona scutellaris*, a Brazilian stingless bee. Biomed Res Int. 2014;2014:927546.

Simon S, Hadrys H. A comparative analysis of complete mitochondrial genomes among Hexapoda. Mol Phylogenet Evol. 2013;69(2):393-403.

Sinden R, Alavi Y, Raine J. Mosquito–malaria interactions: a reappraisal of the concepts of susceptibility and refractoriness. Insect biochemistry and molecular biology. 2004;34(7):625-9.

Sinden RE, Matuschewski K. The sporozoite. 2005; 169-190

Sinden RE. *Plasmodium* differentiation in the mosquito. Parasitology. 1999; 41: 139-148. PM:10697846

Singh K, Reddy B, Patel A, Panchasara H, Parmar N, Patel A, et al. Metagenomic analysis of

buffalo rumen microbiome: Effect of roughage diet on Dormancy and Sporulation genes. Meta gene. 2014;2:252-68.

Sinka ME, Bangs MJ, Manguin S, Rubio-Palis Y, Chareonviriyaphap T, Coetzee M, et al. A global map of dominant malaria vectors. Parasit Vectors. 2012;5(1):69.

Sinka ME, Manguin S. Global distribution of the dominant vector species of malaria. *Anopheles* mosquitoes—new insights into malaria vectors Rijeka, Croatia: InTech p. 2013;36.

Sinka ME, Rubio-Palis Y, Manguin S, Patil AP, Temperley WH, Gething PW, et al. The dominant *Anopheles* vectors of human malaria in the Americas: occurrence data, distribution maps and bionomic précis. Parasit Vectors. 2010;3:72.

Stern A, Mick E, Tirosh I, Sagy O, Sorek R. CRISPR targeting reveals a reservoir of common phages associated with the human gut microbiome. Genome Res. 2012; 22: 1985-1994. PM:22732228

Sternberg ED, Thomas MB. Local adaptation to temperature and the implications for vector-borne diseases. Trends in parasitology. 2014;30(3):115-22.

Stewart JB, Beckenbach AT. Insect mitochondrial genomics: the complete mitochondrial genome sequence of the meadow spittlebug *Philaenus spumarius* (Hemiptera: Auchenorrhyncha: Cercopoidae). Genome. 2005;48(1):46-54.

Stringer C. The Origin of our Species. London: Penguin Books Ltd; 2011.

Sumitani M, Kasashima K, Yamamoto DS, Yagi K, Yuda M, Matsuoka H, Yoshida S. Reduction of malaria transmission by transgenic mosquitoes expressing an antisporozoite antibody in their salivary glands. Insect Mol Biol. 2013; 22: 41-51. PM:23176559

Suttle CA. Viruses in the sea. Nature. 2005; 437: 356-361. PM:16163346

SVS 2013. Sivep malaria. Dados epidemiológicos de malária, por estado. Brasil 2010 e 2011.

MS/SVS. Available at: http://portalsaude.saude.gov.br/portalsaude/arquivos/pdf/2013/Fev/08/dados_malaria_brasil_2010_2011_at2013_svs.pdf.

Takken W, Knols BG. Malaria vector control: current and future strategies. Trends Parasitol. 2009; 25: 101-104. PM:19168392

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 2013;30(12):2725-9.

Tanenbaum DM, Goll J, Murphy S, Kumar P, Zafar N, Thiagarajan M, Madupu R, Davidsen T, Kagan L, Kravitz S. The JCVI standard operating procedure for annotating prokaryotic metagenomic shotgun sequencing data. Stand Genomic Sci. 2010; 2: 229-237. PM:21304707

Tchioffo MT, Boissière A, Churcher TS, Abate L, Gimonneau G, Nsango SE, et al. Modulation of malaria infection in *Anopheles gambiae* mosquitoes exposed to natural midgut bacteria. PloS one. 2013;8(12):e81663.

Team RDC. R Foundation for Statistical Computing. Vienna, Austria. 2013;482.

Thiery I, Nicolas L, Rippka R, de Marsac NT. Selection of cyanobacteria isolated from mosquito breeding sites as a potential food source for mosquito larvae. Applied and environmental microbiology. 1991;57(5):1354-9.

Thomas T, Gilbert J, Meyer F. Metagenomics - a guide from sampling to data analysis. Microb Inform Exp. 2012; 2: 3. PM:22587947

Timmis JN, Ayliffe MA, Huang CY, Martin W. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat Rev Genet. 2004;5(2):123-35.

Tomita K, Weiner AM. Collaboration between CC- and A-adding enzymes to build and repair the 3'-terminal CCA of tRNA in *Aquifex aeolicus*. Science. 2001;294(5545):1334-6.

Tripet F, Aboagye-Antwi F, Hurd H. Ecological immunology of mosquito–malaria

interactions. Trends in parasitology. 2008;24(5):219-27.

Tripet F. Ecological Immunology of mosquito-malaria interactions: Of non-natural versus natural model systems and their inferences. Parasitology. 2009;136(14):1935-42.

Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The human microbiome project. Nature. 2007;449(7164):804-10.

Urbanczyk H, Ast JC, Higgins MJ, Carson J, Dunlap PV. Reclassification of *Vibrio fischeri*, *Vibrio logei*, *Vibrio salmonicida* and *Vibrio wodanis* as *Aliivibrio fischeri* gen. nov., comb. nov., *Aliivibrio logei* comb. nov., *Aliivibrio salmonicida* comb. nov. and *Aliivibrio wodanis* comb. nov. International journal of systematic and evolutionary microbiology. 2007;57(12):2823-9.

Vazquez-Martinez MG, Rodríguez MH, Arredondo-Jiménez JI, Méndez-Sánchez JD, Bond-Compeán JG, Gold-Morgan M. Cyanobacteria associated with *Anopheles albimanus* (Diptera: Culicidae) larval habitats in southern Mexico. Journal of Medical Entomology. 2002;39(6):825-32.

Vezzi A, Campanaro S, D'angelo M, Simonato F, Vitulo N, Lauro F, et al. Life at depth: *Photobacterium profundum* genome sequence and expression analysis. Science. 2005;307(5714):1459-61.

Villegas LM, Pimenta PFP. Metagenomics, paratransgenesis and the *Anopheles* microbiome: a portrait of the geographical distribution of the anopheline microbiota based on a meta-analysis of reported taxa. Memórias do Instituto Oswaldo Cruz. 2014;109(5):672-84.

Wang S, Ghosh AK, Bongio N, Stebbings KA, Lampe DJ, Jacobs-Lorena M. Fighting malaria with engineered symbiotic bacteria from vector mosquitoes. Proceedings of the National Academy of Sciences. 2012;109(31):12734-9.

Wang S, Jacobs-Lorena M. Genetic approaches to interfere with malaria transmission by vector mosquitoes. Trends in biotechnology. 2013;31(3):185-93.

Wang Y, Gilbreath III TM, Kukutla P, Yan G, Xu J. Dynamic gut microbiome across life

history of the malaria mosquito *Anopheles gambiae* in Kenya. PloS one. 2011;6(9):e24767.

Wang Y, Huang XL, Qiao GX. Comparative analysis of mitochondrial genomes of five aphid species (Hemiptera: Aphididae) and phylogenetic implications. PLoS One. 2013;8(10):e77511.

Wei SJ, Shi M, Chen XX, Sharkey MJ, van Achterberg C, Ye GY, et al. New views on strand asymmetry in insect mitochondrial genomes. PLoS One. 2010; 5(9):e12708.

Wei SJ, Shi M, He JH, Sharkey M, Chen XX. The complete mitochondrial genome of *Diadegma semiclausum* (hymenoptera: ichneumonidae) indicates extensive independent evolutionary events. Genome. 2009;52(4):308-19.

Weiss B, Aksoy S. Microbiome influences on insect host vector competence. Trends Parasitol. 2011; 27: 514-522. PM:21697014

Welburn SC, Maudlin I. Tsetse-trypanosome interactions: rites of passage. Parasitol Today. 1999; 15: 399-403. PM:10481151

Whiteley AS, Jenkins S, Waite I, Kresoje N, Payne H, Mullan B, et al. Microbial 16S rRNA Ion Tag and community metagenome sequencing using the Ion Torrent (PGM) Platform. Journal of microbiological methods. 2012;91(1):80-8.

WHO: WHO global malaria programme: world malaria report 2014. Geneva: World Health Organization; 2014.

Widder EA. Bioluminescence in the ocean: origins of biological, chemical, and ecological diversity. Science. 2010;328(5979):704-8.

Williamson SJ, Yooseph S. From bacterial to microbial ecosystems (metagenomics). Methods Mol Biol. 2012; 804: 35-55. PM:22144147

Wolinska J, King KC. Environment can alter selection in host-parasite interactions. Trends Parasitol. 2009; 25: 236-244. PM:19356982

Wolstenholme DR. Animal mitochondrial DNA: structure and evolution. Int Rev Cytol. 1992;141:173-216.

Wolstenholme DR. Genetic novelties in mitochondrial genomes of multicellular animals. Curr Opin Genet Dev. 1992;2(6):918-25.

Wommack KE, Bhavsar J, Ravel J. Metagenomics: read length matters. Appl Environ Microbiol. 2008; 74: 1453-1463. PM:18192407

Wooley JC, Godzik A, Friedberg I. A primer on metagenomics. PLoS Comput Biol. 2010; 6: e1000667. PM:20195499

Yalcindag E, Elguero E, Arnathau C, Durand P, Akiana J, Anderson TJ, et al. Multiple independent introductions of *Plasmodium falciparum* in South America. Proc Natl Acad Sci U S A. 2012;109(2):511-6.

Yu DW, Ji Y, Emerson BC. “Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring,” Methods in Ecology and Evolution, vol. 3, pp. 613–623, 2012.

Zdobnov EM, von Mering C, Letunic I, Torrents D, Suyama M, Copley RR, Christophides GK, Thomasova D, Holt RA, Subramanian GM. Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. Science. 2002; 298: 149-159. PM:12364792

Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008;18(5):821-9.

Zhang Z, Schwartz S, Wagner L, Miller W. A greedy algorithm for aligning DNA sequences. J Comput Biol. 2000;7(1-2):203-14.

Zhao Z, Su TJ, Chesters D, Wang SD, Ho SY, Zhu CD, et al. The mitochondrial genome of

Elodia flavipalpis Aldrich (Diptera: Tachinidae) and the evolutionary timescale of Tachinid flies. PLoS One. 2013;8(4):e61814.

Zimmerman RH. Ecology of malaria vectors in the Americas and future direction. Memórias do Instituto Oswaldo Cruz. 1992;87:371-83.

Appendices

Appendix 1: Additional files; supplementary figures and tables

Additional file 1: Index of bacterial genera identified in anopheline human malaria vectors. An index of bacterial genera identified in anophelines and their habitats is presented according to anopheline species, life stage, sex, nutrition type, sampling locality, population type and taxonomic profiling experimental approach (N/A = Not Applicable). Based on the format and data first published by Minard *et al*. (2013).

Format: XLS Size: 133.1KB
File: Anopheline microbiota Data Bank _ genus level.xls

Additional file 2: Comparison of bacterial composition at a family taxonomic rank between *A. gambiae* and *A. aquasalis* aquatic life stages. Relative abundance matrix (full data set) built with bacterial family rank relative abundance in five aquatic life stages of *A. aquasalis* (generated in this study) and *A. gambiae* data contained in the S9 table from Wang *et al*. (2011).

Format: csv Size: 7.1KB
File: Relative abundance_All bacterial f-OTUs matrix.csv

Additional file 3: Bacterial genera per continent binary matrix (Circos format). Bacterial genera matrix showing the presence or absence of taxa per continent category. This subset of the compiled data refers to bacterial genera that are shared by at least two of the continent categories. Generating a circular ribbons plot with the complete data set was not possible due to matrix size constrains established by the online tool. Matrix shared to show the format required by the Circos online tool.

Format: csv Size: 1.4KB
File:
Bacteria genus vs Continent_with_aquasalis_Common_Circos_no_colorstress_OCT2014.csv

Additional File 4 *Anopheles aquasalis* mitogenome in GenBank format.

Format: csv Size: 32.18KB
File: Aaquasalis_mt.gb

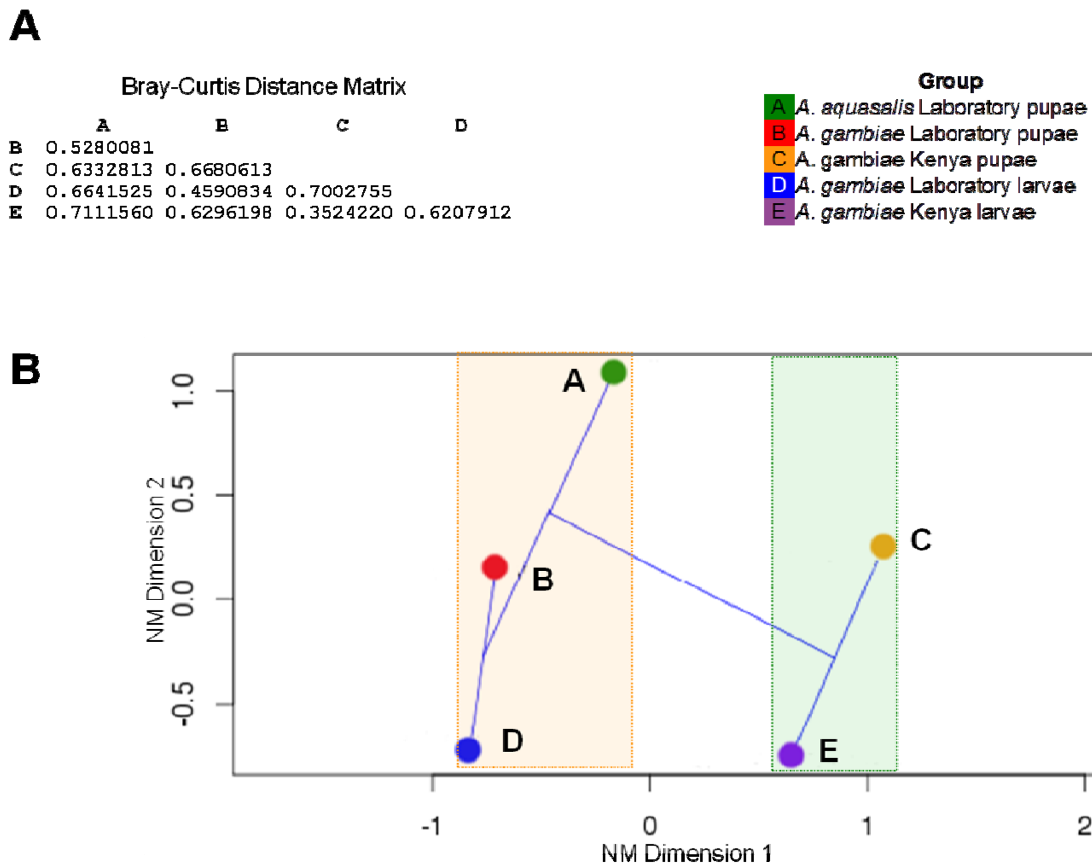


Figure S1 Combined ordination and clustering analyses comparing the bacterial family-rank OTUs abundance profiles of five anopheline aquatic life stages (full dataset). To reduce the impact of data dispersion, the complete bacterial (taxonomic rank = family) relative abundance matrix (**Additional file 2**) was transformed (Arcsin √x) to create a normal distribution of the data. The abundance data used corresponds to five aquatic stage anopheline groups described in the color/letter key. Panel A shows the Bray-Curtis pairwise distance matrix between the anopheline groups tested. Panel B shows the hierarchical clustering dendrogram of the (five) compared anopheline aquatic stages built upon the Bray-Curtis dissimilarities using the "average"method (cophenetic index = 0.92), and the NMDS ordination of the five groups based on the distance rank analysis of their bacterial f-OTUs abundance profiles. The cluster analysis was superimposed to depict the primary connections onto the ordination space. The potential main structure within the dataset is highlighted as rectangular shaded areas resolved along Dimension 1. Regardless of the *Anopheles* species the laboratory-reared groups mapped onto the ordination space shaded in orange, whereas the micro-cosmos reared groups (Kenya) mapped onto the ordination space shaded in green. Thus it appears as if the nurturing aquatic environment acts as a factor driving the ordination of the groups. Additionally, along Dimension 2, a life stage dependant gradient may be separating pupae from larvae (NMDS stress value ≈ 0; iterations = 13).

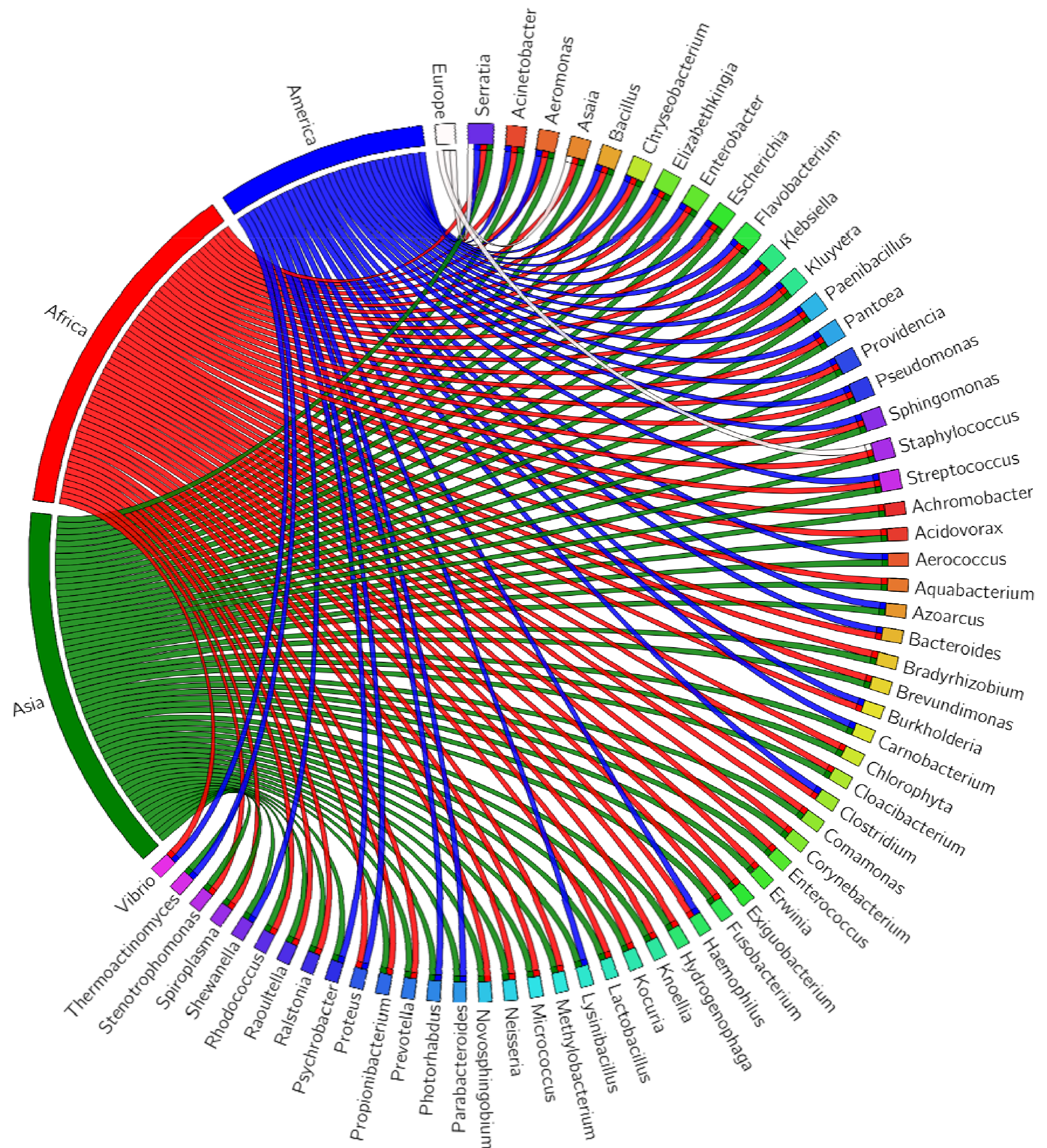


Figure S2 ***Anopheles*-associated microbiota distributed by continent.** Circular ribbon plot built from a subset of the compiled data (**Additional file 3**) depicting the 62 bacterial genera that are shared by at least two continents. Each ribbon represents the presence of the bacterial genus in an anopheline human malaria vector from the continent matching the ribbon color. Thus, American anophelines are represented in blue, African in red, Asian in green and European in white. Notice that *Serratia*, located in the upper segment of the circle, is the only genus reported to date as present in at least one anopheline malaria vector from each Continent. The correlation circular plot was built with the Circos online tool (Krzywinski *et al*., 2009) available at: http://mkweb.bcgsc.ca/tableviewer/visualize/

Table S1 Bacterial community composition at family taxonomic rank associated to *A. aquasalis* laboratory reared pupae. A total of 1028 sequences passed the rRNA based taxonomic profiling criteria applied. They were binned into 34 unique f-OTUs with 1 to 392 sequences per f-OTU and 107 bacterial sequences unclassified at that taxonomic rank.

Family	%
Aerococcaceae	0.10
Aeromonadaceae	28.02
Bacillaceae	0.78
Bacteroidaceae	0.10
Burkholderiaceae	0.10
Carnobacteriaceae	0.10
Clostridiaceae	1.46
Comamonadaceae	0.39
Cytophagaceae	0.39
Ectothiorhodospiraceae	0.10
Enterobacteriaceae	9.14
Erysipelotrichaceae	0.10
Eubacteriaceae	0.68
Flavobacteriaceae	38.13
Francisellaceae	0.10
Helicobacteraceae	0.10
Lachnospiraceae	1.36
Moraxellaceae	1.26
Mycoplasmataceae	0.10
Paenibacillaceae	0.10
Pasteurellaceae	0.88
Porphyromonadaceae	2.24
Pseudoalteromonadaceae	0.10
Pseudomonadaceae	0.29
Rhizobiaceae	0.10
Rhodobacteraceae	0.29
Rhodocyclaceae	0.10
Ruminococcaceae	0.88
Shewanellaceae	0.49
Sphingomonadaceae	0.10
Streptococcaceae	0.19
Thermoactinomycetaceae	0.10
unclassified	10.41
Vibrionaceae	1.17
Victivallaceae	0.10

Table S2 Index of bacterial genera showing their shared or exclusive distribution among anopheline human malaria vectors classified according to the continent in which they exert vectorial activity. A ¨by continent¨ bacterial genera index can be seen in the Additional file 1, spreadsheet 2.

Common elements in "Africa" and "America":
Bacteroides, Burkholderia, Clostridium, Haemophilus, Proteus, Vibrio.

Elements only in "Africa":
Acidovorax, Agromyces, Anaplasma, Aquabacterium, Arcobacter, Arthrobacter, Bacillariophyta, Bradyrhizobium, Cedecea, Chlorophyta, Cloacibacterium, Cobetia, Corynebacterium, Delftia, Ehrlichia, Fusobacterium, Gluconobacter, Janibacter, Knoellia, Methylobacterium, Methylocystis, Methylophilus, Morganella, Mycoplasma, Neisseria, Nocardia, Novosphingobium, Pelagibacter, Phenilobacterium, Porphyrobacter, Prevotella, Propionibacterium, Ralstonia, Rhizobium, Rhodopseudomonas, Roseomonas, Salmonella, Schlegelella, Sediminibacterium, Sphingobium, Spiroplasma, Sulfurospirillum, Thorsellia, Wolbachia, Zymobacter.

Common elements in "Africa", "Asia" and "Europe":
Asaia, Staphylococcus.

Common elements in "Africa", "Asia", "America" and "Europe":
Serratia.

Common elements in "Africa" and "Asia":
Achromobacter, Acidovorax, Aquabacterium, Bradyrhizobium, Brevundimonas, Chlorophyta, Cloacibacterium, Comamonas, Corynebacterium, Enterococcus, Erwinia, Exiguobacterium, Fusobacterium, Hydrogenophaga, Knoellia, Kocuria, Lactobacillus, Methylobacterium, Micrococcus, Neisseria, Novosphingobium, Prevotella, Propionibacterium, Ralstonia, Raoultella, Rhodococcus, Spiroplasma, Stenotrophomonas.

Common elements in "Asia" and "America":
Aerococcus, Azoarcus, Carnobacterium, Lysinibacillus, Parabacteroides, Photorhabdus, Psychrobacter, Shewanella, Thermoactinomyces.

Common elements in "Africa", "Asia" and "America":
Acinetobacter, Aeromonas, Bacillus, Chryseobacterium, Elizabethkingia, Enterobacter, Escherichia, Flavobacterium, Klebsiella, Kluyvera, Paenibacillus, Pantoea, Providencia, Pseudomonas, Sphingomonas, Streptococcus.

Elements only in "Asia":
Aciditerrimonas, Actinomyces, Actinospica, Agrobacterium, Agromonas, Alcaligenes, Anaerococcus, Anaerovorax, Anoxybacillus, Archaeoglobus, Archangium, Arenimonas, Armatimonadetes_gp2, Armatimonadetes_gp5, Asticcacaulis, Balneimonas, Bordetella, Brachybacterium, Brevibacillus, Brevibacterium, Calothrix, Campylobacter, Castellaniella, Cellvibrio, Chromobacterium, Citrobacter, Clostridium IV, Clostridium XIVa, Cohnella, Conexibacter, Cronobacter, Dechloromonas, Demequina, Desemzia, Diaphorobacter, Diplorickettsia, Dolosigranulum, Dyadobacter, Dysgonomonas, Enhydrobacter, Escherichia/Shigella, Facklamia, Falsibacillus, Ferribacterium, Ferruginibacter, Fibrobacter, Flectobacillus, Fontibacillus, Frateuria, Gemella, Gillisia, Gluconacetobacter, Gp1, Gp2, Gp3, Gp4, Haliscomenobacter, Halomonas, Helcococcus, Herbaspirillum, Ignatzschineria, Ignavibacterium, Jeotgalicoccus, Kofleria, Lachnospiracea_incertae_sedis, Lactococcus, Legionella, Leminorella, Leptothrix, Leuconostoc, Luteimonas, Macrococcus, Marmoricola, Massilia, Mesorhizobium, Methermicoccus, Microbacterium, Micromonospora, Moraxella, Myroides, Naxibacter, Nesterenkonia, Nitrospira, Nocardioides, Nocardiopsis, Oceanicola, Ornithinibacillus, Ornithinimicrobium, Paludibacter, Pandoraea, Paracoccus, Parapusillimonas, Peptoniphilus, Pimelobacter, Planomicrobium, Pontibacter, Prauserella, Proteiniphilum, Pseudoxanthomonas, Pusillimonas, Rahnella, Ramlibacter, Rhizomicrobium, Rhodocista, Rhodomicrobium, Rothia, Rubrobacter, Rummeliibacillus, Saccharofermentans, Saccharopolyspora, Salinicoccus, Salinicola, Salinimicrobium, Salisaeta, Sphingobacterium, Steroidobacter, Sulfuricurvum, Sulfuritalea, Telmatospirillum, Terriglobus, Thermicanus, Thermobacillus, Thermomonas, TM7_genera_incertae_sedis, Treponema, Truepera, Tumebacillus, Turicibacter, Variovorax, Wautersiella, Xenorhabdus, Yokenella, Zoogloea.

Elements only in "America":
Actinobacillus, Aggregatibacter, Aliivibrio, Anabaena, Aphanocapsa, Aphanothece, Arsenophonus, Blautia, Bulleidia, Butyrivibrio, Candidatus Sulcia, Capnocytophaga, Curvibacter, Dickeya, Ectothiorhodospira, Edwardsiella, Eubacterium, Ewingella, Faecalibacterium, Francisella, Gaetbulibacter, Geobacillus, Hespellia, Oceanimonas, Odoribacter, Pectobacterium, Phormidium, Photobacterium, Plesiomonas, Pseudoalteromonas, Pseudobutyrivibrio, Riemerella, Robiginitalea, Roseburia, Roseobacter, Ruminococcus, Sarcina, Shigella, Shinella, Shuttleworthia, Spirosoma, Spirulina, Sulfurimonas, Teredinibacter, Ureaplasma, Victivallis, Yersinia.

Table S3 Transformed (Arcsin √x) African-American anopheline core f-OTUs abundance matrix. Based on the compiled information regarding bacterial genera reported in multiple anopheline species (Additional file 1) and the geographic filtering criteria applied, a list of 23 genera shared between American and African anophelines was generated. In order to work with categories that encompassed more data; to be more conservative when regarding the taxonomic profiling prediction; and to reduce the impact of the heterogeneous sources of information used to build the compilation, the g-OTUs were classified into their respective family level taxa (f-OUT's). The relative abundance of each f-OTU from the *A.aquasalis* LAB pupae group were predicted by shotgun metagenomics taxonomic profiling (MG-RAST) whereas the *A. gambiae* relative abundance data were obtained from Wang *et al.* (2011).

Family	*A. aquasalis* LAB pupae	*A. gambiae* LAB pupae	*A. gambiae* Kenya pupae	*A. gambiae* LAB larvae	*A. gambiae* Kenya larvae
Aeromonadaceae	0.55777	0.07217	0.57888	0.04360	0.05294
Bacillaceae	0.08833	0.00000	0.01732	0.01000	0.08009
Bacteroidaceae	0.03119	0.00000	0.00000	0.01000	0.01000
Burkholderiaceae	0.03120	0.01000	0.04692	0.00000	0.01414
Clostridiaceae	0.12109	0.06934	0.04692	0.07077	0.10697
Enterobacteriaceae	0.30720	0.34299	0.10976	0.83187	0.15124
Flavobacteriaceae	0.66558	0.81562	0.05388	0.20263	0.10314
Moraxellaceae	0.11269	0.12360	0.02829	0.04360	0.02000
Paenibacillaceae	0.03120	0.02000	0.00000	0.00000	0.02646
Pasteurellaceae	0.09370	0.00000	0.00000	0.01414	0.00000
Pseudomonadaceae	0.05404	0.06485	0.02450	0.03000	0.03000
Sphingomonadaceae	0.03120	0.08195	0.07287	0.02450	0.10314
Streptococcaceae	0.04412	0.00000	0.00000	0.01414	0.01000
Vibrionaceae	0.10825	0.00000	0.00000	0.00000	0.00000

Table S4 Comparative strand asymmetry analysis based on AT and GC skews between the *A. aquasalis* mtDNA and selected anopheline species.

	A.aquasalis		*A.punctulatus*		*A.gambiae*		*A.darlingi N*		*A.darlingi S*	
Feature	**AT skew**	**GC skew**	**AT skew**	**GC skew**	**AT skew**	**GC skew**	**AT skew**	**GC skew**	**AT skew**	**GC skew**
ND2	-0.0868	-0.1459	-0.1039	-0.1479	-0.0869	-0.1264	-0.0881	-0.1086	-0.0856	-0.1098
COI	-0.1214	-0.0127	-0.1132	0.0000	-0.1128	-0.0166	-0.1305	0.0111	-0.1329	0.0135
COII	-0.0533	-0.0562	-0.0485	-0.0118	-0.0060	-0.0870	-0.0314	-0.0286	-0.0454	0.0112
ATP8	-0.1429	-0.3793	-0.1000	-0.4545	-0.1128	-0.3793	-0.1343	-0.3571	-0.1194	-0.4286
ATPase6	-0.1111	-0.1638	-0.1252	-0.0988	-0.1282	-0.1494	-0.0851	-0.1591	-0.0938	-0.1222
COIII	-0.0736	-0.0783	-0.1281	-0.0133	-0.1408	-0.0472	-0.0948	-0.0439	-0.0971	-0.0303
ND3	-0.0775	-0.1325	-0.0954	-0.0986	-0.0819	-0.1233	-0.0638	-0.1111	-0.0747	-0.0959
ND5	-0.1733	0.2454	-0.1901	0.2110	-0.1657	0.2559	-0.1634	0.2110	-0.1657	0.2240
ND4	-0.2454	0.2476	-0.2479	0.2242	-0.2346	0.1854	-0.2240	0.2218	-0.2255	0.2371
ND4L	-0.2851	0.2549	-0.3468	0.2308	-0.2835	0.3077	-0.2439	0.2593	-0.2470	0.2830
ND6	-0.0562	-0.3250	-0.0512	-0.1842	-0.0583	-0.1899	-0.0857	-0.2000	-0.0813	-0.2000
CYTB	-0.0836	-0.0641	-0.1348	0.0000	-0.1189	-0.0543	-0.0976	-0.0236	-0.0957	-0.0299
ND1	-0.2834	0.3184	-0.2970	0.2607	-0.3122	0.2851	-0.2767	0.2863	-0.2760	0.2978
PCG	-0.1441	0.0354	-0.1592	0.0509	-0.1487	0.0360	-0.1400	0.0479	-0.1420	0.0577
Genome	0.0373	-0.1684	0.0369	-0.1353	0.0322	-0.1540	0.0288	-0.1449	0.0284	-0.1415
lrRNA	-0.0714	0.3362	-0.0672	0.3214	-0.0540	0.3276	-0.0712	0.3471	-0.0696	0.3443
srRNA	-0.0347	0.2704	-0.0608	0.2258	-0.0518	0.2638	-0.0284	0.2327	-0.0316	0.1925
rRNA_concat	-0.0579	0.3095	-0.0648	0.2823	-0.0532	0.3013	-0.0560	0.3017	-0.0561	0.2840
tRNA_concat	0.0103	0.1293	0.0188	0.1607	-0.0026	0.1315	0.0077	0.1188	0.0103	0.0779

Table S5 Comparative compositional analysis represented in AT% between the *A. aquasalis* mtDNA and selected anopheline species.

	A-T %				
Feature	*A.aquasalis*	*A.punctulatus*	*A.gambiae*	*A.darlingi N*	*A.darlingi S*
ND2	81.97	83.53	83.04	82.94	83.14
COI	69.16	69.55	68.64	70.79	70.98
COII	74.01	75.18	73.14	74.45	74.01
ATP8	82.10	86.42	82.10	82.72	82.72
ATP6	74.01	76.21	74.45	74.16	73.57
COIII	70.78	71.41	70.39	71.03	70.65
ND3	76.55	79.94	79.38	79.66	79.38
ND5	78.23	79.06	78.26	79.05	78.99
ND4	76.83	79.06	77.50	78.17	78.32
ND4L	83.00	82.67	83.01	82.00	82.33
ND6	84.76	85.52	84.95	86.67	86.67
CYTB	72.56	75.02	72.47	73.88	73.53
ND1	76.70	77.67	76.61	76.28	76.49
PCG	75.90	77.30	76.00	76.78	76.72
Genome	77.16	78.50	77.56	78.20	78.13
lrRNA	82.48	83.11	82.49	82.64	82.50
srRNA	79.95	80.53	79.63	79.95	79.70
rRNA_concat	81.53	82.14	81.41	81.66	81.48
tRNA_concat	78.54	79.35	77.88	78.51	78.44

Appendix 2: Doctoral Thesis related publications

1) Published scientific paper: literature review and meta-analysis (**First Author**).

¨Metagenomics, paratransgenesis and the *Anopheles* microbiome: a portrait of the geographical distribution of the anopheline microbiota based on a meta-analysis of reported taxa.¨

Mem Inst Oswaldo Cruz. 2014 Aug;109(5):672-84.
PMID:25185007

2) Published scientific paper: literature review (**Collaborating author**)

¨ An overview of malaria transmission from the perspective of Amazon *Anopheles* vectors¨

Mem Inst Oswaldo Cruz. 2015 13/02;110(1):23-47.
PMID: 25742262