A TASTE OF THE GENETICS AND ANATOMY UNDERLYING VERTEBRATE HOST PREFERENCE IN THE *ANOPHELES GAMBIAE* COMPLEX OF MALARIA VECTORS

A Dissertation

by

ZACHARY RAY POPKIN-HALL

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

DOCTOR OF PHILOSOPHY

Chair of Committee,	Michel A. Slotman
Committee Members,	Giridhar Athrey
	J. Spencer Johnston
	Christine Merlin
	Aaron M. Tarone
Head of Department,	Pete D. Teel

August 2020

Major Subject: Entomology

ABSTRACT

In this dissertation, I explore three major areas of inquiry underlying vertebrate host choice in the *Anopheles* (*An.*) *gambiae* complex. *An. coluzzii* is a major vector due to its high anthropophily. Therefore, studying the genetic and anatomical basis of anthropophily is critical prior to designing transgenic control methods for this species.

First, I discuss differential expression of chemosensory genes, which could be involved in host choice, in the labella of *An. coluzzii* and its zoophilic sister species, *An. quadriannulatus*, where I identify 20 differentially expressed genes. Most intriguing are *Obp26*, which shows both female and *An. coluzzii* bias, and *Obp57*, which shows both female and *An. quadriannulatus* bias, along with very high overall expression.

Next, I present data from a series of behavioral experiments conducted on adult female *An. coluzzii* to determine which chemosensory organs are necessary and/or sufficient for mosquitoes to activate host-seeking, respond to host odors, and blood feed from an artificial membrane feeder. The antennae are by far the most important organs for host-seeking activation: their ablation abolishes host odor responses in a dual-port olfactometer, substantially reduces host-seeking activation, and substantially reduces blood feeding. The maxillary palps and CO_2 are not necessary to activate host-seeking, and labella ablation does not significantly reduce host odor response despite generating physiological stress.

Finally, I analyze the molecular evolutionary patterns of the gustatory receptor genes (*Grs*) in six species of the *An. gambiae* complex to identify *Grs* that are

potentially involved in vertebrate host choice. There are sixteen *Grs* with McDonald-Kreitman (MK) test results suggestive of positive selection, twelve with MK test results suggestive of purifying selection, eight with joint Tajima's D/Fay and Wu's H test results suggestive of a recent selective sweep, and four with E test results suggestive of either recovery from a selective sweep or background selection.

In summary, these results identify new directions for the study of the genetic basis of anthropophily in the *An. gambiae* complex, identify the antennae as the primary organ of interest in anthropophily, and provide limited evidence of a major role in either the labella or the *Grs* in vertebrate host choice.

ACKNOWLEDGEMENTS

First and foremost, I express my gratitude to my committee: Dr. Giri Athrey, Dr. J. Spencer Johnston, Dr. Christine Merlin, and Dr. Aaron Tarone; all of whom provided consistently excellent advice and feedback. My advisor, Dr. Michel Slotman was consistently in my corner, and extremely supportive of my development as a scientist, educator, and human. I am very grateful to have had an advisor who supported my own career goals rather than trying to forge me into a miniature copy of himself, as well as one who fully supports taking sick days and vacations.

Many people have been part of the Slotman Lab during my tenure here, but none were more helpful or a better friend to me than Dr. Jake Meyers. I give my thanks as well to the other members of the lab: Dr. Giri Athrey, Dr. Luciano Cosme, Dr. Kevin Deitz, Dr. Sharmila Pathkonda, Isaac McNeely, Fatima Camarillo, Chris Janes, Thomas McGlynn, Xinyue Huang, Brooke Pearce, Sarah Rico, Javier Pellacani, and Dr. Vinaya Shetty.

Chapter 3 took a lot of time and energy to get off the ground, and I wouldn't have been able to do it without the help of many people. Drs. Giri Athrey and Zainulabeuddin Syed gave me lots of advice on how to run an olfactometer, and Isaac McNeely and Dr. Jake Meyers helped me finally get it to actually work, as did Michel. Brooke Pearce and Elí Mondragón were kind enough to generate sweaty, smelly socks for me to use to test things out and eventually to actually run experiments. Several people at the Texas A&M Veterinary Medical Park were instrumental in facilitating the

collection of cow sweat, especially former director Dr. Clay Ashley, manager Destiny Taylor, and Jud Perry. Even more people were kind enough to turn off the light in the behavior room at 5:30 for me: Chris Janes, Thomas McGlynn, Willy Juarez, Brooke Pearce, and Sarah Rico. Finally, two great friends, Dr. Adrian Castellanos and Drew Guinness, helped me figure out the necessary R code to do the statistics I wanted.

Two computer programmer friends, Curtis Cooper and Jenna Hull, helped me figure out why my bash scripts for Chapter 2 and 4 weren't working properly (it was pretty much always because of badly formatted CSV files, pro tip for anyone reading this).

I give my thanks as well to some members of the faculty who, while not on my committee, provided me with invaluable opportunities to expand and deepen my skillsets in science as well as in teaching: Dr. Adrienne Brundage, who fostered my independence as a laboratory instructor and also, among other things, talked me through saving an abandoned tinny kitten; Dr. Craig J. Coates, who gave me multiple opportunities to hone my teaching skills; Dr. Gabe Hamer, for letting me use all kinds of equipment in his lab, and, along with Dr. Sarah Hamer, getting me out into the field to trap mammals, birds, and their associated vectors. Many members of the excellent Department of Entomology staff made things easier for me over the years, and I am also very grateful to the Center for Teaching Excellence, who gave me the chance to spend a year working as a Graduate Teaching Consultant and deepen my own teaching practices in conversation with other graduate students from throughout the university.

I would not have pursued the path I did if not for the incredible opportunities afforded to me during my time as an undergraduate at Macalester College, and I am forever indebted to the great and good Dr. Sarah Boyer, who singlehandedly changed the trajectory of my life by steering me towards the studies of entomology, evolution, and genetics, and who facilitated my first three peer-reviewed publications. While I no longer have the pleasure of working on cyphophthalmids, they will forever hold a place in my heart.

While others may not have realized a PhD was a possibility until later in life, I always assumed I would get one in something, and am pleased to become the fourth Popkin with a PhD. My gratitude goes to the three Drs. Popkin who preceded me for putting the idea in my head (in chronological order): my grandfather, Richard H. Popkin; my uncle, Jeremy D. Popkin; and my mother, Susan J. Popkin. And of course an honorable mention to my late aunt, Margaret L. Popkin, J.D, who along with my mother, showed me the critical importance of doing work that serves the task of tiqqun olam. My thanks also go both of my parents, Susan J. Popkin and Norman K. Hall, for fostering a lifelong fascination with the natural world and for always encouraging me to ask questions and formulate hypotheses.

I would have had a much rougher time as a doctoral student if not for the friendships that I forged along the way. In particular, I am most grateful for Dr. Adrian Castellanos, Dr. Karen Poh, Jocelyn Holt, MacKenzie Kjeldgaard, and Chantal Cough-Schulze. Dr. Castellanos in particular has been a constant support both scientifically and personally.

Lastly, Elí Mondragón kept me grounded, supported, and loved while also pursuing his own PhD in biomedical engineering, which gave me the opportunity to listen to somebody else's grad school complaints for a change. We should all be so lucky to have such a magnificent partner.

CONTRIBUTORS AND FUNDING SOURCES

This work was supervised by a dissertation committee consisting of Dr. Michel A. Slotman, Dr. J. Spencer Johnston, and Dr. Aaron M. Tarone of the Department of Entomology; Dr. Giridhar Athrey of the Department of Poultry Science; and Dr. Christine Merlin of the Department of Biology.

Sequencing for Chapter 2 was performed at the Texas A&M AgriLife Genomics Core. DNA sequences used in Chapter 4 were generated by the *Anopheles* Genome Consortium, led by Michael C. Fontaine of the University of Groningen, Daniel E. Neafsey of the Broad Institute of MIT and Harvard, and Nora J. Besansky of the University of Notre Dame. All other work conducted for the dissertation was completed by the student independently.

RNA sequencing was funded by a Texas A&M AgriLife Insect Vectored Disease Grant awarded to Michel A. Slotman.

This work was partially supported by teaching assistantships through the Department of Entomology; an Excellence Fellowship from the College of Agriculture and Life Sciences; the Herb Dean Endowed Scholarship; the John A. Jackman Endowed Scholarship; a National Science Foundation Closing the GAP Program Fellowship; and a scholarship from the University of Washington to attend the Summer Institute in Statistical Genetics in 2016.

TABLE OF CONTENTS

Page

ABSTRACT ii

ACKNOWLEDGEMENTS iv

CONTRIBUTORS AND FUNDING SOURCES viii

TABLE OF CONTENTS ix

LIST OF FIGURES xii

LIST OF TABLES xv

1. INTRODUCTION AND LITERATURE REVIEW 1

1.1. The *Anopheles gambiae* Species Complex and Malaria Transmission 2
1.2. Insect Sensory Systems and Genetics 5
1.2.1. Insect Chemosensory Systems and Genetics 6
1.3. Studies of Mosquito Host-Seeking Behavior 11
1.3.1. Anatomy of Mosquito Host-Seeking 11
1.3.2. Heat and Thermosensation in Mosquito Host-Seeking 14
1.3.3. Volatile Odorants in Mosquito Host-Seeking 15
1.3.4. Integration of Sensory Modalities in Mosquito Host-Seeking 18
1.3.5. Host Preferences of *An. gambiae* Complex Members 18
1.4. Studies of Chemosensory Gene Expression in Mosquitoes 21
1.4.1. Chemosensory Gene Expression in *Aedes* and *Culex* 21
1.4.2. Chemosensory Gene Expression in *Anopheles* 22
1.5. Evolution of Chemosensory Genes 24
1.5.1. Invertebrates 24
1.5.2. Arthropods 25
1.5.3. Insects 26
1.6. Implications for Novel Mosquito Control Techniques 27
1.7. Dissertation Overview 28

2. SPECIES AND SEX-BIASED EXPRESSION OF CHEMOSENSORY GENES IN THE SISTER SPECIES *ANOPHELES COLUZZII* AND *ANOPHELES QUADRIANNULATUS* 30

2.1. Introduction 30

2.2. Methods 36
2.2.1. Mosquito Rearing 36
2.2.2. Dissections and Sequencing 37
2.2.3. RNA Sequencing Analyses 38
2.2.4. Analysis of Differential Gene Expression 39
2.2.5. Gene Ontology Analysis 39
2.3. Results 40
2.3.1. Differential Gene Expression between Female *An. coluzzii* and *An. quadriannulatus* 41
2.3.2. Differential Gene Expression between Male *An. coluzzii* and *An. quadriannulatus* 49
2.3.3. Sex-Biased Gene Expression in *An. coluzzii* 56
2.3.4. Sex-Biased Gene Expression in *An. quadriannulatus* 59
2.4. Discussion 63
2.4.1. Agreement with Previous Characterization of *An. coluzzii* Labella Transcriptome 64
2.4.2. Female and Species-Biased Genes 65
2.4.3. Male and Species-Biased Genes 68
2.4.4. Species-Biased Genes without Sex Bias 70
2.4.5. Sex-Biased Genes without Species Bias 71
2.4.6. Non-DE Highly Expressed Genes 72
2.4.7. Biological Role of the Labella 75

3. THE ROLE OF THE MAJOR CHEMOSENSORY ORGANS IN ACTIVATION OF HOST-SEEKING AND HOST-DIFFERENTIATION IN *ANOPHELES COLUZZII* 77

3.1. Introduction 77
3.2. Methods 81
3.2.1. Host Choice Experiments 81
3.2.2. Bloodfeeding Experiments 86
3.3. Results 88
3.3.1. Mortality Across Groups 88
3.3.2. Odor Response 89
3.3.3. Bloodfeeding Assay 94
3.4. Discussion 95
3.4.1. Vertebrate Host Preference 95
3.4.2. Activation of Host-Seeking 96
3.4.3. Vertebrate Host Odor Response 101
3.4.4. Bloodfeeding 105
3.5. Conclusion 107

4. MOLECULAR EVOLUTION OF GUSTATORY RECEPTOR GENES IN THE *ANOPHELES GAMBIAE* COMPLEX 111

4.1. Introduction 111
4.1.1. Insect Chemosensory Genes 111
4.1.2. Function and Evolution of Gustatory Receptors 113
4.1.3. *Anopheles gambiae* Complex 117
4.2. Methods 120
4.2.1. Data 120
4.2.2. Variant Discovery Pipeline 121
4.2.3. Data Analysis 122
4.3. Results 125
4.3.1. Fixed Differences Between Species 126
4.3.2. McDonald-Kreitman Test 127
4.3.3. DH Suite of Tests 142
4.4. Discussion 153
4.4.1. McDonald-Kreitman Test 154
4.4.2. Selective Sweeps 160
4.4.3. Comparison to Other Chemosensory Genes 163
4.4.4. Future Directions 164

5. CONCLUSIONS 166

REFERENCES 171

APPENDIX A 204

APPENDIX B 214

APPENDIX C 217

APPENDIX D 250

APPENDIX E 258

LIST OF FIGURES

Page

Figure 1.1 – Schematics of A) Olfactory and B) Gustatory Sensilla. 7

Figure 1.2 – Schematic of Olfactory Sensillum and Associated Receptors and Binding Proteins. 9

Figure 1.3 – Green Fluorescent Protein Labeled Chemosensory Neurons in *An. gambiae* Chemosensory Organs. 12

Figure 2.1 – Overall Differential Gene Expression Profiles of A) Female *An. coluzzii* and Female *An. quadriannulatus*, B) Male *An. coluzzii* and Male *An. quadriannulatus*, C) Male and Female *An. coluzzii*, and D) Male and Female *An. quadriannulatus*.....41

Figure 2.2 – Molecular Function of DE Genes between Female *An. coluzzii* and *An. quadriannulatus*.....43

Figure 2.3 – Expression (CPM) of Major Chemosensory Gene Families in Female *An. coluzzii* and *An. quadriannulatus*: A) Gustatory Receptors, B) Ionotropic Receptors, C) Olfactory Receptors, and D) Odorant Binding Proteins.....45

Figure 2.4 – Molecular Function of DE Genes between Male *An. coluzzii* and *An. quadriannulatus*.....50

Figure 2.5 – Expression (CPM) of Major Chemosensory Gene Families in Male *An. coluzzii* and *An. quadriannulatus*: A) Gustatory Receptors, B) Ionotropic Receptors, C) Olfactory Receptors, and D) Odorant Binding Proteins.....52

Figure 2.6 – Molecular Function of DE Genes between Female and Male *An. coluzzii*. 56

Figure 2.7 – Expression (CPM) of Major Chemosensory Gene Families in Female and Male *An. coluzzii*: A) Gustatory Receptors, B) Ionotropic Receptors, C) Olfactory Receptors, and D) Odorant Binding Proteins. 58

Figure 2.8 – Molecular Function of DE Genes between Female and Male *An. quadriannulatus*.....59

Figure 2.9 – Expression (CPM) of Major Chemosensory Gene Families in Female and Male *An. quadriannulatus*: A) Gustatory Receptors, B) Ionotropic Receptors, C) Olfactory Receptors, and D) Odorant Binding Proteins.....62

Figure 3.1 – Typical examples of ablated mosquitoes: A) a mosquito with ablated antennae, with Johnston's organs visible and a small piece of most proximal antennal segment attached; B) a mosquito with ablated maxillary palps with approximately 10% still attached; C) a mosquito with ablated labella. 82

Figure 3.2 – Dual-Port Olfactometer.....84

Figure 3.3 – Examples of the three categories of mosquitoes recovered following bloodfeeding trials: A) a non-bloodfed mosquito with translucent abdomen; B) a bloodfed but non-engorged mosquito with a small amount of blood visible in the abdomen; C) an engorged mosquito. 87

Figure 3.4 – Proportion of mosquitoes activated to host-seek by condition. 90

Figure 3.5 – Proportion of mosquitoes responding to any host odor. 92

Figure 3.6 – Proportion of activated mosquitoes responding to any host odor.....93

Figure 4.1 – Visual Representation of Selective Sweep/Genetic Hitchhiking and Background Selection.....125

Figure 4.2 – Heat Map showing neutrality indices (NI) generated by McDonald-Kreitman tests between A) the three major vector species and *An. quadriannulatus* and B) *An. arabiensis* and *An. coluzzii* and *An. gambiae* s.s.. 129

Figure 4.3 – Heat Map showing neutrality indices (NI) generated by McDonald-Kreitman tests between A) *An. melas* and the other complex species and B) *An. merus* and the other complex species.....130

Figure 4.4 – TCS Haplotype Networks of *Gr18* in *An. coluzzii* and *An. quadriannulatus*.....132

Figure 4.5 – TCS Haplotype Networks of *Gr48* in *An. gambiae* and *An. quadriannulatus*.....136

Figure 4.6 – TCS Haplotype Networks of *Gr60* in *An. coluzzii* and *An. quadriannulatus*.....137

Figure 4.7 – TCS Haplotype Network of the sugar receptor *Gr17* in *An. coluzzii*. 144

Figure 4.8– TCS Haplotype Network of the sugar receptor *Gr19* in *An. quadriannulatus*.....146

Figure 4.9 – TCS Haplotype Network of the sugar receptor *Gr18* in *An. gambiae* s.s..148

Figure 4.10 – TCS Haplotype Network of the CO_2 receptor *Gr24* in *An. gambiae* s.s. 149

Figure 4.11 – TCS Haplotype Network of *Gr33* in *An. gambiae* s.s. 152

Figure 5.1 – Venn Diagram of Chemosensory Genes of Interest.. 168

LIST OF TABLES

Page

Table 2.1 – Reads obtained from each biological replicate library. 40

Table 2.2 – Differentially expressed chemosensory genes in the labella of female *An. coluzzii* and *An. quadriannulatus*. 42

Table 2.3 – Differentially expressed chemosensory genes in the labella of male *An. coluzzii* and *An. quadriannulatus*. 51

Table 2.4 – Total average expression in CPM of chemosensory genes by family in both sexes of *An. coluzzii* 57

Table 2.5 – Differentially expressed chemosensory genes in the labella of *An. coluzzii* females and males. 58

Table 2.6 – Total average expression in CPM of chemosensory genes by family in both sexes of *An. quadriannulatus*. 60

Table 2.7 – Differentially expressed chemosensory genes in the labella of *An. quadriannulatus* females and males. 61

Table 3.1 – Mortality by Experimental Group in Olfactometer Experiments 88

Table 3.2 – Activation Rate of Mosquitoes by Experimental Group. 89

Table 3.3 – Response Rate of Mosquitoes by Experimental Group. 91

Table 3.4 – Response Rate of Activated Mosquitoes by Experimental Group. 92

Table 3.5 – Degree of Engorgement by Experimental Group. 95

Table 4.1 – Total fixed differences (both synonymous and non-synonymous) identified between species pairs across all *Grs* tested with the McDonald-Kreitman test 126

Table 4.2 – Prevalence of positive and purifying selection based on Neutrality Index between species pairs. 128

Table 4.3 – *Grs* with significantly (or near-significantly) low neutrality indices (NI), suggestive of positive selection. 133

Table 4.4 – *Grs* with significantly (or near-significantly) high neutrality indices, suggestive of purifying selection. 141

Table 4.5 – *Grs* with significant (or near significant) DH values, suggestive of a selective sweep. 143

Table 4.6 – *Grs* with significant (or near significant) E-test values, indicative of recovery from a selective sweep. 147

Table 4.7 – *Grs* with significant (or near significant) Tajima's D or normalized Fay and Wu's H values, indicative of non-neutrality, but selective force unclear. 150

Table 5.1 – Chemosensory Genes of Particular Interest. 168

1. INTRODUCTION AND LITERATURE REVIEW

The *Anopheles* (*An.*) *gambiae* species complex has arguably the largest impact on human health of any animal, as it contains three of the most significant malaria vector species in sub-Saharan Africa: *An. arabiensis*, *An. coluzzii*, and *An. gambiae* sensu stricto (s.s.). These vectors predominantly transmit *Plasmodium falciparum*, or malignant malaria. Despite major advances in recent years, malaria remains one of the leading causes of global childhood mortality[1], with the vast majority of cases (93% in 2018) occurring in sub-Saharan Africa[2]. Of these cases, 99.7% were caused by *P. falciparum*[2].

However, this species complex also presents a unique research opportunity, because in addition to containing these three significant vectors, it also contains two closely related species which are not considered malaria vectors: *An. amharicus* and *An. quadriannulatus*. The complex also contains four other constituent species, about which much less is known: *An. bwambae*, *An. fontenillei*, *An. melas*, and *An. merus*.

The major difference in biology between the malaria vectors in the complex and their non-vector congeners is their host preference: the vectors exhibit a high degree of anthropophily (preference for a human host), whereas the non-vectors are more zoophilic (preferring to feed on non-human mammals). The existence of a genetic basis for vertebrate host preference has been suspected for decades, since a seminal study showing that it was possible to selectively breed mosquitoes from the *An. gambiae* complex to prefer one mammalian host over another[3]. A similar study demonstrated the same possibility in the anthropophilic *Aedes aegypti* and zoophilic *Aedes simpsoni*[4].

More recently, studies on *An. coluzzii* and *An. quadriannulatus* showed that anthropophily in *An. coluzzii* is a dominant trait[5].

While host preference is likely governed by a wide swath of different genes and physiological systems, the most intuitive areas of inquiry are mosquito sensory systems. While vision, thermosensation, and hygroreception all play a role in host-seeking, it is largely a chemosensory process, incorporating both volatile and contact cues[6–10]. As such, the focus of this dissertation is chemosensory-directed behavior, chemosensory anatomy, and chemosensory genes.

Below, I present essential background information for understanding the biological context of the subsequent chapters, beginning first with the *An. gambiae* species complex and its role in malaria transmission, then describing insect sensory and specifically chemosensory systems and the genes that underlie them, behavioral studies on the host-seeking behaviors of mosquitoes, the expression and evolution of chemosensory genes in mosquitoes, and finally the current technological context which provides the potential to use our knowledge of mosquito biology to reduce malaria transmission in new ways. As detailed overviews are also provided in each chapter, the information presented here is of a broader nature. Following this overview, I outline the chapters detailing my doctoral research.

1.1. The *Anopheles gambiae* Species Complex and Malaria Transmission

The *An. gambiae* species complex consists of at least nine closely related and morphologically indistinguishable mosquito species from sub-Saharan Africa, all of which are capable of hybridizing[11–14]. *An. gambiae* s.s. is a strongly anthropophilic

mosquito found throughout both mainland sub-Saharan Africa and Madagascar[14,15]. While *An. gambiae* s.s. was formerly thought to have two distinct molecular forms (M and S), these two forms have subsequently been redescribed as distinct species, with *An. gambiae* s.s. representing the S form, and the M form being described as *An. coluzzii*[14]. *An. coluzzii* shares similar host-seeking behaviors as *An. gambiae* s.s., but has different larval ecology and occupies a more restricted range primarily in West Africa[14,15].

An. arabiensis has more plastic, generalist host-seeking behavior[16], and is also widely distributed, though it prefers more arid habitats than do *An. gambiae* s.s. and *An. coluzzii*[13]. Though it exhibits increased plasticity, it is also a major malaria vector and is also predominantly anthropophilic[16]. The complex also includes two malaria vectors of local importance with much more limited ranges: *An. melas*, which is restricted to brackish water on the west coast of Africa, and *An. merus*, which mirrors *An. melas* on the east coast of Africa[13].

An. quadriannulatus is unlike the species listed above in that it is zoophilic and hence is not a malaria vector[17]. Indeed, while *An. quadriannulatus* feeds on human hosts, it generally prefers bovids[18,19]. Because *An. quadriannulatus* is closely related to the other members of the complex but exhibits such divergent behavior, it is a highly valuable point of contrast for understanding the basis of anthropophilic behavior.

In addition to the species listed above, which are all included in the molecular evolution study, the *An. gambiae* complex also includes the following other species, for which genetic data is largely lacking. *An. amharicus* was formerly considered a separate form of *An. quadriannulatus*, but has since been described as a separate species, and is

primarily distributed in northeastern Africa[14]. Very little is known about *An. bwambae*, although it is presumed to be a locally important malaria vector[20]. *An. fontenillei* was described in 2019, and little is yet known about its biology, although it clusters with *An. coluzzii* and *An. gambiae* s.s.[11]. Furthermore, there are two additional cryptic taxa within the *An. gambiae* complex, which have not been formally described, but are designated as GOUNDRY and TENGRELA[21].

While the constituent species of the *An. gambiae* complex are not the only malaria vectors in sub-Saharan Africa, they nonetheless account for substantial *Plasmodium falciparum* transmission. The other major African malaria vectors outside of the *An. gambiae* complex are also highly anthropophilic, and include *An. funestus*, *An. moucheti*, and *An. nili*[13]. Furthermore, other major malaria vectors from other parts of the world, including *An. stephensi*, *An. maculatus*, *An. culicifacies*, *An. dirus*, *An. farauti*, *An. atroparvus*, *An. sinensis*, and *An. darlingi*, are also known to be anthropophilic, although the degree of this anthropophily varies by species and sometimes more closely resembles the generalist anthropophilic tendencies of *An. arabiensis* rather than the extreme anthropophily of the *An. coluzzii-An. gambiae* s.s. clade[13,22,23].

As such, anthropophily can be understood as the major determinant, along with high vector competence, of the public health importance of a given anopheline mosquito species. Therefore, understanding the genetic, anatomic, and evolutionary basis of anthropophily is crucial for the development of novel control tools not only where the

need is greatest in sub-Saharan Africa, but also in other transmission hotspots, such as the Indian subcontinent and the tropical Americas.

Given widespread insecticide resistance in *Anopheles*[24], drug resistance in *Plasmodium*[25], and behavioral modifications in response to indoor residual spraying and long-lasting insecticide-treated nets[26,27], novel control tools are desperately needed. However, in order to develop such tools, we need a strong understanding of basic mosquito biology, particularly the genetic basis of behaviors that influence vectorial capacity, such as mating, sugar-seeking and nectar-feeding, and anthropophagy. These behaviors rely on a variety of sensory cues, which rely on complex genetics, anatomy, neurobiology, and physiology. Our understanding of the bases of these sensory-driven behaviors has advanced dramatically since the early days of insect physiology, and is rapidly developing.

1.2. Insect Sensory Systems and Genetics

Like vertebrates, insects have various means of detecting a wide variety of stimuli. Well-described insect sensory abilities include vision[28] (including the ability to detect light polarity, as well as ultraviolet and infrared spectra[28,29]), chemosensation (including olfaction, i.e. the detection of volatile cues, and gustation, i.e. the detection of contact chemicals)[30], mechanoreception (touch, which includes both external contact and internal "stretch" reception, as well as audition/hearing, the detection of vibrations)[31–33], thermoreception (the detection of heat)[34,35], hygroreception (the detection of humidity)[34,35], and nociception (the detection of pain)[36].

Most insect sensory systems rely on sensory hairs known as sensilla, which are found throughout the body, and exhibit substantial structural variation depending upon their function[37]. Sensilla perceive external stimuli, and then convey these stimuli via action potentials, which can be tonic (ongoing), phasic (at the beginning or end of a stimulus, or both), or phasic-tonic[37].

The evolution of these sensory systems can be adaptive and thereby enable insects to fill particular niches, which can in turn lead to speciation. In vision, this adaptive evolution can range from the macro-level (e.g. facilitating the predatory lifestyle of Odonata[38]) to the discrimination of exclusively conspecific mating signals in Lampyridae[39]. Hearing facilitates conspecific recognition in Auchenorrhyncha[40] as well as in mosquitoes[41]. Chemosensory genes are known to play a role in host-associated differentiation in butterflies[42], *Drosophila*[43], the yellow fever mosquito *Aedes aegypti*[44], and also show clear signs of species-biased expression in *Anopheles*[45–48].

1.2.1. Insect Chemosensory Systems and Genetics

Chemosensation plays a critical role in the biology of numerous insects. With respect to food alone, chemosensation facilitates the long-range detection of plant and animal hosts or other food sources[49,50], contact assessment of food quality or toxicity[51], and internal nutrition monitoring[52]. In addition to this intuitive role in nutrition, the chemosensory systems also factor in mating biology[53] and oviposition[54].

Chemosensation occurs throughout the insect body, including internally, although it is best understood from the external olfactory and gustatory organs. These organs include the antennae, palps, proboscis, tarsi, and wings[30,55]. Olfactory sensilla

have widespread pores, which allow the penetration of volatile cues, whereas gustatory sensilla have a single terminal pore on the distal tip, which is better suited to contact cues (Figure 1.1)[56]. Olfactory sensilla are classified into three major classes: basiconic, trichoid, and coeloconic. Basiconic sensilla typically house olfactory receptors (ORs), although some gustatory receptors (GRs) are also found there. Trichoid sensilla exclusively house ORs, while coeloconic sensilla typically house ionotropic receptors (IRs)[56]. Olfactory information, which is perceived in peripheral organs is processed in the antennal lobe, subesophageal zone and lateral horn of the brain[57,58].

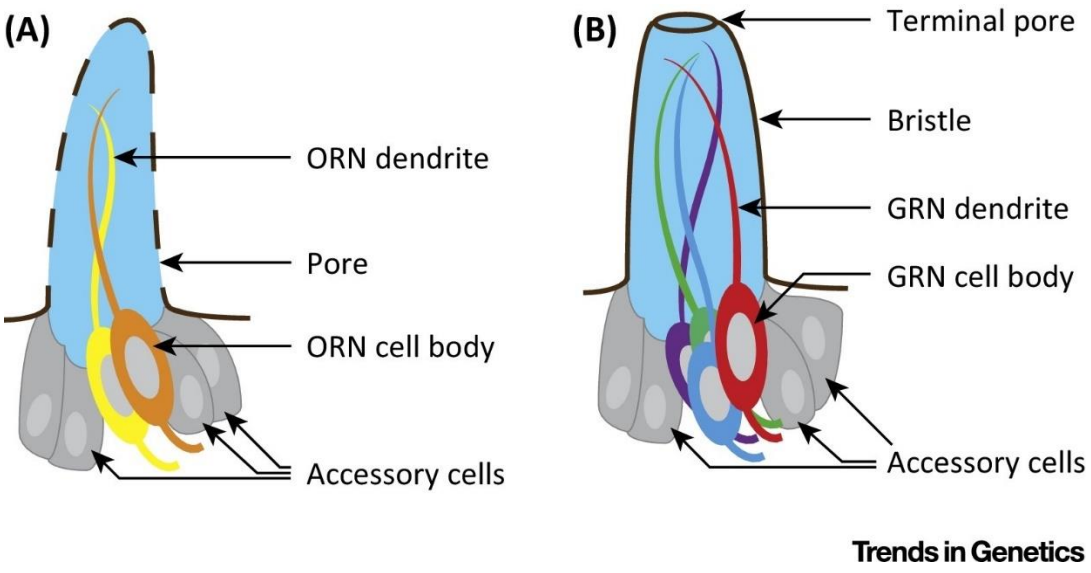

Figure 1.1 – Schematics of A) Olfactory and B) Gustatory Sensilla. Reproduced with permission from Joseph & Carlson, 2015[56].

There are four primary gene families involved in insect chemosensation, with some others playing a lesser role. Of the four primary gene families, three encode receptors: the olfactory receptors (*Ors*)[59], ionotropic receptors (*Irs*)[60], and gustatory receptors (*Grs*)[61]. The fourth primary gene family, the odorant binding proteins (*Obps*), encode carrier proteins. In addition to these four, other gene families that play a role in insect chemosensation include the chemosensory proteins (*Csps*)[62], sensory neuron membrane proteins (*Snmps*)[63], odorant degrading enzymes (ODEs)[63], Pickpockets (PPKs)[56], TRP channels[56], and CheA/CheBs[64,65]. There is also evidence that opsins, which are canonically associated with vision, are required for gustatory responses to certain bitter compounds[66,67].

ORs form heterodimers with the obligate olfactory co-receptor, ORCO, and one specific OR (Figure 1.2)[68]. While some Ors respond exclusively to one ligand, the majority are activated by multiple ligands. Furthermore, ligands are not restricted to any particular receptor, and often activate more than one. Therefore, a combinatorial system relying on input from multiple different receptors is likely the dominant mode of odorant perception in insects[30]. *Ors* are not homologous with the vertebrate olfactory receptors, which are G-protein coupled receptors (GPCRs). Rather, insect *Ors* are also seven-transmembrane domain proteins, but have an inverted topology and lack sequence homology with GPCRs[69].

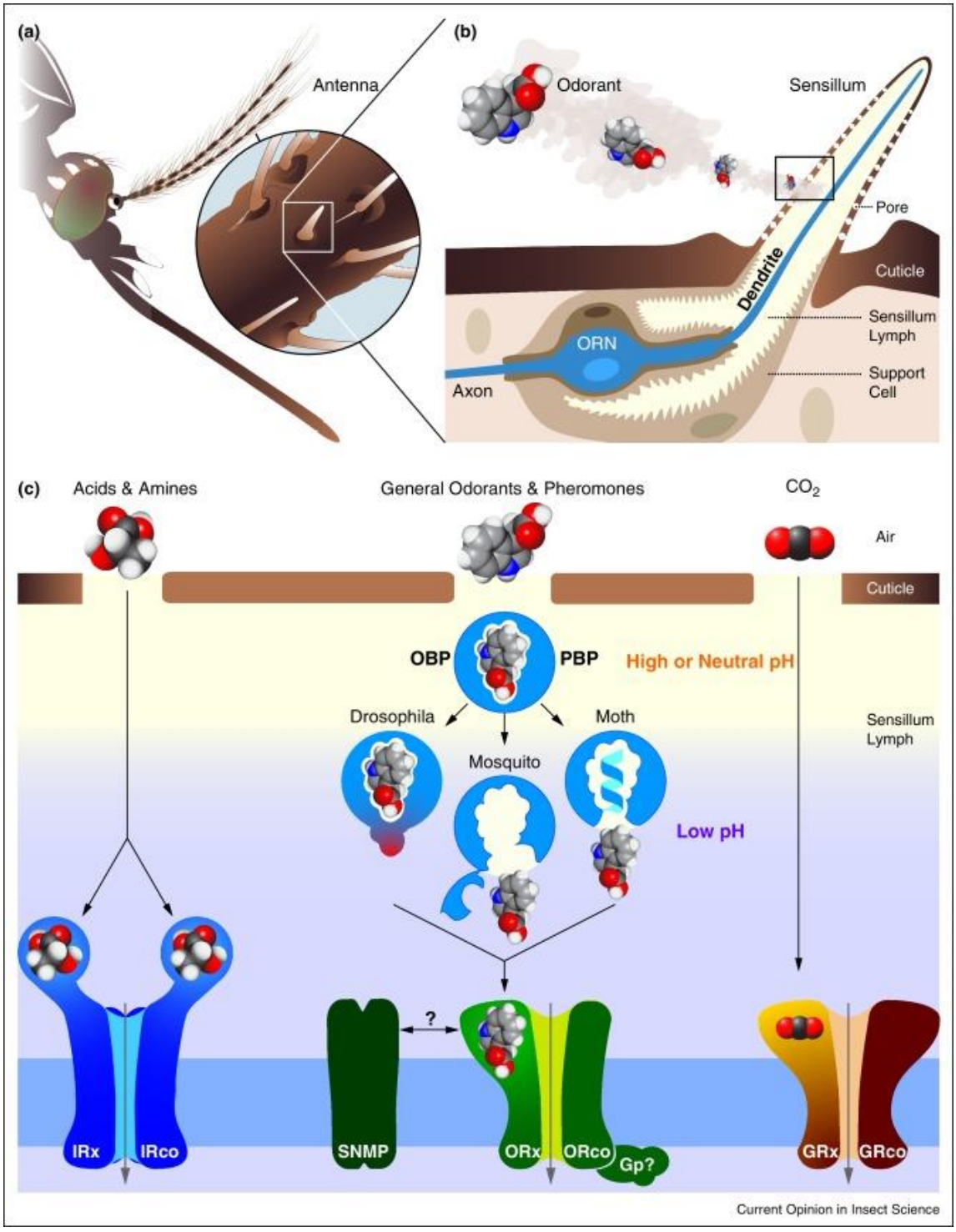

Figure 1.2 – Schematic of Olfactory Sensillum and Associated Receptors and Binding Proteins. Reproduced with Permission from Suh *et al.*, 2014[70].

In contrast to *Ors*, the *Irs* have three different co-receptors: *Ir8a*, *Ir25a*, and *Ir76b*[71]. A given co-receptor can be expressed with up to two other *Irs* or occasionally *Grs*[71,72]. IRs are typically tuned to different chemicals than ORs, such as amines and carboxylic acids[73,74]. While *Ors* are restricted to insects, *Irs* are found throughout the protostomes[75]. Many are also highly conserved across insect orders, in contrast to the rapidly evolving *Grs* and *Ors*[76,77].

Like the *Ors*, the *Grs* are not homologous to vertebrate gustatory receptors. As with the *Ors*, they have an inverted topology[78]. In fact, insect *Ors* form a nested clade within insect *Grs*, meaning that they evolved from *Grs*[79,80]. *Grs* are even more ancient than the *Irs*, predating the evolution of Placozoa[75,81]. Within Arthropoda, they show widespread lineage-specific expansions[75,82,83]. Unlike the *Ors* and *Irs*, *Grs* do not consistently rely on co-receptors, and co-express with between one and four other *Grs*[56,72]. While they canonically respond to contact chemical cues such as sugars, pheromones, and bitter compounds[53,84–90], they also form a highly conserved CO_2 receptor[91,92].

Finally, *Obps* are often, but not exclusively, involved in the transport of hydrophobic chemical cues across the sensillar membrane. Their specific function has been debated[93], and appears to be more variable and complex than originally thought (reviewed by Sun *et al.*[94]). Nonetheless, they are expressed at very high levels in chemosensory tissues[45–48,95,96] and some of them show clear evidence of a chemosensory role[97–99]. Like *Ors*, they are restricted to insects.

1.3. Studies of Mosquito Host-Seeking Behavior

Behavioral studies of mosquitoes have a century-long history, and span a wide range of research questions from functional anatomy to the establishment of vertebrate and plant host preference, as well as attractiveness of constituent host cues. These studies have been run in small cages and Y-tube olfactometers, larger dual-port olfactometers, and in large field experimental apparatus. I present a summary of major findings from these studies, organized around the following larger themes: the anatomy of host-seeking; the role of heat in host attraction; the role of various volatiles in odor-mediated host attraction; the integration of multiple sensory modalities including vision, olfaction, and thermosensation in host seeking; and finally, studies on the inherent host preferences of members of the *An. gambiae* complex.

1.3.1. Anatomy of Mosquito Host-Seeking

In 1951, the organs necessary for attraction to a human hand as well as to a warm object in both *Aedes aegypti* and the North American malaria vector *An. quadrimaculatus* were established. The major chemosensory organs in *An. gambiae* are illustrated in Figure 1.3. In this landmark study, the antennae were shown to be critical for host attraction in both species, although their ablation did not prevent probing activity. Both species could also be attracted and induced to probe with the maxillary palps and proboscis removed, both independently and in combination. Furthermore, mosquito host attraction was reduced in the absence of either antennae or palps, implicating these organs as the major sites of olfactory-based host-seeking activity in both species. However, bilaterally antennectomized and palpectomized *An.*

quadrimaculatus could still be induced to probe, as could *Ae. aegypti* on occasion. Finally, the hind legs were proposed as close-range sense organs to detect either air currents or convection currents[100]. In 1976, both anthropophily and overall host response in *Aedes aegypti* were shown to be reduced after the ablation of antennae[101].

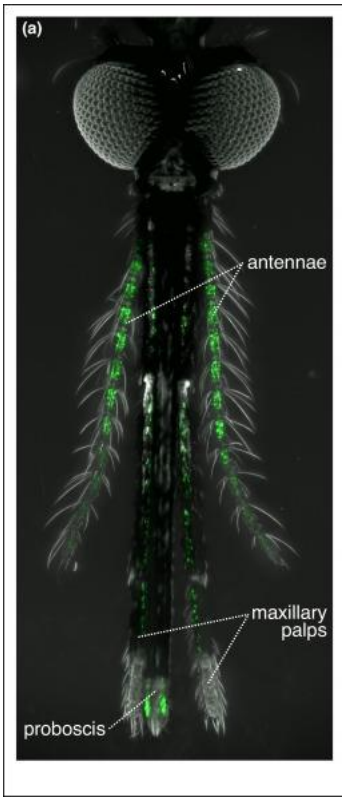

Figure 1.3 – Green Fluorescent Protein Labeled Chemosensory Neurons in *An. gambiae* Chemosensory Organs. Reproduced with permission from Benton, 2017[111].

The antennae were further implicated in host-seeking and thermotaxis in *Aedes aegypti* in 1958 and *An. atroparvus* in 1962, while the maxillary palps were shown to be

unnecessary for long-distance attraction but sufficient to induce biting in antennectomized mosquitoes, indicating chemosensory and/or thermosensory abilities[102,103]. The antennae and tarsi were implicated in hygroreception in 1960[104]. Hygroreceptors were identified in the antennae and carbon dioxide receptors were identified in the maxillary palps of *Aedes aegypti* in 1970[105], while the loss of CO_2 response was demonstrated in palpectomized *Culex quinquefasciatus* in 1971[106]. The CO_2 receptors in the maxillary palps were confirmed via electrophysiology in 1995[107].

Contact chemoreceptors on the tarsi and labella were identified in 1963, along with taste receptors in the cibarium of the winter mosquito *Culiseta inornata*. While the cibarial receptors were sensitive to blood, the receptors on the labella were not, indicating that their role in taste is separate from blood-feeding behavior[108,109]. In *Ae. aegypti*, sense organs on the labrum were found to be sensitive to blood in 1966, while the labella was instead found to be sensitive to sugars and water, and the role of the cibarium was confirmed to be regulation of blood-feeding as well as the ingestion of other liquids[110].

In 2011, Maekawa *et al.* identified the proboscis of *An. stephensi* as a "thermo-antenna", due to its necessity for successfully navigating to an artificial thermal target. They also found that *An. stephensi* with ablated antennae, maxillary palps, and proboscis landed on anesthetized mice in a small box less frequently, while there was no such decline for mosquitoes with ablated hind legs[10]. In addition, Choo *et al.* (2015) demonstrated that the labrum of the proboscis in *Culex quinquefasciatus* plays a major

role in detecting 4-ethylphenol, which is both an oviposition attractant and a blood feeding enhancer[54].

1.3.2. Heat and Thermosensation in Mosquito Host-Seeking

Attraction to warm objects was first established in *Culex quinquefasciatus, Aedes scuttelarius*, and the North American malaria vector *An. punctipennis* in the 1910s[112,113]. A 1949 study showed that host-seeking female *Aedes aegypti* were differentially attracted to objects in the 100-110˚ F (37.8 – 43.3˚ C) range, rather than to objects outside of that range; this study also showed that thermotaxis was specifically directed by convection, which is now widely understood to play a role in inducing biting[114]. Despite this thermotaxis, heat alone was demonstrated to be insufficient to induce biting in *Aedes aegypti* in 1968[115]. Furthermore, the necessity of heat to induce mosquito attraction to host odor was questioned by data in *An. quadrimaculatus* and *Aedes aegypti*[116].

However, biting rates on different parts of the body were correlated with a combination of skin temperature and eccrine gland density in *An. atroparvus* and *An. gambiae* s.s. in 1995[117]. The attractiveness of hand-derived convection currents to *Ae. aegypti*, particularly when combined with water vapor, was established in 1994, although the convection currents were less attractive than sweat[118]. However, further evidence for the role of convection currents was demonstrated in a study in which *An. coluzzii*, *An. gambiae* s.s., *An. arabiensis*, and *An. quadriannulatus* were all found to prefer to bite the feet and legs of a seated volunteer, but bit more randomly along the body in a prostrate volunteer[119]. These laboratory results were confirmed in the field in

2015 with human landing catches, showing that *An. arabiensis*, *An. funestus*, and *An. gambiae* s.s. prefer to feed low to the ground, which typically manifests as a preference for the legs and feet[120]. Furthermore, while deodorant compounds inhibit host-seeking *An. coluzzii* landing, there is no inherent preference for one body part over another in the absence of such compounds, further indicating the importance of convection currents in determining bite location[121].

1.3.3. Volatile Odorants in Mosquito Host-Seeking

In 1922, *Aedes sollicitans* and *Aedes cantator* were found to be strongly activated by carbon dioxide and ammonia, as well as a combination of the two[122]. The first anopheline shown to be attracted to carbon dioxide was the European vector *An. atroparvus* in 1947, after zoophily had previously been demonstrated in this species (although other studies have found it to be more anthropophilic)[123]. The complexity of CO_2 as a host-seeking attractant became apparent in 1952, in a study where *Aedes aegypti* were attracted to a 10% CO_2 stream in a large olfactometer, but repelled in a small olfactometer at a wide range of CO_2 concentrations[124]. Attraction to high concentrations of CO_2 but repellency by lower concentrations was demonstrated in *Aedes aegypti* in 1968, in addition to attraction to blood, urine, and arm sweat, while incubated torso sweat was found to be repellent. Most importantly, this study demonstrated the attractiveness of constituent organic acids in sweat, including lactic (specifically the L(+)-isomer[125]), formic, acetic, and propionic acids. Furthermore, sweat was shown to produce a strong biting response, and some constituent organic acids (acetic, propionic, butyric, and pyruvic) were also found to induce biting. Finally,

glucose promoted a sucking response when mosquitoes were exposed to concentrations found in human blood[115].

In 1990, increased host-seeking activity in the presence of 5-10% CO_2 and high activity around the odor source in the presence of both skin emanations and 5% CO_2 were demonstrated in *An. coluzzii*, suggesting that CO_2 serves as a host-seeking activator while other host kairomones act to draw the mosquitoes to the host location[126]. Human breath was shown to induce biting in the neotropical malaria vector *An. albimanus* in 1994[127]. Furthermore, an activation threshold for CO_2 was demonstrated at 0.01% above background concentrations in 1995, which was replicated with the CO_2 contained in human breath[128].

However, the roles of CO_2, as well as acetone and 1-octen-3-ol, additional generic vertebrate host cues, were shown to vary between different species of *Anopheles* in 1997. While *An. stephensi*, which is more zoophilic than *An. coluzzii*, was highly attracted to CO_2 alone, a combination of highly concentrated acetone and CO_2, and a combination 1-octen-3-ol and CO_2, *An. coluzzii* was repelled by these cues. Conversely, *An. coluzzii* was attracted by a combination of lowly concentrated acetone and CO_2, which did not attract *An. stephensi*[129]. This study pointed to the greater importance of generic host cues for zoophilic than anthropophilic *Anopheles*, which was also more explicitly demonstrated in members of the *An. gambiae* complex (described below). Another study demonstrated that *An. quadriannulatus* was more likely to bite the head of a human host than the other members of the *An. gambiae* complex, likely due to the higher concentration of generic volatiles, such as CO_2 and acetone[119].

An. coluzzii was shown to be attracted to constituent acids in human sweat in 1997, an attraction which was maintained even if sweat was artificially alkalinized, indicating that other constituent sweat chemicals also play a role in host attraction[130]. In addition, both *An. coluzzii* and *An. gambiae* s.s. were shown to be highly attracted to foot odor, as well as to Limburger cheese, notorious for its foot-like smell, in 1995[131]. This attraction was later tied to the similar bacterial compounds produced in the cheese and on human feet[132]. The role of bacteria in generating human kairomones attractive to mosquitoes was further elucidated in 1999, when it was demonstrated that mosquitoes were much more attracted to incubated sweat (where bacteria had grown) than to fresh sweat[133]. Further work has confirmed that bacteria produce most of the skin volatiles that attract mosquitoes[134].

Dekker *et al.* (2005) showed that *Aedes aegypti* became rapidly sensitized to human host odor in the presence of CO_2[135], although other data suggests the same is not true in *An. gambiae*[19]. *Aedes aegypti* is also unable to navigate to a host odor unless there is sustained exposure to it[136]. Once a specific host odor has been identified, host-seeking *Aedes aegypti* do not navigate based on CO_2 and instead rely on the specific host odor[137]. Elevated background levels of CO_2 impede the start of host-seeking in *Aedes aegypti*[138], while a slight increase in CO_2 relative to background levels enhances the landing response of *An. gambiae* in the presence of human skin volatiles[139]. Furthermore, CO_2 is of variable importance at different ranges corresponding to preferred hosts in *An. coluzzii*, *Aedes aegypti*, and *Culex quinquefasciatus*[140]. In addition, acetone and CO_2 have different impacts on host-seeking by *An. coluzzii* and *Aedes*

aegypti. Specifically, acetone is more attractive to *Aedes* than to *Anopheles*, and a higher concentration of acetone reduces *An. coluzzii* attraction to CO_2[141].

1.3.4. Integration of Sensory Modalities in Mosquito Host-Seeking

Color, humidity, and light intensity were all found to influence biting in various species of mosquitoes in the 1910s[122]. More recently, van Breugel *et al.* (2015) demonstrated that host-seeking *Aedes aegypti* associate odor plumes, including CO_2, with visual cues to land on targets with suitable heat and humidity signatures[6]. However, Liu and Vosshall (2019) found that CO_2 does not affect visual responses, but that it is required to unlock heat attraction[7]. Flight studies of host-seeking *Culex quinquefasciatus* have demonstrated attraction to warmed glass beads treated with human foot odor, which was not enhanced in the presence of CO_2. In addition, foot odor was found to modify flight patterns, such that *Culex quinquefasciatus* flew more slowly in the presence of foot odor than in clean air or CO_2 alone[142].

Finally, recent studies on both *An. coluzzii* and *Aedes aegypti* have demonstrated that mosquitoes can learn and modify their behaviors based on exposure to hosts and host odors, and that their past experiences can influence future host choice[143,144]. The rapid learning of host odor compounds in *Aedes aegypti* was found to depend on the dopamine-1 receptor[144]. Furthermore, olfaction-based behavior in *Aedes aegypti* is time-dependent, and individual odorants have specific effects on olfactory sensitivity[145].

1.3.5. Host Preferences of *An. gambiae* Complex Members

Schreck *et al.* (1990) demonstrated a biochemical basis to the differential attractiveness of individual humans to mosquitoes, and also suggested that heat was not

a necessary cue to induce host-seeking[116]. Various studies have demonstrated variable attractiveness of different people to malaria mosquitoes, and have further demonstrated that adults are more attractive than children[17].

In 2004, it was established that both male and female *An. gambiae* prefer honey volatiles to human volatiles one night post-emergence, but that by six nights post-emergence, females prefer human volatiles. This study confirmed field data showing sugar-feeding as a major element of female *An. gambiae* biology[146].

An. coluzzii was shown to consistently choose human odor over clean air or cow odor in a dual-port olfactometer, while *An. quadriannulatus* chose clean air over human odor, thus indicating an innate olfactory basis to their respective host preferences[147]. Furthermore, *An. coluzzii* strongly prefers human odor to both cow odor and CO_2 (pointing to the relative unimportance of this host cue in *An. coluzzii*), and can distinguish human odor even when combined with cow odor[148]. However, *An. quadriannulatus* fed equally on a human or a cow when placed next to one another in an outdoor tent experiment, indicating that this species is less able to finely distinguish vertebrate odor cues than *An. coluzzii*[149]. In a series of dual-choice olfactometer experiments with *An. quadriannulatus*, the mosquitoes were unresponsive to abundant chemicals found in cow odor profiles, specifically CO_2, acetone, and 1-octen-3-ol. The mosquitoes responded to CO_2 in combination with cow odor, but surprisingly preferred human odor, which was attributed to a genetic basis in combination with behavioral plasticity that is less likely to be observed in a natural setting[19].

In a field experiment incorporating bovine-baited and human-baited traps, *An. gambiae* s.l. and *An. funestus* showed a strong preference for human odor. When the *An. gambiae* s.l. in both traps were identified to species, *An. gambiae* s.s. and *An. arabiensis* were relatively equally represented, whereas the cow trap overwhelmingly contained *An. arabiensis*[150]. In another study with *An. arabiensis* and *An. quadriannulatus* where mosquitoes could choose between human odor, bovine odor, and various concentrations of CO_2, *An. quadriannulatus* overwhelmingly chose CO_2 over human odor, while *An. arabiensis* overwhelmingly chose the human odor. *An. quadriannulatus* was further preferentially attracted to higher concentrations of CO_2 than lower, whereas the concentration did not affect host-seeking in *An. arabiensis*, thus indicating that CO_2 plays an increasing role with increasing zoophily[18].

A field study of *An. arabiensis* in southern Ethiopia demonstrated an inherent anthropophagic tendency even in the presence of large numbers of cattle, although this tendency was often counterbalanced by exophagy and opportunistic feeding on abundant hosts[16]. Furthermore, a field study of *An. arabiensis* and *An. quadriannulatus* in Zimbabwe demonstrated increasing *An. quadriannulatus* catch in relation to the size of the bait, with a human and ox together drawing the most mosquitoes. In *An. arabiensis*, three humans attracted more mosquitoes than a single human, and human odor was more attractive than the combination of human and ox odor that attracted *An. quadriannulatus*. However, there was also considerable plasticity in *An. arabiensis* host preference, and CO_2 was ruled out as the main chemical cue inciting *An. arabiensis* to enter human habitations[151].

1.4. Studies of Chemosensory Gene Expression in Mosquitoes

Several studies have been performed to characterize the transcriptome of the major chemosensory organs in mosquitoes, which include the antennae, maxillary palps, labella (the distal-most portion of the proboscis), and tarsi[45–47,95,96,152,153]. Other studies have characterized the expression of chemosensory genes in either the head or the whole body. These studies have predominantly been conducted on the yellow fever mosquito, *Aedes aegypti*, and the *An. gambiae* complex, although expression studies have also incorporated the southern house mosquito, *Culex quinquefasciatus*, and the Asian tiger mosquito, *Aedes albopictus*.

In addition to characterizing the transcriptomes of host-seeking mosquitoes, researchers have also analyzed differential expression between species and sexes, assessed the influence of age and blood-feeding, and demonstrated rhythmic expression of certain chemosensory genes[45–47,95,154–157]. I first summarize these studies in culicine mosquitoes, then describe the existing literature from the *An. gambiae* complex.

1.4.1. Chemosensory Gene Expression in *Aedes* and *Culex*

The chemosensory transcriptome in *Aedes aegypti* was first assessed with RT-PCR of *Ors*, which identified several female-specific *Ors* in the antennae, as well as several larval-specific *Ors*, one *Ors* expressed solely in the maxillary palps, and one *Or* expressed only in the maxillary palps and proboscis[158]. Following that study, there have been comprehensive RNA-Seq studies of the antennae, maxillary palps, labella, tarsi, and internal organs, which have demonstrated expression of multiple chemosensory gene families, along with sex-biased and organ-biased expression patterns[96,153,159,160]. As such,

there is strong evidence that these genes influence sex-specific behaviors, which rely more on some organs than others, such as blood-feeding, mating, and oviposition.

Arguably the most consequential study of chemosensory gene expression in *Aedes aegypti* linked the expression of *AeaegOr4* to human host preference[44]. In addition, chemosensory expression in the antennae has been compared in both sexes of *Aedes aegypti* one, three, and five days post-emergence. This study showed a clear influence of both sex and age on gene expression, particularly in the one-to-three day post-emergence period where attraction to human hosts peaks (unlike *Anopheles*, which mate in swarms further away from hosts, *Aedes aegypti* mates in close proximity to a host, so males are also attracted to human host odors)[157].

Aedes albopictus has been studied much less extensively, largely due to the low quality of the published genome. Nonetheless, a study on male and female antennae, maxillary palps, heads, and full bodies demonstrated differentially expressed chemosensory genes by sex and organ[161].

Significant enrichment of olfactory genes has also been demonstrated in the antennae of *Culex quinquefasciatus*, as has sex-biased expression[162,163]. Furthermore, a small number of *Ors*, *Irs*, and *Csps*, along with a larger number of *Obps* show significant differential expression between host-seeking and blood-fed *Culex quinquefasciatus*[164].

1.4.2. Chemosensory Gene Expression in *Anopheles*

Chemosensory gene expression studies in the *An. gambiae* complex have focused on *An. coluzzii* and *An. gambiae* s.s., although comparative studies have also incorporated *An. quadriannulatus*. An initial study using RT-PCR showed that four

newly identified *Ors* (*AgOrs*1-4) showed enhanced expression in the antennae and maxillary palps of both sexes. In addition, this study demonstrated female-specific expression of *Or1* in the antennae, as well as the down-regulation of *Or1* after blood-feeding[165].

A qRT-PCR study demonstrated sex-biased expression of *Obps* in the antennae, as well as *Obp* down-regulation after blood-feeding[166], while a combined microarray and qRT-PCR study further demonstrated sex-biased expression in the maxillary palps, as well as up-regulation of some male-biased *Obps* in bloodfed females[167]. This study was followed by another RT-PCR and qRT-PCR study that identified sex-biased expression of 80 *Ors* in the antennae and maxillary palps[168].

The first RNA-Seq survey of chemosensory gene expression in *An. coluzzii* analyzed transcriptomes from male and female antennae, maxillary palps, and whole bodies. This study illustrated the strong enhancement of *Ors*, *Irs*, *Obps*, and to a lesser extent, *Grs* in the antennae and palps compared to whole bodies. Furthermore, it showed sex-biased expression in these organs among all four gene families[152]. This study was followed by the first comparative RNA-Seq study with *An. quadriannulatus*, which demonstrated strong species-biased expression of chemosensory genes in the antennae[45]. Another RNA-Seq study more thoroughly analyzed the role of blood-feeding in regulating chemosensory gene expression, demonstrating both decreased and enhanced expression among all four gene families in antennae after a blood meal[156].

Our lab has also demonstrated both sex- and species-biased expression of chemosensory genes in both the antennae and the maxillary palps of *An. coluzzii* and *An.*

quadriannulatus[46–48], while another study characterized the chemosensory gene expression profile of the labella in male and female *An. coluzzii*[95]. Finally, two studies have demonstrated the daily rhythmic expression of *Obp* RNA as well as proteins in whole bodies, heads, and antennae of *An. gambiae* s.s.[154,155].

1.5. Evolution of Chemosensory Genes

The size of chemosensory gene repertoires varies enormously throughout Animalia. Many expansions and contractions of these gene families are likely adaptive and serve to facilitate the exploitation of novel ecological niches. However, much of this evolution is also attributable to drift[169]. A common pattern in both vertebrate and invertebrate chemosensory gene evolution involves the rapid lineage-specific duplication and subsequent pseudogenization of chemoreceptors, known as the "birth-and-death" model[83,169]. However, these patterns vary by gene family and by lineage, with some gene families remaining highly conserved across taxa while others share minimal DNA sequence homology even within genera.

1.5.1. Invertebrates

As mentioned above, there are four major chemosensory gene families in insects, two of which are also shared more broadly among invertebrates: the *Grs* and *Irs*. While *Grs* are absent from sponges, ctenophores, and choanoflagellates, they are found in placozoans, cnidarians, mollusks, nematodes, onychophorans, chelicerates, myriapods, and crustaceans[75]. While they were originally described as G-protein coupled receptors and are also seven-transmembrane proteins, they have an inverted topology and are not related to any vertebrate chemoreceptors[69,170]. *Irs* are less ancient than *Grs*, but are

nonetheless found in protostomes including mollusks, nematodes, onychophorans, and arthropods[75].

The chemosensory repertoire varies substantially between different groups of invertebrates as well as between invertebrates and vertebrates. Arthropods generally have fewer chemosensory genes than both other invertebrates and vertebrates. *Gr* sequences have strikingly low homology between lineages, indicative of their rapid evolution via duplication. *Irs* are more conserved, and the co-receptor *Ir25a* originated prior to the differentiation of protostomes, while the co-receptor *Ir76b* occurs throughout arthropods, and the co-receptor *Ir8a* is found in crustaceans and myriapods, but not chelicerates[75].

1.5.2. Arthropods

As discussed above, the *Grs* and *Irs* are found throughout arthropod lineages, predating the shift to terrestrial life. *Csps* are also found throughout Arthropoda, although their chemosensory-specific role has been questioned[75,171,172]. There are numerous lineage-specific *Gr* expansions and contractions throughout Arthropoda, which mostly occurred after taxonomic diversification had already taken place. However, this is not uniformly the case, as amphipods and copepods have relatively small suites of *Grs*[75].

Several *Irs*, including the co-receptors and some others, are highly conserved across Arthropoda, whereas others are much more divergent and appear to have undergone lineage-specific expansions. A particularly large divergent *Ir* expansion is found in Diptera. *Ir* evolution is thought to occur via gene duplication and retroposition,

as detected by substantial intron loss[76]. Both *Grs* and *Irs* show large lineage-specific expansions and duplications throughout the chelicerates, which likely evolved via sustained birth-and-death[173]. *Irs* are thought to be the major olfactory genes in chelicerates, and there is also evidence of *Obp*-like genes in chelicerate sensory appendages[174].

1.5.3. Insects

Insect *Ors*, which are not related to vertebrate *ORs*, form a clade within the *Grs*, from which they arose[79]. They are thought to have been adaptive to the terrestrial lifestyle of insects, as they facilitate the perception of airborne chemical cues[80]. *Ors*, like *Grs*, have undergone numerous lineage-specific expansions and contractions throughout insects. Most famously, they were lost during the shift to a specialist feeding strategy in *Drosophila erecta* and *D. sechellia*[175,176]. *D. sechellia* also has numerous deleterious fixed mutations due to a bottleneck, although the bottleneck alone does not account for the observed molecular evolutionary patterns[177]. Purifying selection has been implicated as a major force in the evolution of *Ors* in several lineages, although positive selection has also accompanied host shifts[82,178]. The evolution of the *Grs* in insects has been similar, though not identical to that of the *Ors*, and is discussed in detail in Chapter 4.

The *Obps* are also specific to insects[75]. Like the other chemosensory genes, the *Obps* have undergone numerous birth and death events and lineage-specific expansions[179]. There are two groups of *Obps*, the classic and plus-C[83] *Obps*, with the classic *Obps* appearing to be the basal lineage[179]. There is some homology between the *Obps* and *Csps*, suggesting both that they comprise part of a larger family of binding

proteins, as well as that they may mirror the evolution of the *Grs* and *Ors* within Arthropoda. Specifically, *Grs* and *Csps* are both widespread throughout arthropods and represent ancient chemosensory genes, whereas *Ors* and *Obps* are unique to hexapods and may have facilitated the shift to terrestriality[179]. Like *Ors* and *Grs*, *Obps* show signatures of positive selection in parallel to host-associated diversification events[83,180–182].

1.6. Implications for Novel Mosquito Control Techniques

A fuller understanding of the anatomy and genetics underlying vertebrate host preference not only improves our knowledge of the basic biology of an important vector, but also may be instructive for efforts to develop transgenic non-anthropophilic mosquitoes. It is clear from numerous previous studies (outlined above) that the chemosensory system plays a major role in determining anthropophily, and is therefore a promising site for transgenic modification of vertebrate host preference. While the genetic determinants of anthropophily in *An. gambiae* s.l. are complex, there are several *Obps*, *Ors*, *Irs*, and even some *Grs*, which multiple forms of evidence have identified as potentially playing a role in this behavior. Furthermore, previous work has already demonstrated the viability of modifying the vertebrate host preferences of *An. gambiae* s.l. via artificial selection[3,5,182].

CRISPR-*Cas9* is an increasingly tractable technique in mosquitoes, and has been successfully used to modify multiple major vectors including *Aedes aegypti*[183], *Culex quinquefasciatus*[184], and most importantly the malaria vectors *An. stephensi*[185] and *An. gambiae*[186]. As summarized by Adolfi *et al.* (2018), CRISPR, along with other

transgenic techniques, are currently under evaluation to control several species of malaria mosquito. These transgenic approaches are used mostly to either substantially reduce vector population sizes or to enhance vector immunity to *Plasmodium*[187]. These techniques have been enhanced by the recent development of gene drive systems to ensure the rapid spread of desired alleles, which greatly reduce operating costs and are much more sustainable when compared to more frequent transgenic releases[186–189].

Our lab has previously generated a *cycle*-knockout line in *Aedes aegypti*, and current work aims to generate an *AgGr33*-knockout in *An. coluzzii*. Future knockouts of genes identified in both this study and previous ones that show evidence of a role in anthropophily, such as *AgObp26*, have the potential to both further our understanding of the genetic basis of anthropophily, as well as to potentially abolish the inherently anthropophilic nature of *An. coluzzii*.

1.7. Dissertation Overview

I begin in Chapter 2 by discussing the transcriptome of the labella, which is known to be involved in chemosensory activity but has been much less thoroughly studied than the better-known chemosensory organs, the antennae and the maxillary palps. In my study, I analyze pairwise differential gene expression patterns between both males and females of the anthropophilic *An. coluzzii* and zoophilic *An. quadriannulatus*, to identify genes showing evidence of species-biased roles, as well as intraspecific comparisons between sexes, to identify genes underlying sex-specific behaviors such as blood-feeding and mating. While I focus primarily on four known families of chemosensory genes, I also discuss a broader gene ontology analysis.

In Chapter 3, I describe a series of behavioral experiments on the roles of the three major chemosensory organs described above for the activation of host-seeking as well as the ability to successfully discriminate between human and bovine odors. In these experiments, I mechanically ablated each of these organs in different experimental groups and exposed the mosquitoes to a dual-choice olfactometer in order to determine which organs are most important for both vertebrate host-seeking and host preference. In the chapter describing this work, I also describe another set of experiments, wherein I tested the impact of chemosensory organ ablation on blood-feeding success, with the aim of determining the importance of each organ for closer-range host-seeking and probing behaviors.

Finally, in Chapter 4, I present a study of the molecular evolution of one family of chemosensory genes, the gustatory receptors. For this study, I incorporate data from the 16 *Anopheles* genomes project[190] for six constituent species of the *An. gambiae* complex: *An. arabiensis*, *An. coluzzii*, *An. gambiae* s.s., *An. melas*, *An. merus*, and *An. quadriannulatus*. I detail evidence of genes potentially undergoing positive selection, selective sweeps, and purifying selection, which provides us with a sense of the evolutionary forces at work on this gene family in the complex.

2. SPECIES AND SEX-BIASED EXPRESSION OF CHEMOSENSORY GENES IN THE SISTER SPECIES *ANOPHELES COLUZZII* AND *ANOPHELES QUADRIANNULATUS*

2.1. Introduction

Host preferences vary widely within the *Anopheles* (*An.*) *gambiae* sensu lato (s.l.) species complex, and play a predominant role in determining vectorial capacity. The *An. gambiae* complex consists of nine morphologically indistinguishable species, including three major malaria vectors: *An. gambiae* sensu stricto (s.s.), *An. coluzzii*, and *An. arabiensis*. *An. gambiae* s.s. and *An. coluzzii* were formerly considered to be one species[14], and exhibit identical host-seeking behavior and host preference. Both species are highly anthropophilic and predominantly feed on human hosts[148]. While *An. arabiensis* is also anthropophagic, it exhibits more generalist feeding habits[18]. Another species in the complex, *An. quadriannulatus*, is zoophilic and predominantly feeds on bovine hosts[18]. Given the close relationship between *An. coluzzii* and *An. quadriannulatus,* and their strongly divergent host preferences, they are of significant interest in determining the genetic basis of anthropophily, a trait that underlies high vectorial capacity for malaria.

The adaptation to human hosts likely involved genes controlling chemosensory behaviors. Though visual and thermal cues play a role[6,7], host preference is largely mediated by chemosensory systems. The chemosensory sensilla of mosquitoes lie principally on the antennae, maxillary palps, proboscis, and to a lesser degree on the

tarsi[37]. Though chemosensory gene expression has been thoroughly studied in the antennae[45,46,152,191] and the maxillary palps[46,91,152,159], the proboscis has received less attention, particularly in *Anopheles*[95].

Antennae are the primary olfactory organs throughout Insecta and contain many olfactory sensilla. In mosquitoes, the maxillary palps respond to carbon dioxide, as well as several other odorants that may be host cues[91,92,107,152,159]. The proboscis primarily plays a gustatory role in mosquitoes[37,192], although in both *Anopheles* and *Aedes* this organ also expresses olfactory receptors (*Ors*)[193,194]. In *An. stephensi*, the proboscis also plays a role in heat detection[10]. The labella, at the distal tip of the proboscis, contain a chordotonal organ in each lobe, which is surrounded by chemosensory neuron axons, and thought to act as a proprioceptor guiding feeding behavior[37,192]. Three types of hairs are present on the labella, designated T1, T2, and T3. T1 and T2 are externally located on the labella, while T3 are on the oral (i.e. facing the food canal[195]) surface of the labella. Of the five T1 dendrites, four are chemosensory (two are sensitive to salts, one to water, and one to sucrose[196]) and one is mechanosensory. T2 hairs are salt-sensitive and mechanosensory, and T3 hairs perceive sugars[37,108–110,192,197].

Because the proboscis is the feeding organ, its chemosensory role was originally considered exclusively gustatory and mechanosensory prior to the detection of *Orco* and other olfactory gene expression within labellar tissues[37,108,192,193]. Mosquitoes fail to probe when the distal tip of the proboscis is removed, and the proboscis plays a smaller role than other chemosensory organs in the detection of repellents[198]. The proboscis has also been shown to play a role in oviposition in the Southern house mosquito, *Culex*

quinquefasciatus[54].

Chemosensory sensilla contain olfactory sensory neurons (OSNs) or gustatory receptor neurons (GRNs). OSNs are found within olfactory sensilla and contain olfactory receptors (ORs) or ionotropic receptors (IRs). Olfactory sensilla are perforated by several small pores, which facilitate contact with diffuse volatiles[70]. ORs are heterodimers composed of the obligatory olfactory co-receptor (ORCO) and one of 75 specific ORs[68]. *Ors* are unique to insects and are not homologous to vertebrate olfactory receptors, which are G-protein coupled receptors. Although insect ORs are also 7-transmembrane proteins, they have an inverted topology[69,170]. IRs are also multimeric, containing up to three different IRs, including one or two co-receptors (IR25a, IR76b, IR8a)[71]. *Irs* are found throughout the protostomes, with *Ir25a* conserved in all protostome species thus far studied[71]. Although *Irs* are widespread, they were only recently shown to play an olfactory role[60].

Gustatory sensilla have a different morphology than olfactory sensilla, with a single pore at the distal end, which facilitates the perception of contact chemical cues (Figure 1.1)[56]. Within each GRN, multiple gustatory receptors (GRs) are co-expressed: sugar receptors are known to be multimeric, and bitter receptors are also thought to be multimeric[55,72,88]. While most GRs are thought to respond to non-volatile chemicals, three GRs (*Gr*22-24) function as the CO_2 receptor within the maxillary palps[91]. In addition, *Gr33* is among the most highly expressed chemosensory genes in the antennae of male *An. coluzzii* and *An. quadriannulatus*[47]. While GRs predominantly respond to tastants such as sugars and bitter compounds, they also play a role in thermosensation

and light transduction in *Drosophila*[199,200]. Insect *Ors* arose from *Grs* and they form a single clade nested with the *Grs*[79].

Carrier proteins also play an important role in olfaction and gustation, transporting hydrophobic chemicals across the sensillar lymph to the receptors, resulting in the start of the signal transduction cascade. Odorant Binding Proteins (OBPs) are thought to be the main carrier proteins that fill this role in insects[201] and *Obps* are among the most highly expressed genes in insect olfactory organs[70], although there is strong evidence that they fill other roles as well[62,202,203]. Like *Ors,* insect *Obps* are not related to those of vertebrates, but are instead unique to the hexapods[62].

Mosquito host-seeking is primarily a chemosensory-driven process, though other senses (e.g. vision, thermoreception) also play a role. While the host kairomones that *An. gambiae* and *An. coluzzii* use to distinguish between host species have yet to be identified, several volatiles produced by the various hosts are known to attract mosquitoes. For example, host-seeking is activated by carbon dioxide[204], and mosquitoes respond to lactic acid[125], indole[205], and 1-octen-3-ol[129]. Bacterial emanations from the skin are largely responsible for producing the volatile mosquito attractants[206]. While there is some evidence that *An. gambiae* is particularly attracted to the feet, there is also evidence that the host's position (e.g. sitting, standing, or lying down) plays a substantial role in where the mosquito chooses to feed, pointing to the importance of convection currents and widespread volatiles from the skin microbiome[119,120].

Differences in chemosensory gene expression have accompanied variation in host preference among sibling species in other insects. For instance, there are significant

differences in *Or* and *Obp* gene expression between the generalist *Drosophila* species *D. melanogaster* and *D. simulans*, and the specialist *D. sechellia*[207]. *D. sechellia* feeds exclusively on *Morinda citrifolia*, which is toxic to other species, and *Or22a* (a gene which detects a constituent odorant of *M. citrifolia*[208]) is significantly-upregulated in the generalist *D. sechellia* with respect to its generalist sibling species[207]. Expression of *Obp57e* is also involved in inducing female *D. sechellia* attraction to *Morinda* fruit as an oviposition site[209]. Furthermore, *D. sechellia* is losing *Ors* and *Grs* at a rate beyond ten times faster than that of the generalist *D. simulans* (and is generally losing genes due to a bottleneck[177]), and *D. sechellia* shows evidence of positive selection on those receptor loci that it retains[175]. A similar pattern is observed in another *Drosophila* specialist, *D. erecta*[43].

In mosquitoes, the expression of *Or4* in two different subspecies of *Aedes aegypti* is associated with a shift to anthropophily. The domestic mosquito *Aedes aegypti aegypti* has a significantly higher expression level of *Or4* than the sylvatic mosquito *Aedes aegypti formosus*, which corresponds to their relative preference for human hosts; the two forms also demonstrate clear separation in the form of nucleotide substitutions[44].

As part of our efforts to identify candidate genes for host-seeking, we previously identified several significantly differentially-expressed chemosensory genes between the antennae and maxillary palps of female *An. coluzzii* and *An. quadriannulatus*[46]. We have also identified differentially-expressed chemosensory genes between the males of each species (which do not blood-feed), as well as between males and females of both species[47]. By compiling the significantly differentiated genes from the antennae and

maxillary palps with those from the labella, we generate a relatively complete profile of the chemosensory genes most likely to be involved in host preference, which paves the way for a more detailed analysis of the genes that hold particular interest.

The role of the labella as a site of chemosensory gene expression in mosquitoes was originally established in *Aedes aegypti*, where several chemosensory gene families are detected in both male and female mosquitoes: *Grs*, *Irs*, *Ors* (including *Orco*), *Obps*, as well as other gene families that have been implicated in chemosensation in other insects but not *Anopheles*, such as sensory neuron membrane proteins (*Snmps*), TRP and Ppk channels, and *CheA/Bs*[96,153].

The gene expression profile of male and female *An. coluzzii* labella was also recently characterized[95]. There is relatively little differential expression between the sexes, though some significant genes were identified. As expected, the most highly expressed chemosensory genes are *Obps*, but members of all three receptor families are expressed. Thirty-two *Grs* are detected at levels that may be biologically relevant, highlighting their probable role as contact chemoreceptors in this appendage. The labellar role as olfactory organ is confirmed not only by the expression of *Orco* and thirteen other *Ors*, but also by electrophysiological responses to an odorant blend. In addition, seven *Irs* (including the co-receptors *Ir25a* and *Ir76b*) are expressed at biologically-relevant levels[95].

Here I conduct a comparative RNA-Seq analysis of chemosensory genes in the labella of male and female *An. coluzzii* and *An. quadriannulatus* to identify potential candidate genes for host choice in these species. With the inclusion of male data, I am

also able to discriminate between genes that appear to influence behavior unrelated to sex (e.g. sugar-feeding) and those that play female-specific roles and are therefore the strongest candidates for the influence of host-seeking.

2.2. Methods

2.2.1. Mosquito Rearing

The mosquitoes in this study were drawn from laboratory colonies of *An. coluzzii* and *An. quadriannulatus*. The *An. coluzzii* strain (formerly known as *An. gambiae* M form, GASUA) was originally collected in Suakoko, Liberia; the *An. quadriannulatus* strain (SANQUA) was originally collected in Sangwe, Zimbabwe. Both colonies were reared in an insectary at Texas A&M University, College Station, TX, USA. Colonies were maintained at 28º C, 70-80% relative humidity, and a 12 hours light: 12 hours dark photoperiod. Adult mosquitoes were maintained on a 10% sucrose solution and blood-fed with defibrinated bovine blood using a membrane feeding system twice a week. Larvae were maintained at densities of approximately 150 per 2-liter container and fed finely-ground fish food (TetraMin, Blacksburg, VA, USA). Pupae were collected and placed into cages at densities of two cups of 150 pupae per cage.

Cages were checked daily for newly emerged mosquitoes. To maintain consistent age, non-eclosed pupae were transferred to new cages. Mosquitoes were allowed to mate while being maintained for five days on the sucrose solution. The species identity of mosquito colonies was confirmed using the RFLP-PCR method of Fanello *et al*.[210].

2.2.2. Dissections and Sequencing

As anophelines naturally host-seek after dark, and *Orco* expression in *An. coluzzii* peaks after the start of the dark cycle, dissections were performed after the dark cycle had begun. Male and female mosquitoes were immobilized for 10 minutes at 4º C two hours after the beginning of the dark cycle, and were then sorted by sex into separate petri dishes to minimize the possibility of cross-contamination of samples. Dissections were performed under a 10x dissecting microscope on a TissueTek® cold plate (Sakura® FineTek USA, Torrance, CA) chilled to -20º C.

Labella (approximately the lower ¼ of the proboscis) were dissected from both male and female mosquitoes of both species. The labella were immediately placed in RNA*later*™-ICE (Ambion) and subsequently maintained at -20º C prior to extraction. Total RNA was extracted using the Qiagen (Hilden, Germany) RNeasy® mini kit incorporating an on-column DNA digestion. RNA from 100 labella were extracted as a single sample, yielding between 10 to 30 ng of RNA. cDNA libraries were prepared using the Lexogen (Vienna, Austria) QuantSeq 3' mRNA-Seq Library Prep Kit FWD for Illumina, with 8X FS2 buffer and Unique Molecular Identifiers (UMIs) added to allow the use of a larger amount (12 µL) of input RNA.

Six replicate libraries were produced for each sex and species, yielding a total of 24 libraries generated from approximately 2,400 mosquitoes. The libraries were quantified on a Qubit 4 Fluorometer (ThermoFisher Scientific, Waltham, MA, USA) and a quality control analysis was performed using RNA Pico LabChip analysis on an Agilent BioAnalyzer 2100 (Agilent Technologies). Two libraries (one female *An.*

quadriannulatus sample and one male *An. quadriannulatus* sample) failed quality control checks and were not included in the final pool. Libraries were then multiplexed and sequenced over two lanes of an Illumina NovaSeq 6000 platform, generating 100bp single-end reads. Library preparation and quantitation were performed in the Slotman lab at Texas A&M, while quality control analysis, and sequencing were performed at the Texas A&M AgriLife Genomics Core in College Station, TX, USA.

2.2.3. RNA Sequencing Analyses

For each replicate library, between 80 and 120M total reads were generated. The analysis pipeline was adapted from the Genome Analysis Toolkit (GATK) Best Practices for RNA-Seq[211], incorporating some steps specific to library prep with Lexogen UMIs. Following an initial quality check in FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), the UMIs were extracted from the reads into a read index using the Lexogen tool umi2index. New FastQC reports were generated, then libraries were trimmed of adapter sequence, reads with a phred score below 20, and reads shorter than 20 bp using TrimGalore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/).

The trimmed and filtered files were then mapped to the *An. gambiae* genome (PEST v4.12) using STAR[212]. To achieve better mapping quality, a two-pass alignment in STAR was used. Following STAR, reads containing UMIs were collapsed to a single read using the Lexogen tool collapse_UMI_bam. Picard Tools[213] was used to add read groups, sort by read group, and index reads. The two technical replicates of each library were then merged using samtools[214]. The featureCounts tool in the Subread package[215]

was used to generate read counts mapping to each annotated gene, which were then used for differential expression analysis as well as to calculate counts per million (CPM) to estimate expression levels.

2.2.4. Analysis of Differential Gene Expression

Differential expression analysis was performed using the R package edgeR[216], which implements exact tests on count data. Size factors were calculated for each dataset to normalize library size across replicates, while overall means and variances were calculated from a negative binomial distribution model. Genes were significantly differentially expressed if q was < 0.05 (after correcting for multiple testing). Fisher's exact test was used to test significance for each pairwise comparison, and expression was considered significantly different if the false discovery rate (FDR) adjusted p-value (q) was < 0.05. I only consider genes detected at >1 CPM to be expressed in the description and discussion of my results.

CPM values were used to create plots of the relative expression of genes between sexes and species, with significance values from edgeR used to mark genes that exhibit significant differential expression between categories.

2.2.5. Gene Ontology Analysis

All genes that were detected as being significantly differentially expressed between sexes and species according to the adjusted p-value generated by edgeR were used for gene ontology analysis. Gene ontology annotation was performed using Panther (www.pantherdb.org), an online database of gene functions and orthologs. The annotation was performed to identify the molecular function of significantly

differentially expressed genes. Gene IDs of differentially expressed genes were uploaded for each pairwise comparison. Molecular functions of these genes were determined, as were the protein classes, and a Fisher's exact test with an FDR adjustment was performed to check for overrepresentation of chemosensory genes.

2.3. Results

Table 2.1 – Reads obtained from each biological replicate library. CFL = *An. coluzzii* female, QFL = *An. quadriannulatus* female, CML = *An. coluzzii* male, QML = *An. quadriannulatus* male

Library	Reads
CFL1	1,267,090
CFL2	3,082,918
CFL3	2,759,840
CFL4	5,123,204
CFL5	3,548,461
CFL6	4,118,728
QFL1	1,130,395
QFL3	3,457,763
QFL4	4,077,045
QFL5	3,149,548
QFL6	3,321,904
CML1	1,839,759
CML2	2,594,233
CML3	3,572,372
CML4	1,822,594
CML5	3,199,814
CML6	3,060,041
QML1	1,502,331
QML2	2,379,249
QML3	2,778,897
QML5	2,447,419
QML6	2,879,503

Following the processing steps outlined above, between 1,130,395 and 5,123,204 total reads were obtained from each library (Table 2.1). The average overall library size was 2,868,778 reads. The average female *An. coluzzii* library size was 3,316,707 reads,

while the average female *An. quadriannulatus* library size was 3,027,331 reads, the average male *An. coluzzii* library size was 2,681,469 reads, and the average male *An. quadriannulatus* library size was 2,397,480 reads.

2.3.1. Differential Gene Expression between Female *An. coluzzii* and *An. quadriannulatus*

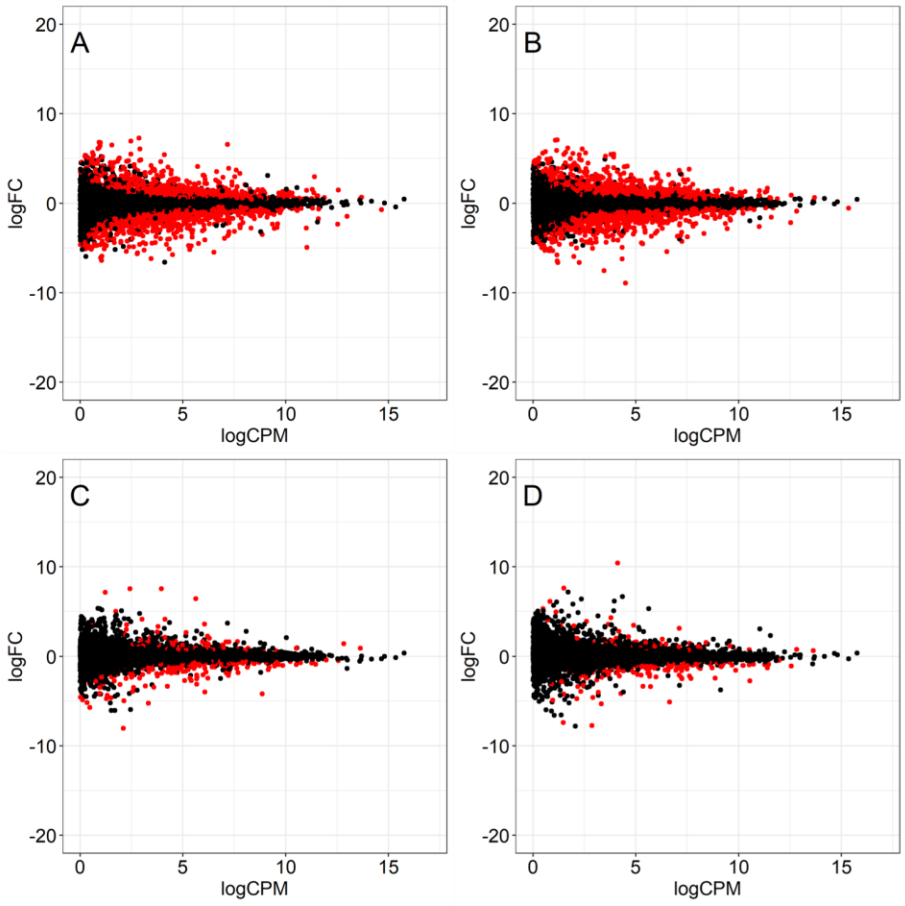

Figure 2.1 – Overall Differential Gene Expression Profiles of A) Female *An. coluzzii* and Female *An. quadriannulatus*, B) Male *An. coluzzii* and Male *An. quadriannulatus*, C) Male and Female *An. coluzzii*, and D) Male and Female *An. quadriannulatus*. Genes shown in black are non-significant, whereas genes shown in red are significant at an FDR-adjusted p-value < 0.05. CPM = counts per million and FC = fold change.

While the overall gene expression profile is highly correlated ($R^2 = 0.881$), a total of 2,138 genes (21.4%) are significantly differentially expressed (DE) between the two species (Figure 2.1A). Of these DE genes, 1,131 are *An. coluzzii*-biased and 1,007 are *An. quadriannulatus*-biased. Seven chemoreceptor genes and seven *Obps* are significantly DE between the females of these two species (Table 2.2).

Table 2.2 – Differentially expressed chemosensory genes in the labella of female *An. coluzzii* and *An. quadriannulatus*. Genes are only included if detected ≥ 1 CPM (receptors) or ≥ 100 CPM (*Obps*). Fold changes were calculated by edgeR and are slightly different than those calculated by hand. They are indicated as positive if *An. coluzzii*-biased and negative if *An. quadriannulatus*-biased.

Gene ID	Gene	*An. coluzzii* (CPM with Standard Error)	*An. quadriannulatus* (CPM with Standard Error)	FC	log2FC	q
AGAP001169	*Gr49*	2.22 ± 0.45	0.41 ± 0.34	8.42	3.07	0.005
AGAP006876	*Gr31*	1.92 ± 0.78	0.16 ± 0.50	8.38	3.07	0.031
AGAP013416	*Ir7w*	4.66 ± 0.79	0.41 ± 0.28	9.06	1.82	<0.001
AGAP002904	*Ir41a*	23.12 ± 1.80	11.28 ± 1.91	2.02	1.02	0.006
AGAP013520	*Ir7x*	1.41 ± 0.36	7.48 ± 0.88	-5.62	-2.49	<0.001
AGAP029499	*Or21*	4.81 ± 0.64	1.24 ± 0.21	3.87	1.95	0.003
AGAP009111	*Or29*	3.20 ± 0.75	7.55 ± 1.71	-2.19	-1.13	0.058
AGAP012321	*Obp26*	1,472.92 ± 484.58	129.10 ± 11.10	11.20	3.48	<0.001
AGAP006080	*Obp54*	1,380.59 ± 178.42	266.44 ± 25.75	4.98	2.32	<0.001
AGAP006077	*Obp51*	394.57 ± 30.45	186.58 ± 19.42	2.06	1.04	<0.001
AGAP012320	*Obp25*	571.57 ± 36.36	382.91 ± 32.62	1.45	0.53	0.042
AGAP006076	*Obp50*	338.53 ± 36.66	646.60 ± 57.78	-1.99	-0.99	<0.001
AGAP002905	*Obp48*	1,603.75 ± 144.93	3,445.31 ± 230.34	-2.22	-1.15	<0.001
AGAP011368	*Obp57*	4,436.63 ± 419.61	11,705.75 ± 656.31	-2.79	-1.48	<0.001

2.3.1.1. Gene Ontology Analysis

The most common molecular functions of DE genes in either species were catalytic activity, binding, and transporter activity (Figure 2.2). The most common protein classes of DE genes were metabolite interconversion enzyme (PC00264), protein modifying enzyme (PC00260), and transporter (PC00227). Olfactory receptor activity was overrepresented (Fisher's exact test: FDR = 0.037).

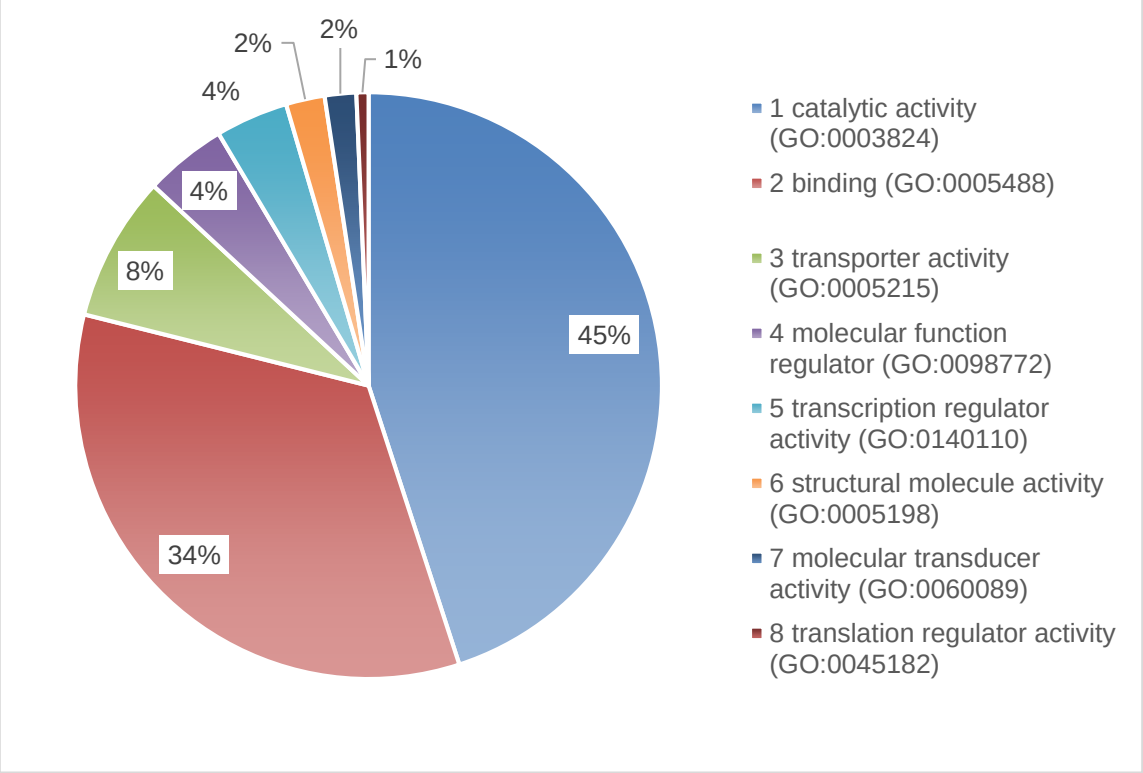

Figure 2.2 – Molecular Function of DE Genes between Female *An. coluzzii* and *An. quadriannulatus*. 749 genes with known functions were identified by PantherDB; percentages refer to these genes. Molecular functions are presented clockwise from right in descending order.

2.3.1.2. Gustatory Receptors (*Grs*)

Of the 61 known *Grs* in *An. gambiae*, eleven are detected >1 CPM in *An. coluzzii* and thirteen are detected >1 CPM in *An. quadriannulatus*. The average overall expression of detected *Grs* in female *An. coluzzii* is 56.74 CPM, while in female *An. quadriannulatus* it is 50.09 CPM. The highest expressed *Gr* in both species is the sugar receptor *Gr17*, which is expressed at 16.55 CPM in *An. coluzzii* and 10.90 CPM in *An. quadriannulatus*. *Gr33* is also relatively highly expressed in both species of female, at 6.68 CPM in *An. coluzzii* and 5.24 CPM in *An. quadriannulatus* (Figure 2.3A). Twelve other *Grs* are expressed > 1 CPM in either species: *Gr8*, *Gr14*, *Gr21*, *Gr29*, *Gr30*, *Gr31*, *Gr43*, *Gr44*, *Gr49*, *Gr51*, *Gr55*, and *Gr62*. Among these genes are two other sugar receptors, *Gr14* and *Gr21*, while the ligands for the other ten *Grs* are unknown.

Overall *Gr* expression between the two species of female is correlated (R^2 = 0.720). However, two genes expressed >1 CPM are significantly enhanced in *An. coluzzii*: *Gr49* and *Gr31* (Table 2.2). *Gr49* is expressed at 2.22 CPM in *An. coluzzii* and 0.41 CPM in *An. quadriannulatus*, an 8.42-fold enhancement (fold changes were generated by edgeR and are slightly different than those calculated by hand). Similarly, *Gr31* is 8.38-fold enhanced from 0.16 CPM in *An. quadriannulatus* to 1.92 CPM in *An. coluzzii*.

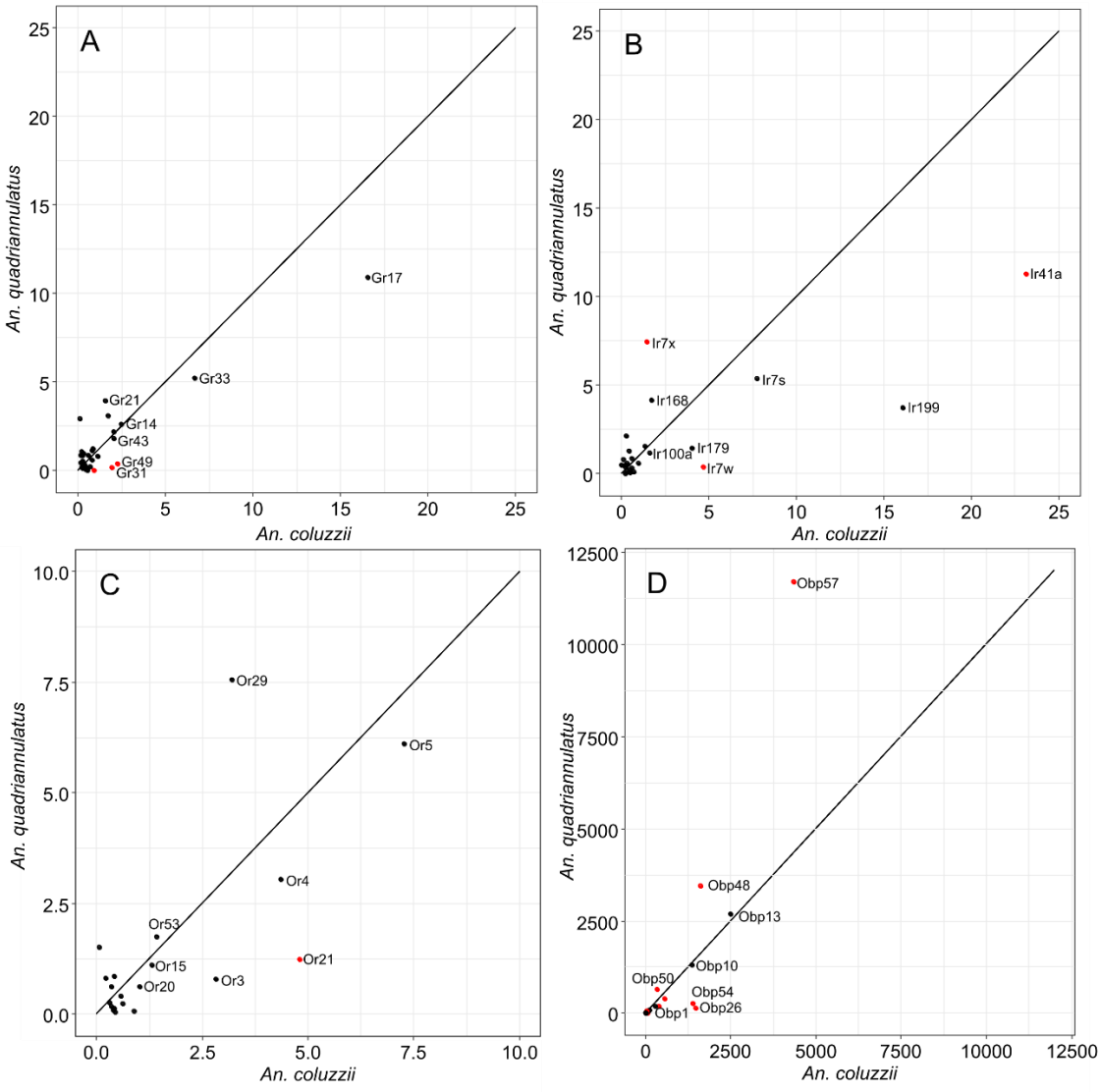

Figure 2.3 – Expression (CPM) of Major Chemosensory Gene Families in Female *An. coluzzii* and *An. quadriannulatus*: A) Gustatory Receptors, B) Ionotropic Receptors, C) Olfactory Receptors, and D) Odorant Binding Proteins. The diagonal line represents equal expression in both species. Genes indicated in red have an FDR-adjusted p-value < 0.05.

2.3.1.3. Ionotropic Receptors (*Irs*)

Of the 65 known *Irs* in *An. gambiae*, eleven are detected >1 CPM in *An. coluzzii* and twelve are detected >1 CPM in *An. quadriannulatus*. The average overall expression of *Irs* in both females is substantially higher than *Grs*, 300.78 CPM in *An. coluzzii* and 276.54 CPM in *An. quadriannulatus*. However, most of that expression is due to the co-receptors *Ir25a* (65.55 CPM in *An. coluzzii* and 59.53 CPM in *An. quadriannulatus* and *Ir76b* (164.84 CPM in *An. coluzzii* and 170.15 CPM in *An. quadriannulatus*). Interestingly, the third co-receptor, *Ir8a* is not detected. With the co-receptors removed, total average *Ir* expression is similar to total average *Gr* expression: 69.92 CPM in *An. coluzzii* and 46.62 CPM in *An. quadriannulatus*. Excluding the co-receptors, the most highly expressed *Ir* in both species is *Ir41a* (Figure 2.3B), with a 2.02-fold DE in *An. coluzzii* (23.12 CPM in *An. coluzzii* versus 11.28 CPM in *An. quadriannulatus*, q=0.006). In addition to *Ir41a*, several other *Irs* were detected > 5 CPM; *Ir199, Ir7s* and *Ir7x*. Seven others are expressed < 5 CPM: *Ir7n*, *Ir7w*, *Ir64a*, *Ir93a*, *Ir100a*, *Ir168,* and *Ir179*.

Ir expression in the labella is highly correlated between female *An. coluzzii* and *An. quadriannulatus* (R^2=0.987). In addition to *Ir41a*, only two other DE genes in this family were observed: *Ir7w* and *Ir7x*. *Ir7w* is nearly *An. coluzzii*-specific, with a 9.06-fold up-regulation from 0.41 CPM in *An. quadriannulatus* to 4.66 CPM in *An. coluzzii*. By contrast, *Ir7x* is strongly *An. quadriannulatus*-biased, showing 5.62-fold enhancement from 1.41 CPM in *An. coluzzii* to 7.48 CPM in *An. quadriannulatus* (Table 2.2).

2.3.1.4. Olfactory Receptors (*Ors*)

Of the 76 known *Ors* in *An. gambiae*, nine are detected >1 CPM in *An. coluzzii* and nine are detected >1 CPM in *An. quadriannulatus*. As with the *Irs*, average overall *Or* expression exceeds that of *Grs*: 125.87 CPM in *An. coluzzii* and 112.35 CPM in *An. quadriannulatus*. However, similarly to *Ir25a* and *Ir76b*, *Orco* is primarily responsible for this high representation, as it is expressed at 90.65 CPM in *An. coluzzii* and 80.61 CPM in *An. quadriannulatus*. When *Orco* is excluded, the total expression of *Ors* is less than that of either the *Grs* or *Irs*, at 35.22 CPM in *An. coluzzii* and 31.73 CPM in *An. quadriannulatus*. The two highest expressed *Ors* are *Or5* and *Or29* (Figure 2.3C). *Or29* appears to be *An. quadriannulatus*-biased (2.19-fold enhanced from 3.20 CPM in *An. coluzzii* to 7.55 CPM) but is only borderline significant (q = 0.058).

Seven other *Ors* were detected at lower levels: *Or3*, *Or4*, *Or15*, *Or20*, *Or21*, *Or40*, and *Or53*. As in the *Irs*, expression between species is highly correlated (R^2 = 0.993), with two genes showing DE; *Or29* (see above) and *Or21*, which is 3.87-fold enhanced from 1.24 CPM in *An. quadriannulatus* to 4.81 CPM in *An. coluzzii* (Table 2.2).

2.3.1.5. Odorant Binding Proteins (*Obps*)

As expected, the *Obps* are expressed at very high levels: an average of 14,891.77 CPM in *An. coluzzii* and 21,461.99 CPM in *An. quadriannulatus*. Eleven *Obps* are expressed > 100 CPM. While *Obp* expression is correlated between the two species (R^2 = 0.798), this correlation is lower than in the other gene families, and a greater proportion of *Obps* are DE. *Obp57* is expressed at a level far surpassing that of the other

Obps: 4,346.63 CPM in *An. coluzzii* and 11,705.75 CPM in *An. quadriannulatus* (the 19th and 9th most expressed gene overall, respectively), which represents a 2.79-fold enrichment in *An. quadriannulatus* (Figure 2.3D, Table 2.2).

In addition to *Obp57*, five other *Obps* are highly expressed (> 1,000 CPM): *Obp10*, *Obp13*, *Obp26*, *Obp48*, and *Obp54*. Of these genes, two are expressed at similar levels in the two species (*Obp10* and *Obp13*). *Obp26* shows the most striking DE pattern of any chemosensory gene in the labella, exhibiting a 11.20-fold up-regulation from 129.10 CPM in *An. quadriannulatus* to 1,472.92 CPM in *An. coluzzii*. *Obp54* is 4.98-fold enhanced in *An. coluzzii*, to 1,380.59 CPM from 266.44 CPM in *An. quadriannulatus*. *Obp48*, by contrast, is 2.2-fold enhanced in *An. quadriannulatus* to 3,445.31 CPM from 1,603.75 CPM in *An. coluzzii*.

In addition, five other *Obps* (*Obp1*, *Obp7*, *Obp25*, *Obp50*, and *Obp51*) are expressed at lower levels (> 100 CPM). *Obp25* shows a 1.45-fold enhancement in *An. coluzzii*. While *Obp50* is 1.99-fold up-regulated in *An. quadriannulatus*, *Obp51* is 2.06-fold up-regulated in *An. coluzzii*.

In addition to these highly expressed *Obps*, five lowly-expressed *Obps* (<100 CPM) show *An. coluzzii* bias and one shows *An. quadriannulatus* bias. Of these lowly-expressed genes, *Obp23* is the most striking, as it is expressed at 55.52 CPM in *An. coluzzii* but nearly absent from *An. quadriannulatus* at 1.99 CPM. The other lowly-expressed *Obps* with *An. coluzzii* bias are *Obp49*, *Obp56*, *Obp69*, and *Obp71*. *Obp6* is *An. quadriannulatus*-biased.

2.3.2. Differential Gene Expression between Male *An. coluzzii* and *An. quadriannulatus*

The overall gene expression profile is largely correlated, somewhat more so than in females ($R^2 = 0.921$), although there are 2,085 genes showing significant differential expression between the two species. Of these DE genes, 1,092 are *An. coluzzii*-biased and 993 are *An. quadriannulatus*-biased (Figure 2.1B). In males, nine chemosensory receptor genes and six *Obps* are significantly DE.

2.3.2.1. Gene Ontology Analysis

The most common molecular functions of DE genes in either species were catalytic activity, binding, and transporter activity (Figure 2.4). The most common protein classes of DE genes were metabolite interconversion enzyme (PC00264), protein modifying enzyme (PC00260), and transporter (PC00227). Olfactory receptor activity was not overrepresented.

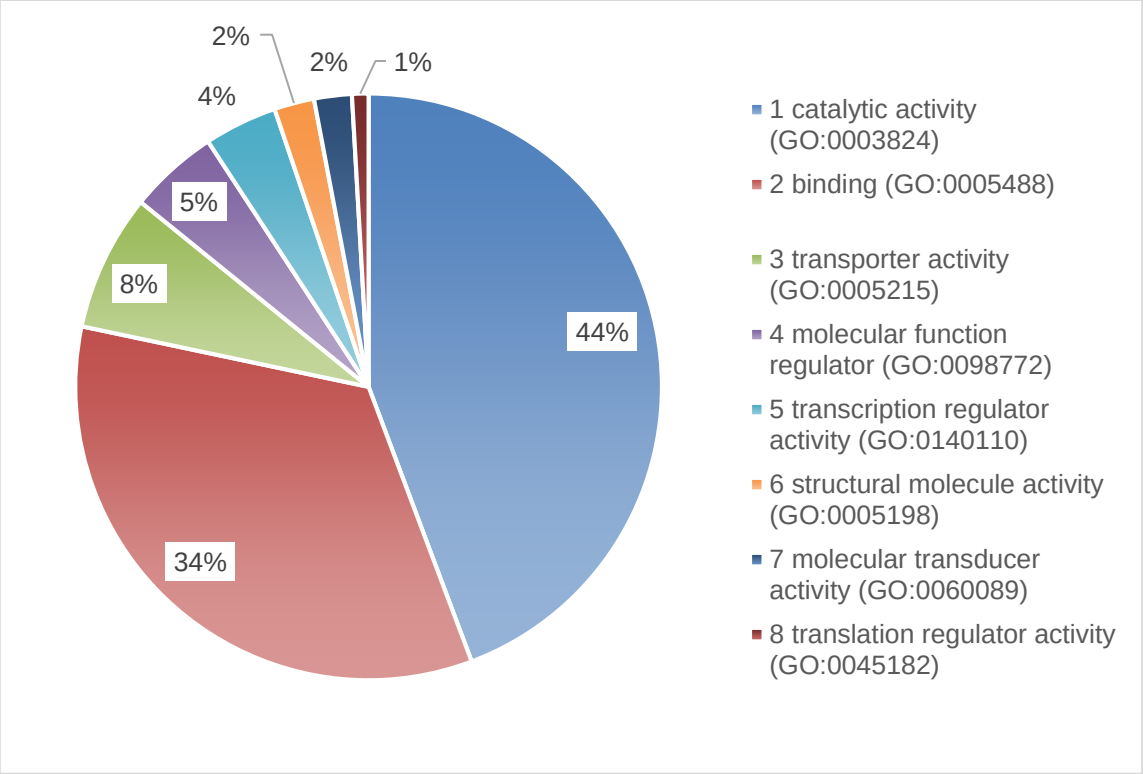

Figure 2.4 – Molecular Function of DE Genes between Male *An. coluzzii* and *An. quadriannulatus*. 770 genes with known functions were identified by PantherDB; percentages refer to these genes. Molecular functions are presented clockwise from right in descending order.

Table 2.3 – Differentially expressed chemosensory genes in the labella of male *An. coluzzii* and *An. quadriannulatus*. Genes are only included if detected ≥ 1 CPM (receptors) or ≥ 100 CPM (Obps). Fold changes were calculated by edgeR and are slightly different than those calculated by hand. Fold changes are indicated as positive if *An. coluzzii*-biased and negative if *An. quadriannulatus*-biased.

Gene ID	Gene	*An. coluzzii* (CPM with Standard Error)	*An. quadriannulatus* (CPM with Standard Error)	FC	log2FC	q
AGAP001169	*Gr49*	2.48 ± 0.29	0.28 ± 0.28	6.61	2.73	0.003
AGAP003260	*Gr21*	1.15 ± 0.37	6.31 ± 1.81	-5.25	-2.39	<0.001
AGAP006877	*Gr32*	0.29 ± 0.19	1.79 ± 0.61	-6.16	-2.62	0.045
AGAP001170	*Gr48*	0.40 ± 0.21	4.70 ± 3.84	-11.8	-3.56	0.031
AGAP009857	*Gr4*	0.60 ± 0.37	9.99 ± 9.46	-17.4	-4.11	0.012
AGAP013416	*Ir7w*	4.03 ± 1.07	1.28 ± 0.23	3.15	1.65	0.031
AGAP013520	*Ir7x*	0.43 ± 0.16	4.29 ± 0.90	-8.52	-3.09	<0.001
AGAP029232	*Ir179*	0.19 ± 0.19	15.01 ± 9.35	-66.3	-6.05	0.011
AGAP029499	*Or21*	4.52 ± 0.84	1.73 ± 0.80	2.93	1.55	0.038
AGAP006080	*Obp54*	1,057.87 ± 53.67	276.81 ± 43.11	3.79	1.92	<0.001
AGAP006077	*Obp51*	274.38 ± 8.52	172.63 ± 23.29	1.55	0.63	0.013
AGAP001189	*Obp10*	1,698.48 ± 124.86	1,176.92 ± 88.20	1.41	0.49	0.048
AGAP006076	*Obp50*	338.89 ± 28.56	497.53 ± 80.08	-1.51	-0.60	0.046
AGAP011368	*Obp57*	3,134.65 ± 104.67	5,679.23 ± 341.67	-1.85	-0.88	<0.001
AGAP007286	*Obp48*	1,693.07 ± 137.31	4,281.55 ± 397.82	-2.55	-1.35	<0.001

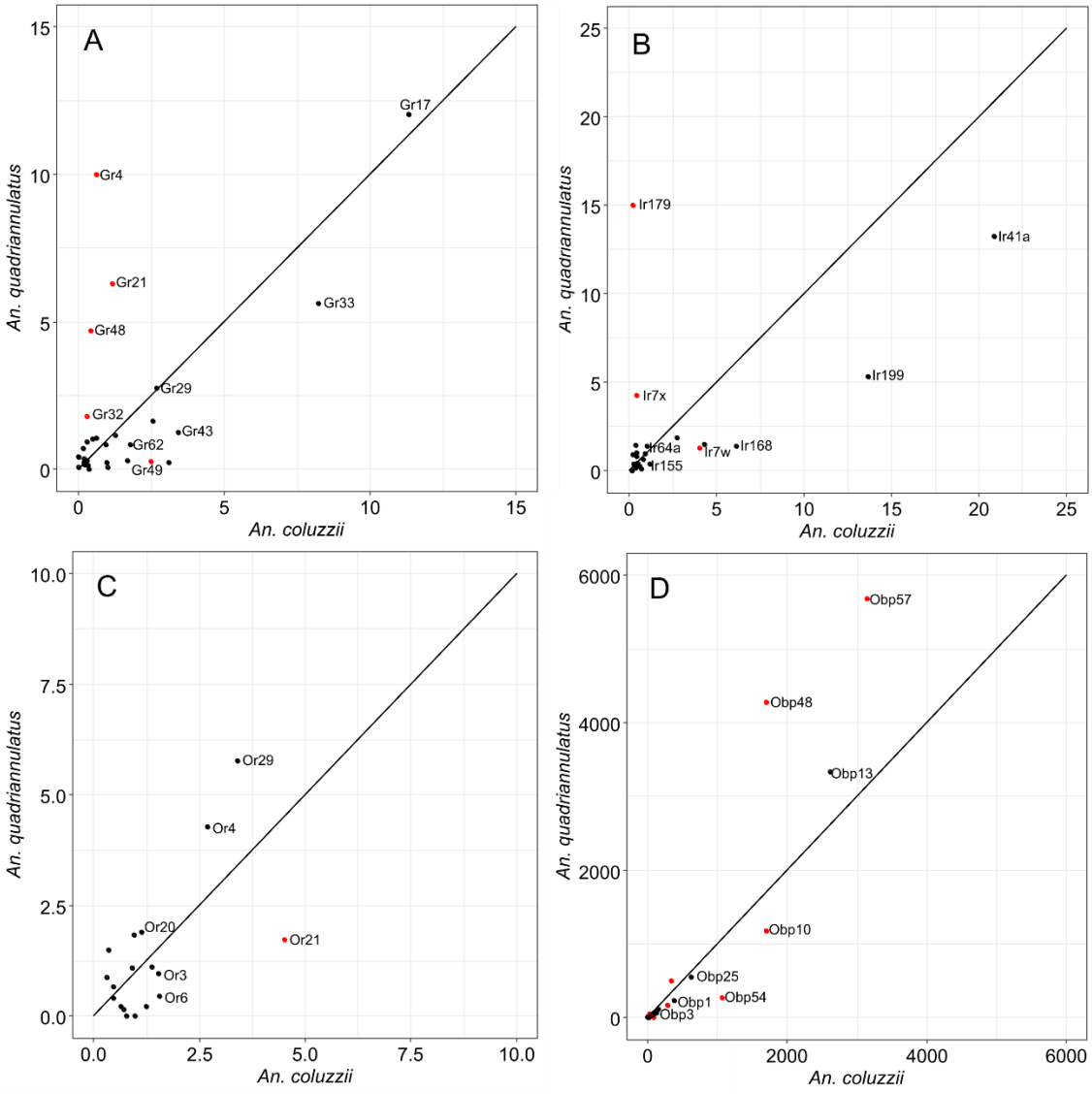

Figure 2.5 – Expression (CPM) of Major Chemosensory Gene Families in Male *An. coluzzii* and *An. quadriannulatus*: A) Gustatory Receptors, B) Ionotropic Receptors, C) Olfactory Receptors, and D) Odorant Binding Proteins. The diagonal line represents equal expression in both species. Genes indicated in red have an FDR-adjusted p-value < 0.05.

2.3.2.2. Gustatory Receptors (*Grs*)

Average overall expression of *Grs* is comparable to that of females: in *An. coluzzii* males, total average *Gr* expression is 49.56 CPM, while it is 59.87 CPM in *An. quadriannulatus* males. Unlike in females, however, *Gr* expression is lowly correlated ($R^2 = 0.472$). In males, sixteen *Grs* are expressed ≥1 CPM (Figure 2.5A). As in females, the most highly expressed *Gr* in both species is the sugar receptor, *Gr17*. Following *Gr17*, *Gr4* is the most highly expressed gene in *An. quadriannulatus* (9.99 CPM), while it is nearly absent from male *An. coluzzii* (Table 2.3). Following *Gr17* and *Gr4*, the most highly expressed gene is *Gr33*, which expressed at similar levels in males of both species. The sugar receptor *Gr21* is also among the more highly expressed *Grs*, with a 5.25-fold DE in *An. quadriannulatus* males, where it is expressed at 6.31 CPM.

Twelve *Grs* were detected at lower levels: *Gr8*, *Gr10*, *Gr14*, *Gr29*, *Gr31*, *Gr32*, *Gr43*, *Gr44*, *Gr48*, *Gr49*, *Gr51*, and *Gr62*. Three of these genes are DE; as in females, *Gr49* is highly *An. coluzzii*-biased. In addition to *Gr4* and *Gr21* (see above), *Gr32* and *Gr48* are also strongly *An. quadriannulatus*-biased. *Gr48* is expressed at 4.70 CPM in *An. quadriannulatus* and <1 CPM in *An. coluzzii*, while *Gr32* is expressed at 1.79 CPM in *An. quadriannulatus* and <1 CPM in *An. coluzzii*. Neither *Gr48* nor *Gr32* have known functions or clear *Drosophila* orthologs.

2.3.2.3. Ionotropic Receptors (*Irs*)

As with the *Grs*, the average total expression level of the *Irs* is similar to that of females, 262.60 CPM in *An. coluzzii* and 309.72 CPM in *An. quadriannulatus*. Unlike the *Grs*, expression of *Irs* is highly correlated, as in females ($R^2 = 0.979$). With the co-

receptors removed, total *Ir* expression is 65.88 CPM and 58.06 CPM, respectively. Interestingly, *Ir25a* is marginally significantly (q = 0.050) *An. quadriannulatus*-biased (1.47-fold, Figure 2.5B). Twelve specific *Irs* were detected in one or both species. As in females, *Ir41a* is the most highly expressed specific *Ir*, although in males its expression is similar in the two species. *Ir179* is the second-most highly expressed specific *Ir* in *An. quadriannulatus* males (15.01 CPM) but is not detected in *An. coluzzii* (Table 2.3). *Ir199* is also expressed at relatively high but similar level in the two species.

In addition to *Ir179*, two other *Irs* are significantly DE between the males. As in females, *Ir7w* is *An. coluzzii*-biased (3.15-fold), and *Ir7x* is nearly exclusive to *An. quadriannulatus* (4.29 CPM in male *An. quadriannulatus* but <0.5 CPM in *An. coluzzii*). The remaining six *Irs* detected are *Ir7n*, *Ir64a*, *Ir100a*, *Ir155*, *Ir168*, and *Ir176*.

2.3.2.4. Olfactory Receptors (*Ors*)

Or expression is highly correlated in males ($R^2 = 0.975$). As expected, *Orco* is the most highly expressed *Or* in both species. With *Orco* included, average total *Or* expression is 157.47 CPM in male *An. coluzzii* and 154.44 CPM in male *An. quadriannulatus*. When *Orco* is excluded, total *Or* expression drops to 53.80 CPM in *An. coluzzii* and 40.46 CPM in *An. quadriannulatus*. The most highly expressed specific *Or* in *An. coluzzii* is *Or40*, at 16.43 CPM, followed by *Or5*, which is expressed at 9.23 CPM in *An. coluzzii* and 11.06 CPM in *An. quadriannulatus* (Figure 2.5C). The third-most highly expressed *Or*, *Or21*, is the only DE *Or* between the two males (Table 2.3). As in females, *Or21* is 2.93-fold enhanced in male *An. coluzzii*. Ten additional *Ors* are detected: *Or3*, *Or4*, *Or6*, *Or15*, *Or16*, *Or20*, *Or27*, *Or29*, *Or50*, and *Or53*.

2.3.2.5. Odorant Binding Proteins (*Obps*)

Obp expression in males is more correlated than in females ($R^2 = 0.862$). As in females, *Obp57* is by far the most highly expressed *Obp* in both males: in *An. coluzzii*, it is expressed at 3,134.65 CPM, while it is expressed at 5,679.23 CPM in *An. quadriannulatus*, a 1.85-fold enhancement (Figure 2.5D). Four other *Obps* are also expressed >1,000 CPM, all of which are also very highly expressed in females: *Obp10*, *Obp13*, *Obp48*, and *Obp54*. While *Obp13* is again not DE, *Obp10* is 1.41-fold enhanced in male *An. coluzzii* from 1,176.92 CPM to 1,698.48 CPM (Table 2.3). As in females, *Obp48* is 2.55-fold enhanced in *An. quadriannulatus* and *Obp54* is 3.79-fold enhanced in *An. coluzzii*.

Six other *Ops* are relatively highly expressed (>100 CPM): *Obp1*, *Obp3*, *Obp7*, *Obp25*, *Obp50*, and *Obp51*. As in females, *Obp50* is 1.51-fold upregulated in *An. quadriannulatus* and *Obp51* is 1.55-fold upregulated in *An. coluzzii*. The other highly expressed *Obps* are not DE. In addition to the six *Obps* listed above, six more lowly expressed *Obps* (<100 CPM) are DE. As in females, *Obp26* is *An. coluzzii*-biased in males, although it is expressed at a low level of 81.47 CPM in male *An. coluzzii* and 24.49 CPM in male *An. quadriannulatus*. Again mirroring female differential expression, *Obp23* is nearly exclusive to *An. coluzzii* (80.48 CPM versus 3.46 CPM). *Obp6*, *Obp19*, *Obp38*, and *Obp46* all show *An. quadriannulatus* bias, but are expressed at very low (< 20 CPM) levels, excluding *Obp6*, which is 2.17-fold enhanced from 26.85 CPM in *An. coluzzii* to 56.75 CPM.

2.3.3. Sex-Biased Gene Expression in *An. coluzzii*

Overall gene expression between the two sexes of *An. coluzzii* is very highly correlated ($R^2 = 0.941$), with only 295 significantly differentially expressed genes (Figure 2.1C). 206 genes are female-biased, compared to 89 male-biased genes.

2.3.3.1. Gene Ontology Analysis

The most common molecular functions of DE genes in either sex were catalytic activity, binding, and transcription regulator activity (Figure 2.6). The most common protein classes of DE genes were metabolite interconversion enzyme (PC00264), protein modifying enzyme (PC00260), and nucleic acid binding protein (PC00171). Olfactory receptor activity was not overrepresented.

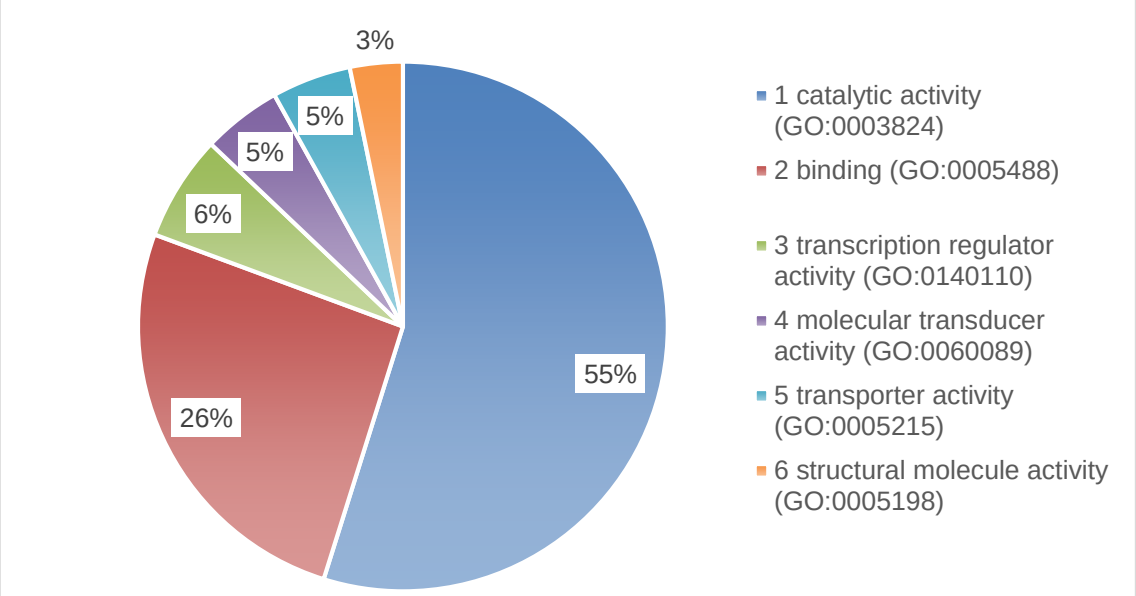

Figure 2.6 – Molecular Function of DE Genes between Female and Male *An. coluzzii*. 62 genes with known functions were identified by PantherDB; percentages refer to these genes. Molecular functions are presented clockwise from right in descending order.

2.3.3.2. Chemosensory Gene Expression

Total average expression of chemosensory genes in females is 15,375.16 CPM, while in males it is 12,697.77 (Table 2.4). A chi-square test of independence indicated that this overall female bias was significant (X^2 (4,2) = 10.36, p (3 df) = 0.016).

Table 2.4 – Total average expression in CPM of chemosensory genes by family in both sexes of *An. coluzzii*.

Gene Family	Female (CPM with Standard Error)	Male (CPM with Standard Error)
Grs	56.74 ± 6.81	49.56 ± 6.68
Irs	300.78 ± 16.40	262.60 ± 15.69
Ors	125.87 ± 9.52	157.47 ± 14.69
Obps	14,891.77 ± 1,112.17	12,698.14 ± 211.48
Total	15,375.16 ± 1,134.88	13,167.77 ± 212.03

While one *Or* and two *Obps* are sex-biased (Table 2.5), the vast majority of chemosensory genes are not DE. *Gr* expression between the sexes is correlated (R^2 = 0.753, Figure 2.7A), as is *Ir* expression (R^2 = 0.997) (Figure 2.7B). All *Ors* are expressed at similar levels between the sexes (R^2 = 0.974), apart from *Or40*, which is male-specific (Figure 2.7C). Both DE *Obps* are female-biased, most notably *Obp26*, which is 18.57-fold enhanced in females (Figure 2.7D). The most highly-expressed *Obp* in both sexes, *Obp57*, is 1.41-fold enhanced in females. Despite these female-biased *Obps*, overall *Obp* expression is highly correlated between sexes (R^2 = 0.895).

Table 2.5 – Differentially expressed chemosensory genes in the labella of *An. coluzzii* females and males. Genes are only included if detected ≥ 1 CPM (receptors) or ≥ 100 CPM (Obps). Fold changes were calculated by edgeR and are slightly different than those calculated by hand. Fold changes are indicated as positive if female-biased and negative if male-biased.

Gene ID	Gene	Female (CPM with Standard Error)	Male (CPM with Standard Error)	FC	log2FC	q
AGAP002558	*Or40*	0.05 ± 0.05	16.43 ± 13.66	-186.16	-7.54	0.037
AGAP012321	*Obp26*	1,472.92 ± 484.58	81.47 ± 3.43	18.57	4.22	<0.001
AGAP011368	*Obp57*	4,346.63 ± 419.61	3,134.65 ± 104.67	1.41	0.50	0.048

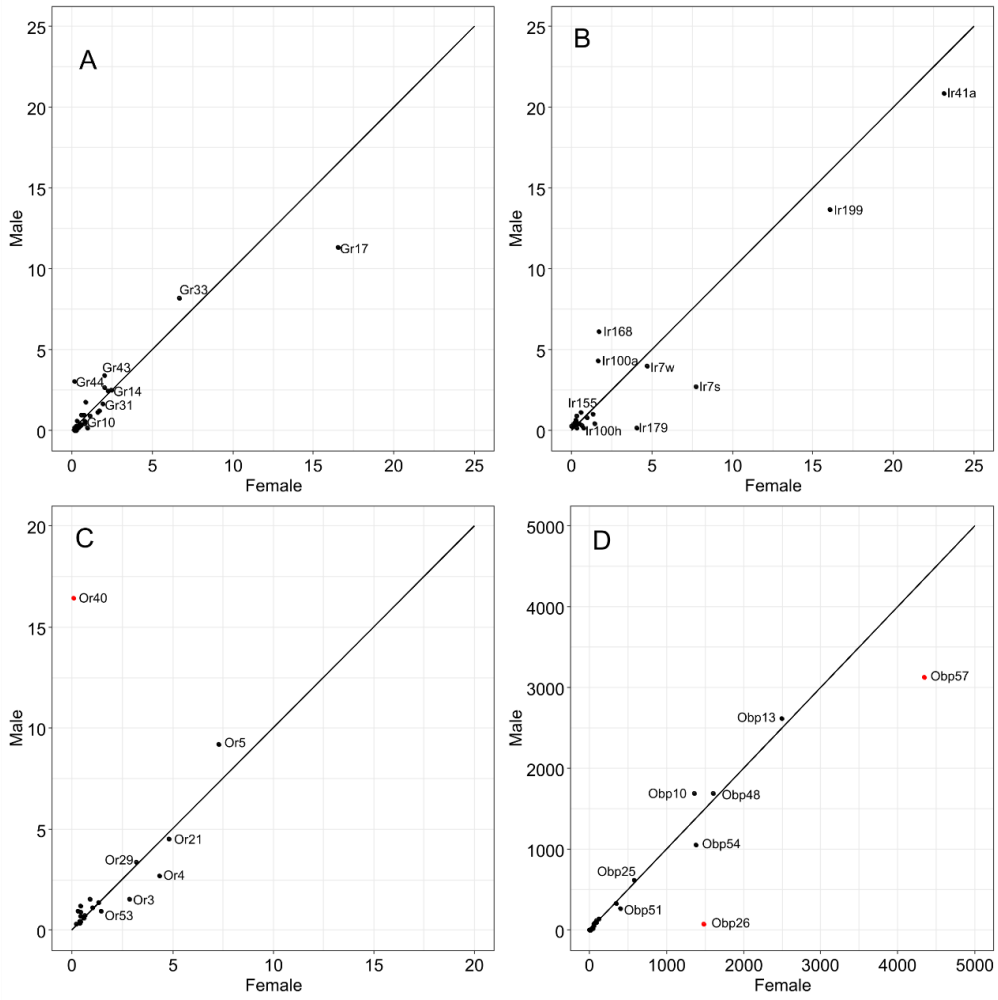

Figure 2.7 – Expression (CPM) of Major Chemosensory Gene Families in Female and Male *An. coluzzii*: A) Gustatory Receptors, B) Ionotropic Receptors, C) Olfactory Receptors, and D) Odorant Binding Proteins. The diagonal line represents equal expression in both sexes. Genes indicated in red have an FDR-adjusted p-value < 0.05.

2.3.4. Sex-Biased Gene Expression in *An. quadriannulatus*

Similarly to *An. coluzzii*, overall gene expression between the two sexes of *An. quadriannulatus* is very highly correlated ($R^2 = 0.949$), with only 273 significantly differentially expressed genes (Figure 2.1D). 186 genes are female-biased, compared to 87 male-biased genes.

2.3.4.1. Gene Ontology Analysis

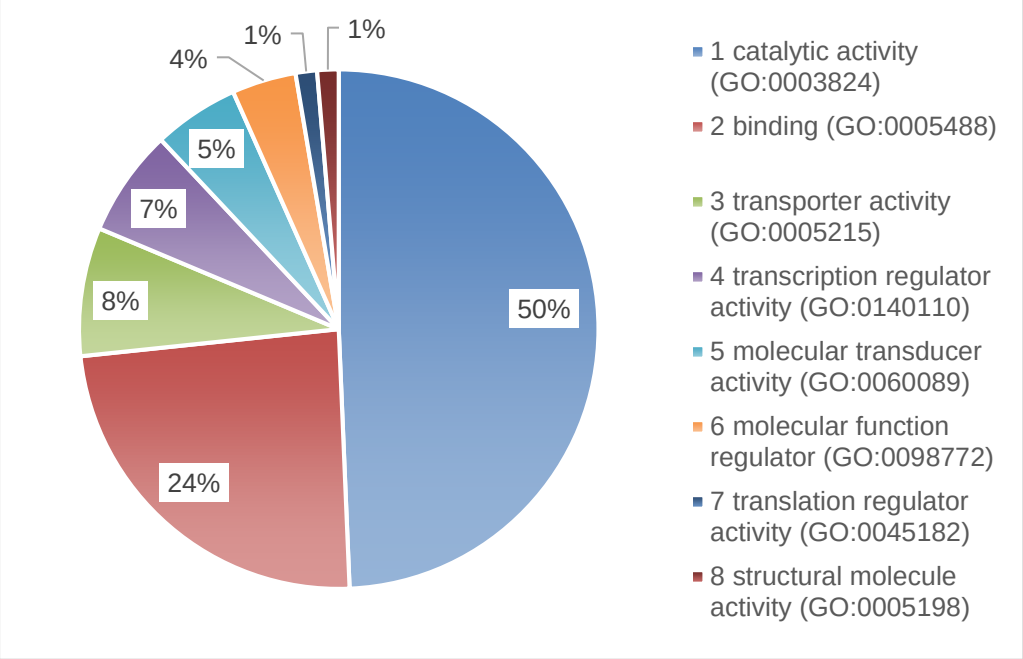

Figure 2.8 – Molecular Function of DE Genes between Female and Male *An. quadriannulatus*. 75 genes with known functions were identified by PantherDB; percentages refer to these genes. Molecular functions are presented clockwise from right in descending order.

The most common molecular functions of DE genes in either sex were catalytic activity, binding, and transporter activity. The most common protein classes of DE genes were metabolite interconversion enzyme (PC00264), protein modifying enzyme (PC00260), and extracellular matrix protein (PC00102). Olfactory receptor activity was not overrepresented.

2.3.4.2. Chemosensory Gene Expression

Average total expression of chemosensory genes in females is 21,900.97 CPM, while in males it is 17,358.74 (Table 2.6). A chi-square test of independence indicated that this overall female bias was significant (X^2 (4,2) = 43.55, p (3 df) < 0.001).

Table 2.6 – Total average expression in CPM of chemosensory genes by family in both sexes of *An. quadriannulatus*.

Gene Family	Female (CPM with Standard Error)	Male (CPM with Standard Error)
Grs	50.09 ± 3.39	59.87 ± 14.50
Irs	276.54 ± 17.83	309.72 ± 30.42
Ors	112.35 ± 6.91	154.44 ± 36.55
Obps	21,461.99 ± 1,158.83	16,834.71 ± 1,060.55
Total	21,900.97 ± 1,145.62	17,358.74 ± 1,022.72

Table 2.7 – Differentially expressed chemosensory genes in the labella of *An. quadriannulatus* females and males. Genes are only included if detected ≥ 1 CPM (receptors) or ≥ 100 CPM (Obps). Fold changes were calculated by edgeR and are slightly different than those calculated by hand. Fold changes are indicated as positive if female-biased and negative if male-biased.

Gene ID	Gene	Female (CPM with Standard Error)	Male (CPM with Standard Error)	FC	log2FC	q
AGAP012321	*Obp26*	129.10 ± 11.10	24.49 ± 3.55	5.40	2.43	<0.001
AGAP011368	*Obp57*	11,705.75 ± 656.31	5,679.23 ± 341.67	2.13	1.09	<0.001

Among the chemosensory genes, only two *Obps* are DE: *Obp57* and *Obp26.* Both are female-biased (Figure 2.9D, Table 2.7). *Obp57*, which is the most highly expressed *Obp* across both species and sexes, is 2.13-fold enhanced in female *An. quadriannulatus*. While *Obp26* is expressed at much lower levels than in female *An. coluzzii*, there is a 5.40-fold up-regulation in female *An. quadriannulatus*. Despite these *Obps*, *Obp* gene expression is generally correlated between the sexes ($R^2 = 0.838$). *Gr* expression between the sexes is lowly correlated ($R^2 = 0.565$), but no *Grs* are significantly ($q < 0.05$) DE (Figure 2.9A). *Ir* expression is very highly correlated ($R^2 = 0.988$, Figure 2.9B), as is *Or* expression ($R^2 = 0.997$, Figure 2.9C).

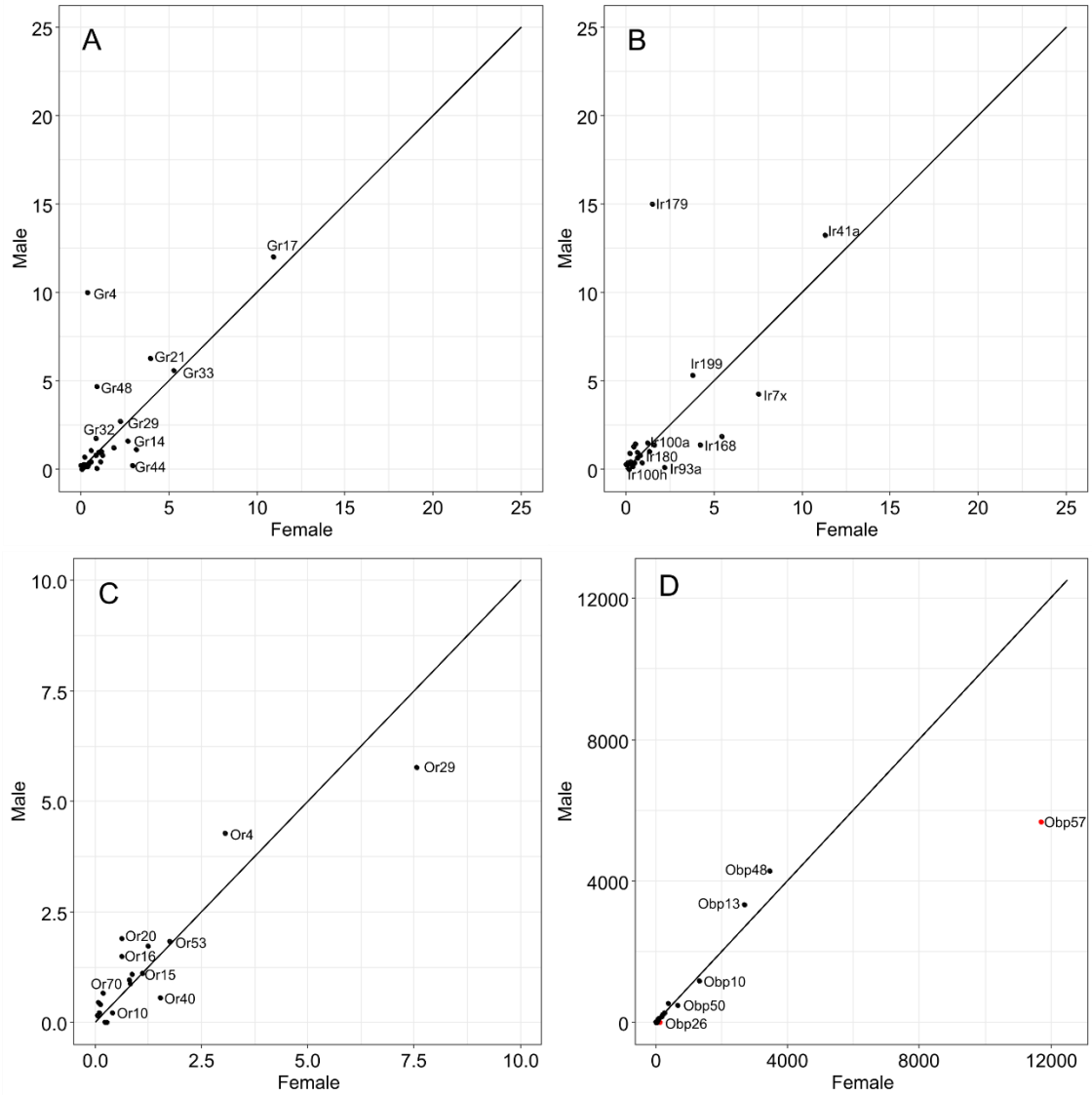

Figure 2.9 – Expression (CPM) of Major Chemosensory Gene Families in Female and Male *An. quadriannulatus*: A) Gustatory Receptors, B) Ionotropic Receptors, C) Olfactory Receptors, and D) Odorant Binding Proteins. The diagonal line represents equal expression in both sexes. Genes indicated in red have an FDR-adjusted p-value < 0.05.

2.4. Discussion

I have identified several genes that are differentially expressed in four pairwise comparisons of male and female *An. coluzzii* and *An. quadriannulatus*, including some genes that are strong candidates for controlling behaviors that are not only species-specific, but also female-specific, which could include vertebrate host choice. I have also identified genes whose expression patterns suggest they are of greater importance in males, suggesting a potential role for the labella in male-specific behaviors, which could include swarming or mating. In addition, there are several genes which show evidence of species bias, but not sex bias, indicating a probable role in sex-neutral behaviors, such as sugar-feeding.

Beyond the differential expression results, I also contribute the first characterization of the chemosensory gene expression profile of the labella in *An. quadriannulatus*, and offer a comparison to a previous profile in *An. coluzzii*, which was also recently characterized by Saveer *et al.*[95] (discussed below). In assessing highly expressed genes which are not DE, I identify genes which show evidence of an important role in this organ, and discuss comparisons with the far better understood system at work in *Drosophila* where possible, as well as previous work characterizing the gene expression profile of the labella in *Aedes aegypti*. Finally, I consider the sum total of the expression data to evaluate the potential role of the labella in host choice.

2.4.1. Agreement with Previous Characterization of *An. coluzzii* Labella Transcriptome

There are four major differences in my study methodology and that of Saveer *et al.*[95]. Most importantly, their mosquitoes were dissected during the light cycle, which could explain many discrepancies between our data. In addition, while I produced six replicate libraries of six-day-post-emergence labella, they produced only two replicate libraries of mosquitoes ranging in age from four to six days post-emergence. Finally, there are several differences in RNA extraction and library preparation methods.

Nevertheless, the overall trends of our data are largely concordant, suggesting that most chemosensory gene expression in the labella does not vary significantly by time of day. Even so, there are some discrepancies between the two datasets, which could stem from daily gene expression patterns, differences in age structure, or from differences in RNA and cDNA preparation. In the *Grs*, the most interesting area of disagreement is in the candidate sugar receptors. While I only detected very high expression in *Gr17* and relatively high expression in *Grs* 14 and 21, they also detected very high expression in *Gr15*, and relatively high expression in *Grs* 18 and 25.

In the *Irs*, they detect *Ir41a* above even the co-receptor *Ir25a*, whereas this is not the case in my data. While the only members of the *Ir7* clade expressed at relatively high levels in my data are *Ir7s*, *Ir7x*, and *Ir7w*, Saveer *et al.* also identified relatively high expression of *Ir7t*. *Or6*, *Or13*, and *Or55* were all relatively abundant in both sexes in their analysis, but none are detected at high levels in mine.

Obp expression is largely similar between the two studies, with the major exception of *Obp13*, which while one of the most highly expressed *Obps* in my data, is completely absent from theirs. This *Obp*'s daily expression pattern has not been characterized, but the major discrepancy could reflect a similar daily rhythm as those found in other *Obps*, differences in age structure, or the different library preparation methods[155]. Interestingly, other *Obps* which have previously been characterized as having these daily rhythms in the antennae do not appear to differ between the two studies, most notably *Obps* 10, 25, and 26. However, another of these *Obps*, *Obp7*, is highly expressed in my data but not that of Saveer *et al.*[95,155].

2.4.2. Female and Species-Biased Genes

Obp26 is a clear candidate for further study, particularly of its ligand(s) and its impacts on host-seeking behavior. It is by far most highly expressed in female *An. coluzzii* labella, where it shows significant enhancement compared against male *An. coluzzii* as well as female *An. quadriannulatus*. This pattern suggests a female-specific biological role, which is further supported by its significant enhancement in female *An. quadriannulatus* versus males of the same species. In addition, there is evidence for a species-specific role, given that it is significantly enhanced in both female and male *An. coluzzii* when compared to *An. quadriannulatus*. As such, *Obp26* has strong potential to be involved in behaviors that distinguish female *An. coluzzii* from their conspecific males and the females of their more zoophilic sister species. Interestingly, *Obp26* is located within a QTL for human host preference in *An. coluzzii*, where it also shows evidence of a selective sweep[182]. While the specific function is not known, an RNAi-

mediated knockdown had no impact on blood-feeding[217]. Host preference of the RNAi-mediated knockdown mosquitoes was not tested. However, *Obp26* shows evidence of increased expression during scotophase[154], and is also *An. coluzzii*-biased in the female antennae and maxillary palps[45,46]. However, *Obp26* is male-biased in *An. coluzzii* antennae[47], and is not significantly DE between the two species of males in the maxillary palps[48], which could reflect differing functions between the three organs.

The opposite pattern is observed in *Obp57*, which is expressed at higher levels in female *An. quadriannulatus* than both *An. coluzzii* females and males of both species. As with *Obp26*, it shows evidence of both female bias and species bias. However, unlike *Obp26*, it is expressed at very high levels (>1,000 CPM) in all mosquitoes analyzed. The biological function of *Obp57* is also unknown. However, it shows evidence of increased expression in mosquitoes which are exposed to agricultural pesticides[218]. While it may be involved in some aspect of biology specific to female *An. quadriannulatus*, it also appears to be important for both conspecific males and both sexes of its anthropophilic sister species. However, like *Obp26*, its expression patterns are different in other organs: it is *An. coluzzii*-biased in female maxillary palps, and not significantly DE in male maxillary palps or the antennae of either species[46–48].

Previous work in *An. gambiae* has shown that *Obps* have complex co-expression patterns, and act not simply as transporters, but play more complicated roles in odorant detection[219]. Their ligand-binding abilities are broad and overlapping[220], as are the antennal sensilla where many of them are highly expressed[49], and as such it has been proposed that any individual *Obp* may play a role in the recognition of multiple

chemicals, as well as that multiple *Obps* may interact in order to recognize any particular chemical[219]. Given this complexity, as well as the conflicting patterns in different organs, it is possible that *Obp26* and *Obp57* play multiple roles, bind to multiple chemicals, and/or interact with each other and/or other *Obps* to facilitate either odorant or tastant recognition in the anopheline labella, and that these roles may be different in different organs. *Obps* have been studied extensively *in vitro*, particularly in *Drosophila*, but their roles *in vivo* are still largely unknown and may not necessarily correlate with *in vitro* manipulations[94]. As such, further study on these *Obps* is warranted.

While *Obp26* and *Obp57* are the only genes in this study which show conclusive patterns of both female-biased and species-biased expression, four other genes are significantly DE between females, but not males: *Gr31*, *Ir41a*, *Obp25* (all *An. coluzzii*-biased), and *Or29* (*An. quadriannulatus*-biased). *Gr31* has no known ligand or orthologs beyond *Anopheles*. By contrast, *Ir41a* is known to elicit amine and imine compound responses, is responsive to a wide range of odorants, and is particularly responsive to pyrrolidine and 2-methyl-2-thiazoline[73]. In addition, the co-receptors *Ir25a* and *Ir76b*, which are both highly expressed in the labella of all groups, are obligatory for *Ir41a* responsiveness[73]. Furthermore, its homolog in *Drosophila melanogaster* is necessary to elicit cadaverine and putrescine attraction. Since these chemicals are unlikely to be involved in vertebrate host preference, *Ir41a* could be involved in assessing plant host quality.

Obp25 has no known ligand, but is up-regulated following mating in female *An. gambiae* s.s., although not in *An. coluzzii*, and may play a role in promoting assortative

mating in mixed-species swarms[221]. Like *Obp26*, *Obp25* is up-regulated during scotophase[154]. Finally, *Or29* is activated by linalool, which may be an odor cue for nectar-seeking mosquitoes[222]. These genes may either be DE in other comparisons (i.e. female-biased intraspecifically as well or species-biased in males as well) but not detected due to low power, or they may indeed only be DE in females.

2.4.3. Male and Species-Biased Genes

No genes in the four major chemosensory gene families show evidence of both male and species bias. However, several genes show evidence of species bias in males, but not females. Those genes include: *Gr4*, *Gr21*, *Gr32*, *Gr48*, *Ir179*, and *Obp10*. While *Obp10* is *An. coluzzii*-biased in males, the other five genes are all *An. quadriannulatus*-biased. In addition, *Or40* shows male bias in *An. coluzzii* and is lowly expressed in both sexes of *An. quadriannulatus*. No ligand is known for *Obp10*, although its expression has been characterized as irregular with respect to rhythms between RNA and protein abundance when compared with other *An. gambiae Obps*[155]. *Or40* is expressed in larval *An. gambiae*[223], and is known to be activated by DEET[224] (which was not known to be present in the lab during dissection). However, its *Drosophila* ortholog, *DmOr83a*, with which it shares 26% DNA sequence similarity, is thought to be nonfunctional[225]. Despite its 26-fold higher expression in male *An. coluzzii* than male *An. quadriannulatus*, this difference is not significant ($q = 0.116$). *Or40* expression is highly variable between male *An. coluzzii* replicates, ranging from 0 CPM in two replicates to 84.3 CPM in another, which likely explains these results. It is difficult to explain why it would be

absent in some replicates, but extremely abundant in others, and as such biological significance to this pattern is elusive.

Gr21 is a sugar receptor[95], while *Grs* 4, 32, and 48 have no known ligands. Furthermore, none of these three *Grs* have known orthologs in either *Drosophila* or *Aedes aegypti*. In addition, *Gr32* is alternatively spliced, with two transcripts. Unfortunately, our chosen sequencing method is unable to determine from which transcript the enhancement in male *An. quadriannulatus* stems, although the *Gr32*-RA transcript was expressed at slightly higher levels in male *An. coluzzii* in a previous study[95]. In addition to being DE between males, *Gr48* also shows evidence of positive selection in *An. quadriannulatus* (Chapter 4). As with the genes discussed above, these genes could either be male-biased but not detected as such, species-biased but not detected as such, or simply DE in males but not other comparisons. There is precedent for male-biased *Grs* in *Drosophila*, where *Grs* expressed in the tarsi perceive female pheromones and stimulate or suppress courtship behavior[53,84]. Ongoing work in our lab on chemosensory gene expression in the tarsi may determine if *Grs* are male-biased in *Anopheles* tarsi as well.

Finally, *Ir179*, like *Or40*, is highly variable between replicates: its expression ranges from 0 CPM to 44.5 CPM. There is also some variation in female *An. quadriannulatus* expression, but of a lesser degree, with the highest expression at 5.2 CPM. *Ir179* expression is also highly variable in *An. coluzzii*. This gene was only recently identified and has no known ligand or non-anopheline orthologs.

2.4.4. Species-Biased Genes without Sex Bias

Eight chemosensory genes are significantly DE between species in both sexes. *Gr49*, *Ir7w*, *Or21*, *Obp51*, and *Obp54* show consistent *An. coluzzii*-biased expression; while *Ir7x*, *Obp48*, and *Obp50* show consistent *An. quadriannulatus*-biased expression. Furthermore, while *Obp26* and *Obp57* show their most pronounced expression in females, they are both also significantly DE in males of *An. coluzzii* and *An. quadriannulatus*, respectively. None of these genes have known ligands, although their equal expression between sexes suggests that they mediate species-specific but not sex-specific behaviors. The perception of plant-derived chemicals is a likely role, as both sexes feed on plant sugars.

Ir7w is the best understood of the *An. coluzzii*-biased genes, although its specific ligand is unknown. The *Ir7* clade (which includes both *Ir7w* and *Ir7x*) is also expressed in the *Drosophila* and *Aedes aegypti* labella[71,76,96], as well as more broadly across Insecta, where it also shows evidence of substantial gene duplication in parallel with gustatory adaptation to new environments[226,227]. However, the *Anopheles* (*Ag*)*Ir7* genes only share between 10-15% DNA sequence homology with the *DmIr7* genes. *Gr49* has no clear *Drosophila* ortholog but its *Aedes aegypti* ortholog, *AaegGr43*, shows male bias in the labella and tarsi[153]. In addition, *Gr49* shows marginally significant ($p = 0.075$) evidence of a selective sweep in *An. gambiae* s.s. (Chapter 4). Like several other *Grs*, it is alternatively spliced with two transcripts, and while it is unclear which transcript is behind the differential expression pattern, only *Gr49-RA* shows evidence of selection.

The RA transcript was also more highly expressed than the RB transcript in a previous study[95].

Or21 has no known orthologs. Despite its *An. coluzzii* bias in the labella, it is *An. quadriannulatus*-biased in female antennae[46], which suggests it may play different roles in the two organs. Indeed, there is evidence in *Drosophila* that a single OR may bind with multiple different ligands[59], which raises the possibility that *Or21* responds to a different chemical in *An. quadriannulatus* antennae than in *An. coluzzii* labella. Neither *Obp51* nor *Obp54* have clear orthologs, but both genes show enhanced transcription following exposure to agricultural pesticides, similarly to *Obp57*[218].

Ir7x is part of the same *Ir7* clade as *Ir7w*, so may play a parallel role in *An. quadriannulatus*. Neither *Obp48* nor *Obp50* have known orthologs beyond *Anopheles*, despite several studies on *Obp48*[219,228,229]. *Obp48* is highly expressed in the antennae and is down-regulated following a blood meal[229], so likely plays a role in host-seeking there, although the expression pattern is not consistent with such a role in the labella. *Obp50*, like *Obps* 51, 54, and 57, is enhanced in the presence of agricultural pesticides[218].

2.4.5. Sex-Biased Genes without Species Bias

Unlike in other organs like the antennae, there are no chemosensory genes that show a sex-biased but not species-biased expression pattern in the labella. While both *Obp26* and *Obp57* show clear evidence of female bias, and multiple genes show less conclusive evidence of male bias, all of these genes also show species bias. However, there is a clear pattern in both species that overall chemosensory gene expression is

higher in female labella than in male. This likely stems from the female labellar role as a blood-feeding organ or from volatile-driven host-seeking behavior.

2.4.6. Non-DE Highly Expressed Genes

In addition to those genes listed above which show evidence of differential expression, I have also identified several genes which are highly expressed but not DE in the labella. Among the *Grs*, the clear standout is the sugar receptor *Gr17*, which is the most highly expressed in all four mosquito groups. Interestingly, despite its similarly high expression in both sexes of both species, *Gr17* shows evidence of a selective sweep in *An. coluzzii* (Chapter 4), pointing to its biological importance. Its specific ligand is unknown, but the perception of sugars in *Drosophila* is quite complex[87], and it is entirely possible that *Gr17* may play a combinatorial role in the perception of a variety of different sugars. The other sugar receptors (*Grs* 14-21 and *Gr25*) are mostly expressed at much lower levels, with only *Gr14* and *Gr21* detected >1 CPM. However, as mentioned above, expression in multiple sugar receptors was much higher in Saveer *et al*. (2018)[95]. These *Grs* are classified as sugar receptors due to sequence homology with *Aedes aegypti* candidate sugar receptors rather than from deorphanization studies in *Anopheles*, so it is possible both that other *Grs* perceive sugars or that some of these *Grs* are not in fact functional sugar receptors in *Anopheles*. It is also possible, although unlikely, that sugar perception works differently in *Anopheles* than *Drosophila*. *Gr33* is also relatively highly expressed in males and females of both species, in contrast to its male-specific expression in the antennae[47]. Ongoing work generating a CRISPR-*Cas9 Gr33* knockout

strain of *An. coluzzii* may clarify the role of this gene in both male antennae and the labella of both sexes.

The two most highly expressed *Irs* in both species and sexes are the co-receptors *Ir76b* and *Ir25a*, which is expected. While *Ir25a* is more highly expressed in the *Aedes aegypti* labellum[96], our results confirm that *Ir76b* is the more highly expressed co-receptor in *An. coluzzii* labella[95]. The fact that the third co-receptor, *Ir8a*, was not detected in any samples suggests that it plays no role in the labella. It was similarly undetected by Saveer *et al.* in their study of the *An. coluzzii* labella[95], but was abundantly detected in the antennae[46,47]. Its *Aedes aegypti* ortholog is required for acid-sensing[8], but given that labellar neurons have robust responses to acids[95], another co-receptor likely mediates acid perception in *Anopheles* labella. In addition to the DE *Ir7w* and *Ir7x*, *Ir7s* is also relatively highly expressed in both sexes of both species.

In both sexes of both species, *Orco* is by far the most highly-expressed *Or*, which was also true of previous studies in both *An. coluzzii* and *Aedes aegypti*[95,96]. Thus, my results concur with several previous studies that showed olfactory activity in the labella. Apart from *Or21*, *Or29*, and *Or40*, all of which are DE, there are two other notable *Ors* which are not DE: *Or4* and *Or5*. Neither of these *Ors* has a known ligand or any clear orthologs outside of *Anopheles*.

Lastly, there are three *Obps* which are abundantly expressed (>100 CPM) but not significantly DE: *Obp1*, *Obp7*, and *Obp13*. *Obp1* has been characterized as an indole mediator[228], and is the most highly expressed *Obp* in female *An. coluzzii* antennae, where it shows an *An. coluzzii* bias[46]. Indole is an oviposition attractant, and work in

Culex quinquefasciatus has demonstrated the role of the proboscis in oviposition: specifically, the labrum is required for the long-distance perception of the oviposition attractant 4-ethylphenol, although neither the labrum nor the labella have a role in detecting the close-range oviposition attractant skatole[54]. Indole stimulates the antennae of female *An. gambiae*[205], and is detected by a conserved *Or2*/*Or10* clade across mosquitoes[230]. However, neither *Or* is detected >1 CPM in the labella, so it is unlikely that the role of *Obp1* in the labella is to mediate indole perception. Indeed, other work has shown that *Obp1* is co-expressed with other *Obps*, such as *Obp4* in the antennae of *An. gambiae*[220,231], suggesting that it could play other important roles in the labella.

Obp7 shows *An. coluzzii* bias in the antennae[45], is not required for indole response[228], and its protein abundance peaks during the dark cycle[155]. Its relative abundance in our data set stands in contrast to its apparent absence in the previous characterization of labellar expression in *An. coluzzii*[95]. Its *Drosophila* ortholog, *DmObp69a*, plays a role in cis-vaccenyl acetate (cVA) mediated mating behavior. Specifically, its abundance is reduced by cVA in males and increased by cVA in females. cVA makes males more aggressive and females more sexually receptive, so *DmObp69a* mediates these behavioral changes by binding with cVA[232]. No cVA equivalent is known in *Anopheles*, so the direct relevance of this orthologous relationship is unclear, but the equivalent expression in males and females could support a similar role in mating. However, it should be noted that all mosquitoes included in this study were presumed to have already mated. RNA-Seq from unmated mosquito labella might better elucidate the potential role of this *Obp* in anopheline mating. *Obp13* shows

a strong *An. coluzzii* bias in the female antennae[46], and like several other highly-expressed *Obps* shows increased abundance after pesticide exposure[218]. Despite its extreme abundance in all of our samples, it was not detected by Saveer *et al*.[95]. Its *Aedes aegypti* ortholog, *AaegObp57,* is also one of the most abundant *Obps* in the labella[96]. However, no ligands or non-mosquito orthologs are known.

2.4.7. Biological Role of the Labella

Two genes, *Obp26* and *Obp57*, present the best evidence for a possible role of the labella in female-specific and species-specific behaviors in the *An. gambiae* complex. Given both the high expression levels and the differential expression patterns of these two genes, they clearly play an important role in differentiating the behavior of these two females. However, it cannot conclusively be determined whether they play a role in host choice or in another female-specific behavior, such as oviposition or mate recognition. If indeed they play a role in host choice, it is unclear which other genes they would interact with to do so, given that no receptors show the same pattern in the labella. As such, it is difficult to conclude that the labella are of great importance to vertebrate host choice, though they may indeed play a lesser role. It is also possible that genes responsive to vertebrate host odor are equally expressed in both sexes, but used to different ends in males than in females.

Nonetheless, it is clear that the labella detect both long-range and contact chemical cues. The high expression levels of known sugar receptors suggest that the labella are important for the detection and/or evaluation of nectar sources, and the high expression levels of other chemosensory receptors concord with anatomical and

electrophysiological observations to conclude that the labella respond to a variety of other chemical cues as well. Indeed, previous work has shown that labellar neurons respond to a variety of chemical cues that are present both in human sweat and oviposition sites, bolstering the evidence that the labella play roles in both host-seeking and oviposition[95].

At this point, the chemosensory gene expression patterns of *An. coluzzii* and *An. quadriannulatus* have been extensively analyzed. However, with the advent of CRISPR-*Cas9* technologies in *Anopheles*[185,186,188], it is now feasible to attempt a knock-out of the genes identified in this study, which when combined with a bioassay such as a dual-port olfactometer, could help to more conclusively elucidate their role in female host-choice, if any, as well as other behaviors, such as oviposition, sugar-feeding, and mating.

3. THE ROLE OF THE MAJOR CHEMOSENSORY ORGANS IN ACTIVATION OF HOST-SEEKING AND HOST-DIFFERENTIATION IN *ANOPHELES COLUZZII*

3.1. Introduction

Many of the major mosquito vectors of human diseases like malaria and dengue fever, have high vectorial capacities due to their anthropophilic tendencies. That is, these mosquitoes strongly prefer to feed on a human host as opposed to other non-human vertebrates. Although visual and thermal cues contribute to host-seeking and host preference[6,7,10,112,113,122,233], the mosquito's chemosensory systems are critical to these behaviors[17,122,129,130,132,134,223,234,235]. Despite significant study of the expression and evolution of chemosensory genes in anthropophilic mosquitoes, relatively few studies exist to determine conclusively which chemosensory organs facilitate vertebrate host preference[4,100].

Olfaction predominantly occurs in the mosquito antennae, along with the maxillary palps and to a smaller extent in the proboscis, where *Orco* expression was previously demonstrated and a recent study identified olfactory sensilla that are responsive to several odorant blends[95,193,194,236]. The maxillary palps are host to the carbon dioxide receptors, which play an important role in host-seeking[91,106,107]. The proboscis is the main site of gustation, with the tarsi also playing a gustatory role[96,153]. In addition to their roles in the detection of both volatile and contact chemical cues, the chemosensory organs also play important non-chemosensory roles. The antennae are involved in hygroreception[104,105] and thermoreception[237], while the maxillary palps are

also involved in both thermosensation and mechanosensation[159]. In addition, the proboscis also plays a crucial role in thermosensation[10].

The focus of this study is the *Anopheles* (*An.*) *gambiae* complex, which contains three of the dominant vectors of *Plasmodium falciparum* malaria in sub-Saharan Africa, including *An. coluzzii*. *An. coluzzii* and its sister species *An. gambiae* s.s. are highly anthropophilic[17,130,147,148,150,233,238], which contributes greatly to their vectorial capacity. *An. coluzzii* and *An. gambiae* s.s. were originally considered two different forms (M and S, respectively) of the same species, but were recently split into separate species[14]. Another member of the *An. gambiae* complex, *An. quadriannulatus*, is zoophilic, preferring to feed on cattle[18,19,147,149]. While previous work has established that there are significant differences in chemosensory gene expression in these sibling species[45–47,239], no one has yet assessed the role of chemosensory organs in the host-feeding decisions made by the two species.

Previous work has demonstrated that the antennae are critical for activation of host-seeking in both *Aedes aegypti* and the North American and European malaria vectors *An. quadrimaculatus* and *An. atroparvus*[4,100,102,103]. Furthermore, Mukwaya (1976) showed that *Aedes aegypti* went from preferring a human arm over a live rat 72% of the time when intact to only 58% of the time when antennectomized, indicating that the antennae play a clear role in host preference in this species[101]. While antennectomized *Aedes aegypti* and *An. quadrimaculatus* show a clear reduction in close-range host-seeking activity when presented with a human arm, *An. quadrimaculatus* still exhibit probing behavior (i.e. they push their head towards the

skin) when their maxillary palps and/or proboscis are removed, and even when their antennae are removed[100]. Roth proposed that the hind legs might play a thermosensory role in the detection of host convection currents, which might have been sufficient to initiate probing.

Work in *Culex pipiens fatigans* has also shown that palpectomized mosquitoes are much less likely to leave a release chamber when exposed to either human hand odor or CO_2[106], although the maxillary palps can be removed without abolishing probing behavior in both *Aedes aegypti* and *An. quadrimaculatus*[100,102]. Finally, Maekawa *et al*. (2011) demonstrated that the South Asian malaria vector *An. stephensi* is less responsive to a murine host at close-range when the antennae, maxillary palps, or proboscis are removed, with the proboscis causing the greatest reduction in host landings, which they attributed to its thermosensory role[10]. They also showed that while mosquitoes with ablated antennae or maxillary palps could still respond to a hot plate serving as an artificial host, those with ablated proboscis could not, which led them to consider the proboscis as a "thermo-antenna" necessary to host-seeking behavior.

The antennae clearly play a critical role in host preference and host-seeking activation, but other organs are sufficient to initiate probing. The maxillary palps are well known as the site of CO_2 perception, but their importance for odor detection at either close or long-range is less clear. Furthermore, the role of CO_2 is dependent on mosquito host preference, and work comparing the generalist feeder *Culex quinquefasciatus* with the anthropophilic *Aedes aegypti* and *An. coluzzii* has shown that these mosquitoes are responsive to different concentrations of CO_2[140]. Similarly, CO_2

increases in importance in the *An. gambiae* complex with increasing zoophily[18], so the maxillary palps could potentially vary in importance with host preference if their major role in host-seeking is CO_2 perception. However, if they play a significant olfactory role aside from CO_2, as expression and electrophysiology analyses suggest[46,48,152,159,240], they could still be important in anthropophilic mosquitoes. As mosquitoes still probe with these chemosensory organs removed, it is possible that they rely on contact chemicals detected by the tarsi or on heat or other cues to initiate probing. While the role of the proboscis in distance thermosensation and close-range host response has been demonstrated[10], as has its olfactory capabilities[95], its behavioral role in olfaction has not been clearly identified.

In this study, I sought to determine which of the major chemosensory organs of *An. coluzzii* are necessary for initiating host-seeking and for responding to a vertebrate host odor. My aim was also to determine which olfactory organs are necessary and/or sufficient for human host preference. I removed the olfactory organs of female *An. coluzzii*, which were provided with a choice between human and bovine odor in a dual-port olfactometer. In addition, to assess the roles of these organs in closer-range host-seeking and blood-feeding, I replicated older ablation probing experiments by removing the same organs and giving the mosquitoes access to an artificial membrane blood feeder.

3.2. Methods

3.2.1. Host Choice Experiments

3.2.1.1. Mosquito Rearing

Mosquitoes for this study were drawn from a lab colony of *An. coluzzii* (SUA2LA strain, originally collected from Suakoko, Liberia). Mosquitoes were reared in an insectary at Texas A&M University (College Station, TX, USA). Colonies were maintained at 28º C and 80% relative humidity (RH), with a light:dark photoperiod of 12 hours incorporating 30 minutes of low light at the beginning and end of the light cycle to simulate dawn and dusk, respectively. Adult mosquitoes were maintained on a 10% sucrose solution and blood-fed with defibrinated bovine blood using a membrane-feeding system twice a week. Larvae were maintained at densities of approximately 60 per 2-liter container and fed finely-ground fish food (TetraMin, Blacksburg, VA, USA). Pupae were collected and placed into cages at a density of approximately 150-400 pupae per cage.

Cages were checked daily for newly emerged mosquitoes. To keep the age of mosquitoes used in the behavioral studies consistent, non-eclosed pupae were transferred to new cages. Mosquitoes were maintained in mixed-sex cages for five days prior to the experiments, but were not blood-feed.

3.2.1.2. Pre-Experiment Mosquito Processing

Mosquitoes were knocked down one day before experiments at 4º C for ten minutes, and kept cool while they were separated by sex. The females were then separated into either the control group (manipulated, but not exposed to ablation), or one

of the following experimental groups: antennae ablated, maxillary palps ablated, labella ablated, hind leg ablated. In addition, one experimental group consisted of intact mosquitoes that were not exposed to carbon dioxide during the host-choice experiment.

Antenna, maxillary palp, and leg ablation were conducted by pinching a sharp, very fine-tipped pair of forceps as close to the base of the organ as possible. While it was possible to remove nearly the entire antennae distal to the Johnston's organ, the maxillary palps were more challenging to remove in this fashion, so approximately 10% of the most proximal portion remained attached (Figure 3.1). The labella were ablated using a scalpel, with care taken to minimize removal of non-labellar proboscis tissue. Mosquitoes were then transferred to an experimental release cage and provided a 10% sucrose solution under insectary conditions.

Figure 3.1 – Typical examples of ablated mosquitoes: A) a mosquito with ablated antennae, with Johnston's organs visible and a small piece of most proximal antennal segment attached; B) a mosquito with ablated maxillary palps with approximately 10% still attached; C) a mosquito with ablated labella.

Eight hours prior to the experiment, the sucrose solution was replaced by distilled water, and mosquitoes were transferred to the behavior room, also maintained at 28º C

and 80% RH. Two hours prior to the experiment and concurrent with the light:dark schedule in the insectary, the room was transitioned to low, diffuse light of approximately 2.5 lux.

3.2.1.3. Odor Sources

Human odor was obtained from a male volunteer, who wore a pair of cotton socks for a period of 48 hours. Bovine odor was obtained by tying a pair of nylon pantyhose around a cow's hind leg for a period of 24 hours. Cows were located at the Texas A&M Veterinary Medical Research Park in College Station, Texas, USA.

Both human and bovine odor samples were stored in plastic sandwich bags at -20º C immediately following collection. 30 minutes prior to the start of experiments, the odor samples were placed in a 37º incubator in an attempt to replicate vertebrate body temperature and stimulate odor development.

3.2.1.4. Olfactometer Design

Experiments took place in a Plexiglass olfactometer with dimensions 185 cm long X 78 cm wide X 75 cm tall (Figure 3.2), modeled after Braks and Takken (1999) and Knols et al. (1994)[133,241]. While five of the six sides of the olfactometer were solid Plexiglass, the sixth was mostly comprised of mesh, in order to allow odors to dissipate and create a natural gradient. A release cage was attached to an 8 cm diameter entrance port located in the center of the mesh wall of the olfactometer.

On the opposite wall, two odor ports of diameter 3 cm were attached at a distance of 18 cm from the side and 34 cm from each other. Each odor port contained either a human or bovine odor sample, which were alternated every night experiments occurred.

Warmed (approximately 37º C), humidified air was pumped through each odor port at a rate of approximately 0.4 m/s and 5% carbon dioxide was pumped through a separate tube roughly equidistant from both odor ports at the same approximate rate.

Figure 3.2 – Dual-Port Olfactometer. Release cage is on left, while odor ports are on right.

3.2.1.5. Experimental Protocol

While a set of control mosquitoes was run through the olfactometer every night, the order of each night's trials was randomized. A maximum of six trials were conducted each night. Once the odor samples were installed and the carbon dioxide was running, the release cage door was opened and the experimenter left the room for a period of 20

minutes. At the conclusion of the 20 minute trial, the release cage door was closed and sealed with tape. Odor ports were checked for responding mosquitoes and sealed with tape. Visibly dead mosquitoes in the release cage were counted at this stage to facilitate accurate counts at a later stage. A vacuum was used to clear the olfactometer of mosquitoes between trials. Mosquitoes in the release and odor port collection cages were knocked down between -4 and -20º C, then counted. Mosquitoes were examined during the counting phase to verify their experimental group (i.e. which organs, if any, had been ablated).

3.2.1.6. Data Processing

The following information was recorded for each trial: the total number of starting mosquitoes, the total number of dead mosquitoes at time of experiment run, the total number of mosquitoes in each odor port, and the total number of mosquitoes remaining in the release cage post-run. Dead mosquitoes in the release cage were subtracted from the total. The remaining total was scored as follows: active, responsive, or inactive. Inactive mosquitoes were those which were still in the release cage at the end of the experiment: they were considered non-activated and non-responsive. Active mosquitoes were those which left the release cage (including both those in the main olfactometer chamber and those in an odor port), while responsive mosquitoes included only those which ended the experiment in one of the odor ports. The proportions of active, responsive, and non-activated/non-responsive mosquitoes were calculated for each trial and averaged for each condition. Mortality was also calculated, and the location of the human odor sample port was recorded for each experiment.

3.2.1.7. Statistical Analysis

The data distribution was analyzed with the R[242] packages car[243] and MASS[244]. The data were found to best fit a negative binomial distribution, and therefore further analysis was performed with the R package glmmTMB[245]. The effect of experimental condition on activation rate and response rate was analyzed by a generalized linear mixed model with date as a random variable to account for variability in atmospheric pressure, heat, humidity, and other abiotic factors. A model was also tested incorporating port identity as a random variable, but the Akaike Information Criterion was lower for the model incorporating only date as a random variable, thus indicating a better fit. A Tukey post-hoc test was performed using the emmeans package[246]. In addition, a chi-square test of independence was performed to determine if there were differences in the proportion of activated mosquitoes responding to host odor. This chi-square test was followed by a Bonferroni-corrected posthoc test in rcompanion[247] to determine which pairs differed significantly. The same procedure was performed to compare mortality among experimental groups. All experimental mosquitoes exclusively responded to human host odor, therefore no statistical test for vertebrate host preference were necessary.

3.2.2. Bloodfeeding Experiments

3.2.2.1. Experimental Protocol

Mosquito rearing and ablations were performed as above. No mosquitoes in the bloodfeeding trials were exposed to carbon dioxide, and no labellar ablations were performed, as this would prevent mosquitoes from blood-feeding. As with the

olfactometer experiments, sucrose solution was replaced with distilled water eight hours prior to the start of the experiment. Mosquitoes were maintained in the rearing insectary under normal conditions. Approximately two hours into the dark cycle, mosquitoes were given access to an artificial membrane feeder filled with approximately 2 mL of defibrinated bovine blood for a duration of thirty minutes. At the end of the feeding period, the number of dead mosquitoes in the cage was recorded and mosquitoes were knocked down at -20º C scored into one of three categories (Figure 3.3): non-bloodfed (i.e. non-engorged with no blood visible in the abdomen), bloodfed but not engorged (i.e. blood was visible in the abdomen, but no abdominal distention was observed), or engorged (i.e. blood was visible in the abdomen and the abdomen was clearly distended).

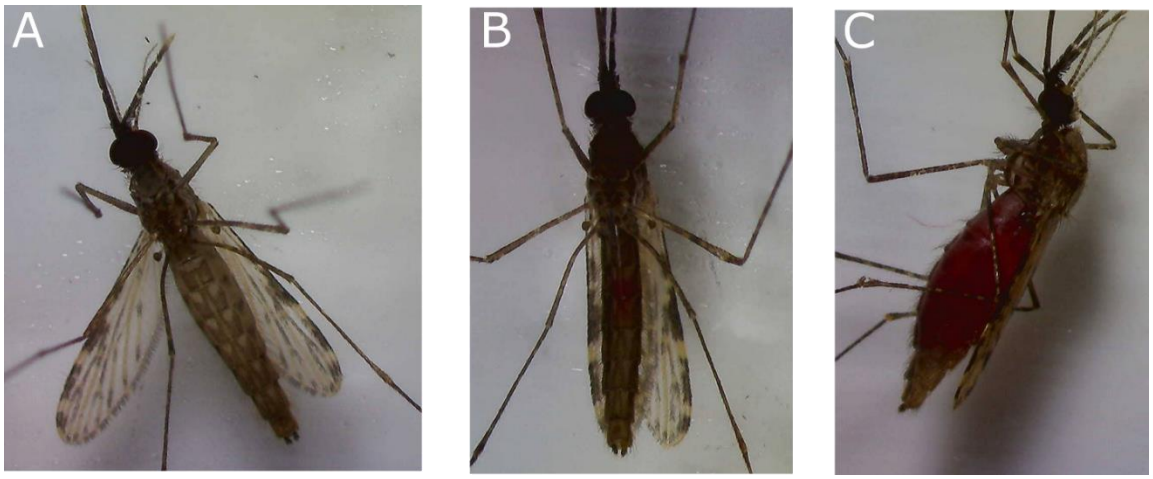

Figure 3.3 – Examples of the three categories of mosquitoes recovered following bloodfeeding trials: A) a non-bloodfed mosquito with translucent abdomen; B) a bloodfed but non-engorged mosquito with a small amount of blood visible in the abdomen; C) an engorged mosquito.

3.2.2.2. Statistical Analysis

A chi-square test of independence was performed to determine if mechanical ablation of each organ had a significant impact on degree of bloodfeeding and engorgement. This chi-square test was followed by a Bonferroni-corrected post-hoc test in rcompanion[247] to determine which pairwise comparisons differed significantly.

3.3. Results

3.3.1. Mortality Across Groups

Mortality for the long-range host-seeking experiments is recorded for each experimental group (Table 3.1). The labella-ablated group had the highest mortality (46%), while the no CO_2 group had the lowest (3.31%). While the antennae-ablated group had a significantly ($p < 0.05$) higher mortality rate than the intact, palp-ablated, leg-ablated and no CO_2 groups prior to Bonferroni adjustment, none of these differences were significant after adjustment ($p > 0.05$). The labella-ablated group had a significantly ($p < 0.01$) higher mortality rate than all other experimental groups.

Table 3.1 – Mortality by Experimental Group in Olfactometer Experiments. Experimental groups in the same statistical group do not differ significantly.

Experimental Group	Total Alive	Starting Number	Mortality Rate	Statistical Group	Replicate Runs
Intact	1,584	1,760	10%	A	30
Antennae Ablated	360	475	24.21%	A	8
Palps Ablated	418	454	7.93%	A	8
Labella Ablated	351	650	46%	B	10
Hind Leg Ablated	412	442	6.79%	A	8
No CO_2	467	483	3.31%	A	8

3.3.2. Odor Response

3.3.2.1. Activation of Host-Seeking

Intact mosquitoes were consistently activated at the highest rate (83.33%), although there was considerable variance by experimental trial (50% - 96.67%), as is typical of studies in a dual-port olfactometer (Table 3.2, Figure 3.4). Ablation of a hind leg or the maxillary palps did not significantly reduce the activation rate (70.87%, $p=0.528$ and 76.89%, $p=0.957$, respectively). In contrast, mosquitoes with antennae ablated had a drastically reduced activation rate, with only 36.67% activated ($p<0.0001$). Similarly, mosquitoes with labella ablated were also much less active, with 51.57% activated ($p<0.0001$). The absence of CO_2 also reduced activation rate, to 61.89% ($p=0.019$).

Table 3.2 – Activation Rate of Mosquitoes by Experimental Group. Experimental groups in the same statistical group do not differ significantly.

Experimental Group	Active Mosquitoes	Total Live Mosquitoes	Activation Rate	Statistical Group	Replicate Runs
Intact	1,320	1,584	83.33%	A	30
Antennae Ablated	132	360	36.67%	B	8
Palps Ablated	321	418	76.79%	A, C	8
Labella Ablated	181	351	51.57%	B	10
Hind Leg Ablated	292	412	70.87%	A, C	8
No CO_2	289	467	61.89%	C	8

Mosquitoes with antennae ablated had activation rates significantly lower than all but those with labella ablated, which had a similarly low activation rate and also had a significantly lower activation rate than all other treatment groups. The activation rate of mosquitoes with leg ablated, maxillary palps ablated, or with no exposure to CO_2 did not differ significantly.

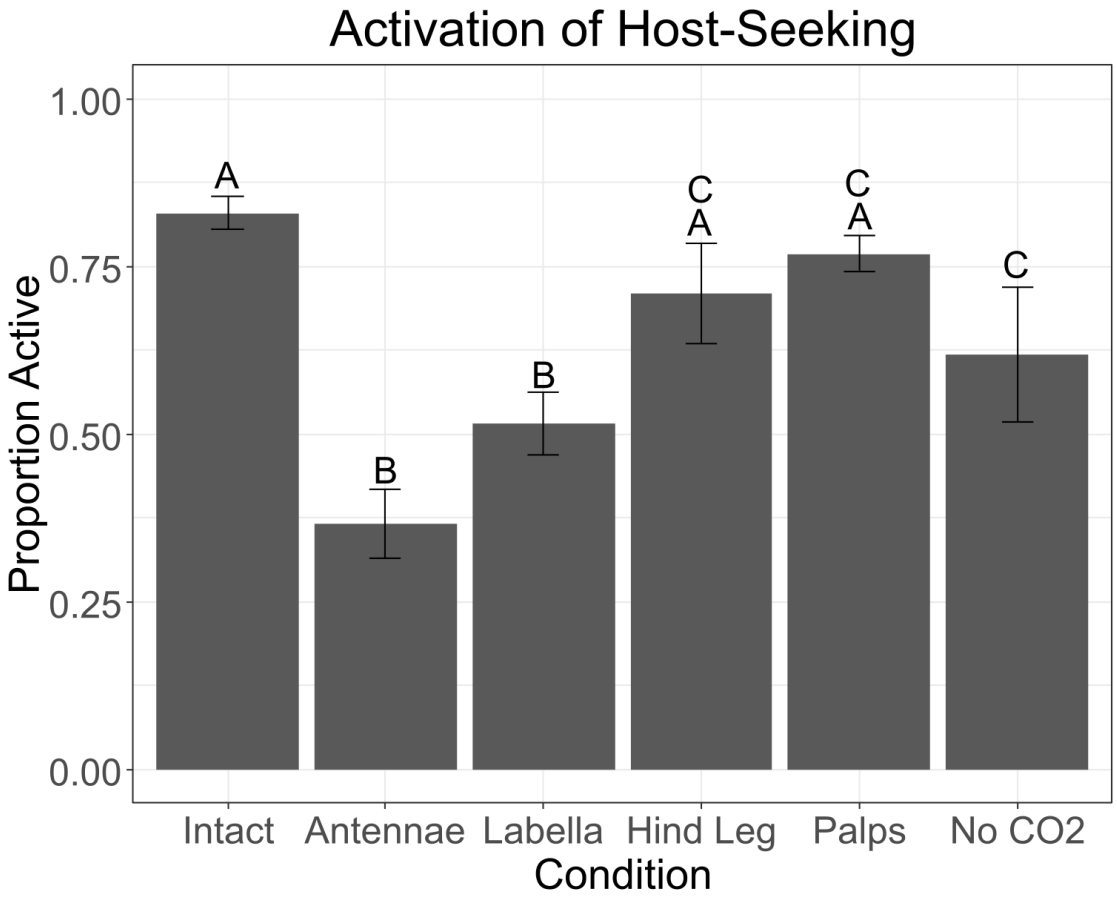

Figure 3.4 – Proportion of mosquitoes activated to host-seek by condition. Error bars represent standard error. Bars sharing the same letter do not differ significantly ($p > 0.05$).

3.3.2.2. Vertebrate Odor Response

Strong anthropophily was retained across responders of all experimental groups. Whereas responsive control group mosquitoes chose human odor 97.27% of the time (a chi-squared test determined that this differed significantly from a random choice – $p < 0.001$), all responsive mosquitoes in all experimental groups only chose human odor. Therefore, responsiveness is analyzed solely as a binary condition.

Among all live intact mosquitoes the overall response rate was 23.1% (Table 3.3, Figure 3.5). In most experimental groups, the response rate was significantly lower, with ablation of the antennae reducing the response rate to 0%. Ablation of the maxillary palps and labella significantly reduced the response rate (6.7%, p=0.0001 and 9.12%, p=0.002, respectively), as did the removal of CO_2 (3.21%, p=0.0001). Leg ablation did not did significantly reduce the response rate (11.17%, p=0.162). However, when comparing only non-intact mosquitoes, excluding those with antennae ablated, no significant differences were found.

Table 3.3 – Response Rate of Mosquitoes by Experimental Group. Experimental groups in the same statistical group do not differ significantly.

Experimental Group	Responsive Mosquitoes	Total Live Mosquitoes	Response Rate	Statistical Group	Replicate Runs
Intact	366	1,584	23.11%	A	30
Antennae Ablated	0	360	0%	-	8
Palps Ablated	28	418	6.7%	B	8
Labella Ablated	32	351	9.12%	B	10
Hind Leg Ablated	46	412	11.17%	A, B	8
No CO_2	15	467	3.21%	B	8

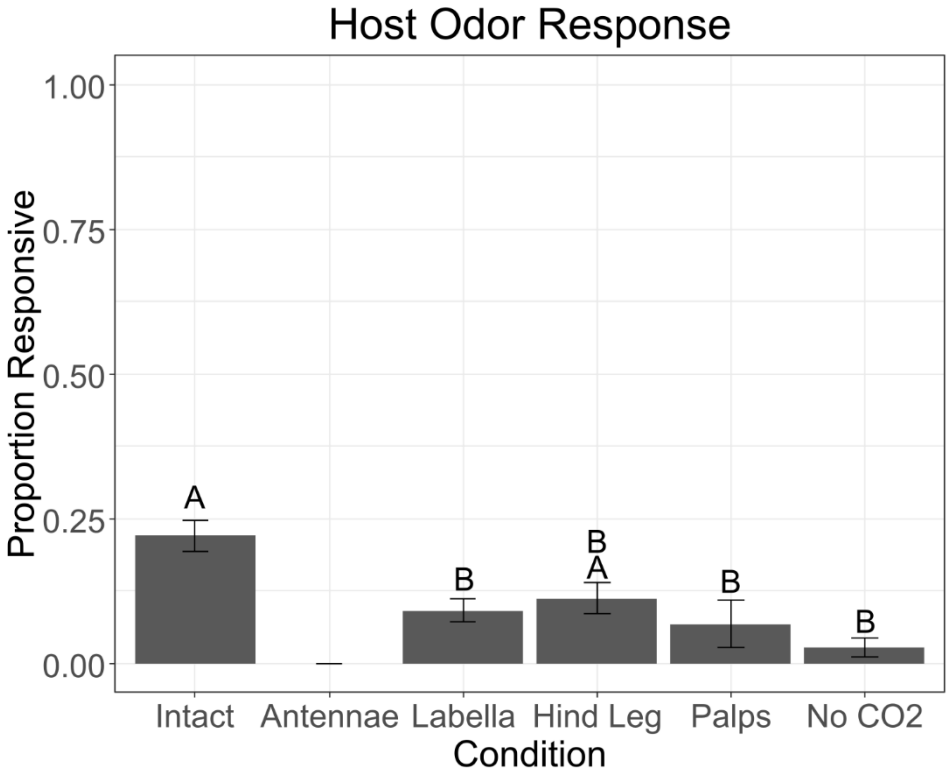

Figure 3.5 – Proportion of mosquitoes responding to any host odor. Error bars represent standard error. Bars sharing the same letter do not differ significantly ($p > 0.05$).

Table 3.4 – Response Rate of Activated Mosquitoes by Experimental Group. Experimental groups in the same statistical group do not differ significantly.

Experimental Group	Responsive Mosquitoes	Non-Responsive Mosquitoes	Total Active Mosquitoes	Response Rate	Statistical Group	Replicate Runs
Intact	366	954	1,320	27.73%	A	30
Antennae Ablated	0	132	132	0%	-	8
Palps Ablated	28	293	321	8.72%	C, D	8
Labella Ablated	32	149	181	17.68%	A, B	10
Hind Leg Ablated	46	246	292	15.75%	B, C	8
No CO_2	15	274	289	5.19%	D	8

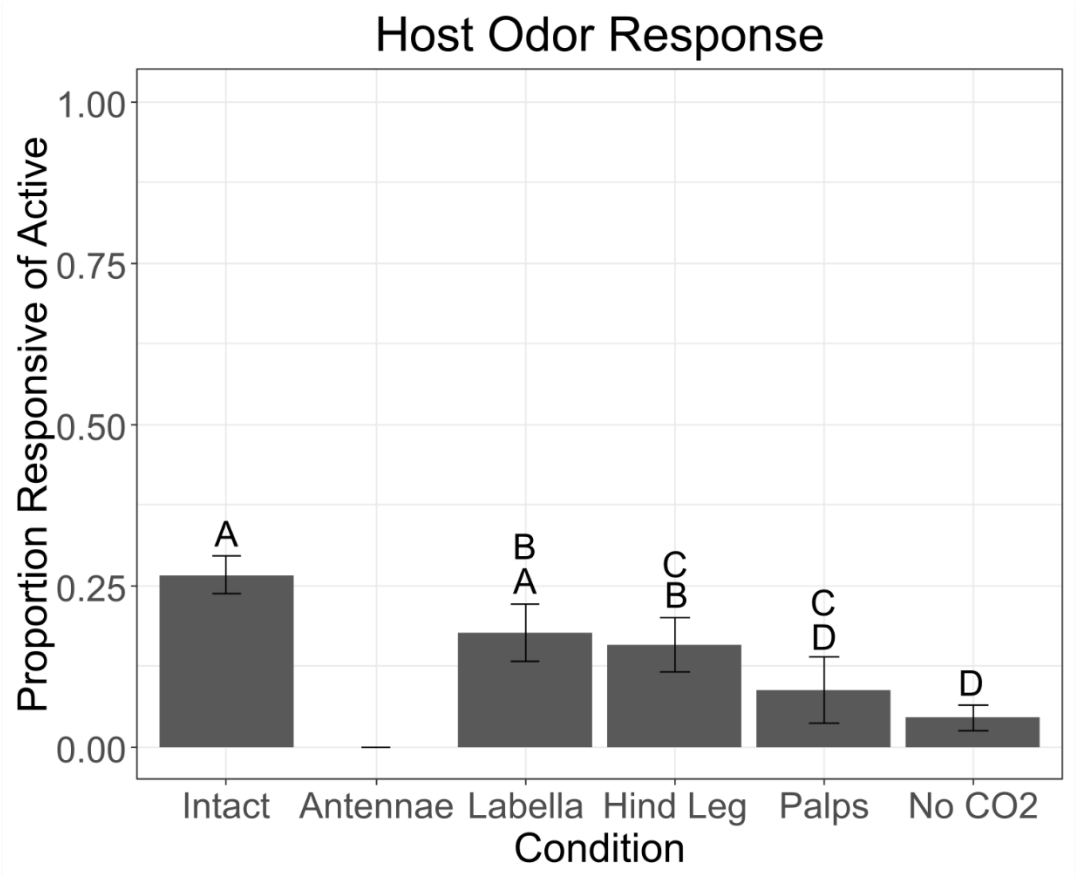

Figure 3.6 – Proportion of activated mosquitoes responding to any host odor. Error bars represent standard deviations. Bars sharing the same letter do not differ significantly (p > 0.05).

The proportion of responsive mosquitoes among activated mosquitoes was analyzed next. A chi-square test of independence determined that the proportion varied significantly by group (X^2 (5,2) = 117.02, p (4 df) < 0.001). The response rate of activated intact mosquitoes was 27.73% (Table 3.4, Figure 3.6). In most experimental groups, this proportion was significantly reduced. As ablation of the antennae eliminated host response, the response rate of these activated mosquitoes was 0. Ablation of the maxillary palps and hind leg significantly reduced the response rate of activated

mosquitoes (8.72%, $p < 0.001$ and 15.75%, $p < 0.001$, respectively), as did the removal of CO_2 (5.19%, $p < 0.001$). Labella ablation reduced the response rate to 17.68%, which was a significant difference ($p=0.005$) prior to Bonferroni correction, but no longer significant ($p=0.054$) after the correction. When comparing only non-intact mosquitoes, no significant differences were found between mosquitoes with a hind leg ablated and either labella or maxillary palps ablated. However, there was a significant difference ($p < 0.001$) between the group with CO_2 removed and all other mosquitoes except those with maxillary palps ablated, as well as between mosquitoes with maxillary palps ablated and labella ablated ($p = 0.047$).

3.3.3. Bloodfeeding Assay

A total of 209 intact mosquitoes, 221 leg-ablated mosquitoes, 197 antenna-ablated mosquitoes, and 197 maxillary palp-ablated mosquitoes were included in experiments conducted across three nights. Counts of mosquitoes in the non-bloodfed, bloodfed non-engorged, and bloodfed engorged are given in Table 3.5. A chi-square test of independence determined that engorgement varied significantly ($p < 0.001$) between groups. Intact mosquitoes were the most likely to be engorged (44.98%) and the least likely not to feed (47.37%). By contrast, mosquitoes with antennae ablated were the least likely to be engorged (8.12%) and the most likely not to feed (84.77%). Feeding patterns vary highly significantly (X^2 (4,3) = 73.3, p (6 df) ≤ 0.001) between all groups, except mosquitoes with a hind leg or maxillary palps ablated, which do not vary significantly ($p = 1$).

Table 3.5 – Degree of Engorgement by Experimental Group. Experimental groups in the same statistical group do not differ significantly.

Condition	Non-Bloodfed	Bloodfed, Non-Engorged	Bloodfed, Engorged	Total Mosquitoes	Statistical Group
Intact	99 (47.37%)	16 (7.66%)	94 (44.98%)	209	A
Antenna	167 (84.77%)	14 (7.11%)	16 (8.12%)	197	B
Leg	147 (74.62%)	14 (7.11%)	60 (30.46%)	221	C
Maxillary Palp	131 (66.5%)	15 (7.61%)	51 (25.89%)	197	C

3.4. Discussion

This study clarifies the specific roles of the different olfactory organs in *An. coluzzii* as they pertain to multiple facets of vertebrate host-seeking, both from a distance in a dual-port olfactometer, and close-range attraction to and probing of an artificial membrane feeder. The study does not explicitly determine which chemicals or senses are most important for these behaviors, but in combination with previous studies, allows us to assess the roles of particular sensory organs in their perception. Finally, the study reinforces the high anthropophilic tendencies of *An. coluzzii*.

3.4.1. Vertebrate Host Preference

In this study, experimentally manipulated mosquitoes responded exclusively to human host odor, and intact mosquitoes overwhelmingly did the same. These results concur with those of previous studies[130,148,150,182] to reaffirm the highly anthropophilic nature of *An. coluzzii*. In fact, the anthropophilic tendency was even higher in this study than in previous ones. While the human and bovine odors appeared equally pungent to

me, it is possible that the human odor was in fact stronger from the mosquitoes' perspective. Another possibility for this especially strong anthropophily could be their long history in laboratory culture, where for many generations they have been exposed to human kairomones, but not bovine, which could have selected for a higher degree of anthropophily in the cultured strain, particularly as human kairomones (from sweaty socks or breath) are often used to promote bloodfeeding. As such, those mosquitoes that are more likely to bloodfeed in the presence of human kairomones could be overrepresented in subsequent generations.

3.4.2. Activation of Host-Seeking

The studies in the dual-port olfactometer allow us to determine the necessity of the given chemosensory organs (or chemical cues) to activate host-seeking in the presence of human and bovine volatiles, heat, and humidity during the early stages of the dark cycle when anopheline host-seeking is most common. Organs and cues are discussed individually, followed by a summation of the overall pattern.

3.4.2.1. Antennae

When compared to intact mosquitoes, fewer than half of the mosquitoes without antennae entered the olfactometer, suggesting that while activation of host-seeking can occur without this organ, it plays a critical role. Further evidence for this conclusion comes from the fact that the mosquitoes without antennae were also significantly less active than all other experimental groups, excluding those without labella, which were similarly lowly activated but had by far the highest mortality. Like mosquitoes with ablated labella, mosquitoes with ablated antennae had higher mortality than the other

ablation and intact groups, but these differences were not significant after correcting for multiple tests. As such, poor physiological condition (which is presumably correlated with mortality rate) is not a sufficient explanation for the significant reduction in activation rate observed in these mosquitoes, particularly as mosquitoes with maxillary palps ablated, an injury of similar degree, were highly activated. Rather, the antennae play a critical role in vertebrate host odor detection, as has been previously demonstrated[100–102].

The critical role of the antennae in responding to host stimuli from a distance is expected, given the density of chemosensory sensilla, expression of chemosensory receptor genes and odorant binding proteins, and evidence of a role in detecting host heat signatures and humidity as well as odor[37,45–47,58,104,105,152,191,198,248]. Previous studies on antennectomized *Aedes aegypti* and the North American malaria vector *An. quadrimaculatus* showed an inability to locate hosts at closer range[100,102], which is consistent with my results. A study on the European malaria vector *An. atroparvus* also showed a reduction in response to host odors with antennal ablation, as well as reduced temperature sensitivity[103], which could also account for the reduced activation of the *An. coluzzii* mosquitoes in this study. While the proboscis plays a major role as a "thermo-antenna" in *An. stephensi,* where mosquitoes with antennae ablated can still respond to a hot plate[10], the heat in my olfactometer is much more diffuse, so it is conceivable that the loss of the thermosensory abilities of the antennae in combination with the loss of their olfactory abilities made it particularly difficult for these mosquitoes to localize the "host".

Furthermore, in *Orco* mutant *Aedes aegypti* and *Aedes albopictus*, strong attraction to vertebrate hosts is retained in the presence of CO_2, although anthropophily is lost[249,250]. This result indicates that the elimination of olfactory receptors on the antennae should not completely eliminate host responses, due to the presence of ionotropic receptors not dependent on *Orco*, whose ligands are amines and carboxylic acids and include various host kairomones[251]. However, the loss of all chemoreceptors on the antennae, including both olfactory and ionotropic, as well as the subset of gustatory receptors expressed there, would likely have a much more profound impact.

3.4.2.2. Maxillary Palps

In contrast to mosquitoes with ablated antennae, mosquitoes with ablated maxillary palps did not significantly differ in activation rate from intact mosquitoes. The maxillary palps are best understood as the site of CO_2 perception, but they also express other chemosensory receptors and odorant binding proteins, and are responsive to a variety of vertebrate and plant volatiles[46,48,58,91,92,105,107,152,159,236,240,252]. The lack of impact of maxillary palp ablation on host-seeking activation suggests two main possibilities. The first possibility is that the functions of the maxillary palps are sufficiently redundant with those of the antennae and proboscis that these organs are sufficient to respond to host kairomones in their absence. The second possibility is that the small amount of maxillary palp tissue left attached to the head (Figure 3.1) was sufficient to perceive CO_2 and/or other host kairomones and thereby activate host-seeking. However, given that the vast majority of the maxillary palps were ablated, it is highly unlikely that the function of this small piece would be as strong as in intact

mosquitoes. While no trials were run on mosquitoes with palps ablated in the absence of CO_2 to establish the total abolition of CO_2 perception, the trials run on intact mosquitoes in its absence provide the best available comparison to determine whether CO_2 perception is strictly necessary to activate host-seeking in *An. coluzzii*. If the dominant role of the maxillary palps is indeed CO_2 detection, my results suggest that CO_2 is not in fact a necessary activator in this species.

3.4.2.3. Carbon Dioxide

As expected, host-seeking activation was significantly reduced in the mosquitoes which were not exposed to CO_2. However, it is interesting to note that while mosquitoes without access to CO_2 were less activated than the intact mosquitoes with CO_2 present, they were still more activated than those without antennae. CO_2 has been determined to be a host-seeking activator in anophelines as well as the anthropophilic *Aedes aegypti*[128,135]; however, our data suggests that while its presence undoubtedly enhances the activation of host-seeking at this range, it is not strictly necessary, and other host odorants are sufficient to attract host-seeking *An. coluzzii*. A synergistic effect between CO_2 and other host volatiles has been reported in *Aedes aegypti*, where mutants without CO_2 receptors exhibited reduced but not abolished host-seeking behavior[253]. However, it should be noted that while no additional CO_2 was added during these experiments and had not been added to the room for at least 24 hours prior to the experiment, there may have been residual CO_2 left over from the experimenter. In addition, the air flow in the experimental room is intentionally poor to maintain consistent conditions and prevent

mosquitoes from escaping, which may also have resulted in higher levels of ambient CO_2 than those present outside the room.

Activation of host-seeking did not differ significantly between mosquitoes with ablated maxillary palps and those who were not exposed to CO_2, suggesting that at least at this range, the other chemosensory organs are sufficient. These results also suggest that CO_2 is not strictly necessary to activate host-seeking in *An. coluzzii*, which aligns with previous studies showing that it is more important in the host seeking of zoophilic species[18]. While a functional role of the remaining attached segments of the maxillary palps cannot be ruled out, the lack of a significant difference between mosquitoes with ablated maxillary palps and those not exposed to CO_2 suggests that the reduction in activation among the maxillary palp group most likely corresponds to a loss in CO_2 perception. Even if the remaining maxillary palp pieces were functional, it is highly likely that these mosquitoes perceived much less CO_2 than intact ones.

3.4.2.4. Labella

Like mosquitoes with ablated antennae, mosquitoes with ablated labella were much less likely than intact mosquitoes or the other experimental groups to activate host-seeking behavior. The poor condition of mosquitoes with ablated labella is unsurprising, given their impeded sugar-feeding and hydration. Indeed, the mosquitoes with ablated labella had nearly double the mortality rate of mosquitoes with ablated antennae, and quadruple that of the other treatment groups. Consequently, it is quite likely that their lowered activation rate stemmed largely from a poor physiological state rather than an inability to detect the necessary host cues. Further evidence for this hypothesis comes

from the relatively high response rates of the mosquitoes with ablated labella which were activated.

3.4.2.5. Hind Leg

The activation rate of mosquitoes with a hind leg removed was not significantly reduced from that of intact mosquitoes, indicating that injury effect alone was not sufficient to inhibit host-seeking. There was also no significant difference between these mosquitoes and those with ablated maxillary palps, which in combination with the similarity to mosquitoes with no CO_2 exposure, suggests that at least at this range, the maxillary palps and CO_2 are both dispensable. Furthermore, the dramatic reduction in activation rate when ablating the antennae combined with the consistent activation rates among the other experimental groups (excluding those with ablated labella) points to the primacy of the antennae. Specifically, the antennae are the most indispensable chemosensory organ for the activation of host-seeking at the two-meter range, and are sufficient to initiate host-seeking except where mosquitoes are severely dehydrated/starved (i.e. those with ablated labella).

3.4.3. Vertebrate Host Odor Response

In addition to recognizing the presence of any host cues, the mosquitoes also need to be able to distinguish between odors sufficiently to make a choice between the human or bovine odor port. While recognizing generic mammalian volatiles and distinguishing between human and bovine odors are undoubtedly related abilities, the impact of each modification was not necessarily identical on both abilities. That is, host-seeking mosquitoes may have been able to detect mammalian cues, but not determine if

they were an acceptable concentration; alternatively, they may have been able to detect some human or bovine kairomones but not the full suite.

3.4.3.1. Antennae

As zero mosquitoes with ablated antennae responded to any odor, despite 37% of them entering the olfactometer, it is clear that the antennae play an indispensable role in the response behavior as well as their critical role in activation. There are four possible explanations for why some mosquitoes with ablated antennae were activated but not responsive: either they were able to detect some host odors but not enough of the other host odors to locate the odor ports, they lacked sufficient olfactory signals to make a decision between the two odors, other organs may be sufficient to activate host-seeking but not to respond to a host odor without input from the receptors on the antennae (which would make sense in the context of thermal attraction driven by the proboscis but lacking sufficient olfactory cues to actually respond), or activation could be somewhat random and not entirely driven by chemosensation. Experiments run in the absence of any host cues could determine the extent to which chemotaxis drives olfactometer entry.

Injury effect alone is not sufficient to explain this discrepancy, as all the other modified mosquitoes, including those with ablated labella, were still able to enter the human odor port. Video tracking of the mosquitoes in the olfactometer might clarify the process of odor evaluation and point to the specific role of the antennae.

3.4.3.2. Maxillary Palps

Although mosquitoes with ablated maxillary palps did not differ significantly from intact mosquitoes in activation rates, they were much less likely to respond to a

host odor, both when looking at total numbers and when looking exclusively at active mosquitoes. However, given that their response rates did not significantly differ from that of leg-ablated mosquitoes in either case, it is impossible to rule out injury effect as an explanation for this reduction. Similarly, their response rate did not significantly differ from that of the mosquitoes that were not exposed to CO_2, so I was unable to determine a role of the maxillary palps in host-seeking behavior. However, several studies have demonstrated that the maxillary palps express a variety of chemosensory genes and respond to a variety of different plant and vertebrate volatiles, as well as that they play mechanosensory and thermosensory roles[46,58,91,152,159,240]. This discrepancy could indicate that the maxillary palps play a complementary role to the antennae rather than an independent one. Alternatively, they may matter more for assessment of host quality once contact has been made rather than at a distance, though their known sensitivity to volatile odors would seem to suggest the latter. It is also possible that their loss would have a more significant effect if the ablated mosquitoes were exposed to real hosts rather than merely heated and humidified host odors.

3.4.3.3. Carbon Dioxide

As mentioned, mosquitoes without CO_2 exposure were significantly less likely than intact mosquitoes to respond to human host odor. While this reduction in odor response was not significantly different from any of the other experimental groups when comparing all live mosquitoes, it was significant when comparing only the response rates of activated mosquitoes, indicating a potential role for CO_2 between activation of host-seeking and landing on a host. As mentioned above, the response rate of mosquitoes

without CO_2 exposure did not differ significantly from that of mosquitoes with ablated palps when comparing both all live mosquitoes and only activated mosquitoes. In addition, activated mosquitoes without CO_2 exposure were significantly less likely than those with ablated labella or an ablated hind leg to respond to host odor. This difference is likely due to the synergism between CO_2 and other host kairomones[253].

3.4.3.4. Labella

In sharp contrast to their significantly lowered activation rate, the response rate of mosquitoes with ablated labella did not significantly differ from any experimental group other than the completely non-responsive mosquitoes with ablated antennae. This result suggests that the removal of the labella did not have any more significant effect than a generic injury effect. Further evidence for this hypothesis comes from the response rates of only activated mosquitoes, wherein the mosquitoes with ablated labella did not differ significantly from the intact mosquitoes, or those with an ablated hind leg.

While my data suggest that many of them may have been in too poor a physiological state to enter the olfactometer, those that did were generally as capable of responding to human host odor as other mosquitoes. Therefore, while there is clearly a variety of chemosensory activity localized in the labella (Chapter 2), this organ does not appear necessary for odor-mediated host localization. The study that established the proboscis as a "thermo-antenna" focused on the entire proboscis[10], so it is possible that the rest of the organ was sufficient to maintain thermotaxis while the antennae were sufficient to maintain chemotaxis. Similarly, the role of the proboscis in oviposition and

bloodfeeding appears to reside in the labrum rather than in the labella[54]. As such, the behavioral importance of labellar olfactory abilities remains elusive.

3.4.3.5. Hind Leg

When comparing the total number of responsive mosquitoes, there was no significant reduction in response rates from intact mosquitoes for mosquitoes with a hind leg removed. However, when comparing the proportion of activated mosquitoes that responded to host odor, there was a significant reduction. Nonetheless, this reduction was relatively small compared to the mosquitoes with ablated palps and antennae (who did not respond at all). As such, there is likely an injury effect which interferes with host-seeking in activated mosquitoes. However, it is difficult to explain why this effect was not observed in mosquitoes with ablated labella, although the sample size in that case was much smaller (292 vs. 181 activated mosquitoes, respectively). One possibility could be the potential thermosensory role of the hind legs proposed by Roth (1951)[100]. However, there was no significant difference in the ability of *An. stephensi* with ablated hind legs to bloodfeed on an anesthetized mouse host[10], so if such a role exists it is likely minimal. Regardless, while it appears likely that a reduction in odor and heat perception occurred in the mosquitoes with maxillary palps ablated, it is impossible to completely rule out injury effect as an explanation for their reduced odor response, just as it was impossible to completely rule out for their reduced activation rates.

3.4.4. Bloodfeeding

While the olfactometer studies approximate host-seeking from a distance, the blood-feeding studies examine the ability of modified mosquitoes to detect some host

cues from close range, as well as their proclivity to feed when injured. The artificial membrane feeder lacks host kairomones, but instead relies primarily on heat, although the mosquitoes may also be able to detect some chemical cues from the blood itself or increased humidity associated with liquid food.

In this study, all modified mosquitoes were significantly less likely than intact mosquitoes to blood-feed, and mosquitoes with their antennae ablated were the least likely to blood-feed. The antennae have been established as a major site of both thermoreception and hygroreception[100,102–104,124]. As such, the dramatic reduction in blood-feeding in mosquitoes with ablated antennae likely stems in large part from a severely reduced ability to detect the heat and the humidity that would normally drive them to feed. However, mosquitoes with a hind leg removed also fed significantly less than intact mosquitoes, as did those with maxillary palps ablated. Therefore, an injury effect likely also played a role in reducing feeding behavior. *An. stephensi* with hind legs ablated did not feed less on an anesthetized mouse than intact *An. stephensi*[10], so it is possible that this injury effect would have been less apparent on a host rather than the artificial membrane feeder.

Nonetheless, a previous study on *Aedes aegypti* and *An. quadrimaculatus* established that mosquitoes with ablated antennae, maxillary palps, and even proboscis, still probed a host, even if they were physically unable to feed[100]. While antennectomized *Aedes aegypti* in that study only fed if they were physically placed upon a human hand and generally failed to respond to a heated thermos, antennectomized *An. quadrimaculatus* readily probed a hand, heated thermos, or even a

warm current of air[100]. However, a major difference between Roth's study and my own is the presence of the experimenter, meaning that human odor cues were present at all times, which likely acted synergistically with thermal cues to induce probing. Lacking those cues, the *An. coluzzii* in this study were much less likely to blood-feed without antennae. Nonetheless, it is possible that probing still occurred even though feeding did not, as the trials were not captured on video.

As the mosquitoes with ablated maxillary palps did not significantly differ from those with a hind leg removed, I conclude that the maxillary palps are not necessary to induce this behavior. That is, the reduction in blood-feeding among mosquitoes with ablated maxillary palps most likely stemmed from injury rather than an inability to locate the membrane feeder. This result is consistent with a previous study on *Aedes aegypti*, which showed that the amputation of the maxillary palps did not significantly reduce attraction to either a hand or a warm flask[102]. However, in *An. stephensi*, mosquitoes with ablated palps were also significantly less likely than intact mosquitoes to feed on an anesthetized mouse[10], so it is possible that the reduction in feeding among the *An. coluzzii* with maxillary palps ablated could have a basis beyond injury effects.

3.5. Conclusion

Other studies have addressed the question of which organs lead mosquito host-seeking[10,100,102,106], and their role in host preference has also been addressed[101], but these studies have largely occurred at close range, with a limited number of host cues presented to the mosquitoes, and/or with live hosts where it is difficult to dissociate the roles of chemical, thermal, and visual cues. In addition, studies have focused either on

Aedes aegypti or less anthropophilic anophelines. *Aedes aegypti* and *An. coluzzii* are not closely related, have very different life histories and daily habits, and the maxillary palps of culicines and anophelines are different shapes and sizes from one another. As such, it does not necessarily follow that the anatomy underlying host-seeking and host preference in *Aedes aegypti* will match that in *Anopheles*.

While *An. atroparvus*, *An. quadrimaculatus* and *An. stephensi* are all competent malaria vectors which readily feed on humans, all are much more zoophilic than *An. coluzzii*[13,22,23]. Given that CO_2 is known to play a larger role in the host-seeking of zoophilic members of the *An. gambiae* complex than the anthropophilic species[18], and that anthropophilic mosquitoes may be more sensitive to specific human kairomones rather than generic mammalian cues, the organs which host the receptors for these chemicals may matter more in one species than the other. As such, this is the first study to assess the anatomy underlying host-seeking in a highly anthropophilic anopheline.

Unfortunately, the anatomy controlling vertebrate host preference is still unidentified, and the exact host-seeking functions of the maxillary palps and labella are unclear. Nonetheless, I have determined the impact of ablating the major chemosensory organs most likely involved with host-seeking in some way on the host-seeking process. Specifically, I looked at the impact of these ablations on activation of host-seeking and response to a host odor after initiation of host-seeking. Furthermore, I analyzed the importance of the antennae and maxillary palps in bloodfeeding.

The antennae are critical for activation of host-seeking, response to a vertebrate host odor, as well as feeding on an artificial membrane feeder in the African malaria

vector *An. coluzzii*. They are likely indispensable for these behaviors due to their dominant roles in olfaction, thermosensation, and hygroreception. While the other mosquito sensory organs, the maxillary palps and labella, also are known to play roles in all three senses[10,54,95], they do not seem to be able to compensate for the loss of the antennae, indicating that they likely play complementary, synergistic, and/or redundant roles with the antennae to locate host odor. However, it is also possible that their sensory roles are more important in other behavioral arenas, such as mating or oviposition, where the labrum of the proboscis has been shown to be critically important[54].

While this study allows us to draw some conclusions about the relative importance of the different *An. coluzzii* olfactory organs to host-seeking and blood-feeding behavior, additional studies would further clarify their role. Most obviously, in this study, only intact mosquitoes were deprived of CO_2. Similarly depriving the other experimental groups of CO_2 would help clarify the precise roles of these organs by decoupling them from this best-known host volatile, as well as the inherent necessity of this chemical cue in combination with other host kairomones.

In addition, while these results indicate the importance of these organs in terms of general activation and responsiveness, I was not able to answer the question of which organs contribute to host preference, since the mosquitoes have the third choice of not responding at all, instead of choosing one or the other host odor, as occurred with all of the activated mosquitoes with ablated antennae.

While our dual-port olfactometer provides a closer proxy for the actual host-seeking environment, a simpler Y-tube assay could better elucidate the mechanisms of

host-preference by forcing the mosquitoes to make a choice. Lastly, while many cues are provided to mimic a host, a similar study with the actual vertebrate hosts would provide the full set of cues, and thereby better indicate the importance of each organ.

4. MOLECULAR EVOLUTION OF GUSTATORY RECEPTOR GENES IN THE *ANOPHELES GAMBIAE* COMPLEX

4.1. Introduction

Mosquitoes in the *Anopheles* (*An.*) *gambiae* complex are the major vectors of malaria in sub-Saharan Africa, and consequently are responsible for the deaths of hundreds of thousands of children every year. One reason for this high mortality is that two species in this complex (*An. coluzzii* and *An. gambiae* s.s.) are highly anthropophilic and feed almost exclusively on human hosts. As such, understanding the genetic basis of human host preference is critically important to the possible development of transgenic mosquitoes. The chemosensory genes are a promising candidate for elucidating this basis, as there is clear evidence that the organs where they are highly expressed control host preference. Furthermore, there is significant species and sex-biased gene expression in these gene families, with some of these genes known to detect human kairomones.

4.1.1. Insect Chemosensory Genes

Insects have four major chemosensory gene families: the gustatory receptors (*Grs*), ionotropic receptors (*Irs*), odorant binding proteins (*Obps*), and olfactory receptors (*Ors*). Several other gene families also play minor roles in insect chemosensation, including chemosensory proteins (*Csp*s), sensory neuron membrane proteins (SNMPs), Pickpockets, TRP channels, opsins, and CheA/CheBs. *Grs* and *Ors* have an inverted topology from that of G-protein coupled receptors, which form the basis of vertebrate chemosensory systems[69,78]. ORs are heterodimers comprised of the obligate olfactory co-receptor (ORCO) and one specific OR[68], while both GRs and IRs

can be multimeric. There are three IR co-receptors which may be expressed with up to two other IRs, and sometimes GRs[71,72]. By contrast, GRs do not typically rely on co-receptors but are co-expressed with multiple other GRs[56]. Gustatory sensilla usually contain three or four taste neurons, although some contain only two[85,90]. A single gustatory receptor neuron may contain between two and five co-expressed GRs[72]. Finally, OBPs are thought to act as binding proteins carrying hydrophobic odorant molecules across the sensillar lymph[254], although recent work has challenged their widespread necessity in mediating olfactory responses[93].

While *Ors* and *Obps* are restricted to hexapods, *Irs* are present throughout Protostomia, and *Grs* are even more ancient, present in such basal animals as placozoans, though they are absent from ctenophores, sponges, and choanoflagellates[75,81]. The *Grs* are the basal member of a larger insect chemosensory receptor superfamily that also includes the *Ors*, which are derived from within the *Grs*[79,80]. While *Irs* are largely conserved across insect orders[76,77], *Grs* and *Ors* are characterized by large numbers of lineage-specific expansions throughout Arthropoda[75,82,83]. The *Gr* repertoire tends to be smallest in specialists and largest in generalists[255], with extremes exemplified by the human body louse, *Pediculus humanus*, with only six *Grs*[256], and the German cockroach, *Blattella germanica*, with 431[257].

Gr repertoires can also vary substantially within orders and even genera. Within the Brachycera suborder of flies, *Gr* repertoires vary widely, from 14 in the obligate bloodfeeder, *Glossina morsitans*[258,259] to 76 in the generalist, *Musca domestica*[260]. Examples of pseudogenizations and expansions abound in *Drosophila* (reviewed in

Robertson (2019)[255] and discussed in greater detail below), while the *Gr* repertoires of two hematophagous fly genera, *Glossina* and *Anopheles* are much more stable[190,258].

4.1.2. Function and Evolution of Gustatory Receptors

Grs encode receptors which mediate perception of a wide variety of chemical cues. These receptors are found on sensilla with a single distal pore, which facilitates the perception of concentrated tastants rather than airborne volatile chemicals[261]. However, three *Grs*(*AgGrs22-24*) together encode the carbon dioxide receptor in mosquitoes[91], allowing them to perceive a volatile cue that plays a critical function to varying degrees in different mosquito genera and species[140,141]. The *Gr* CO_2 receptors are highly conserved in *Drosophila*, *Aedes*, *Anopheles*, *Culex*, *Bombyx*, and *Tribolium*, but appear absent from Hymenoptera, Phthiraptera, Hemiptera, *Daphnia*, and *Ixodes*, despite the role of CO_2 in *Ixodes* questing behavior[262].

Other *Grs* respond to sugars, salts, pheromones, and bitter compounds in complex ways, with different combinations of *Grs* eliciting specific responses to particular concentrations of chemicals, and sometimes inhibiting one another[53,84–90].

Grs are mostly expressed in the well-known taste organs of the labellum and tarsi, as well as the wings of *Drosophila*[263]. However, there is also evidence of *Grs* playing important roles in the brain, such as *DmGr43a*, which acts as an internal fructose receptor dictating feeding behavior[52,87]. *Grs*, including both CO_2 and sugar receptors, are also expressed in olfactory neurons[87,92,264,265], and other *Grs* are expressed in the digestive tract[52,266] and non-chemosensory neurons[267], in which they can fulfill non-gustatory roles, such as avoidance of heat and light[199,200].

Although many *Grs* have been deorphanized in *Drosophila*, the vast majority of anopheline *Grs* have unknown ligands. The functional characterization of *Grs* has been more difficult than for other receptors, due to the inability to heterologously express them[268]. As such, while the carbon dioxide receptors and some sugar receptors have been identified in *Anopheles*, the biological function of most anopheline *Grs* remains unclear.

The evolution of the insect chemosensory gene superfamily has been extensively studied both broadly within the arthropods/insects as well as in great detail in some taxa. The evolution of these genes both in insects and in chelicerates tends to follow a birth-and-death model with numerous lineage-specific duplications, followed by pseudogenization and eventual gene loss[82,173]. While *Ors* are derived from *Grs*, the evolutionary dynamics of these two gene families are not identical: *Grs* have increased replacement divergence (i.e. fixed amino acid changes between species) relative to *Ors*, as well as lower neutrality indices (suggestive of positive selection, as discussed below), which could stem from either stronger positive selection (the accumulation of fixed amino acid changes) in *Grs* than *Ors* or from weaker purifying selection (the purging of mutations causing amino acid changes) in *Grs* than *Ors*[43,269]. *Grs* likely underwent gene duplications followed by subsequent differentiation during the process of speciation, and exhibit minimal sequence similarities to one another both within and between species[75,79].

Positive selection potentially associated with host shifts leading to speciation has been detected in the *Grs* in several other insect taxa. There is evidence of positive selection (although relaxed purifying selection cannot be ruled out) in *Grs* of the

specialist *Drosophila sechellia*, which may have facilitated its ecological shift[175]. In addition, McBride & Arguello (2007)[43] identified evidence of positive selection in the *Grs* of five *Drosophila* species, two specialists and three generalists. In this study, they also showed that the specialists *D. sechellia* and *D. erecta* are losing bitter receptor-encoding *Grs* significantly faster than their generalist congeners[43]. There is also evidence of positive selection acting in concert with differential gene expression in one *Gr* and three *Ors* to facilitate both the multiple rapid specializations of *D. mojavensis* and adaptation to multiple different hosts of *D. arizonae*, its sister species[270]. A similar pattern is observed in the butterfly *Heliconius melpomene*, where female oviposition behavior appears to drive both the expression and evolution of *Grs*, particularly those which are differentially expressed on female forelegs[271].

In the pea aphid, *Acyrthosiphon pisum*, most *Grs* and *Ors* are the products of rapid and recent duplications, and are not found in other lineages. Furthermore, many of these recently duplicated genes have evolved under positive selection, perhaps facilitating ecological adaptations[272]. In the nymphalid butterfly *Vanessa cardui*, frequent *Gr* duplications occurred in the transition from a specialist to a generalist lifestyle[273].

The plasticity of *Grs* in responding to environmental change is embodied by an adaptive change, thought to be caused by one or more mutations, in a population of German cockroaches, *Blattella germanica*, which has caused them to be glucose-averse and thus avoid bait stations. Specifically, a bitter gustatory neuron has become modified such that it will accept glucose rather than its original ligand[274].

Further evidence for the role of *Grs* in modulating species-specific behaviors comes from a study on *DmGr64e*, which showed that flies from eight different *Drosophila* species with functional copies of the gene are highly responsive to glycerol, while *D. pseudoobscura* and *D. persimilis*, which have pseudogenized copies, are much less responsive to glycerol[275]. Another example is found in the butterfly *Papilio xuthus*, which relies on *PxutGr1* to detect synephrine, an oviposition stimulant found in its limited range of host plants[276]. While these examples are atypical in the sense that *Grs* generally co-express with multiple other *Grs* to perceive and respond to tastants, they nonetheless show that divergent behaviors can be determined by changes in *Gr* sequence.

However, positive selection on the whole is thought to play a relatively minor role in the evolution of *Grs*, with strong purifying selection acting as the dominant evolutionary force in *Drosophila*[51,82]. When Gardiner *et al.* (2009)[269] applied a more stringent model to the chemosensory genes of 12 *Drosophila* species, they identified only six significantly diversifying genes (four *Ors* and two *Grs*). Furthermore, most diversification, particularly in the *Grs*, appears to be due to relaxed purifying selection, rather than positive selection[269]. Purifying selection also played a major role in the evolution of *Grs* and *Ors* in Lepidoptera, although there are also some genes under positive selection, which in combination with gene family expansion, especially in bitter receptors[277], could have facilitated adaptive radiation in response to that of angiosperms[42].

In addition, selection on chemosensory gene expression may be responsible for altering the sensitivity to certain odors, rather than changes in the protein structure resulting in changes in ligand binding affinity[278]. Numerous chemosensory genes are differentially expressed between the anthropophilic *An. coluzzii* and zoophilic *An. quadriannulatus*[45–47], which could reflect either differential sensitivity to odors or changes in ligand binding affinity due to protein structure changes.

There is also some evidence that changes in chemosensory genes influence mosquito host preference. Previous work in *Aedes aegypti* has demonstrated that differential expression as well as nucleotide substitutions in a single *Or* can influence host preference between two closely related subspecies, the zoophilic *Aedes aegypti formosus* and the highly anthropophilic *Aedes aegypti aegypti*[44]. Our lab also previously performed QTL mapping and analyzed the molecular evolution of several *Ors, Irs,* and *Obps* in the *An. gambiae* complex (discussed below)[182].

4.1.3. *Anopheles gambiae* Complex

The *An. gambiae* complex consists of nine cryptic species with varying host preferences and distributions: *An. amharicus*, *An. arabiensis*, *An. bwambae*, *An. coluzzii*, *An. fontenillei*, *An. gambiae* sensu stricto, *An. melas*, *An. merus*, and *An. quadriannulatus*[11,14]. Due to unavailable genomic data, *An. amharicus*, *An. bwambae*, and *An. fontenillei* are excluded from this study.

An. coluzzii and *An. gambiae* s.s. are major malaria vectors due to their levels of anthropophily[130,147,148,150,279–281], a strong preference to feed on a human host. *An. coluzzii* is predominantly found in West Africa, while the distribution of *An. gambiae*

s.s. is more widespread[15]. They were formerly referred to as the M and S molecular forms, respectively, of *An. gambiae* s.s., but have subsequently been split into two separate species[14].

An. arabiensis is also a major malaria vector with anthropophilic tendencies but exhibits more plasticity in host choice between humans and other large mammals than either *An. coluzzii* or *An. gambiae* s.s.[16,18,281,282]. Its range overlaps with that of *An. coluzzii* and *An. gambiae* s.s., but it tends to prefer dryer habitats[13]. *An. arabiensis*'s wider host feeding behaviors have led to its use in multiple studies on the genetic basis of human host preference. In particular, genes within inversions on chromosomes 2R and 3R are associated with differences in vertebrate host preference[283–285]. There is also substantial introgression between *An. arabiensis*, *An. coluzzii*, and *An. gambiae* s.s [286,287], as well as between *An. merus* and *An. quadriannulatus*[288]. This introgression is primarily found in the autosomes, such that autosomal sequences from *An. arabiensis* cluster with the *An. coluzzii* and *An. gambiae* s.s. clade, while X chromosome sequences instead cluster with *An. quadriannulatus*[288]. These autosomal regions are home to the entire suite of *Grs*, as well as all of the QTL chemosensory genes that were previously identified by our lab[182]. Selection maintains the 2La inversion in *An. arabiensis*[289,290], and it is likely that the other genes which have introgressed into *An. arabiensis* are adaptive for its shift to anthropophily[288].

An. melas and *An. merus* are both locally important malaria vectors breeding in brackish coastal waters on the western and eastern coasts of Africa, respectively, reflective of a greater salt tolerance than other species, as well as potentially different

varieties of host plants reflective of this habitat. While both species are less anthropophilic than *An. coluzzii* and *An. gambiae* s.s., they can occur in sufficient abundance to cause significant malaria transmission[13,291–297]. Finally, *An. quadriannulatus* is not considered a malaria vector, due to its propensity to feed primarily on non-human animals, particularly cattle[19,147,298–300].

Part of the biology underlying differential vertebrate host preference among the species of the *An. gambiae* complex is a different level of responsiveness to mammalian host cues, such as carbon dioxide, acetone, lactic acid, and 1-octen-3-ol. 1-octen-3-ol, which is detected by *AgOr8*, works synergistically with carbon dioxide to attract both *An. gambiae* s.s. and *Aedes aegypti* to human hosts, so much so that *An. gambiae* can be made to respond to chicken odor if it is supplemented with a human-like concentration of 1-octen-3-ol[268,301]. High concentrations of carbon dioxide and acetone fail to elicit host-seeking in *An. gambiae* s.s.[129], and CO_2 is also relatively unimportant in the host-seeking behaviors of *An. arabiensis*[18,151]. *An. coluzzii* and *Aedes aegypti*, which is also highly anthropophilic, show relatively little activation of CO_2 receptors at relatively high concentrations, indicating its secondary role to other vertebrate host volatiles in the host-seeking process[140]. CO_2 plays a much larger role in the host-seeking behavior of *An. quadriannulatus*, which shows a high degree of attraction to CO_2 even in the absence of other bovine volatiles[18].

Previous work has focused on differential expression of chemosensory genes in the major mosquito sensory organs between the anthropophilic *An. coluzzii* and zoophilic *An. quadriannulatus*[45–47], as well as the physical ablation of said organs[100–

[103,106]. In addition, our lab previously analyzed the evolution of some *An. gambiae* s.l. *Irs*, *Obps*, and *Ors*. In this study, six host preference QTL were identified that explain nearly half of the phenotypic variation between anthropophilic and non-zoophilic mosquitoes. This study identified 19 *Ors*, 2 *Irs*, and 11 *Obps* within the QTL, five of which also show evidence of positive selection. In addition, the CO_2 receptor *Gr24* was also identified in the QTL[182].

In this study, I survey the molecular evolutionary patterns of the *Grs* in the *An. gambiae* complex in the context of species-specific host-seeking behaviors. Most *Grs* in the *An. gambiae* complex have unknown ligands, with the exception of the CO_2 receptors listed above, the putative sugar receptors *Grs* 14-21[302], and the sugar receptor *Gr25*[303]. I am particularly interested in identifying *Grs* that diverged between the anthropophilic *An. coluzzii* and *An. gambiae* s.s. clade and the zoophilic *An. quadriannulatus*. Furthermore, I describe evidence of positive selection between other lineages, as well as evidence of purifying selection and recovery from selective sweeps. We focus especially on genes which are relatively highly expressed in the chemosensory organs of *An. coluzzii* and *An. quadriannulatus*, for which the biological relevance of these evolutionary patterns is most readily apparent.

4.2. Methods

4.2.1. Data

Whole genome sequences for six members of the *An. gambiae* complex (*An. arabiensis*, *An. coluzzii*, *An. gambiae* s.s., *An. melas*, *An. merus*, and *An. quadriannulatus*) from the 16 Genomes Project[190,304] were downloaded from NCBI. No

comparable data exist for the remaining constituent species of the complex, none of which has been extensively studied. Genome data for a total of 96 *An. arabiensis*, 12 *An. coluzzii*, 26 *An. gambiae*, 65 *An. melas*, 72 *An. merus*, and 72 *An. quadriannulatus* individuals were used for analysis.

4.2.2. Variant Discovery Pipeline

Whole genome sequences were processed according to the GATK 4 Best Practices Workflow for DNA Variant Discovery[211]. Reads were aligned to their respective genomes in BWA[305] except in the case of *An. coluzzii*, which was aligned to the *An. gambiae* genome due to the low quality of the *An. coluzzii* genome and the very close evolutionary relationship between the two species. All genomes were downloaded from VectorBase[306].

Following mapping in BWA, Picard Tools (http://broadinstitute.github.io/picard) was used to add read group information and mark duplicate reads. A bed file was created to mark intervals corresponding to the genomic coordinates of the genes of interest, which were determined by running a local BLAST search[307] using the *An. gambiae* gene sequences. In cases where more than one nonconsecutive match was recovered, regions were prioritized by length, e-value, and bit score.

This interval file was used to restrict the use of HaplotypeCaller to only the genes of interest, thereby increasing processing speed. Haplotype Caller was run using gVCF mode in order to improve the accuracy of physical phasing where possible, and all samples from a given species were processed together through the remaining steps. Indels were excluded from further analysis, as were any reads not meeting the following

quality standards: QD < 2.0 || FS >55.0 || MQ < 40.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0.

A custom bash script incorporating bcftools[308] was used to screen for low coverage sequences by counting the number of variants genotyped in the entire species dataset, and then comparing individual sequences for the presence of those genotypes (whether variant or non-variant). All individuals with an average of at least 50% of variants successfully genotyped across all genes were used for McDonald-Kreitman analysis, while sequences with missing genotypes were excluded from DH test analyses. Bcftools was also used to extract two phased fasta sequences for each individual. BEDTools[309] was used to extract individual genes from each genome sequence. Two data sets were generated using the masked gene data sets: one with full gene sequences including introns and another including only coding sequence (CDS). In the case of genes with multiple splice variants, multiple CDS data sets were generated.

4.2.3. Data Analysis

CDS data sets including all individuals were imported into DnaSP version 6.12.03[310], which was then used to perform a McDonald-Kreitman test[311] between each pair of species to detect signs of positive selection. All sequences were used for this test, as it relies solely on comparisons between polymorphic sites in coding sequences. Different splice variants were treated as unique genes for this test, but not for any others which do not rely on coding sequences only. *Grs* were considered to be undergoing positive selection if a significant p-value was calculated in addition to a neutrality index (NI) < 1, whereas they were considered to be undergoing purifying selection if a

significant p-value was combined with NI > 1. The NI is calculated by comparing the ratios of amino acid changes to synonymous substitutions between two species and within each species. Specifically, it is calculated as $NI = \frac{\frac{pN}{pS}}{\frac{dN}{dS}}$, where dN refers to the number of fixed replacement substitutions between species, dS refers to the number of fixed synonymous substitutions between species, pN refers to the number of polymorphic replacement substitutions within a species, and pS refers to the number of polymorphic synonymous substitutions within a species. DnaSP was also used to generate basic population parameters for each species by the use of a Fu and Li D test[312], from which only the population parameters and not the significance values were considered.

Data sets including the full gene sequence were produced only if at least ten sequences with known genotypes for each variant site existed in a given species. In cases of alternatively spliced genes, the longest possible sequence was used for the DH suite of tests. As a result of these more stringent criteria, these data sets were primarily produced for *An. coluzzii* and *An. gambiae* s.s., which were sequenced at a higher depth of coverage than the other species in this study. These data sets were exported in fasta format with one outgroup sequence. Where possible, data sets were produced with *An. coluzzii* and *An. gambiae* s.s. analyzed both as individual species and as a single clade. These fasta files were loaded in the *Readms* module of DH (available from https://zeng-lab.group.shef.ac.uk/wordpress/?page_id=28), where the following tests were

implemented with 10,000 coalescent simulations: D[313], normalized H[314,315], DH[315], and E[315].

While all of these tests can detect directional selection, DH is unique in its insensitivity to other population genetic forces, and is designed to detect evidence of selective sweeps (illustrated in Figure 4.1). Tajima's D can detect balancing or purifying selection, but is also sensitive to changes in population size. Fay and Wu's H is used primarily to detect genetic hitchhiking, but is also sensitive to reductions in population size and the presence of population structure. The E test identifies the recovery of genetic diversity following its loss (e.g. following a selective sweep) and is robust to population structure, but sensitive to both background selection (illustrated in Figure 4.1) and increases in population size[315].

Finally, TCS haplotype networks were generated in POPART[316].

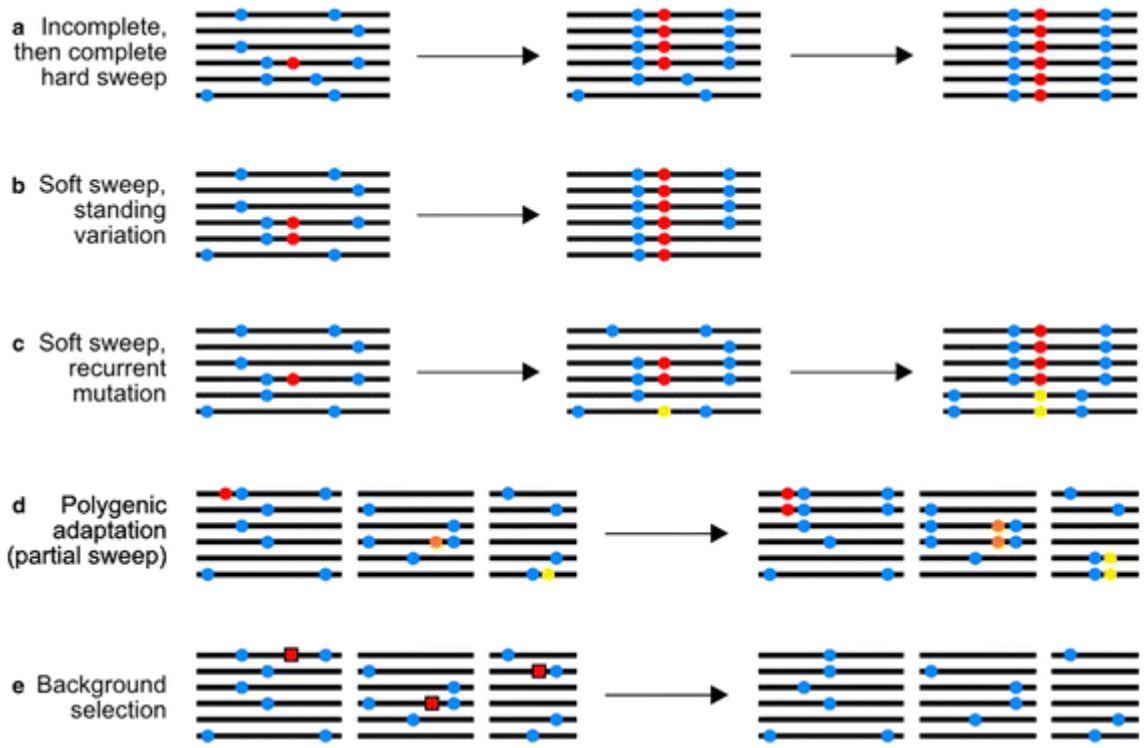

Figure 4.1 – Visual Representation of Selective Sweep/Genetic Hitchhiking and Background Selection. A) a classic or hard sweep wherein a single advantageous mutation goes quickly to fixation and brings linked neutral loci with it. B) a soft sweep wherein the advantageous allele is linked to multiple neutral haplotypes. C) a soft sweep wherein the advantageous allele arises via two separate mutations with different associated neutral haplotypes. D) an incomplete or partial sweep where variation in linked neutral loci has reduced but fixation has not yet occurred. E) background selection wherein both deleterious alleles and their linked neutral alleles are purged. Reproduced from Booker *et al.*, 2017[317] under the terms of a Creative Commons Attribution 4.0 International License.

4.3. Results

Sequences were obtained for all species for *AgGrs1-60* with the following exceptions: a complete *Gr6* sequence was not found in the *An. arabiensis* genome; complete *Gr2*, *Gr10*, *Gr44*, *Gr56*, and *Gr57* sequences were not found in the *An. melas* genome; complete *Gr44*, *Gr49*, *Gr53* and *Gr57* sequences were not found in the *An.*

merus genome. *Gr5*, *Gr9*, and *Gr10* were not analyzed due to a low number of available sequences and excessive premature stop codons.

4.3.1. Fixed Differences Between Species

No fixed differences were identified between *An. coluzzii* and *An. gambiae* in any *Gr* (Table 4.1). Similarly, no fixed differences were found in most *Grs* between either of these species and *An. arabiensis* (53 out of 75 for *An. coluzzii*, and 60 out of 75 for *An. gambiae*). Most *Grs* (between 64 and 71) have fixed differences between these three species and *An. quadriannulatus*. Fixed differences are found at almost every locus between every species pair that includes either *An. melas* or *An. merus*.

Table 4.1 – Total fixed differences (both synonymous and non-synonymous) identified between species pairs across all *Grs* tested with the McDonald-Kreitman test

Species Pair	*Grs* with Fixed Differences	*Grs* without Fixed Differences	Percent *Grs* with Fixed Differences
arabiensis-coluzzii	22	53	29.3%
arabiensis-gambiae	15	60	20.0%
arabiensis-melas	62	0	100%
arabiensis-merus	66	1	98.5%
arabiensis-quadriannulatus	71	4	94.7%
coluzzii-gambiae	0	76	0%
coluzzii-melas	63	0	100%
coluzzii-merus	67	1	98.5%
coluzzii-quadriannulatus	64	12	84.2%
gambiae-melas	63	0	100%
gambiae-merus	67	1	98.5%
gambiae-quadriannulatus	65	11	85.5%
melas-merus	60	0	100%
melas-quadriannulatus	63	0	100%
merus-quadriannulatus	67	0	100%

4.3.2. McDonald-Kreitman Test

With the exclusion of the missing *Grs* listed above, all *Grs* were analyzed using the McDonald-Kreitman (MK) test between every species pair. It should be noted that the multiple testing problem associated with the large number of MK tests performed in my study precludes the conclusive identification of positive selection. While an adjusted p-value produced by the Benjamini-Hochberg procedure is frequently used to reduce the type I error rate, this procedure also substantially decreases power, to the point where no genes were significant after this procedure in a previous study from the Slotman lab[182]. As such, the *Grs* I identify with significant MK tests should be considered to be potentially, but not conclusively, under the influence of selection.

When comparing the anthropophilic *An. arabiensis* and *An. gambiae* s.s. with the zoophilic *An. quadriannulatus*, the majority of *Grs* have a NI < 1, suggesting the action of positive selection. By contrast, when comparing the other highly anthropophilic species, *An. coluzzii*, to *An. quadriannulatus*, the majority of *Grs* have a NI > 1, suggesting the action of purifying selection (Figure 4.2A, Table 4.2).

Table 4.2 – Prevalence of positive and purifying selection based on Neutrality Index between species pairs.

Species Pair	***Grs* with NI < 1**	***Grs* with NI >1**
arabiensis-coluzzii	10	3
arabiensis-gambiae	8	2
arabiensis-melas	22	36
arabiensis-merus	31	32
arabiensis-quadriannulatus	34	23
coluzzii-melas	33	29
coluzzii-merus	25	39
coluzzii-quadriannulatus	24	27
gambiae-melas	36	26
gambiae-merus	25	38
gambiae-quadriannulatus	28	23
melas-merus	12	45
melas-quadriannulatus	28	33
merus-quadriannulatus	30	35

As indicated in Table 4.1, there are few fixed differences between *An. arabiensis* and either *An. coluzzii* or *An. gambiae* s.s. In both comparisons, the *Grs* with fixed differences overwhelmingly show negative NIs consistent with positive selection (Table 4.2, Figure 4.2B).

When comparing *An. melas* and *An. merus* to each other, the vast majority of NI values are consistent with purifying selection, whereas NIs are more evenly distributed when comparing these species to the others (Figure 4.3), although in most cases the majority of *Grs* have a NI > 1, suggestive of purifying selection. However, when comparing both *An. coluzzii* and *An. gambiae* s.s. to *An. melas*, most *Grs* have a NI < 1, suggestive of positive selection (Table 4.2).

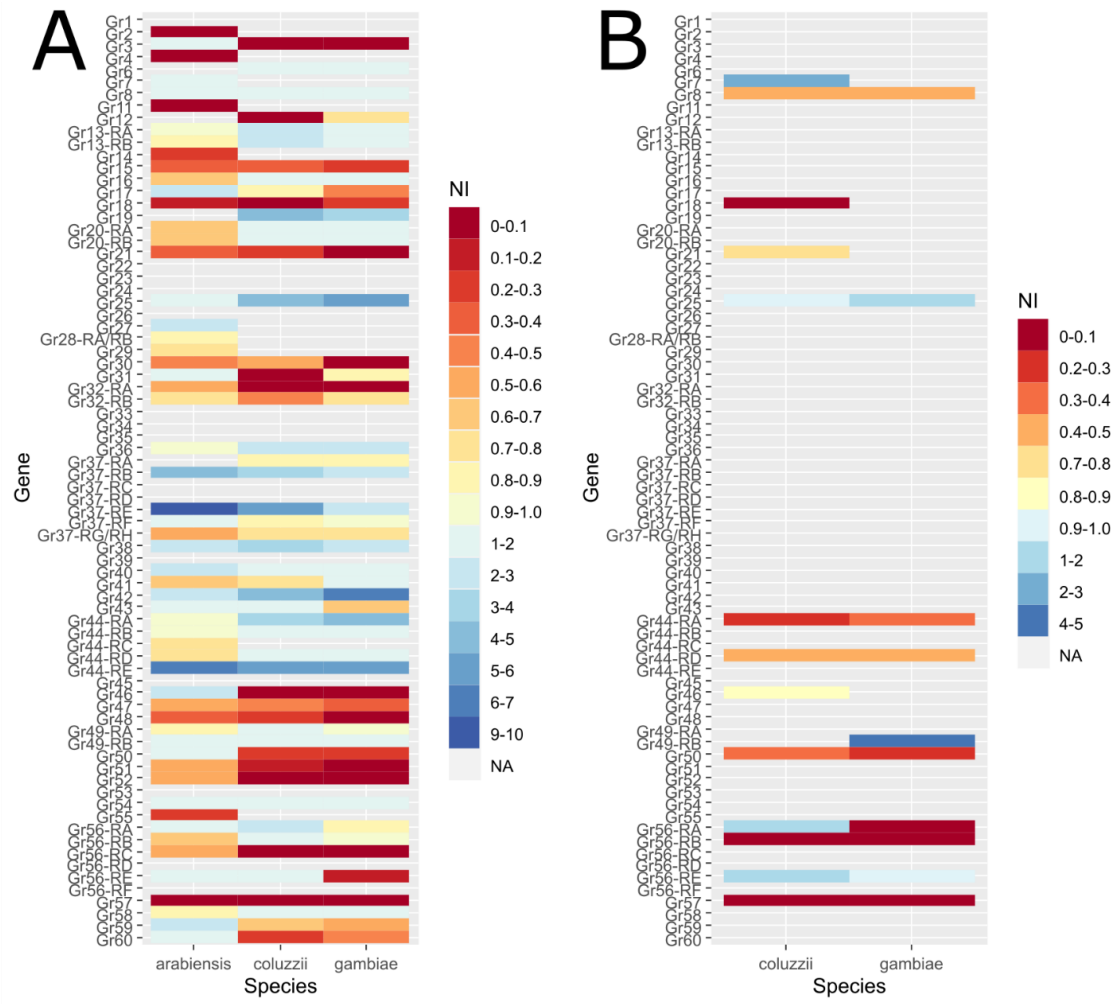

Figure 4.2 – Heat Map showing neutrality indices (NI) generated by McDonald-Kreitman tests between A) the three major vector species and *An. quadriannulatus* and B) *An. arabiensis* and *An. coluzzii* and *An. gambiae* s.s.. *Grs* with a NI < 1 are suggestive of positive selection and are represented in yellow to red shades with increasing selection strength. *Grs* with a NI > 1 are suggestive of purifying selection and are represented in green to blue shades with increasing selection strength. *Grs* where NI could not be computed are represented in grey.

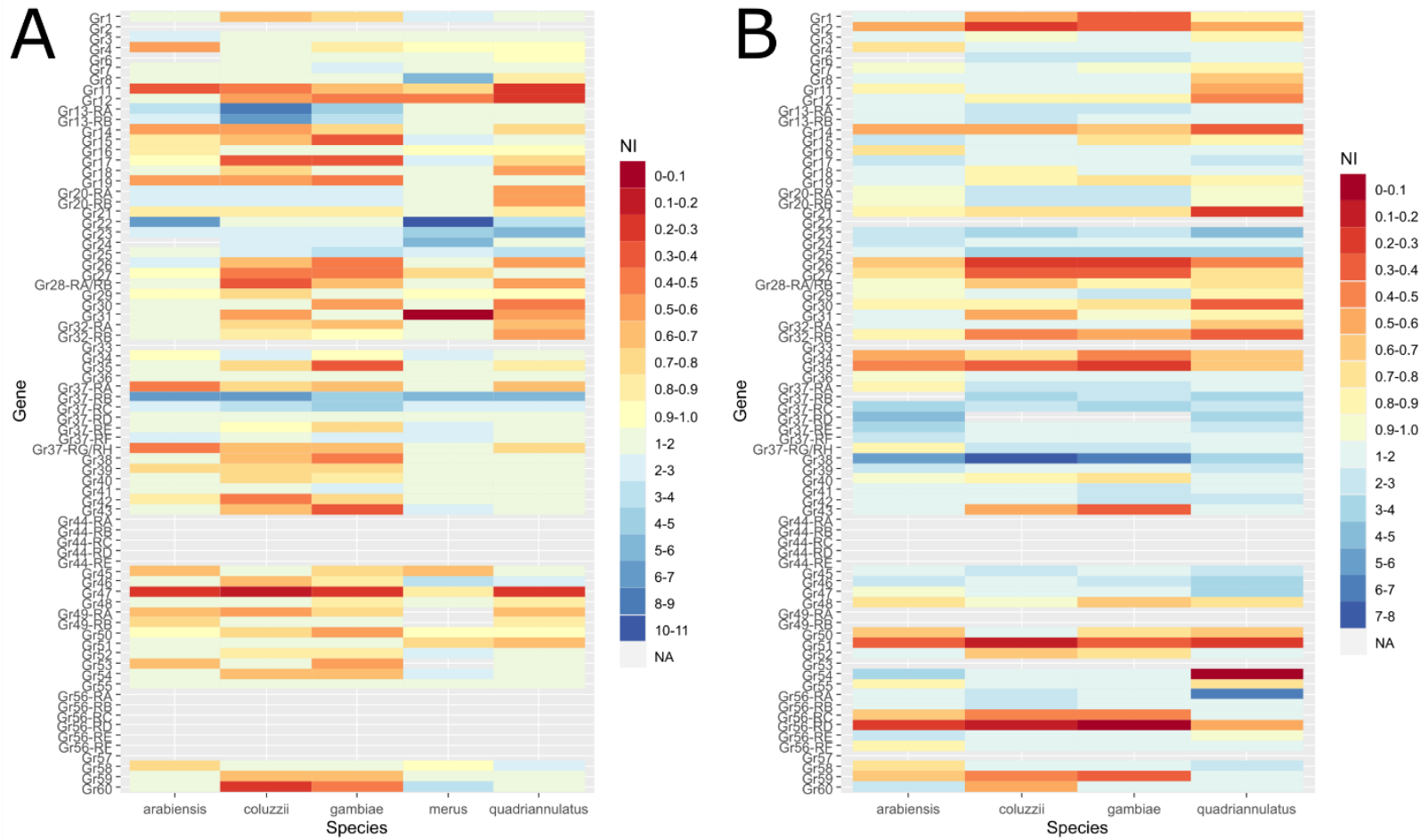

Figure 4.3 – Heat Map showing neutrality indices (NI) generated by McDonald-Kreitman tests between A) *An. melas* and the other complex species and B) *An. merus* and the other complex species. *Grs* with a NI < 1 are suggestive of positive selection and are represented in yellow to red shades with increasing selection strength. *Grs* with a NI > 1 are suggestive of purifying selection and are represented in green to blue shades with increasing selection strength. *Grs* where NI could not be computed are represented in grey.

4.3.2.1. Positive Selection

I am primarily interested in genes showing evidence of positive selection between the anthropophilic *An. coluzzii* and *An. gambiae* s.s. and the zoophilic *An. quadriannulatus*. In one of the 76 *Gr* coding sequences included in the MK test, a significant excess of fixed replacement substitutions was observed between these species pairs: the sugar receptor *Gr18* in the *An. coluzzii – An. quadriannulatus* comparison (Figure 4.4). This excess is defined by a significant p-value (< 0.05) as well as a neutrality index (NI) < 1. Four additional *Grs* showed marginally significant ($0.05 < p < 0.07$) excesses of fixed replacement substitutions: *Gr21, Gr48, Gr50, and Gr60* (Table 4.3).

An. arabiensis is also of interest as a relatively anthropophilic close relative of *An. quadriannulatus* with extensive signatures of introgression with *An. coluzzii* and *An. gambiae* s.s. Two *Grs* showed a significant excess of fixed replacement substitutions between *An. arabiensis* and *An. quadriannulatus*: *Gr4* and *Gr18* (Table 4.3). In addition, *Gr48* showed a marginally significant excess of fixed replacement substitutions.

In addition to the significant excess of fixed replacement substitutions identified in *Gr18* when comparing *An. quadriannulatus* to *An. arabiensis* and *An. coluzzii*, the NI is also low (0.237) between *An. gambiae* s.s. and *An. quadriannulatus*, although the MK test was not significant for this comparison. However, *Gr18* shows evidence of recovery from a selective sweep in *An. gambiae* s.s. (discussed below). In addition, haplotype diversity (HD) for *Gr18* is lower in *An. coluzzii* than in *An. quadriannulatus*, although nucleotide diversity (π) is similar. The TCS networks for *Gr18* in both species show

numerous low-frequency haplotypes with no obvious central high-frequency node, although the TCS network in *An. coluzzii* appears slightly more constrained around a central set of haplotypes. Both networks are slightly star-shaped, but without a central high-frequency haplotype (Figure 4.4). Despite its known ligand, *Gr18* is lowly expressed in both *An. coluzzii* and *An. quadriannulatus* chemosensory tissues, though it is less than 100 bp from the highly expressed *Gr17*.

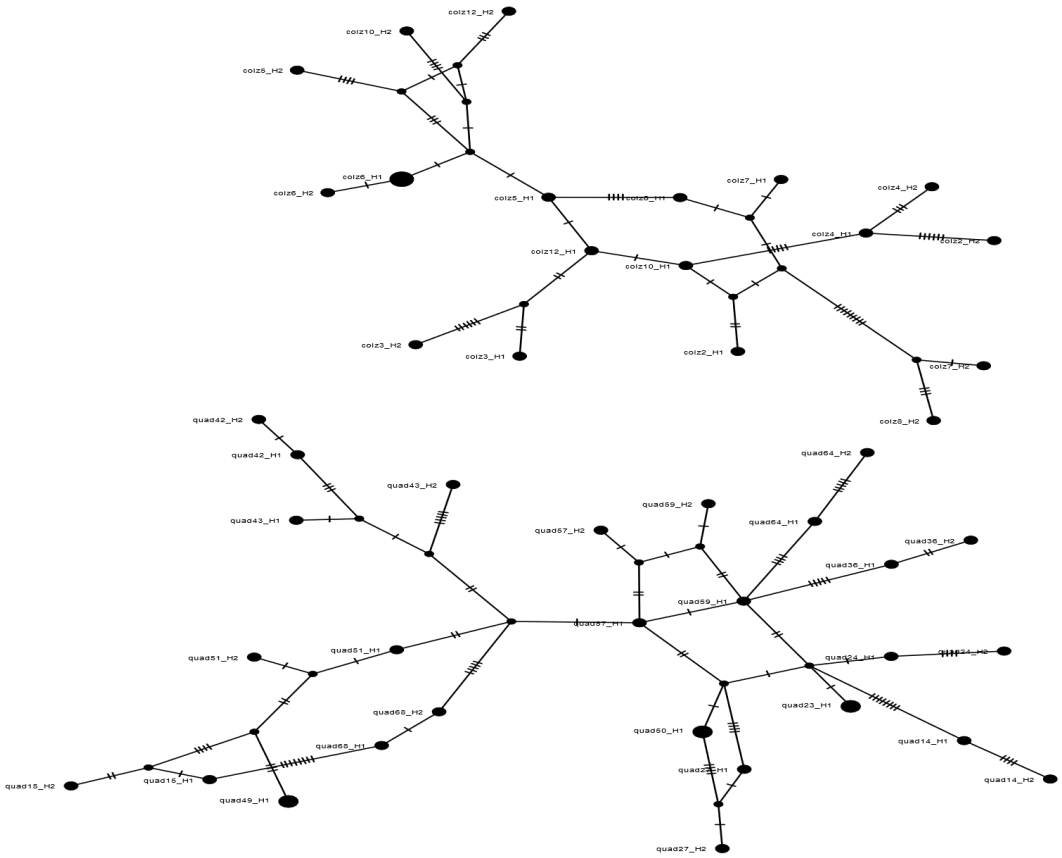

Figure 4.4 – TCS Haplotype Networks of *Gr18* in *An. coluzzii* and *An. quadriannulatus*. Each hashed line represents one nucleotide substitution and haplotype nodes are weighted by frequency. In *An. coluzzii*, there are 18 unique haplotypes spread among 20 sequences (π = 0.005 and HD = 0.984). In *An. quadriannulatus*, there are 27 unique haplotypes among 30 sequences (π = 0.006 and HD = 0.993).

Table 4.3 – *Grs* with significantly (or near-significantly) low neutrality indices (NI), suggestive of positive selection.

Gene	Species Pair	dN	dS	dN/dS	pN	pS	pN/pS	p-value	NI
Gr2	*coluzzii-merus*	8	7	1.14	18	60	0.3	0.027	0.263
Gr4	*arabiensis-quadriannulatus*	5	0	NA	57	67	0.851	0.024	0.000
Gr11	*arabiensis-melas*	16	11	1.45	17	33	0.515	0.053	0.354
Gr11	*melas-quadriannulatus*	15	12	1.25	9	33	0.273	0.005	0.218
Gr12	*melas-quadriannulatus*	18	7	2.57	22	35	0.629	0.008	0.244
Gr18	*arabiensis-quadriannulatus*	6	2	3.00	20	51	0.392	0.014	0.131
Gr18	*coluzzii-quadriannulatus*	5	1	5.00	15	52	0.288	0.005	0.058
Gr21	*gambiae-quadriannulatus*	2	0	NA	32	98	0.327	0.065	0.000
Gr21	*merus-quadriannulatus*	7	8	0.875	12	49	0.245	0.045	0.280
Gr26	*coluzzii-merus*	10	6	1.67	31	84	0.369	0.008	0.221
Gr26	*gambiae-merus*	9	6	1.50	28	88	0.318	0.011	0.212
Gr27	*coluzzii-merus*	15	5	3.00	56	58	0.966	0.050	0.322
Gr35	*gambiae-melas*	9	15	0.6	21	98	0.214	0.051	0.357
Gr35	*gambiae-merus*	5	8	0.625	17	106	0.160	0.037	0.257
Gr47	*coluzzii-melas*	5	3	1.67	17	62	0.274	0.023	0.165
Gr47	*gambiae-melas*	5	3	1.67	32	35	0.914	0.049	0.226
Gr47	*melas-quadriannulatus*	7	7	1.00	12	42	0.286	0.051	0.286
Gr48	*arabiensis-quadriannulatus*	12	6	2.00	26	40	0.650	0.060	0.325
Gr48	*gambiae-quadriannulatus*	3	0	NA	56	95	0.589	0.054	0.000
Gr50	*arabiensis-gambiae*	7	3	2.33	41	73	0.562	0.045	0.241
Gr50	*gambiae-quadriannulatus*	7	4	1.75	42	88	0.477	0.049	0.273
Gr51	*arabiensis-merus*	19	7	2.71	33	36	0.917	0.037	0.338
Gr51	*coluzzii-merus*	16	3	5.33	54	66	0.818	0.002	0.153
Gr51	*merus-quadriannulatus*	11	4	2.75	30	49	0.612	0.021	0.223
Gr56-RD	*arabiensis-merus*	10	5	2.00	17	38	0.447	0.017	0.224
Gr56-RD	*coluzzii-merus*	10	4	2.50	23	57	0.404	0.005	0.161
Gr56-RD	*gambiae-merus*	11	3	3.67	30	83	0.361	0.000	0.099
Gr59	*gambiae-merus*	13	11	1.18	22	49	0.449	0.052	0.380
Gr60	*coluzzii-melas*	13	7	1.86	16	34	0.471	0.016	0.253
Gr60	*coluzzii-quadriannulatus*	8	4	2.00	22	42	0.524	0.053	0.262

Gr50 has a significant excess of fixed replacement substitutions between *An. gambiae* and *An. quadriannulatus*, and a non-significant (p = 0.092) excess between *An. coluzzii* and *An. quadriannulatus*. By contrast, there is also a significant excess of fixed replacement substitutions between *An. gambiae* and *An. arabiensis*, and a non-significant (p = 0.182) excess between *An. arabiensis* and *An. coluzzii*. Furthermore, there is a non-significant (p = 1) lack of fixed replacement substitutions between *An. arabiensis* and *An. quadriannulatus*. *Gr50* has no known ligand and is not detected in chemosensory tissues.

Gr48 has a marginally significant (p = 0.054) excess of fixed replacement substitutions between *An. gambiae* and *An. quadriannulatus*, a nonsignificant excess between *An. coluzzii* and *An. quadriannulatus*, and a marginally significant excess (p = 0.06) between *An. arabiensis* and *An. quadriannulatus*. *Gr48* has no known ligand, but is relatively highly expressed in male *An. quadriannulatus* labella. Furthermore, π and HD are both lower in *An. quadriannulatus* than in *An. gambiae* s.s. The TCS network of *Gr48* in *An. gambiae* s.s. shows almost entirely low-frequency haplotypes, which have no apparent central node. By contrast, while there are also numerous low-frequency haplotypes in *An. quadriannulatus*, there is also a central cluster of haplotypes separated by low numbers of nucleotide substitutions. However, the network is only somewhat star-shaped and has no central high-frequency haplotype (Figure 4.5).

Gr60 has a marginally significant (p = 0.053) excess of fixed replacement substitutions between *An. coluzzii* and *An. quadriannulatus* and a nonsignificant excess between *An. gambiae* and *An. quadriannulatus*, but a nonsignificant lack of fixed

replacement substitutions between *An. arabiensis* and *An. quadriannulatus*. *Gr60* has no known ligand, but is relatively highly expressed in male *An. quadriannulatus* maxillary palps. Like *Gr48*, both π and HD are lower than in *An. coluzzii*. The TCS network of *Gr60* in *An. coluzzii* is devoid of clustering, while that in *An. quadriannulatus* is clustered around a central set of relatively high-frequency haplotypes, but not any one high-frequency haplotype (Figure 4.6).

Finally, *Gr4* has a significant excess of fixed replacement substitutions between *An. arabiensis* and *An. quadriannulatus*, but no fixed replacement substitutions between the latter and either *An. coluzzii* or *An. gambiae*. This gene is lowly expressed in both male and female *An. coluzzii* labella, as well as female *An. quadriannulatus* labella, but is highly expressed in male *An. quadriannulatus* labella. In addition, *Gr4*, along with *Grs* 3 and 5-8, is located relatively close to *ABCG7*, an ATP-binding cassette transporter.

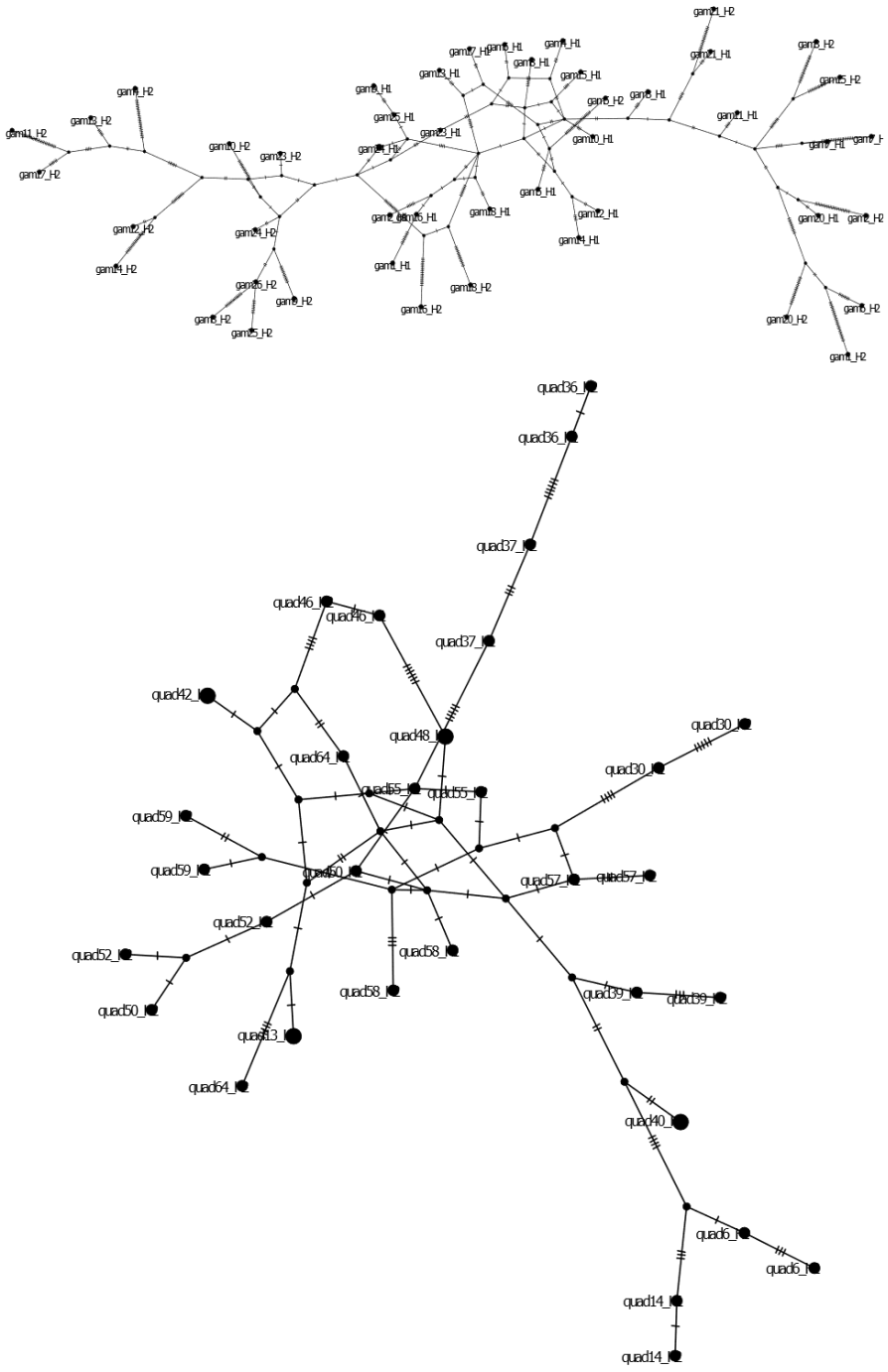

Figure 4.5 – TCS Haplotype Networks of *Gr48* in *An. gambiae* and *An. quadriannulatus*. Each hashed line represents one nucleotide substitution and haplotype nodes are weighted by frequency. In *An. gambiae*, there are 47 unique haplotypes spread among 48 sequences (π = 0.02 and HD = 0.999). In *An. quadriannulatus*, there are 31 unique haplotypes among 36 sequences (π = 0.007 and HD = 0.994).

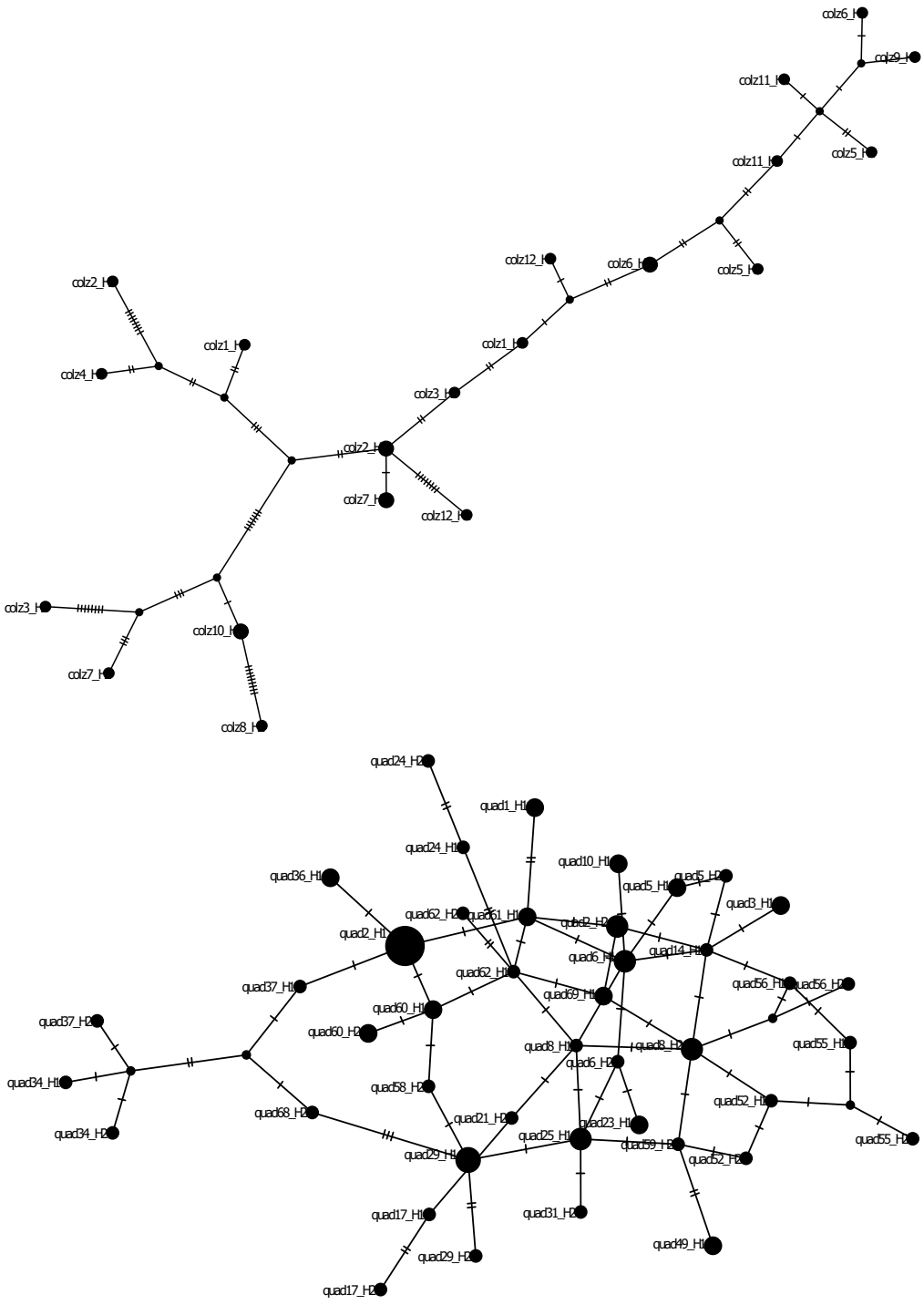

Figure 4.6 – TCS Haplotype Networks of *Gr60* in *An. coluzzii* and *An. quadriannulatus*. Each hashed line represents one nucleotide substitution and haplotype nodes are weighted by frequency. In *An. coluzzii*, there are 20 unique haplotypes spread among 24 sequences (π = 0.009 and HD = 0.986). In *An. quadriannulatus*, there are 43 unique haplotypes among 74 sequences (π = 0.003 and HD = 0.973).

Multiple *Grs* have MK test results and NIs suggestive of positive selection when comparing other species pairs (Table 4.3). Two *Grs* show a significant excess of fixed replacement substitutions between *An. coluzzii* and *An. melas*: *Gr47* and *Gr60*, both of which are lowly expressed in *An. coluzzii* labella. *Gr47* shows the same pattern between *An. gambiae* s.s. and *An. melas*, as well as a marginally significant ($p = 0.051$) excess between *An. melas* and *An. quadriannulatus*.

Gr35, an undetected gene in any examined tissues of either *An. coluzzii* or *An. quadriannulatus*, has a marginally significant ($p = 0.051$) excess of fixed replacement substitutions and an NI suggestive of positive selection when comparing *An. gambiae* and *An. melas* and a significant excess when comparing *An. gambiae* and *An. merus*. *Gr11* and *Gr12* both show MK test results and NIs suggestive of positive selection between *An. melas* and *An. quadriannulatus*. *Gr11* is also marginally significant ($p = 0.053$) between *An. arabiensis* and *An.* melas (Table 4.3). Both *Grs* are expressed at low levels in *An. quadriannulatus* chemosensory organs.

Five *Grs* have significant excesses of fixed replacement substitutions between *An. coluzzii* and *An. merus*: *Gr2*, which is relatively highly expressed in female *An. coluzzii* labella; *Gr26*, which is lowly expressed in *An. coluzzii* chemosensory organs but highly expressed in *An. quadriannulatus* antennae; *Gr27*, *Gr51*, and *Gr56*, all of which are expressed at very low levels (Table 4.3). *Gr56* has six splice isoforms in *An. gambiae* s.s., and signs of positive selection as detected by the MK test and NI are only observed in sequences aligned to the *Gr56-RD* transcript. *Gr56-RD* also shows potential evidence of positive selection between both *An. arabiensis* and *An. gambiae* s.s.

compared to *An. merus*, while *Gr26* shows the same pattern between *An. gambiae* and *An. merus*. *Gr51* also shows this pattern between *An. arabiensis* and *An. merus*, as well as between *An. merus* and *An. quadriannulatus*.

In addition, *Gr59* shows a marginally significant ($p = 0.052$) excess between *An. gambiae* and *An. merus*. This *Gr* is expressed at low levels in the chemosensory tissues of both *An. coluzzii* and *An. quadriannulatus*. Finally, the sugar receptor *Gr21* has an excess of fixed replacement substitutions between *An. merus* and *An. quadriannulatus* (Table 4.3). As expected, given the overwhelming influence of purifying selection between these species, no *Grs* show a significant excess of nonsynonymous fixed differences between *An. melas* and *An. merus*.

4.3.2.2. Purifying Selection

Ten *Grs* in the *An. gambiae* complex show evidence of purifying selection, as determined by the MK test identifying a significant lack of fixed replacement substitutions (Table 4.4). In addition, two other *Grs* show marginally significant evidence of purifying selection as determined by the MK test.

Between *An. coluzzii* and *An. quadriannulatus*, three *Grs* show a significant lack of fixed replacement substitutions: *Gr39*, *Gr44*-RE, and *Gr45*. While *Gr39* and *Gr45* are lowly expressed in both *An. coluzzii* and *An. quadriannulatus*, *Gr44* is relatively highly expressed in female *An. quadriannulatus*. These three *Grs* show the same pattern between *An. gambiae* s.s. and *An. quadriannulatus*, and all but *Gr45* show purifying selection based on the MK test between *An. arabiensis* and *An. quadriannulatus*. *Gr19*

and *Gr37-RE* also show purifying selection between *An. arabiensis* and *An. quadriannulatus*.

Gr13-RA shows a marginally significant lack of fixed replacement substitutions between *An. coluzzii* and *An. melas*. *Gr38* shows a significant lack between *An. merus* and all three major vectors. In addition, *Gr54* shows significant evidence of purifying selection based on the MK test between *An. arabiensis* and *An. merus*. Three *Grs* show significant evidence of purifying selection between *An. melas* and *An. merus*: *Gr8*, the CO_2 receptor *Gr22*, and *Gr3*. Finally, *Gr60* shows marginally significant evidence of purifying selection based on the MK test between *An. melas* and *An. merus*.

Table 4.4 – *Grs* with significantly (or near-significantly) high neutrality indices, suggestive of purifying selection.

Gene	Species Pair	dN	dS	dN/dS	pN	pS	pN/pS	p-value	Neutrality Index
Gr8	*melas-merus*	8	29	0.276	14	10	1.40	0.006	5.075
Gr13-RA	*coluzzii-melas*	1	7	0.143	32	27	1.19	0.054	8.296
Gr19	*arabiensis-quadriannulatus*	0	7	0.000	28	46	0.609	0.050	NA*
Gr22	*melas-merus*	1	17	0.059	7	11	0.636	0.041	10.818
Gr33	*melas-merus*	0	19	0.000	3	9	0.333	0.049	NA*
Gr37-RE	*arabiensis-quadriannulatus*	1	9	0.111	21	21	1.00	0.032	9.000
Gr38	*arabiensis-merus*	4	21	0.190	17	17	1.00	0.012	5.250
Gr38	*coluzzii-merus*	2	17	0.118	25	28	0.893	0.005	7.589
Gr38	*gambiae-merus*	2	16	0.125	29	38	0.763	0.013	6.105
Gr39	*arabiensis-quadriannulatus*	0	14	0.000	16	25	0.640	0.005	NA*
Gr39	*coluzzii-quadriannulatus*	0	10	0.000	25	39	0.641	0.013	NA*
Gr39	*gambiae-quadriannulatus*	0	8	0.000	35	47	0.745	0.021	NA*
Gr44-RE	*arabiensis-quadriannulatus*	2	12	0.167	56	52	1.08	0.010	6.462
Gr44-RE	*coluzzii-quadriannulatus*	2	14	0.143	36	45	0.800	0.023	5.600
Gr44-RE	*gambiae-quadriannulatus*	2	12	0.167	56	66	0.848	0.025	5.091
Gr45	*coluzzii-quadriannulatus*	0	4	0.000	34	28	1.21	0.050	NA*
Gr54	*arabiensis-merus*	11	22	0.500	23	13	1.77	0.016	3.538
Gr60	*melas-merus*	16	22	0.727	15	6	2.50	0.055	3.438

**No NI was calculated for these* Grs *due to 0 fixed replacement substitutions.*

4.3.3. DH Suite of Tests

Selective sweeps can be detected by significantly negative Tajima's D values, as well as significantly negative Fay and Wu's H values, although both tests are subject to biases from demographic forces. As such, it cannot be conclusively determined that genes are undergoing a selective sweep from the result of a single test. I therefore only consider genes with significantly negative values on the joint DH test, which is a combination of the two tests and was designed to be robust to the influence of demographic factors[315], to show strong evidence of sweep. Since sweep may be detected in linked loci, I also discuss genes with known functions located near the *Grs* of interest, as these other genes could be responsible for the detected patterns. Furthermore, the sheer number of statistical tests performed in this analysis greatly reduces power, such that no significant result may be taken as definitive evidence of a selective sweep.

Since the DH suite of tests was only run on fully genotyped sequences, far fewer sequences were available for analysis than for the MK test. *An. coluzzii* and *An. gambiae* s.s. were available for most *Grs*: 50 and 53 respectively. *An. quadriannulatus* was available for 36 *Grs*, while the remaining three species were rarely available. The DH suite of tests was only able to be run on three *Grs* in *An. arabiensis*, seven in *An. melas*, and six in *An. merus*.

Grs which showed significant deviations from neutrality based on this suite of tests are presented in the following order: *Grs* with significant joint DH test results, suggestive of selective sweep; followed by *Grs* with significant E test results, suggestive of recovery from sweep or background selection; followed by those *Grs* that showed

significance on either the Tajima's D or Fay and Wu's H test but not the joint DH test, where demographic factors cannot be ruled out as an explanation.

4.3.3.1. Joint DH Test (Potential Selective Sweeps)

Table 4.5 – *Grs* with significant (or near significant) DH values, suggestive of a selective sweep.

Gene	Species	DH	π	Haplotype Diversity (HD)
Gr3	*An. melas*	0.044	0.001	0.632
Gr11	*An. melas*	0.046	0.001	0.556
Gr12	*An. melas*	0.052	0.002	0.667
Gr17	*An. coluzzii*	0.047	0.008	0.938
Gr19	*An. quadriannulatus*	0.007	0.002	0.943
Gr36	*An. gambiae*	0.056	0.008	0.998
Gr41	*An. quadriannulatus*	0.019	0.002	0.94
Gr59	*An. gambiae*	0.054	0.008	0.988

A total of five *Grs* show significant evidence of a potential selective sweep based on the joint DH test, while an additional three *Grs* show near-significant ($0.05 \geq p \geq 0.06$) evidence thereof (Table 4.5). In *An. coluzzii*, the highly expressed sugar receptor, *Gr17*, is the only *Gr* showing significant evidence of sweep. The TCS network of *Gr17* in *An. coluzzii* is somewhat star-shaped, but there is no central high-frequency haplotype, which could be a consequence of the length of time following the sweep (Figure 4.7). In *An. gambiae*, *Gr36* and *Gr59* both show marginally significant evidence of sweep. Both genes are undetected in *An. coluzzii* and *An. quadriannulatus* chemosensory tissues, although *Gr59* is located within 200 bp of *ATPsynB*, F-type H+-transporting ATPase subunit B, which plays a critical role in cellular respiration. *Gr36* is approximately 3,000

bp away from *Gr60*, and is on the opposite strand of *Gr41* and *AGAP001124*, an aminomethyltransferase. *Gr60* shows evidence of positive selection based on the MK test between *An. coluzzii* and *An. quadriannulatus* and *Gr41* shows evidence of a selective sweep based on the joint DH test in *An. quadriannulatus*.

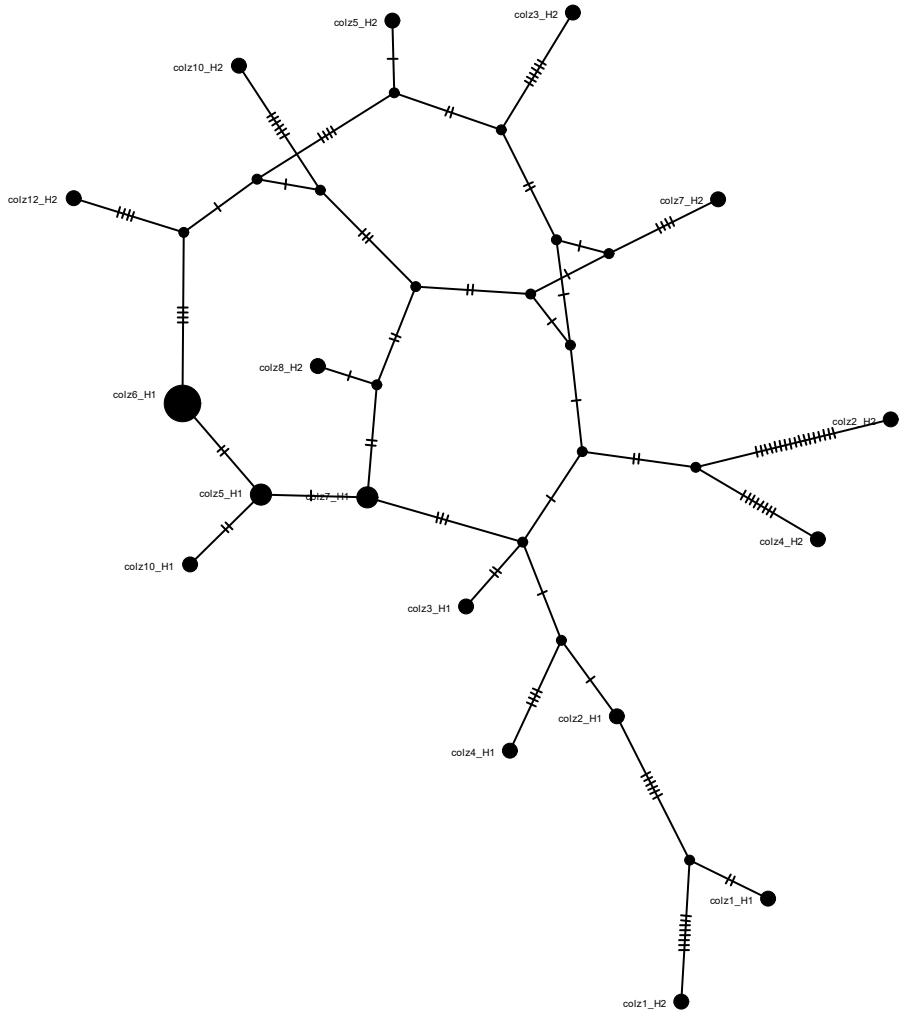

Figure 4.7 – TCS Haplotype Network of the sugar receptor *Gr17* in *An. coluzzii*. Each hashed line represents one nucleotide substitution and haplotype nodes are weighted by frequency. There are 17 haplotypes among 24 sequences (π = 0.008 and HD = 0.938).

In *An. quadriannulatus*, the lowly expressed sugar receptor *Gr19* shows significant evidence of a potential selective sweep. The TCS network features a clear star shape with central high-frequency haplotype surrounded by lower-frequency haplotypes, which is a hallmark of sweep (Figure 4.8). *Gr41* also shows significant evidence of potential sweep (Table 4.5). *Gr41* is very lowly expressed or not detected in all chemosensory organs, as well as the full body of *An. coluzzii* [152]. However, as mentioned above, it is also located near other *Grs* which could be undergoing selection, as well as multiple non-chemosensory genes that fulfill basic biological roles.

Finally, three of the seven *An. melas Grs* analyzed show evidence of potential selective sweep based on the joint DH test (Table 4.5). *Gr3* and *Gr11* are both significant, while *Gr12* is near-significant ($p = 0.052$). *Gr3* also shows evidence of recovery from a selective sweep in *An. gambiae* s.s. (Table 4.6), but is lowly expressed in female *An. coluzzii* labella and not detected elsewhere. However, *Gr3*, along with *Grs* 4-8, is located relatively close to *ABCG7*, an ATP-binding cassette transporter.

Gr11 also shows evidence of selection based on the MK test between *An. melas* and *An. arabiensis*, as well as between *An. melas* and *An. quadriannulatus* (Figure 4.3,Table 4.3). *Gr11* is lowly expressed in chemosensory tissues of *An. coluzzii* and *An. quadriannulatus*, but is within 300 bp of *Gr12*. *Gr12* is also lowly expressed in *An. coluzzii* and *An. quadriannulatus* chemosensory tissues.

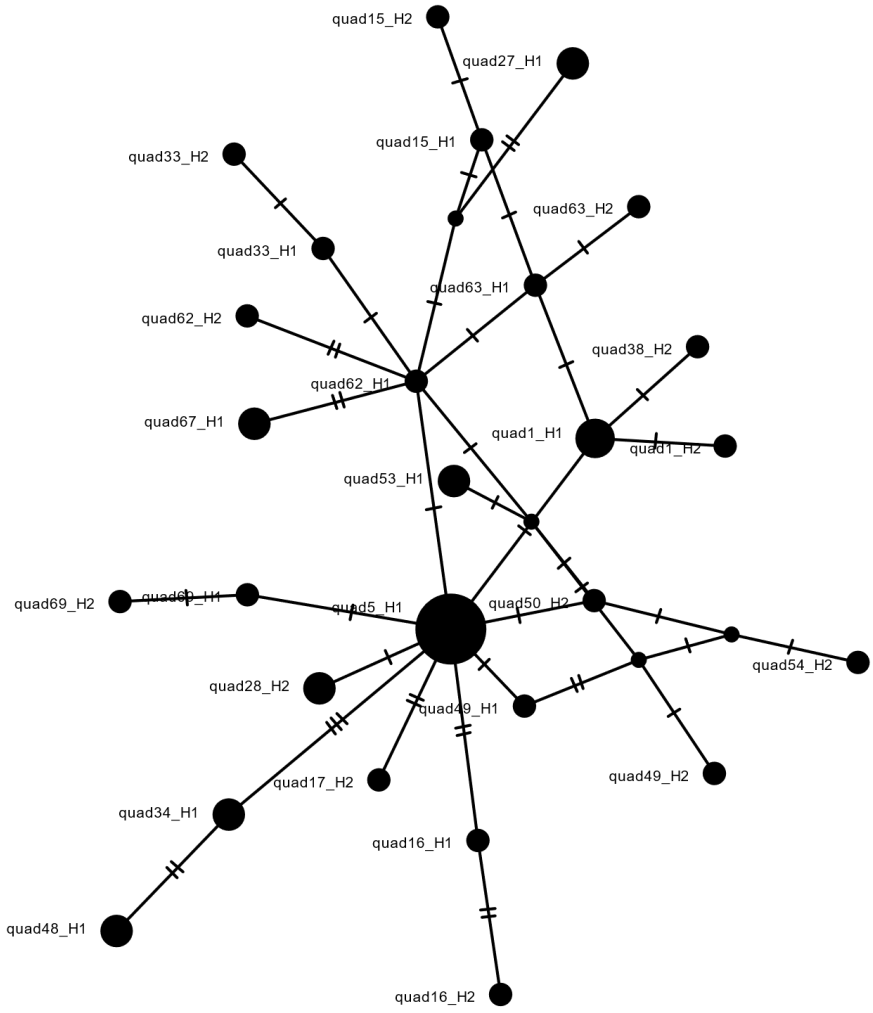

Figure 4.8– TCS Haplotype Network of the sugar receptor *Gr19* in *An. quadriannulatus*. Each hashed line represents one nucleotide substitution and haplotype nodes are weighted by frequency. There are 16 haplotypes among 44 sequences (π = 0.002 and HD = 0.943).

4.3.3.2. E Test (Recovery from Selective Sweep or Background Selection)

Table 4.6 – *Grs* with significant (or near significant) E-test values, indicative of recovery from a selective sweep.

Gene	Species	E	p-value	π	Haplotype Diversity
Gr3	*An. gambiae*	-1.51	0.047	0.011	0.974
Gr18	*An. gambiae*	-1.82	0.01	0.008	0.995
Gr24	*An. gambiae*	-1.68	0.025	0.011	0.998
Gr57	*An. coluzzii*	-1.59	0.052	0.004	0.979

Three *Grs* in *An. gambiae* s.s. show significant E test values, which detect an excess of low-frequency polymorphisms suggestive of recovery from a selective sweep or background selection: *Gr3*, *Gr18*, and *Gr24* (Table 4.6). In *An. coluzzii*, *Gr57* is near significant (p = 0.052). This *Gr* is very lowly expressed in chemosensory tissues, but is located approximately 4,000 bp from *AGAP004715*, which encodes a pyruvate dehydrogenase phosphatase regulatory subunit. As discussed above, *Gr3* is lowly expressed but is near an ATP-binding cassette transporter gene. *Gr18* encodes a lowly-expressed sugar receptor, but is adjacent to *Gr17*, which is a very highly expressed sugar receptor. In addition to its excess of low-frequency variants in *An. gambiae* (Figure 4.9), it has an excess of fixed replacement substitutions between *An. quadriannulatus* and both of the other major vectors, *An. arabiensis* and *An. coluzzii*. Finally, the CO_2 receptor *Gr24* also has a significant excess of low-frequent variants (Figure 4.10), which is surprising due to the high sequence conservation of the CO_2 receptors and the importance of CO_2 as a host cue. However, *Gr24*'s closest neighbor on the 2R

chromosome is *AGAP001914*, which is approximately 1,000 bp away and encodes a cell division cycle protein.

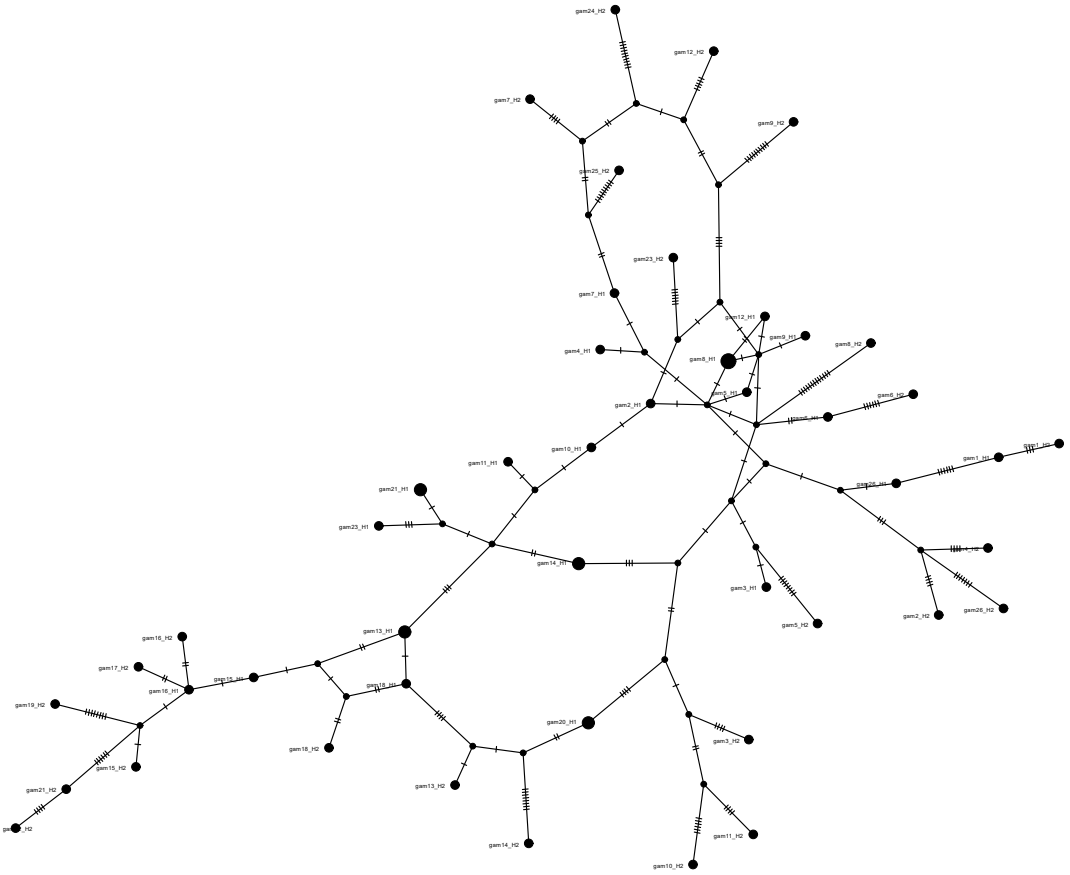

Figure 4.9 – TCS Haplotype Network of the sugar receptor *Gr18* in *An. gambiae* s.s. Each hashed line represents one nucleotide substitution and haplotype nodes are weighted by frequency. The E test detects recovery from a selective sweep or background selection based on an excess of low-frequency variants. Tajima's D can also detect this phenomenon. E = -1.82 (p = 0.01) and D = -1.45 (p = 0.048).

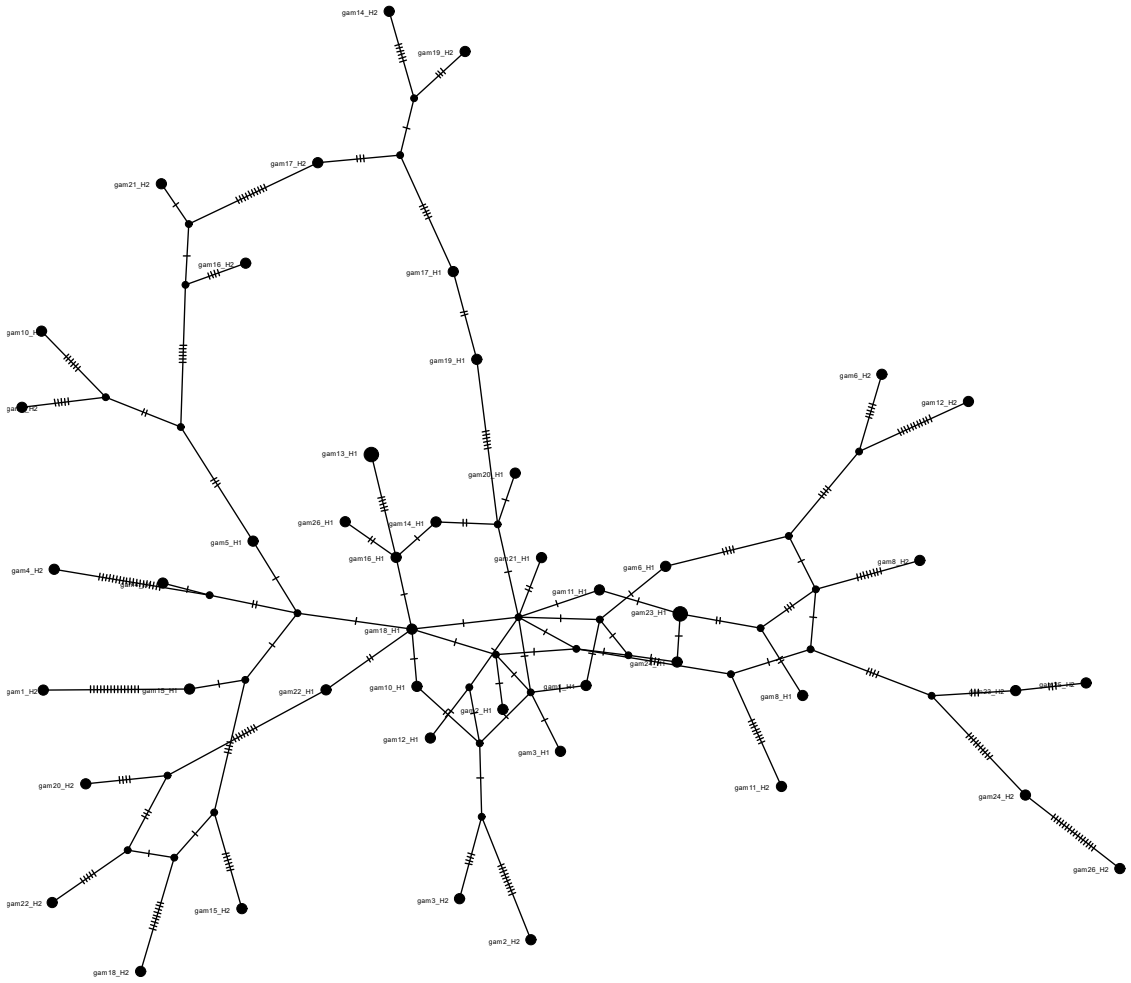

Figure 4.10 – TCS Haplotype Network of the CO_2 receptor *Gr24* in *An. gambiae* s.s. Each hashed line represents one nucleotide substitution and haplotype nodes are weighted by frequency. The E test detects recovery from a selective sweep or background selection based on an excess of low-frequency variants. Tajima's D can also detect this phenomenon. E = -1.68 (p = 0.025) and D = -1.39 (p = 0.058).

4.3.3.3. Tajima's D or Fay and Wu's H (Selective Force Unclear)

In addition to the *Grs* described above with significant MK test, joint DH test, or E test results, there are also ten *Grs* that have significantly negative Tajima's D or normalized Fay and Wu's H value (Table 4.7). Both test results could indicate recovery

from selective sweeps, but both tests are also susceptible to demographic factors. Specifically, a significantly negative Tajima's D could reflect an increase in population size, and a significantly negative Fay and Wu's H could reflect a reduction in population size or the presence of population structure. A significantly negative Fay and Wu's H result could also indicate strong purifying selection. Since multiple possibilities arise with these significant results, it is not possible to determine the actual selective force at work.

Table 4.7 – *Grs* with significant (or near significant) Tajima's D or normalized Fay and Wu's H values, indicative of non-neutrality, but selective force unclear.

Gene	Species	D	p-val	H	p-val	π	HD
Gr1	*An. quadriannulatus*	-0.034	0.565	-1.77	0.048	0.008	0.996
Gr7	*An. gambiae*	-1.49	0.041	-0.034	0.316	0.013	1.0
Gr8	*An. gambiae*	-1.59	0.03	-0.53	0.185	0.01	0.995
Gr15	*An. gambiae*	-1.53	0.039	-0.129	0.282	0.01	0.998
Gr15	*An. quadriannulatus*	-0.203	0.489	-2.85	0.014	0.005	0.985
Gr22	*An. melas*	-0.658	0.277	-3.07	0.01	0.002	0.803
Gr33	*An. gambiae*	-1.48	0.043	-0.017	0.321	0.006	0.988
Gr33	*An. merus*	0.096	0.56	-2.46	0.04	0.0003	0.4
Gr43	*An. quadriannulatus*	-0.703	0.26	-1.61	0.058	0.002	0.94
Gr46	*An. gambiae*	-1.42	0.06	-0.815	0.141	0.006	1
Gr51	*An. quadriannulatus*	-0.788	0.234	-3.02	0.009	0.008	0.978
Gr59	*An. gambiae*	-1.21	0.059	-0.881	0.127	0.008	0.988
Gr59	*An. quadriannulatus*	-0.124	0.529	-1.79	0.047	0.004	0.99

Four *Grs* have significantly negative Tajima's D values in *An. gambiae* s.s.: *Gr7*, *Gr8*, *Gr15*, and *Gr33* (Table 4.7). In addition, *Gr46* and *Gr59* have marginally significantly negative Tajima's D values. As mentioned above, *Grs* 7 and 8 are located near *ABCG7*, which is an ATP-binding cassette transporter. *Gr15* is a lowly-expressed

sugar receptor, but is adjacent to (about 1,500 bp from) *CLIPB36*, a CLIP-domain serine protease, which is involved in innate immune response and embryonic development [318].

Gr33, in addition to its significant Tajima's D value and excess of rare alleles (Figure 4.11), is male-biased in the antennae of both *An. coluzzii* and *An. quadriannulatus* [47] and relatively highly expressed in the labella of both sexes (Chapter 2). It also is highly conserved between *An. arabiensis*, *An. coluzzii*, *An. gambiae* s.s., and *An. quadriannulatus*, with no fixed replacement substitutions between any combination of these species. Furthermore, there is only one fixed synonymous substitution between *An. arabiensis* and both *An. coluzzii* and *An. gambiae* s.s., two between *An. quadriannulatus* and both *An. coluzzii* and *An. gambiae* s.s., and five between *An. arabiensis* and *An. quadriannulatus*. It also shows signs of purifying selection based on the MK test between *An. melas* and *An. merus* (Table 4.4).

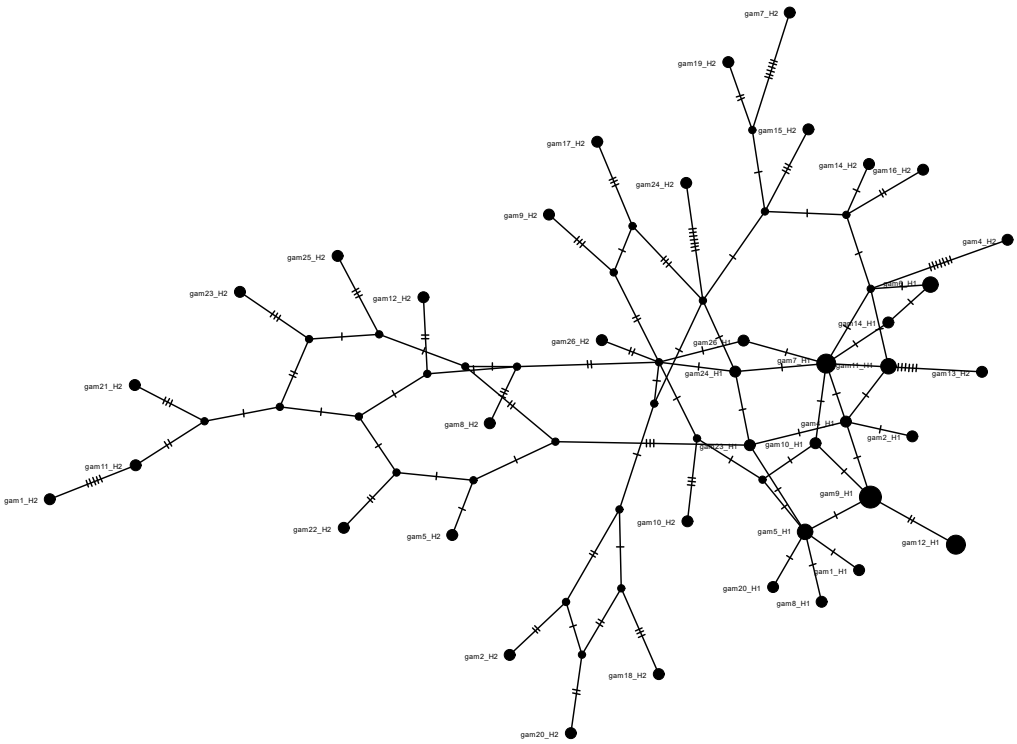

Figure 4.11 – TCS Haplotype Network of *Gr33* in *An. gambiae* s.s. Each hashed line represents one nucleotide substitution and haplotype nodes are weighted by frequency. There are 40 haplotypes among 50 sequences (π = 0.006 and HD = 0.988).

Gr46 is lowly expressed in chemosensory tissues, but is on the opposite strand of *LRIM16A*, leucine-rich immune protein 16A, which is involved in innate immunity and downregulated following infection with *Plasmodium berghei*[319]. As mentioned above, *Gr59* is located near *ATPsynB*, which plays a role in cellular respiration.

Four *Grs* have significantly negative normalized Fay and Wu's H values (indicative of an excess of high-frequency variants) in *An. quadriannulatus*: *Gr1*, *Gr15*, *Gr51*, and *Gr59* (Table 4.7). In addition, *Gr43* is marginally significant on this test. *Gr1* is *An. quadriannulatus*-biased in antennae of both sexes[46], but is also adjacent to two different cystinosin genes, which are responsible for exporting cystine molecules from

the lysosome. As mentioned above, *Gr15* is adjacent to *CLIPB36*. *Gr43* is relatively highly expressed in the labella, but is also located on the opposite strand of *RpL22*, a 60S ribosomal protein which may be involved in regulating deltamethrin resistance[320]. *Gr51* is expressed in the labella and is adjacent to *Gr50*, which showed evidence of positive selection based on the MK test, as well as *Gr52*, which is *An. coluzzii*-specific in the female maxillary palps[46].

Finally, *An. melas* and *An. merus* have one *Gr* each that shows an excess of high-frequency variants based on Fay and Wu's H. In *An. melas*, the CO_2 receptor *Gr22* shows this pattern, in addition to the evidence of purifying selection between *An. melas* and *An. merus* in this *Gr*. However, *Gr22* is located across from *RpS11*, which is a 40S ribosomal protein. In *An. merus*, *Gr33*, which also shows an excess of fixed replacement substitutions between *An. melas* and *An. merus* has an excess of high-frequency variants.

4.4. Discussion

In this study, I examined the selective forces acting on *Grs* in the *An. gambiae* complex. Specifically, I have identified potential evidence of positive selection in sixteen *Grs* and purifying selection in twelve *Grs* based on the McDonald-Kreitman test. I have also identified eight *Grs* potentially under the influence of sweep via the joint DH test and four *Grs* recovering from either sweep or background selection via the E test. Finally, I have identified ten *Grs* that show deviations from neutrality from less clear forces via Tajima's D and Fay and Wu's H tests. This is the first evolutionary study on the *Grs* in the *An. gambiae* complex, as well as in *Anopheles* more broadly, whereas the

evolution of *Drosophila Grs* and its relationship to ecological adaptations has been extensively studied[43,82,175,269,270].

In addition, there have also been several studies on the evolution of *Grs* in Lepidoptera[42,271,273,277,321]. By contrast, the mosquito literature on the evolution of chemosensory genes is limited to one seminal paper linking the evolution of an *Or* to human host preference in *Aedes aegypti*[44] and a paper from our lab on the *Ors*, *Irs*, and *Obps* in the *An. gambiae* complex[182]. As such, I can compare what we know of *Gr* evolution in other lineages while incorporating the evolutionary and ecological contexts of the six *An. gambiae* s.l. species and comparing the observed patterns in the *Grs* to those in the other chemoreceptors. There is evidence that *Grs* are under both positive and purifying selection in the *An. gambiae* complex, which may have contributed to the development of species-specific behavioral ecology, as is seen in other taxa. However, the support for a *Gr* role in the evolution of anthropophily is much sparser than that for other chemosensory genes.

4.4.1. McDonald-Kreitman Test

4.4.1.1. Positive Selection

In addition to detecting the specific *Grs* undergoing selection in the *An. gambiae* complex, the MK test provides a general picture of the dominant selective force at work by generating Neutrality Index (NI) values. Previous studies on *Drosophila* and Lepidoptera have emphasized the role of purifying selection on *Grs* over that of positive selection[51,82,269,277]. However, positive selection on *Grs* has also played a role in specific

lineages, potentially facilitating adaptative differentiation in both lineages, as well as in the pea aphid[42,43,175,270–272,277].

I have demonstrated variability in the relative importance of these selective forces by lineage and ecology. The strongest trend towards positive selection is between *An. arabiensis* and both *An. coluzzii* and *An. gambiae* s.s. (Table 4.2, Figure 4.2B). There were a maximum of thirteen *Gr* transcripts where the NI could be computed between these species, which reflects their sequence homology and history of introgression[288]. Most of the transcripts with NIs consistent with positive selection between both species pairs have unknown ligands, with the exception of two sugar receptors (*Grs* 18 and 21) between *An. arabiensis* and *An. coluzzii*. As such, biological significance is unclear. Nonetheless, most *Gr* transcripts that diverge between these species appear to be advantageous, which is reflected by their NI values. In addition, while *An. arabiensis* is more of a generalist feeder than *An. gambiae* s.s. and *An. coluzzii*, it is nevertheless more anthropophilic than the remaining complex members[18], which would suggest that those genes which confer a selective advantage to *An. coluzzii* and *An. gambiae* would have the same effect in *An. arabiensis*.

By contrast, the strongest trend towards purifying selection is between *An. melas* and *An. merus* (Figure 4.3A, Table 4.2). The three CO_2 receptor *Grs* have high NI values, which is expected, given the importance of this cue in host-seeking, particularly for more zoophilic species such as *An. melas* and *An. merus*[13]. These two species also share similar ecologies, wherein both are found in brackish coastal wetlands[13]. As such, the other *Grs* showing evidence of purifying selection are most likely important for their

survival in these habitats. It is conceivable that some of them perceive salts, as some gustatory receptor neurons are responsive to salt[89,322], although as their ligands are mostly unknown, more study is needed. It is also possible that these *Grs* facilitate feeding on nectar sources that are more common in coastal habitats than elsewhere.

In the other species pairs, *Grs* are more evenly distributed between NIs consistent with purifying or positive selection. However, the average NIs consistently favor purifying selection, although to a lesser degree than between *An. melas* and *An. merus* (Table 4.2). This finding suggests that like in *Drosophila* and Lepidoptera, *Gr* evolution is primarily characterized by purifying selection. Similarly, a QTL for human host preference identified by our lab contains mostly *Ors* and *Obps*. While six *Grs* were located inside the QTL, only one was expressed at sufficient levels in female antennae and maxillary palps to merit further investigation, compared to 18 *Ors*, 12 *Obps*, and three *Irs*[182]. In sum, these results indicate a low likelihood that most *Grs* play a major role in vertebrate host preference; instead, they likely contribute to conserved behaviors across the complex.

Nonetheless, sixteen *Grs* have significant MK test results consistent with positive selection. While most of them have unknown ligands, two sugar receptors show evidence of positive selection when comparing *An. quadriannulatus* to the major vector species: *Gr18* and *Gr21*. *Gr21* is expressed at relatively high levels in the labella, although *Gr18* is not. However, the sugar receptors are in close proximity on the 2R chromosome arm. *Drosophila* sugar receptors are co-expressed, so *Anopheles* sugar receptors may be as well. As such, *Gr18 per se* may not play a critical role in the

detection of an important sugar resource, but may either have hitchhiked with an important gene (such as *Gr17*, described below) or may be co-expressed with other sugar receptors. Regardless, its expression is not sex-biased, so any ecologically important role is almost certainly related to evaluating nectar sources.

Two of the *Grs* showing evidence of positive selection between *An. quadriannulatus* and *An. coluzzii* (*Gr60*) and *An. gambiae* s.s. (*Gr48*) are male-biased and *An. quadriannulatus*-biased in the maxillary palps and labella, respectively. Neither of these *Grs* has a known ligand. While females engage in several sex-specific behaviors such as blood-feeding, host-seeking, and oviposition, the only male-specific behaviors are swarming[323] and mating. The antennal fibrillae are known to play an important role in male detection of auditory cues in both swarming and close-range mating behaviors in the *An. gambiae* complex, and the male claspers recognize if females have mated, but other organs have not been implicated in male mating[323,324].

While neither the maxillary palps nor labella have been established as playing a role in mating biology in *Anopheles*, there is evidence that the maxillary palp detects female inhibitory cues in *Drosophila*[325], which raises the possibility that a similar phenomenon might occur in *Anopheles*, particularly as *Drosophila Grs* expressed in the labellum and tarsi are known to play a role in inhibiting male-male courtship[326]. The labellum is well-established in a male mating role in *Drosophila*[55].

Ongoing work in the Slotman Lab explores the mating behavior of male and female mosquitoes when encountering both conspecific and heterospecific mates, which may define a role (or lack thereof) for the tarsi, male maxillary palps and labella in

mating behavior. There is evidence that females in mixed-species swarms mate assortatively[327], although males are not believed to do so[324], despite the ability to detect females who have already mated.

Aside from *Grs* 18, 21, 48, and 60, most other *Grs* with MK test results consistent with positive selection are lowly expressed in chemosensory organs or show evidence of selection in species with uncharacterized transcriptomes. As such, it is more difficult to explain their significance. In the case of lowly expressed *Grs*, it is impossible to know which organs they might be expressed in, if any, so their biological roles are unknown. Most anopheline *Gr* ligands (other than the CO_2 and sugar receptors) are unknown, in contrast to *Drosophila*, where they are known to also perceive bitter compounds and cuticular hydrocarbons, as well as mediate light and heat avoidance[53,84–90,199,200]. Furthermore, *Grs* are difficult to heterologously express and thereby deorphanize[268], and *Grs* also evolve rapidly via gene duplication along lineages, meaning that most anopheline *Grs* do not have known homologs in *Drosophila*. In the case of *Grs* showing significance in species other than the *An. coluzzii-An. gambiae* s.s. clade or *An. quadriannulatus*, there is a complete absence of transcriptomic data to suggest where these *Grs* might be expressed or whether they show species-biased expression patterns.

It is possible that the lowly expressed *Grs* are expressed at higher levels in either the tarsi or internal organs (such as the brain or the gut), or even differentially in larvae. Ongoing work in the Slotman Lab explores the chemosensory gene expression profile of each of the three tarsi in male and female *An. coluzzii* and *An. arabiensis*, which may

grant more clarity on the *Grs* undergoing selection. Furthermore, since the tarsi have an established role in *An. gambiae* mating behaviors[324], these gene expression patterns may help to explain the evolutionary patterns I detect in *Gr48* and *Gr60*, although *An. quadriannulatus* is not included in this study.

The McDonald-Kreitman test is susceptible to demographic forces such as an expanding population combined with slightly deleterious mutations[311] and it tends to underestimate adaptive evolution in the presence of slightly deleterious mutations[328,329]. The joint DH test is much more robust to forces other than directional selection, but the two tests are not interchangeable, and it is anticipated that *Grs* showing significant evidence of positive selection based on one test will not necessarily show significant evidence based on the other. As such, we can tentatively accept the *Grs* that were identified by the MK test as potentially adaptive, pending further investigation, although as mentioned above, the multiple testing problem precludes the conclusive identification of positive selection.

4.4.1.2. Purifying Selection

Twelve *Grs* show evidence of purifying selection based on the MK test. As with those *Grs* showing evidence of positive selection, the most interesting are those which are conserved between *An. quadriannulatus* and highly anthropophilic species. There are three such *Grs* when including only *An. coluzzii* and *An. gambiae* s.s.: *Gr39*, *Gr44-RE*, and *Gr45* (Table 4.4). If including *An. arabiensis*, the sugar receptor *Gr19* and *Gr37-RE* are also of interest. Of these *Grs*, only *Gr44* is expressed at relatively high levels in chemosensory tissues, and our study design precludes identifying the specific transcript

expressed therein. As such, and given the lack of homologs as mentioned previously, it is difficult to determine why these *Grs* would be conserved.

4.4.2. Selective Sweeps

4.4.2.1. Joint DH Test

Eight *Grs* have joint DH test results consistent with recent selective sweeps (Table 4.5). As discussed above, it is easiest to interpret these significant results in *Grs* which are highly expressed in either *An. coluzzii* or *An. quadriannulatus* chemosensory organs. Of the eight identified, only the highly-expressed sugar receptor *Gr17* fits these criteria. In addition, *Gr17* ends approximately 350 bp before *Gr18*, which shows evidence of positive selection. Like *Gr18*, the expression of *Gr17* is not sex-biased, so it presumably does not underly any sex-biased behaviors. One possible role of *Gr18* is nectar-feeding, which grants both males and females access to critical nutritional resources, the only resource fed on by adult males and a resource which has been shown to prolong the lives and reduce blood-feeding frequency of adult females[330].

While the other *Grs* with significant joint DH test results consistent with selective sweeps are lowly expressed, many of them are located near genes encoding critical cellular functions or other *Grs* which are more highly expressed and also show evidence of positive selection. *Gr59*, which shows evidence of sweep in *An. gambiae* is located within 200 bp of *ATPsynB*. The *Drosophila melanogaster* ortholog of this gene is required for the development of male reproductive organs[331]. If there was an advantageous mutation at this locus in *An. gambiae*, it could explain the evidence of

sweep in *Gr59,* as the joint DH test is not sensitive to background selection against deleterious mutations[315].

Gr41 is lowly expressed in *An. quadriannulatus*, but is on the opposite strand of *Gr60,* and next to *AGAP001124*, an aminomethyltransferase. *Gr60* shows significant evidence of positive selection on the MK test as well as male *An. quadriannulatus* bias in the maxillary palps. While it is possible that *Gr41* plays a role elsewhere in the body, an advantageous mutation in one of these other genes could also have caused this significant result. *Gr36,* which is also lowly expressed and shows evidence of sweep in *An. gambiae*, is also in this region of chromosome 2R.

Grs 11 and 12, both of which show evidence of sweep in *An. melas*, are adjacent to one another and also near *Grs* 9 and 10 on the 3R chromosome. However, all of these *Grs* are lowly expressed in *An. coluzzii* or *An. quadriannulatus*, and there are no genes other than *Grs* nearby with known functions. *An. melas* transcriptomic data might show evidence of their expression to clarify this result. *Gr3*, the other *Gr* showing evidence of sweep in *An. melas* is located near *Grs* 4-8 on the 3R chromosome, and is located relatively close to *ABCG7*, an ATP-binding cassette transporter. While these genes are broadly known as transporters, the specific function of its *Drosophila* ortholog, which shares approximately 75% sequence homology, is unknown[332]. Therefore, the source of the selective sweep in *Gr3* is elusive.

4.4.2.2. E Test

Four *Grs* show E test results consistent with recovery from a selective sweep, although the E test is also sensitive to background selection (

Table 4.6). All three *Grs* with significant E test results occur in *An. gambiae* s.s., while the nearly significant *Gr57* occurs in *An. coluzzii*. This *Gr* is very lowly expressed in chemosensory tissues, but is located approximately 4,000 bp from *AGAP004715* on the 2L chromosome. This gene encodes a pyruvate dehydrogenase phosphatase regulatory subunit, although a more specific function is unknown. Nonetheless, background selection against deleterious mutations in this gene could potentially account for the E test result observed in *Gr57*.

In *An. gambiae*, two of the three *Grs* with significant E test results (*Gr3* and the sugar receptor *Gr18*) are discussed above. The other *Gr* with a significant E test value is the CO_2 receptor *Gr24*. *Gr24* was also included in the QTL for human host preference, although it was thought to be unlikely to actually contribute to host preference[182]. The CO_2 receptors are highly conserved across insects[333], and there are no fixed differences in *Gr24* between *An. arabiensis*, *An. coluzzii*, *An. gambiae* s.s., and *An. quadriannulatus*, although there are six to eight nonsynonymous and one to two synonymous fixed differences when comparing these species to *An. melas* or *An. merus*. CO_2 is thought to matter less as a host-seeking cue in the anthropophilic members of the complex than in more zoophilic or generalist species in the complex[18]. However, the lack of fixed differences between *An. gambiae* and *An. quadriannulatus* suggests background selection against deleterious mutations, potentially with impaired CO_2 perception and therefore difficulty finding hosts, as a reason for this significant E test result.

Nonetheless, this result paired with its presence in the QTL suggests further consideration of the evolution of *Gr24*.

4.4.2.3. Tajima's D and Fay and Wu's H

Fourteen *Grs* were significant on one of these tests but not the DH test or E test (Table 4.7). Most of them have unknown functions and low expression, with the exceptions of the sugar receptor *Gr15*, CO_2 receptor *Gr22*, and *Gr33*, which is highly expressed and male-biased in the antennae[47]. Both Tajima's D and Fay and Wu's H can detect positive selection and purifying selection, although H is somewhat better at the latter. Tajima's D can also detect background selection and population growth, while Fay and Wu's H is better at detecting population shrinkage and subdivision[315]. Given their wide capabilities and the lack of significant results on other tests in these *Grs*, it is impossible with the currently available data to identify the selective or demographic forces at work.

4.4.3. Comparison to Other Chemosensory Genes

Athrey *et al.* (in review)[182] identified a total of five *Obps*, four *Ors*, and three *Irs* which showed evidence of positive selection in combination with presence in a human host preference QTL or differential expression in *An. coluzzii*. By contrast, no *Grs* show similarly clear patterns. The *Grs* which show evidence of positive selection in comparisons including either *An. coluzzii* or *An. gambiae* and *An. quadriannulatus* are lowly-expressed in *An. coluzzii*. However, two *Grs* which are *An. quadriannulatus* biased in males (*Gr48* and *Gr60*) show evidence of positive selection in these species

pairs, indicating that they could potentially be involved in species-specific male behaviors.

4.4.4. Future Directions

While we have high-quality variant data from *An. coluzzii* and *An. gambiae* s.s. for these 57 *Grs*, data for the other four species are of a much lower quality, as evidenced by the overall sequencing coverage, the number of variants detected, and the number of complete sequences. While we have sufficient data to draw conclusions about evolutionary patterns, particularly with respect to the zoophilic non-vector *An. quadriannulatus*, a more conclusive and thorough analysis would require either deeper coverage of the whole genome sequences of the remaining species, or a more targeted sequencing approach, such as Linked Target Capture[334], which would allow precise sequencing of only the *Grs* at a much greater depth.

In addition to the relatively low quality of *Gr* sequences in the other species in the complex, there is also no expression data from *An. arabiensis*, *An. melas*, and *An. merus*. As such, it is very difficult to assess which of the genes that show significant deviations from neutrality in these species are most important, and what their biological roles might be. However, given the high degree of introgression between *An. arabiensis* and the other anthropophilic species, there are likely to be fewer overlooked *Grs* in this species than in the coastal species.

While there are no obvious candidate genes for anthropophily identified in this study, there are several avenues meriting significant further study. The sugar receptors *Grs* 14-21 are closely clustered on the 2R chromosome, and three of them show

evidence of directional selection in either an anthropophilic species or *An. quadriannulatus*. The means by which they detect different sugars are not understood to nearly the same degree as in *Drosophila*, but the potential presence of positive selection in them merits a better understanding, as it could relate to the ability to successfully exploit preferred nectar resources, or to evaluate sugar sources which are found in greater abundance near preferred vertebrate hosts. Duplications and pseudogenizations are known to have occurred in other sugar receptor lineages[335], so the functionality of the entire suite of sugar receptors in the *An. gambiae* complex also bears further investigation.

In addition, the potential positive selection in *Grs* 48 and 60 in *An. quadriannulatus*, both of which are male-biased in chemosensory organs other than the antennae, suggest the need for further study of male mating behavior to determine what roles these organs might play in the process, whether gustatory or olfactory. Finally, in addition to the lack of transcriptomic data for *An. arabiensis*, *An. melas*, and *An. merus*, there is a lack of *Gr* expression data outside of the adult chemosensory organs. In particular, it is unknown to what extent *Anopheles Grs* are expressed in other organs, such as the brain and gut. Furthermore, study of *Gr* expression in larvae might clarify some of the *Grs* showing evidence of selection in this study. Research on *Aedes aegypti* larvae has shown that they rely on chemokinesis to navigate chemical gradients, and has further suggested that they likely rely on *Grs* and *Irs* rather than *Ors* to do so[336]. As such, characterization of *Gr* expression profiles in larval *Anopheles* could potentially illuminate biological meaning for some lowly expressed *Grs* in adults.

5. CONCLUSIONS

Anthropophily is a major contributor to malaria vectorial capacity in the *Anopheles* (*An.*) *gambiae* complex. The *An. gambiae* complex is a uniquely valuable model to understand the genetic basis of anthropophily, as it contains two strongly anthropophilic members, *An. coluzzii* and *An. gambiae* s.s. as well as the zoophilic *An. quadriannulatus*. Other complex members have more variable and opportunistic host preferences[13]. By characterizing the genes and anatomy associated with anthropophily in this context, we may be able to determine its basis.

The mosquito chemosensory system plays a major role in determining host preference[17,251]. By identifying differential gene expression (DGE) between *An. coluzzii* and *An. quadriannulatus* in the chemosensory organs, we can determine which genes show evidence of species bias, sex bias, and most importantly, sex and species bias. Genes which are both *An. coluzzii*- and female-biased have the greatest likelihood of influencing anthropophily. By contrast, genes which are both *An. quadriannulatus* and female-biased could influence zoophily, although evidence supports a greater reliance on general vertebrate host cues in this mosquito than on bovine-specific volatiles[18,19,151].

Furthermore, by identifying DGE in all chemosensory organs, we may be able to determine which organs are most likely to play a role in determining host preference, and thereby, determine which genes are candidates meriting evaluation for transgenic mosquito control aimed at reducing contact between vector and host. Previous work has identified DGE in the antennae and maxillary palps of *An. coluzzii* and *An.*

quadriannulatus[45–48]. While the labella transcriptome of *An. coluzzii* was characterized by another group during the course of my PhD[95], I am the first to analyze it during scotophase, the first to analyze it in *An. quadriannulatus*, and the first to identify DGE between species in this organ. While I identify numerous differentially expressed (DE) genes in this organ, the vast majority are not both female- and species-biased. As such, the transcriptomic evidence for the labella having a role in host choice is less compelling than that for the antennae and maxillary palps. However, there are nevertheless two highly expressed *Obps* which bear further examination. *Obp26* is strongly biased towards female *An. coluzzii* in the labella, and also appears in a human host preference QTL[182]. By contrast, *Obp57* is strongly biased towards female *An. quadriannulatus* in the labella. Figure 5.1 and Table 5.1 represent the intersections of chemosensory genes which are DE in the labella, those which show evidence of positive selection, and those which in addition to being in one of the previous categories, are also included in the human host preference QTL.

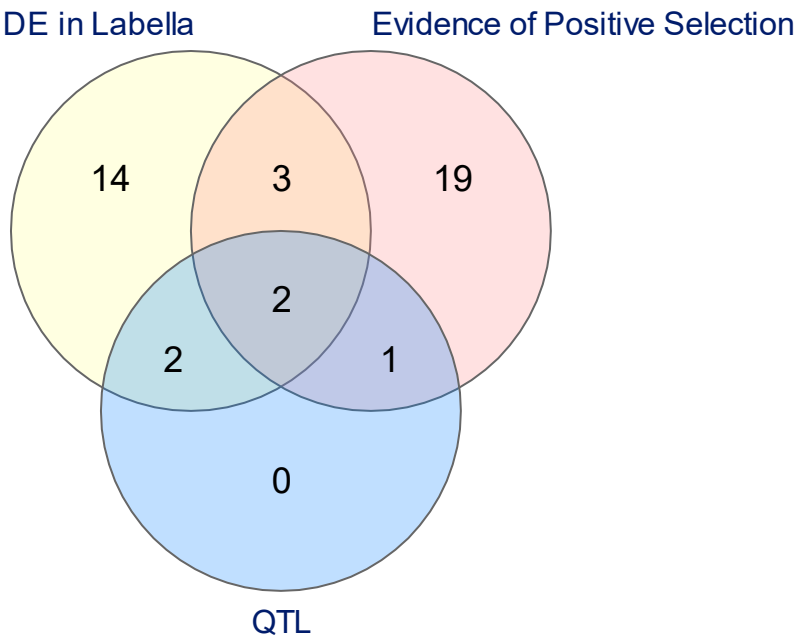

Figure 5.1 – Venn Diagram of Chemosensory Genes of Interest. Genes are included if identified as DE in any comparison in Chapter 2, showing evidence of positive selection based on MK, DH, or E tests in Chapter 4, or identified in the QTL[182] and also either of the above categories.

Table 5.1 – Chemosensory Genes of Particular Interest. Genes are included if two or more criteria met: DE in labella, showing evidence of positive selection, present in human host preference QTL[182].

Gene	Occurrences	Present in:
Obp26	3	DE in Labella, Evidence of Positive Selection, QTL
Or29	3	DE in Labella, Evidence of Positive Selection, QTL
Gr4	2	DE in Labella, Evidence of Positive Selection
Gr21	2	DE in Labella, Evidence of Positive Selection
Gr24	2	Evidence of Positive Selection, QTL
Gr48	2	DE in Labella, Evidence of Positive Selection
Obp25	2	DE in Labella, QTL
Obp57	2	DE in Labella, QTL

Understanding the anatomy underlying host choice allows us to prioritize those genes which are DE in the organs that control it. As such, I ablated the antennae, maxillary palps, and labella of female *An. coluzzii* of host-seeking age and compared

their activation and response rates to those of intact mosquitoes. Human host preference was never completely abolished, but mosquitoes with ablated antennae had a greatly reduced activation rate, and failed to respond to any host odor. Therefore, the antennae are clearly the major organ dictating host choice, as has also been determined in other mosquito species[100–102]. While labella ablation significantly reduced activation, it did not reduce responsiveness, further suggesting that the labella do not play a major role in host choice, at least before landing on a potential host. The maxillary palps also do not appear to play a major role in their own right, given the relatively small impact of their ablation on activation and responsiveness. However, the presence of female and *An. coluzzii*-biased chemosensory gene expression in the palps suggests they are important, although it is possible their role is complementary to that of the antennae, rather than independent.

Finally, understanding the molecular evolutionary patterns of chemosensory genes, in combination with expression data, complements DGE studies by identifying those genes which show signs of selection in anthropophilic species, as well as those which appear conserved between species based on evidence of purifying selection. Previous study has identified multiple *Ors*, *Obps*, and *Irs* which are present in a QTL for human host preference, show evidence of positive selection or selective sweep, and/or show signs of female and *An. coluzzii*-biased expression in one or more organs[182]. By contrast, while I have identified multiple *Grs* which could be under positive selection, none of these *Grs* are both female and *An. coluzzii*-biased. As such, while *Grs* may have facilitated adaptation to nectar resources, they do not appear to have played a major role in the adaptation to human hosts in the *An. gambiae* complex.

We can conceptualize our understanding of anthropophily in the context of Nikolaas Tinbergen's four questions: 1) adaptive function, 2) evolutionary basis, 3) mechanistic basis, and 4) developmental basis[337]. In this study, the adaptive function is the anthropophilic behavior and the chemosensory genes are one proposed evolutionary basis for this behavior. While the behavioral experiments and comparisons between expression profiles in the different chemosensory organs provide some clues of the mechanisms and developmental pathways involved in anthropophily, these areas of inquiry still await substantial additional study.

REFERENCES

1. UNICEF. Levels & Trends in Child Mortality: Report 2019-Estimates developed by the UN Inter-agency Group for Child Mortality Estimation. *Unicef/Who/Wb/Un*. 2019:1-32.

2. World Health Organization. *World Malaria Report 2019*.; 2019. https://www.who.int/publications-detail/world-malaria-report-2019.

3. Gillies MT. Selection for Host Preference in *Anopheles gambiae*. *Nature*. 1964;203:852-854. doi:10.1038/203852a0.

4. Mukwaya LG. Genetic control of feeding preferences in the mosquitoes *Aedes (Stegomyia) simpsoni* and *aegypti*. *Physiol Entomol*. 1977;2(2):133-145. doi:10.1111/j.1365-3032.1977.tb00091.x.

5. Pates H V., Curtis CF, Takken W. Hybridization studies to modify the host preference of *Anopheles gambiae*. *Med Vet Entomol*. 2014;28(S1):68-74. doi:10.1111/mve.12070.

6. van Breugel F, Riffell J, Fairhall A, Dickinson MH. Mosquitoes use vision to associate odor plumes with thermal targets. *Curr Biol*. 2015;25(16):2123-2129. doi:10.1016/j.cub.2015.06.046.

7. Liu MZ, Vosshall LB. General Visual and Contingent Thermal Cues Interact to Elicit Attraction in Female *Aedes aegypti* Mosquitoes. *Curr Biol*. 2019. doi:10.1016/j.cub.2019.06.001.

8. Raji JI, Gonzalez S, DeGennaro M. *Aedes aegypti Ir8a* mutant female mosquitoes show increased attraction to standing water. *Commun Integr Biol*. 2019;12(1):181-186. doi:10.1080/19420889.2019.1681063.

9. Takken W. The Role of Olfaction in Host-Seeking of Mosquitoes: A Review. *Insect Sci its Appl*. 1991;12(1/2/3):287-295.

10. Maekawa E, Aonuma H, Nelson B, Yoshimura A, Tokunaga F, Fukumoto S, Kanuka H. The role of proboscis of the malaria vector mosquito *Anopheles stephensi* in host-seeking behavior. *Parasit Vectors*. 2011;4(1):10. doi:10.1186/1756-3305-4-10.

11. Barrón MG, Paupy C, Rahola N, Akone-Ella O, Ngangue MF, Wilson-Bahun TA, Pombi M, Kengne P, Costantini C, Simard F, González J, Ayala D. A new species in the major malaria vector complex sheds light on reticulated species evolution. *Sci Rep*. 2019;9(1):14753. doi:10.1038/s41598-019-49065-5.

12. Davidson G. *Anopheles gambiae* Complex. *Nature*. 1962;196(4857):907. https://www.nature.com/articles/196907a0.

13. Sinka ME, Bangs MJ, Manguin S, Coetzee M, Mbogo CM, Hemingway J, Patil AP, Temperley WH, Gething PW, Kabaria CW, Okara RM, Van Boeckel T, Godfray HCJ, Harbach RE, Hay SI. The dominant *Anopheles* vectors of human malaria in Africa, Europe and the Middle East: occurrence data, distribution maps and bionomic précis. *Parasit Vectors*. 2010;3(1):117. doi:10.1186/1756-3305-3-117.

14. Coetzee M, Hunt RH, Wilkerson R, della Torre A, Coulibaly MB, Besansky NJ. *Anopheles coluzzii* and *Anopheles amharicus*, new members of the *Anopheles gambiae* complex. *Zootaxa*. 2013;3619(3):246-274. doi:10.11646/zootaxa.3619.3.2.

15. Della Torre A, Tu Z, Petrarca V. On the distribution and genetic differentiation of *Anopheles gambiae* s.s. molecular forms. *Insect Biochem Mol Biol*. 2005;35(7):755-769. doi:10.1016/j.ibmb.2005.02.006.

16. Tirados I, Costantini C, Gibnson G, Torr SJ. Blood-feeding behaviour of the malarial mosquito *Anopheles arabiensis*: implications for vector control. *Med Vet Entomol*. 2006;20(4):425-437. doi:10.1111/j.1365-2915.2006.652.x.

17. Takken W, Knols BGJ. Odor-Mediated Behavior of Afrotropical Malaria Mosquitoes. *Annu Rev Entomol*. 1999;44(1):131-157. doi:10.1146/annurev.ento.44.1.131.

18. Dekker T, Takken W. Differential responses of mosquito sibling species *Anopheles arabiensis* and *An. quadriannulatus* to carbon dioxide, a man or a calf. *Med Vet Entomol*. 1998;12(2):136-140. doi:10.1046/j.1365-2915.1998.00073.x.

19. Pates HV, Takken W, Curtis CF. Laboratory studies on the olfactory behaviour of *Anopheles quadriannulatus*. *Entomol Exp Appl*. 2005;114(2):153-159. doi:10.1111/j.1570-7458.2005.00249.x.

20. White GB. *Anopheles bwambae* sp.n., a malaria vector in the Semliki Valley, Uganda, and its relationships with other sibling species of the *An.gambiae* complex (Diptera: Culicidae). *Syst Entomol*. 1985;10(4):501-522. doi:10.1111/j.1365-3113.1985.tb00155.x.

21. Tennessen JA, Ingham VA, Toé KH, Guelbéogo WM, Sagnon N, Kuzma R, Ranson H, Neafsey DE. A population genomic unveiling of a new cryptic mosquito taxon within the malaria-transmitting *Anopheles gambiae* complex. *bioRxiv*. May 2020:2020.05.26.116988. doi:10.1101/2020.05.26.116988.

22. Sinka ME, Bangs MJ, Manguin S, Chareonviriyaphap T, Patil AP, Temperley WH, Gething PW, Elyazar IR, Kabaria CW, Harbach RE, Hay SI. The dominant *Anopheles* vectors of human malaria in the Asia-Pacific region: occurrence data, distribution maps and bionomic précis. *Parasit Vectors*. 2011;4(1):89. doi:10.1186/1756-3305-4-89.

23. Sinka ME, Rubio-Palis Y, Manguin S, Patil AP, Temperley WH, Gething PW, Van Boeckel T, Kabaria CW, Harbach RE, Hay SI. The dominant *Anopheles* vectors of human malaria in the Americas: occurrence data, distribution maps and bionomic précis. *Parasit Vectors*. 2010;3(1):72. doi:10.1186/1756-3305-3-72.

24. Ranson H, Lissenden N. Insecticide Resistance in African *Anopheles* Mosquitoes: A Worsening Situation that Needs Urgent Action to Maintain Malaria Control. *Trends Parasitol*. 2016;32(3):187-196. doi:10.1016/j.pt.2015.11.010.

25. Haldar K, Bhattacharjee S, Safeukui I. Drug resistance in *Plasmodium*. *Nat Rev Microbiol*. 2018;16(3):156-170. doi:10.1038/nrmicro.2017.161.

26. Carrasco D, Lefèvre T, Moiroux N, Pennetier C, Chandre F, Cohuet A. Behavioural adaptations of mosquito vectors to insecticide control. *Curr Opin Insect Sci*. 2019;34(March):48-54. doi:10.1016/j.cois.2019.03.005.

27. Sherrard-Smith E, Skarp JE, Beale AD, Fornadel C, Norris LC, Moore SJ, Mihreteab S, Charlwood JD, Bhatt S, Winskill P, Griffin JT, Churcher TS. Mosquito feeding behavior and how it influences residual malaria transmission across Africa. *Proc Natl Acad Sci U S A*. 2019;116(30):15086-15096. doi:10.1073/pnas.1820646116.

28. Borst A. *Drosophila*'s View on Insect Vision. *Curr Biol*. 2009;19(1):R36-R47. doi:10.1016/j.cub.2008.11.001.

29. Buschbeck EK, Friedrich M. Evolution of Insect Eyes: Tales of Ancient Heritage, Deconstruction, Reconstruction, Remodeling, and Recycling. *Evol Educ Outreach*. 2008;1(4):448-462. doi:10.1007/s12052-008-0086-z.

30. Fleischer J, Pregitzer P, Breer H, Krieger J. Access to the odor world: olfactory receptors and their role for signal transduction in insects. *Cell Mol Life Sci*. 2018;75(3):485-508. doi:10.1007/s00018-017-2627-5.

31. Keil TA. Functional Morphology of Insect Mechanoreceptors. *Microsc Res Tech*. 1997;39(6):506-531. doi:10.1002/(SICI)1097-0029(19971215)39:6<506::AID-JEMT5>3.0.CO;2-B.

32. Duggan A, García-Añoveros J, Corey DP. Insect mechanoreception: What a long, strange TRP it's been. *Curr Biol*. 2000;10(10):384-387. doi:10.1016/S0960-

9822(00)00478-4.

33. Yack JE. The Structure and Function of Auditory Chordotonal Organs in Insects. *Microsc Res Tech*. 2004;63(6):315-337. doi:10.1002/jemt.20051.

34. Steinbrecht RA. Chemo-, Hygro-, and Thermoreceptors. In: Bereiter-Hahn J, Matoltsy AG, Richards KS, eds. *Biology of the Integument*. New York: Springer Berlin Heidelberg; 1984:523-553. doi:10.1007/978-3-642-51593-4_28.

35. Tichy H, Gingl E. Problems in Hygro- and Thermoreception. In: Barth FC, Schmid A, eds. *Ecology of Sensing*. New York: Springer Berlin Heidelberg; 2001:271-287. doi:10.1007/978-3-662-22644-5_14.

36. Tracey WD, Wilson RI, Laurent G, Benzer S. *painless*, a *Drosophila* gene essential for nociception. *Cell*. 2003;113(2):261-273. doi:10.1016/S0092-8674(03)00272-1.

37. McIver SB. Sensilla of Mosquitoes (Diptera: Culicidae). *J Med Entomol*. 1982;19(5):489-535. doi:10.1093/jmedent/19.5.489.

38. Bybee SM, Johnson KK, Gering EJ, Whiting MF, Crandall KA. All the better to see you with: A review of odonate color vision with transcriptomic insight into the odonate eye. *Org Divers Evol*. 2012;12(3):241-250. doi:10.1007/s13127-012-0090-6.

39. Cronin TW, Järvilehto M, Weckström M, Lall AB. Tuning of photoreceptor spectral sensitivity in fireflies (Coleoptera: Lampyridae). *J Comp Physiol - A Sensory, Neural, Behav Physiol*. 2000;186(1):1-12. doi:10.1007/s003590050001.

40. Henry CS. Singing and cryptic speciation in insects. *Trends Ecol Evol*. 1994;9(10):388-392. doi:10.1016/0169-5347(94)90061-2.

41. Pennetier C, Warren B, Dabiré KR, Russell IJ, Gibson G. "Singing on the Wing" as a Mechanism for Species Recognition in the Malarial Mosquito *Anopheles gambiae*. *Curr Biol*. 2010;20(2):131-136. doi:10.1016/j.cub.2009.11.040.

42. Engsontia P, Sangket U, Chotigeat W, Satasook C. Molecular Evolution of the Odorant and Gustatory Receptor Genes in Lepidopteran Insects: Implications for Their Adaptation and Speciation. *J Mol Evol*. 2014;79(1-2):21-39. doi:10.1007/s00239-014-9633-0.

43. McBride CS, Arguello JR. Five Drosophila Genomes Reveal Nonneutral Evolution and the Signature of Host Specialization in the Chemoreceptor Superfamily. *Genetics*. 2007;177(November 2007):1395-1416. doi:10.1534/genetics.107.078683.

44. McBride CS, Baier F, Omondi AB, Spitzer SA, Lutomiah J, Sang R, Ignell R, Vosshall LB. Evolution of mosquito preference for humans linked to an odorant receptor. *Nature*. 2014;515(7526):222-227. doi:10.1038/nature13964.

45. Rinker DC, Zhou X, Pitts RJ, Rokas A, Zwiebel LJ. Antennal transcriptome profiles of anopheline mosquitoes reveal human host olfactory specialization in *Anopheles gambiae*. *BMC Genomics*. 2013;14(1):749. doi:10.1186/1471-2164-14-749.

46. Athrey G, Cosme L V., Popkin-Hall Z, Pathikonda S, Takken W, Slotman MA. Chemosensory gene expression in olfactory organs of the anthropophilic *Anopheles coluzzii* and zoophilic *Anopheles quadriannulatus*. *BMC Genomics*. 2017;18(1):751. doi:10.1186/s12864-017-4122-7.

47. Athrey G, Popkin-Hall Z, Cosme LV, Takken W, Slotman MA. Species and sex-specific chemosensory gene expression in *Anopheles coluzzii* and *An. quadriannulatus* antennae. *Parasit Vectors*. 2020;13(1):212. doi:10.1186/s13071-020-04085-3.

48. Athrey G, Popkin-Hall ZR, Takken W, Slotman MA. The Expression of Chemosensory Genes in Male Maxillary Palps of *Anopheles coluzzii* and *Anopheles quadriannulatus*. In review, *J Med Entomol*.

49. Qiu YT, van Loon JJA, Takken W, Meijerink J, Smid HM. Olfactory coding in antennal neurons of the malaria mosquito, *Anopheles gambiae*. *Chem Senses*. 2006;31(9):845-863. doi:10.1093/chemse/bjl027.

50. Nikbakhtzadeh MR, Terbot II JW, Otienoburu PE, Foster WA. Olfactory basis of floral preference of the malaria vector *Anopheles gambiae* (Diptera: Culicidae) among common African plants. *J Vector Ecol*. 2014;39(2):372-383. doi:10.3376/i1081-1710-39-372.

51. Whiteman NK, Pierce NE. Delicious poison: genetics of *Drosophila* host plant preference. *Trends Ecol Evol*. 2008;23(9):473-478. doi:10.1016/j.tree.2008.05.010.

52. Miyamoto T, Slone J, Song X, Amrein H. A Fructose Receptor Functions as a Nutrient Sensor in the *Drosophila* Brain. *Cell*. 2012;151(5):1113-1125. doi:10.1016/j.cell.2012.10.024.

53. Bray S, Amrein H. A putative *Drosophila* pheromone receptor expressed in male-specific taste neurons is required for efficient courtship. *Neuron*. 2003;39(6):1019-1029. doi:10.1016/S0896-6273(03)00542-7.

54. Choo Y-M, Buss GK, Tan K, Leal WS. Multitasking roles of mosquito labrum in

oviposition and blood feeding. *Front Physiol*. 2015;6(OCT):1-11. doi:10.3389/fphys.2015.00306.

55. Montell C. A taste of the *Drosophila* gustatory receptors. *Curr Opin Neurobiol*. 2009;19(4):345-353. doi:10.1016/j.conb.2009.07.001.

56. Joseph RM, Carlson JR. *Drosophila* Chemoreceptors: A Molecular Interface Between the Chemical World and the Brain. *Trends Genet*. 2015;31(12):683-695. doi:10.1016/j.tig.2015.09.005.

57. Schultzhaus JN, Saleem S, Iftikhar H, Carney GE. The role of the *Drosophila* lateral horn in olfactory information processing and behavioral response. *J Insect Physiol*. 2017;98(10-11):29-37. doi:10.1016/j.jinsphys.2016.11.007.

58. Riabinina O, Task D, Marr E, Lin C-C, Alford R, O'Brochta DA, Potter CJ. Organization of olfactory centres in the malaria mosquito *Anopheles gambiae*. *Nat Commun*. 2016;7:13010. doi:10.1038/ncomms13010.

59. Hallem EA, Ho MG, Carlson JR. The Molecular Basis of Odor Coding in the *Drosophila* Antenna. *Cell*. 2004;117(7):965-979. doi:10.1016/j.cell.2004.05.012.

60. Benton R, Vannice KS, Gomez-Diaz C, Vosshall LB. Variant Ionotropic Glutamate Receptors as Chemosensory Receptors in *Drosophila*. *Cell*. 2009;136(1):149-162. doi:10.1016/j.cell.2008.12.001.

61. Clyne PJ. Candidate Taste Receptors in *Drosophila*. *Science (80-)*. 2000;287(5459):1830-1834. doi:10.1126/science.287.5459.1830.

62. Pelosi P, Iovinella I, Felicioli A, Dani FR. Soluble proteins of chemical communication: an overview across arthropods. *Front Physiol*. 2014;5(August):320. doi:10.3389/fphys.2014.00320.

63. Vogt RG. Molecular Basis of Pheromone Detection in Insects. In: Gilbert LI, Iatro K, Gill S, eds. *Comprehensive Insect Physiology, Biochemistry, Pharmacology and Molecular Biology*. Vol 3. London: Elsevier; 2005:753-804.

64. Xu A, Park S-K, D'Mello S, Kim E, Wang Q, Pikielny CW. Novel genes expressed in subsets of chemosensory sensilla on the front legs of male *Drosophila melanogaster*. *Cell Tissue Res*. 2002;307(3):381-392. doi:10.1007/s00441-002-0524-0.

65. Park SK, Mann KJ, Lin H, Starostina E, Kolski-Andreaco A, Pikielny CW. A *Drosophila* Protein Specific to Pheromone-Sensing Gustatory Hairs Delays Males' Copulation Attempts. *Curr Biol*. 2006;16(11):1154-1159. doi:10.1016/j.cub.2006.04.028.

66. Leung NY, Thakur DP, Gurav AS, Kim SH, Di Pizio A, Niv MY, Montell C. Functions of Opsins in *Drosophila* Taste. *Curr Biol*. 2020;30(8):1367-1379.e6. doi:10.1016/j.cub.2020.01.068.

67. Li Q, DeBeaubien NA, Sokabe T, Montell C. Temperature and Sweet Taste Integration in *Drosophila*. *Curr Biol*. 2020;30(11):2051-2067.e5. doi:10.1016/j.cub.2020.03.066.

68. Larsson MC, Domingos AI, Jones WD, Chiappe ME, Amrein H, Vosshall LB. *Or83b* Encodes a Broadly Expressed Odorant Receptor Essential for *Drosophila* Olfaction. *Neuron*. 2004;43(5):703-714. doi:10.1016/j.neuron.2004.08.019.

69. Benton R, Sachse S, Michnick SW, Vosshall LB. Atypical Membrane Topology and Heteromeric Function of *Drosophila* Odorant Receptors In Vivo. Bjorkman P, ed. *PLoS Biol*. 2006;4(2):e20. doi:10.1371/journal.pbio.0040020.

70. Suh E, Bohbot JD, Zwiebel LJ. Peripheral olfactory signaling in insects. *Curr Opin Insect Sci*. 2014;6:86-92. doi:10.1016/j.cois.2014.10.006.

71. Rytz R, Croset V, Benton R. Ionotropic Receptors (IRs): Chemosensory ionotropic glutamate receptors in *Drosophila* and beyond. *Insect Biochem Mol Biol*. 2013;43(9):888-897. doi:10.1016/j.ibmb.2013.02.007.

72. Amrein H. Mechanism of Taste Perception in *Drosophila*. In: Zufall F, Munger SD, eds. *Chemosensory Transduction: The Detection of Odors, Tastes, and Other Chemostimuli*. Elsevier; 2016:245-269. doi:10.1016/B978-0-12-801694-7.00014-7.

73. Pitts RJ, Derryberry SL, Zhang Z, Zwiebel LJ. Variant Ionotropic Receptors in the Malaria Vector Mosquito *Anopheles gambiae* Tuned to Amines and Carboxylic Acids. *Sci Rep*. 2017;7(January):40297. doi:10.1038/srep40297.

74. Hussain A, Zhang M, Üçpunar HK, Svensson T, Quillery E, Gompel N, Ignell R, Grunwald Kadow IC. Ionotropic Chemosensory Receptors Mediate the Taste and Smell of Polyamines. *PLoS Biol*. 2016;14(5):1-30. doi:10.1371/journal.pbio.1002454.

75. Eyun S, Soh HY, Posavi M, Munro JB, Hughes DST, Murali SC, Qu J, Dugan S, Lee SL, Chao H, Dinh H, Han Y, Doddapaneni H, Worley KC, Muzny DM, Park E-O, Silva JC, Gibbs RA, Richards S, Lee CE. Evolutionary History of Chemosensory-Related Gene Families across the Arthropoda. *Mol Biol Evol*. 2017;34(8):1838-1862. doi:10.1093/molbev/msx147.

76. Croset V, Rytz R, Cummins SF, Budd A, Brawand D, Kaessmann H, Gibson TJ, Benton R. Ancient protostome origin of chemosensory ionotropic glutamate

receptors and the evolution of insect taste and olfaction. *PLoS Genet*. 2010;6(8). doi:10.1371/journal.pgen.1001064.

77. Silbering AF, Rytz R, Grosjean Y, Abuin L, Ramdya P, Jefferis GSXE, Benton R. Complementary function and integrated wiring of the evolutionarily distinct Drosophila olfactory subsystems. *J Neurosci*. 2011;31(38):13357-13375. doi:10.1523/JNEUROSCI.2360-11.2011.

78. Zhang HJ, Anderson AR, Trowell SC, Luo AR, Xiang ZH, Xia QY. Topological and functional characterization of an insect gustatory receptor. *PLoS One*. 2011;6(8):1-10. doi:10.1371/journal.pone.0024111.

79. Robertson HM, Warr CG, Carlson JR. Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A*. 2003;100 Suppl:14537-14542. doi:10.1073/pnas.2335847100.

80. Missbach C, Dweck HKM, Vogel H, Vilcinskas A, Stensmyr MC, Hansson BS, Grosse-Wilde E. Evolution of insect olfactory receptors. *Elife*. 2014;2014(3):1-22. doi:10.7554/eLife.02115.

81. Robertson HM. The insect chemoreceptor superfamily is ancient in animals. *Chem Senses*. 2015;40(9):609-614. doi:10.1093/chemse/bjv046.

82. Gardiner A, Barker D, Butlin RK, Jordan WC, Ritchie MG. Drosophila chemoreceptor gene evolution: Selection, specialization and genome size. *Mol Ecol*. 2008;17(7):1648-1657. doi:10.1111/j.1365-294X.2008.03713.x.

83. Sánchez-Gracia A, Vieira FG, Rozas J. Molecular evolution of the major chemosensory gene families in insects. *Heredity (Edinb)*. 2009;103(3):208-216. doi:10.1038/hdy.2009.55.

84. Miyamoto T, Amrein H. Suppression of male courtship by a *Drosophila* pheromone receptor. *Nat Neurosci*. 2008;11(8):874-876. doi:10.1038/nn.2161.

85. Weiss LA, Dahanukar A, Kwon JY, Banerjee D, Carlson JR. The Molecular and Cellular Basis of Bitter Taste in *Drosophila*. *Neuron*. 2011;69(2):258-272. doi:10.1016/j.neuron.2011.01.001.

86. Delventhal R, Carlson JR. Bitter taste receptors confer diverse functions to neurons. *Elife*. 2016;5(FEBRUARY2016):1-23. doi:10.7554/eLife.11181.

87. Fujii S, Yavuz A, Slone J, Jagge C, Song X, Amrein H. *Drosophila* sugar receptors in sweet taste perception, olfaction, and internal nutrient sensing. *Curr Biol*. 2015;25(5):621-627. doi:10.1016/j.cub.2014.12.058.

88. Sung HY, Jeong YT, Lim JY, Kim H, Oh SM, Hwang SW, Kwon JY, Moon SJ. Heterogeneity in the *Drosophila* gustatory receptor complexes that detect aversive compounds. *Nat Commun*. 2017;8(1):1484. doi:10.1038/s41467-017-01639-5.

89. Shim J, Lee Y, Jeong YT, Kim Y, Lee MG, Montell C, Moon SJ. The full repertoire of *Drosophila* gustatory receptors for detecting an aversive compound. *Nat Commun*. 2015;6(1):8867. doi:10.1038/ncomms9867.

90. Dweck HKM, Carlson JR. Molecular Logic and Evolution of Bitter Taste in *Drosophila*. *Curr Biol*. 2020;30(1):17-30.e3. doi:10.1016/j.cub.2019.11.005.

91. Lu T, Qiu YT, Wang G, Kwon JY, Rutzler M, Kwon H-W, Pitts RJ, van Loon JJA, Takken W, Carlson JR, Zwiebel LJ. Odor Coding in the Maxillary Palp of the Malaria Vector Mosquito *Anopheles gambiae*. *Curr Biol*. 2007;17(18):1533-1544. doi:10.1016/j.cub.2007.07.062.

92. Jones WD, Cayirlioglu P, Kadow IG, Vosshall LB. Two chemosensory receptors together mediate carbon dioxide detection in *Drosophila*. *Nature*. 2007;445(7123):86-90. doi:10.1038/nature05466.

93. Xiao S, Sun JS, Carlson JR. Robust olfactory responses in the absence of odorant binding proteins. *Elife*. 2019;8:1-17. doi:10.7554/eLife.51040.

94. Sun JS, Xiao S, Carlson JR. The diverse small proteins called odorant-binding proteins. *Open Biol*. 2018;8(12):180208. doi:10.1098/rsob.180208.

95. Saveer AM, Pitts RJ, Ferguson ST, Zwiebel LJ. Characterization of Chemosensory Responses on the Labellum of the Malaria Vector Mosquito, *Anopheles coluzzii*. *Sci Rep*. 2018;8(1):5656. doi:10.1038/s41598-018-23987-y.

96. Sparks JT, Bohbot JD, Dickens JC. The genetics of chemoreception in the labella and tarsi of *Aedes aegypti*. *Insect Biochem Mol Biol*. 2014;48(1):8-16. doi:10.1016/j.ibmb.2014.02.004.

97. Kim MS, Repp A, Smith DP. LUSH odorant-binding protein mediates chemosensory responses to alcohols in *Drosophila melanogaster*. *Genetics*. 1998;150(2):711-721.

98. Xu P, Atkinson R, Jones DNM, Smith DP. *Drosophila* OBP LUSH Is Required for Activity of Pheromone-Sensitive Neurons. *Neuron*. 2005;45(2):193-200. doi:10.1016/j.neuron.2004.12.031.

99. Scheuermann EA, Smith DP. Odor-Specific Deactivation Defects in a *Drosophila* Odorant-Binding Protein Mutant. *Genetics*. 2019;213(3):897-909. doi:10.1534/genetics.119.302629.

100. Roth LM. Loci of sensory end-organs used by mosquitoes (*Aedes aegypti* (L.) and *Anopheles quadrimaculatus* Say) in receiving host stimuli. *Ann Entomol Soc Am*. 1951;44(1):59–74. doi:10.1093/aesa/44.1.59.

101. Mukwaya LG. The role of olfaction in host preference by *Aedes (Stegomyia) simpsoni* and *Ae. aegypti*. *Physiol Entomol*. 1976;1(4):271-276. doi:10.1111/j.1365-3032.1976.tb00976.x.

102. Rahm U. Die Funktion der Antennen, Palpen und Tarsen von *Aedes aegypti* L. beim Aufsuchen des Wirtes. *Rev suisse Zool*. 1958;65(37):779-792. doi:10.5962/bhl.part.75553.

103. Ismail IAH. Sense organs in the antennae of *Anopheles maculipennis atroparvus* (v. Thiel), and their possible function in relation to the attraction of female mosquito to man. *Acta Trop*. 1962;19(1):1-58. https://www.cabdirect.org/cabdirect/abstract/19641000234. Accessed May 18, 2020.

104. Bar-Zeev M. The Location of Hygroreceptors and Moisture Receptors in *Aedes aegypti (L.)*. *Entomol Exp Appl*. 1960;3:251-256. doi:10.1017/CBO9781107415324.004.

105. Kellogg FE. Water vapour and carbon dioxide receptors in *Aedes aegypti*. *J Insect Physiol*. 1970;16(1):99-108. doi:10.1016/0022-1910(70)90117-4.

106. Omer SM, Gillies MT. Loss of response to carbon dioxide in palpectomized female mosquitoes. *Entomol Exp Appl*. 1971;14(2):251-252. doi:10.1111/j.1570-7458.1971.tb00162.x.

107. Grant AJ, Aghajanian JG, O'Connell RJ, Wigton BE. Electrophysiological responses of receptor neurons in mosquito maxillary palp sensilla to carbon dioxide. *J Comp Physiol A*. 1995;177(4):389-396. doi:10.1007/BF00187475.

108. Owen WB. The contact chemoreceptor organs of the mosquito and their function in feeding behaviour. *J Insect Physiol*. 1963;9(1):73-87. doi:10.1016/0022-1910(63)90085-4.

109. Larsen JR, Owen WB. Structure and Function of the Ligula of the Mosquito *Culiseta inornata* (Williston). *Trans Am Microsc Soc*. 1971;90(3):294. doi:10.2307/3225189.

110. Salama HS. The function of mosquito taste receptors. *J Insect Physiol*. 1966;12(9):1051-1060. doi:10.1016/0022-1910(66)90120-X.

111. Benton R. The neurobiology of gustation in insect disease vectors: progress and

potential. *Curr Opin Insect Sci*. 2017;20:19-27. doi:10.1016/j.cois.2017.02.003.

112. Howlett FM. The Influence of Temperature upon the Biting of Mosquitoes. *Parasitology*. 1910;3(4):479-484.

113. Marchand W. First Account of a Thermotropism in *Anopheles punctipennis* with Bionomic Observations. *Psyche (Stuttg)*. 1916;V(December):130-135.

114. Peterson DG, Brown AWA. Studies of the Responses of the Female *Aëdes* Mosquito. Part III. The Response of *Aëdes aegypti* (L.) to a Warm Body and its Radiation. *Bull Entomol Res*. 1951;42(3):535-541. doi:10.1017/S0007485300028935.

115. Müller W. Die Distanz- und Kontakt-Orientierung der Stechmücken (*Aedes aegypti*) (Wirtsfindung, Stechverhalten und Blutmahlzeit). *Z Vgl Physiol*. 1968;58(3):241-303.

116. Schreck CE, Kline DL, Carlson DA. Mosquito Attraction to Substances from the Skin of Different Humans. *J Am Mosq Control Assoc*. 1990;6(3):406-410.

117. De Jong R, Knols BGJ. Selection of biting sites on man by two malaria mosquito species. *Experientia*. 1995;51(1):80-84. doi:10.1007/BF01964925.

118. Eiras AE, Jepson PC. Responses of female *Aedes aegypti* (Diptera: Culicidae) to host odours and convection currents using an olfactometer bioassay. *Bull Entomol Res*. 1994;84(02):207. doi:10.1017/S0007485300039705.

119. Dekker T, Takken W, Knols BGJ, Bouman E, Van de Laak S, de Bever A, Huisman PWT. Selection of biting sites on a human host by *Anopheles gambiae* s.s., *An. arabiensis* and *An. quadriannulatus*. *Entomol Exp Appl*. 1998;87:295-300. doi:10.1046/j.1570-7458.1998.00334.x.

120. Braack L, Hunt R, Koekemoer LL, Gericke A, Munhenga G, Haddow AD, Becker P, Okia M, Kimera I, Coetzee M. Biting behaviour of African malaria vectors: 1. where do the main vector species bite on the human body? *Parasit Vectors*. 2015;8(1):76. doi:10.1186/s13071-015-0677-9.

121. Verhulst NO, Weldegergis BT, Menger D, Takken W. Attractiveness of volatiles from different body parts to the malaria mosquito *Anopheles coluzzii* is affected by deodorant compounds. *Sci Rep*. 2016;6(January):27141. doi:10.1038/srep27141.

122. Rudolfs W. Chemotropism of Mosquitoes. *New Jersey Agric Exp Station*. 1922;367.

123. Thiel PH van, Weurman C. L'attraction exercée sur *Anopheles maculipennis atroparvus* par l'acide carbonique dans l'aparreil de choix II. *Acta Trop*. 1947;IV(1):1-9.

124. Willis ER, Roth LM. Reactions of *Aedes aegypti (L.)* to carbon dioxide. *J Exp Zool*. 1952;121(1):149-179. doi:10.1002/jez.1401210107.

125. Acree Jr. F, Turner RB, Gouck HK, Beroza M, Smith N. L-Lactic Acid: A Mosquito Attractant Isolated from Humans. *Science (80-)*. 1968;161(3):1346-1347.

126. Takken W, Knols BGJ. Flight Behaviour of *Anopheles gambiae* Giles (Diptera: Culicidae) in Response to Host Stimuli: A Windtunnel Study. *Exp Appl Entomol Proc Netherlands Entomol Soc (N E V)*. 1990;1:121-128.

127. Knols BGJ, Takken W, de Jong R. Influence of human breath on selection of biting sites by *Anopheles albimanus*. *J Am Mosq Control Assoc*. 1994;10(3):423-426.

128. Healy TP, Copland MJ. Activation of *Anopheles gambiae* mosquitoes by carbon dioxide and human breath. *Med Vet Entomol*. 1995;9(3):331-336. doi:10.1111/j.1365-2915.1995.tb00143.x.

129. Takken W, Dekker T, Wijnholds YG. Odor-Mediated Flight Behavior of *Anopheles gambiae* Giles Sensu Stricto and *An. stephensi* Liston in Response to CO2, Acetone, and 1-Octen-3-ol (Diptera: Culicidae). *J Insect Behav*. 1997;10(3):395-407. doi:10.1007/BF02765606.

130. Braks M, Cork A, Takken W. Olfactometer Studies on the Attraction of *Anopheles gambiae* sensu stricto (Diptera: Culicidae) to Human Sweat. *Exp Appl Entomol Proc Netherlands Entomol Soc (N E V)*. 1997;8:99-104.

131. de Jong R, Knols BGJ. Olfactory responses of host-seeking *Anopheles gambiae s. s.* Giles (Diptera: Culicidae). *Acta Trop*. 1995;59:333-335.

132. Knols BGJ, van Loon JJA, Cork A, Robinson RD, Adam W, Meijerink J, Jong R De, Takken W. Behavioural and electrophysiological responses of the female malaria mosquito *Anopheles gambiae* (Diptera: Culicidae) to Limburger cheese volatiles. *Bull Entomol Res*. 1997;87(2):151-159. doi:10.1017/S0007485300027292.

133. Braks MAH, Takken W. Incubated Sweat but Not Fresh Sweat Attracts the Malaria Mosquito *Anopheles gambiae* sensu stricto. *J Chem Ecol*. 1999;25(25):663-672. doi:10.1023/A:1020970307748.

134. Takken W, Verhulst NO. Chemical signaling in mosquito–host interactions: the role of human skin microbiota. *Curr Opin Insect Sci*. 2017;20:68-74. doi:10.1016/j.cois.2017.03.011.

135. Dekker T, Geier M, Cardé RT. Carbon dioxide instantly sensitizes female yellow fever mosquitoes to human skin odours. *J Exp Biol*. 2005;208(15):2963-2972. doi:10.1242/jeb.01736.

136. Dekker T, Carde RT. Moment-to-moment flight manoeuvres of the female yellow fever mosquito (*Aedes aegypti* L.) in response to plumes of carbon dioxide and human skin odour. *J Exp Biol*. 2011;214(20):3480-3494. doi:10.1242/jeb.055186.

137. Lacey ES, Ray A, Cardé RT. Close encounters: contributions of carbon dioxide and human skin odour to finding and landing on a host in *Aedes aegypti*. *Physiol Entomol*. 2014;39(1):60-68. doi:10.1111/phen.12048.

138. Majeed S, Hill SR, Ignell R. Impact of elevated CO2 background levels on the host-seeking behaviour of *Aedes aegypti*. *J Exp Biol*. 2014;217(Pt 4):598-604. doi:10.1242/jeb.092718.

139. Webster B, Lacey ES, Cardé RT. Waiting with Bated Breath: Opportunistic Orientation to Human Odor in the Malaria Mosquito, *Anopheles gambiae*, is Modulated by Minute Changes in Carbon Dioxide Concentration. *J Chem Ecol*. 2015;41(1):59-66. doi:10.1007/s10886-014-0542-x.

140. Majeed S, Hill SR, Dekker T, Ignell R. Detection and perception of generic host volatiles by mosquitoes: responses to CO2 constrains host-seeking behaviour. *R Soc Open Sci*. 2017;4(5):170189. doi:10.1098/rsos.170189.

141. Ghaninia M, Majeed S, Dekker T, Hill SR, Ignell R. Hold your breath – Differential behavioral and sensory acuity of mosquitoes to acetone and carbon dioxide. Dickens JC, ed. *PLoS One*. 2019;14(12):e0226815. doi:10.1371/journal.pone.0226815.

142. Lacey ES, Cardé RT. Activation, orientation and landing of female *Culex quinquefasciatus* in response to carbon dioxide and odour from human feet: 3-D flight analysis in a wind tunnel. *Med Vet Entomol*. 2011;25(1):94-103. doi:10.1111/j.1365-2915.2010.00921.x.

143. Vantaux A, Lefèvre T, Dabiré KR, Cohuet A. Individual experience affects host choice in malaria vector mosquitoes. *Parasit Vectors*. 2014;7:249. doi:10.1186/1756-3305-7-249.

144. Vinauger C, Lahondère C, Wolff GH, Locke LT, Liaw JE, Parrish JZ, Akbari OS, Dickinson MH, Riffell JA. Modulation of Host Learning in *Aedes aegypti*

Mosquitoes. *Curr Biol*. 2018;28(3):333-344.e8. doi:10.1016/j.cub.2017.12.015.

145. Eilerts D, VanderGiessen M, Bose E, Broxton K, Vinauger C. Odor-Specific Daily Rhythms in the Olfactory Sensitivity and Behavior of *Aedes aegypti* Mosquitoes. *Insects*. 2018;9(4):147. doi:10.3390/insects9040147.

146. Foster WA, Takken W. Nectar-related vs. human-related volatiles: behavioural response and choice by female and male *Anopheles gambiae* (Diptera: Culicidae) between emergence and first feeding. *Bull Entomol Res*. 2004;94(2):145-157. doi:10.1079/BER2003288.

147. Dekker T, Takken W, Braks AH. Innate preference for host-odour blends modulates degree of anthropophagy of *Anopheles gambiae* s.l. (Diptera: Culicidae). *J Med Entomol*. 2001;38:868-871.

148. Pates HV, Takken W, Stuke K, Curtis CF. Differential behaviour of *Anopheles gambiae* sensu stricto (Diptera: Culicidae) to human and cow odours in the laboratory. *Bull Entomol Res*. 2001;91(04):289-296. doi:10.1079/BER200198.

149. Pates H V., Takken W, Curtis CF, Huisman PW, Akinpelu O, Gill GS. Unexpected anthropophagic behaviour in *Anopheles quadriannulatus*. *Med Vet Entomol*. 2001;15(3):293-298. doi:10.1046/j.0269-283X.2001.00310.x.

150. Costantini C, Sagnon N, Torre A Della, Diallo M, Brady J, Gibson G, Coluzzi M. Odor-mediated host preferences of West African mosquitoes, with particular reference to malaria vectors. *Am J Trop Med Hyg*. 1998;58(1):56-63. doi:10.4269/ajtmh.1998.58.56.

151. Torr SJ, Torre A della, Calzetta M, Costantini C, Vale GA. Towards a fuller understanding of mosquito behaviour: use of electrocuting grids to compare the odour-orientated responses of *Anopheles arabiensis* and *An. quadriannulatus* in the field. *Med Vet Entomol*. 2008;22(2):93-108. doi:10.1111/j.1365-2915.2008.00723.x.

152. Pitts RJ, Rinker DC, Jones PL, Rokas A, Zwiebel LJ. Transcriptome profiling of chemosensory appendages in the malaria vector *Anopheles gambiae* reveals tissue- and sex-specific signatures of odor coding. *BMC Genomics*. 2011;12(1):271. doi:10.1186/1471-2164-12-271.

153. Sparks JT, Vinyard BT, Dickens JC. Gustatory receptor expression in the labella and tarsi of *Aedes aegypti*. *Insect Biochem Mol Biol*. 2013;43(12):1161-1171. doi:10.1016/j.ibmb.2013.10.005.

154. Rund SSC, Hou TY, Ward SM, Collins FH, Duffield GE. Genome-wide profiling of diel and circadian gene expression in the malaria vector *Anopheles gambiae*.

Proc Natl Acad Sci. 2011;108(32):E421-E430. doi:10.1073/pnas.1100584108.

155. Rund SSC, Bonar NA, Champion MM, Ghazi JP, Houk CM, Leming MT, Syed Z, Duffield GE. Daily rhythms in antennal protein and olfactory sensitivity in the malaria mosquito *Anopheles gambiae*. *Sci Rep*. 2013;3(1):2494. doi:10.1038/srep02494.

156. Rinker DC, Pitts RJ, Zhou X, Suh E, Rokas A, Zwiebel LJ. Blood meal-induced changes to antennal transcriptome profiles reveal shifts in odor sensitivities in *Anopheles gambiae*. *Proc Natl Acad Sci*. 2013;110(20):8260-8265. doi:10.1073/pnas.1302562110.

157. Tallon AK, Hill SR, Ignell R. Sex and age modulate antennal chemosensory-related genes linked to the onset of host seeking in the yellow-fever mosquito, *Aedes aegypti*. *Sci Rep*. 2019;9(1):43. doi:10.1038/s41598-018-36550-6.

158. Bohbot J, Pitts RJ, Kwon H-W, Rützler M, Robertson HM, Zwiebel LJ. Molecular characterization of the *Aedes aegypti* odorant receptor gene family. *Insect Mol Biol*. 2007;16(5):525-537. doi:10.1111/j.1365-2583.2007.00748.x.

159. Bohbot JD, Sparks JT, Dickens JC. The maxillary palp of *Aedes aegypti*, a model of multisensory integration. *Insect Biochem Mol Biol*. 2014;48(1):29-39. doi:10.1016/j.ibmb.2014.02.007.

160. Matthews BJ, McBride CS, DeGennaro M, Despo O, Vosshall LB. The neurotranscriptome of the *Aedes aegypti* mosquito. *BMC Genomics*. 2016;17(1):32. doi:10.1186/s12864-015-2239-0.

161. Lombardo F, Salvemini M, Fiorillo C, Nolan T, Zwiebel LJ, Ribeiro JM, Arcà B. Deciphering the olfactory repertoire of the tiger mosquito *Aedes albopictus*. *BMC Genomics*. 2017;18(1):770. doi:10.1186/s12864-017-4144-1.

162. Leal WS, Choo Y-M, Xu P, da Silva CSB, Ueira-Vieira C. Differential expression of olfactory genes in the southern house mosquito and insights into unique odorant receptor gene isoforms. *Proc Natl Acad Sci*. 2013;110(46):18704-18709. doi:10.1073/pnas.1316059110.

163. Pelletier J, Leal WS. Characterization of olfactory genes in the antennae of the Southern house mosquito, *Culex quinquefasciatus*. *J Insect Physiol*. 2011;57(7):915-929. doi:10.1016/j.jinsphys.2011.04.003.

164. Taparia T, Ignell R, Hill SR. Blood meal induced regulation of the chemosensory gene repertoire in the southern house mosquito. *BMC Genomics*. 2017;18(1):393. doi:10.1186/s12864-017-3779-2.

165. Fox AN, Pitts RJ, Robertson HM, Carlson JR, Zwiebel LJ. Candidate odorant receptors from the malaria vector mosquito *Anopheles gambiae* and evidence of down-regulation in response to blood feeding. *Proc Natl Acad Sci U S A*. 2001;98(25):14693-14697. doi:10.1073/pnas.261432998.

166. Justice RW, Dimitratos S, Walter MF, Woods DF, Biessmann H. Sexual dimorphic expression of putative antennal carrier protein genes in the malaria vector *Anopheles gambiae*. *Insect Mol Biol*. 2003;12(6):581-594.

167. Biessmann H, Nguyen QK, Le D, Walter MF. Microarray-based survey of a subset of putative olfactory genes in the mosquito *Anopheles gambiae*. *Insect Mol Biol*. 2005;14(6):575-589. doi:10.1111/j.1365-2583.2005.00590.x.

168. Iatrou K, Biessmann H. Sex-biased expression of odorant receptors in antennae and palps of the African malaria vector *Anopheles gambiae*. *Insect Biochem Mol Biol*. 2008;38(2):268-274. doi:10.1016/j.ibmb.2007.11.008.

169. Nei M, Niimura Y, Nozawa M. The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. *Nat Rev Genet*. 2008;9(12):951-963. doi:10.1038/nrg2480.

170. Vosshall LB, Amrein H, Morozov PS, Rzhetsky A, Axel R. A Spatial Map of Olfactory Receptor Expression in the *Drosophila* Antenna. *Cell*. 1999;96(5):725-736. doi:10.1016/S0092-8674(00)80582-6.

171. Pelosi P, Iovinella I, Zhu J, Wang G, Dani FR. Beyond chemoreception: Diverse tasks of soluble olfactory proteins in insects. *Biol Rev*. 2017. doi:10.1111/brv.12339.

172. Mei T, Fu W-B, Li B, He Z-B, Chen B. Comparative genomics of chemosensory protein genes (CSPs) in twenty-two mosquito species (Diptera: Culicidae): Identification, characterization, and evolution. Hull JJ, ed. *PLoS One*. 2018;13(1):e0190412. doi:10.1371/journal.pone.0190412.

173. Vizueta J, Rozas J, Sánchez-Gracia A. Comparative Genomics Reveals Thousands of Novel Chemosensory Genes and Massive Changes in Chemoreceptor Repertories across Chelicerates. *Genome Biol Evol*. 2018;10(5):1221-1236. doi:10.1093/gbe/evy081.

174. Vizueta J, Frías-López C, Macías-Hernández N, Arnedo MA, Sánchez-Gracia A, Rozas J. Evolution of chemosensory gene families in arthropods: Insight from the first inclusive comparative transcriptome analysis across spider appendages. *Genome Biol Evol*. 2016;9(1):evw296. doi:10.1093/gbe/evw296.

175. McBride CS. Rapid evolution of smell and taste receptor genes during host

specialization in *Drosophila sechellia*. *Proc Natl Acad Sci*. 2007;104(12):4996-5001. doi:10.1073/pnas.0608424104.

176. Dworkin I, Jones CD. Genetic Changes Accompanying the Evolution of Host Specialization in *Drosophila sechellia*. *Genetics*. 2009;181(2):721-736. doi:10.1534/genetics.108.093419.

177. Legrand D, Tenaillon MI, Matyot P, Gerlach J, Lachaise D, Cariou M-L. Species-Wide Genetic Variation and Demographic History of *Drosophila sechellia*, a Species Lacking Population Structure. *Genetics*. 2009;182(4):1197-1206. doi:10.1534/genetics.108.092080.

178. Tunstall NE, Sirey T, Newcomb RD, Warr CG. Selective Pressures on *Drosophila* Chemosensory Receptor Genes. *J Mol Evol*. 2007;64(6):628-636. doi:10.1007/s00239-006-0151-6.

179. Vieira FG, Rozas J. Comparative genomics of the odorant-binding and chemosensory protein gene families across the arthropoda: Origin and evolutionary history of the chemosensory system. *Genome Biol Evol*. 2011;3(1):476-490. doi:10.1093/gbe/evr033.

180. Foret S, Maleszka R. Function and evolution of a gene family encoding odorant binding-like proteins in a social insect, the honey bee (*Apis mellifera*). *Genome Res*. 2006;16(11):1404-1413. doi:10.1101/gr.5075706.

181. Campanini EB, De Brito RA. Molecular evolution of Odorant-binding proteins gene family in two closely related *Anastrepha* fruit flies. *BMC Evol Biol*. 2016;16(1):1-16. doi:10.1186/s12862-016-0775-0.

182. Athrey G, Takken W, Slotman MA. Genetic Analysis of Human Host Preference in the Malaria Vector *Anopheles coluzzii*. In review.

183. Kistler KE, Vosshall LB, Matthews BJ. Genome Engineering with CRISPR-Cas9 in the Mosquito *Aedes aegypti*. *Cell Rep*. 2015;11(1):51-60. doi:10.1016/j.celrep.2015.03.009.

184. Anderson ME, Mavica J, Shackleford L, Flis I, Fochler S, Basu S, Alphey L. CRISPR/Cas9 gene editing in the West Nile Virus vector, *Culex quinquefasciatus* Say. Moreira LA, ed. *PLoS One*. 2019;14(11):e0224857. doi:10.1371/journal.pone.0224857.

185. Macias VM, McKeand S, Chaverra-Rodriguez D, Hughes GL, Fazekas A, Pujhari S, Jasinskiene N, James AA, Rasgon JL. Cas9-Mediated Gene-Editing in the Malaria Mosquito *Anopheles stephensi* by ReMOT Control. *G3*. 2020;10:g3.401133.2020. doi:10.1534/g3.120.401133.

186. Hammond A, Galizi R, Kyrou K, Simoni A, Siniscalchi C, Katsanos D, Gribble M, Baker D, Marois E, Russell S, Burt A, Windbichler N, Crisanti A, Nolan T. A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector *Anopheles gambiae*. *Nat Biotechnol*. 2016;34(1):78-83. doi:10.1038/nbt.3439.

187. Adolfi A, Lycett GJ. Opening the toolkit for genetic analysis and control of Anopheles mosquito vectors. *Curr Opin Insect Sci*. 2018;30:8-18. doi:10.1016/j.cois.2018.07.014.

188. Kyrou K, Hammond AM, Galizi R, Kranjc N, Burt A, Beaghton AK, Nolan T, Crisanti A. A CRISPR–Cas9 gene drive targeting *doublesex* causes complete population suppression in caged *Anopheles gambiae* mosquitoes. *Nat Biotechnol*. 2018;36(11):1062-1071. doi:10.1038/nbt.4245.

189. Simoni A, Hammond AM, Beaghton AK, Galizi R, Taxiarchi C, Kyrou K, Meacci D, Gribble M, Morselli G, Burt A, Nolan T, Crisanti A. A male-biased sex-distorter gene drive for the human malaria vector *Anopheles gambiae*. *Nat Biotechnol*. May 2020. doi:10.1038/s41587-020-0508-1.

190. Neafsey DE, Waterhouse RM, Abai MR, Aganezov SS, Alekseyev MA, Allen JE, Amon J, Arca B, Arensburger P, Artemov G, Assour LA, Basseri H, Berlin A, Birren BW, Blandin SA, Brockman AI, Burkot TR, Burt A, Chan CS, Chauve C, Chiu JC, Christensen M, Costantini C, Davidson VLM, Deligianni E, Dottorini T, Dritsou V, Gabriel SB, Guelbeogo WM, Hall AB, Han M V., Hlaing T, Hughes DST, Jenkins AM, Jiang X, Jungreis I, Kakani EG, Kamali M, Kemppainen P, Kennedy RC, Kirmitzoglou IK, Koekemoer LL, Laban N, Langridge N, Lawniczak MKN, Lirakis M, Lobo NF, Lowy E, MacCallum RM, Mao C, Maslen G, Mbogo C, McCarthy J, Michel K, Mitchell SN, Moore W, Murphy KA, Naumenko AN, Nolan T, Novoa EM, O'Loughlin S, Oringanje C, Oshaghi MA, Pakpour N, Papathanos PA, Peery AN, Povelones M, Prakash A, Price DP, Rajaraman A, Reimer LJ, Rinker DC, Rokas A, Russell TL, Sagnon N, Sharakhova M V., Shea T, Simao FA, Simard F, Slotman MA, Somboon P, Stegniy V, Struchiner CJ, Thomas GWC, Tojo M, Topalis P, Tubio JMC, Unger MF, Vontas J, Walton C, Wilding CS, Willis JH, Wu Y-C, Yan G, Zdobnov EM, Zhou X, Catteruccia F, Christophides GK, et al. Highly evolvable malaria vectors: The genomes of 16 *Anopheles* mosquitoes. *Science (80-)*. 2014;347(6217):1258522-. doi:10.1126/science.1258522.

191. Pitts RJ, Zwiebel LJ. Antennal sensilla of two female anopheline sibling species with differing host ranges. *Malar J*. 2006;5(1):26. doi:10.1186/1475-2875-5-26.

192. Pappas LG, Larsen JR. Gustatory hairs on the mosquito, *Culiseta inornata*. *J Exp Zool*. 1976;196(3):351-360. doi:10.1002/jez.1401960309.

193. Kwon H-W, Lu T, Rützler M, Zwiebel LJ. Olfactory responses in a gustatory organ of the malaria vector mosquito *Anopheles gambiae*. *Proc Natl Acad Sci*. 2006;103(36):13526-13531. doi:10.1073/pnas.0601107103.

194. Melo ACA, Rützler M, Pitts RJ, Zwiebel LJ. Identification of a Chemosensory Receptor from the Yellow Fever Mosquito, *Aedes aegypti*, that is Highly Conserved and Expressed in Olfactory and Gustatory Organs. *Chem Senses*. 2004;29(5):403-410. doi:10.1093/chemse/bjh041.

195. Lee RMKW, Craig DA. Fine Structure of the Sense Organs on the Labella and Labium of the Mosquito *Aedes aegypti* (L.). *Open Entomol J*. 2009;3(1):7-17. doi:10.2174/1874407900903010007.

196. Owen WB, Larsen JR, Pappas LG. Functional units in the labellar chemosensory hairs of the mosquito *Culiseta inornata* (Williston). *J Exp Zool*. 1974;188(2):235-247. doi:10.1002/jez.1401880211.

197. Sinitsyna EE. Bioelectric activity of the contact chemoreceptive sensilla of *Aedes aegypti* L. (Diptera, Culicidae) and food reactions to two-component solutions. *Entomol Rev*. 1971;50:151-155.

198. Amer A, Mehlhorn H. The sensilla of *Aedes* and *Anopheles* mosquitoes and their importance in repellency. *Parasitol Res*. 2006;99(4):491-499. doi:10.1007/s00436-006-0185-0.

199. Ni L, Bronk P, Chang EC, Lowell AM, Flam JO, Panzano VC, Theobald DL, Griffith LC, Garrity PA. A gustatory receptor paralogue controls rapid warmth avoidance in *Drosophila*. *Nature*. 2013;500(7464):580-584. doi:10.1038/nature12390.

200. Xiang Y, Yuan Q, Vogt N, Looger LL, Jan LY, Jan YN. Light-avoidance-mediating photoreceptors tile the *Drosophila* larval body wall. *Nature*. 2010;468(7326):921-926. doi:10.1038/nature09576.

201. Leal WS. Odorant Reception in Insects: Roles of Receptors, Binding Proteins, and Degrading Enzymes. *Annu Rev Entomol*. 2013;58(1):373-391. doi:10.1146/annurev-ento-120811-153635.

202. Larter NK, Sun JS, Carlson JR. Organization and function of *Drosophila* odorant binding proteins. *Elife*. 2016;5:1-22. doi:10.7554/eLife.20242.

203. Gong D-P, Zhang H-J, Zhao P, Xia Q-Y, Xiang Z-H. The odorant binding protein gene family from the genome of silkworm, *Bombyx mori*. *BMC Genomics*. 2009;10:332. doi:10.1186/1471-2164-10-332.

204. Gillies MT. The role of carbon dioxide in host-finding by mosquitoes (Diptera: Culicidae): a review. *Bull Entomol Res*. 1980;70(1940):525-532. http://journals.cambridge.org/production/action/cjoGetFulltext?fulltextid=2389524.

205. Meijerink J, Braks MAH, Brack AA, Adam W, Dekker T, Posthumus MA, van Beek TA, van Loon JJA. Identification of Olfactory Stimulants for *Anopheles gambiae* from Human Sweat Samples. *J Chem Ecol*. 2000;26(6):1367-1382.

206. Verhulst NO, Qiu YT, Beijleveld H, Maliepaard C, Knights D, Schulz S, Berg-Lyons D, Lauber CL, Verduijn W, Haasnoot GW, Mumm R, Bouwmeester HJ, Claas FHJ, Dicke M, van Loon JJA, Takken W, Knight R, Smallegange RC. Composition of Human Skin Microbiota Affects Attractiveness to Malaria Mosquitoes. *PLoS One*. 2011;6(12). doi:10.1371/journal.pone.0028991.

207. Kopp A, Barmina O, Hamilton AM, Higgins L, McIntyre LM, Jones CD. Evolution of Gene Expression in the *Drosophila* Olfactory System. *Mol Biol Evol*. 2008;25(6):1081-1092. doi:10.1093/molbev/msn055.

208. Dekker T, Ibba I, Siju KP, Stensmyr MC, Hansson BS. Olfactory Shifts Parallel Superspecialism for Toxic Fruit in *Drosophila melanogaster* Sibling, *D. sechellia*. *Curr Biol*. 2006;16(1):101-109. doi:10.1016/j.cub.2005.11.075.

209. Matsuo T, Sugaya S, Yasukawa J, Aigaki T, Fuyama Y. Odorant-binding proteins OBP57d and OBP57e affect taste perception and host-plant preference in *Drosophila sechellia*. *PLoS Biol*. 2007;5(5):0985-0996. doi:10.1371/journal.pbio.0050118.

210. Fanello C, Santolamazza F, della Torre A. Simultaneous identification of species and molecular forms of the *Anopheles gambiae* complex by PCR-RFLP. *Med Vet Entomol*. 2002;16:461-464.

211. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella K V., Altshuler D, Gabriel S, DePristo MA. *From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline*.; 2013. doi:10.1002/0471250953.bi1110s43.

212. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21. doi:10.1093/bioinformatics/bts635.

213. Picard. http://broadinstitute.github.io/picard.

214. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis

G, Durbin R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-2079. doi:10.1093/bioinformatics/btp352.

215. Liao Y, Smyth GK, Shi W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923-930. doi:10.1093/bioinformatics/btt656.

216. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2009;26(1):139-140. doi:10.1093/bioinformatics/btp616.

217. Das S, Dimopoulos G. Molecular analysis of photic inhibition of blood-feeding in *Anopheles gambiae*. *BMC Physiol*. 2008;8:23. doi:10.1186/1472-6793-8-23.

218. Nkya TE, Poupardin R, Laporte F, Akhouayri I, Mosha F, Magesa S, Kisinza W, David J-P. Impact of agriculture on the selection of insecticide resistance in the malaria vector *Anopheles gambiae*: a multigenerational study in controlled conditions. *Parasit Vectors*. 2014;7(1):480. doi:10.1186/s13071-014-0480-z.

219. Schultze A, Pregitzer P, Walter MF, Woods DF, Marinotti O, Breer H, Krieger J. The Co-Expression Pattern of Odorant Binding Proteins and Olfactory Receptors Identify Distinct Trichoid Sensilla on the Antenna of the Malaria Mosquito *Anopheles gambiae*. Newcomb RD, ed. *PLoS One*. 2013;8(7):e69412. doi:10.1371/journal.pone.0069412.

220. Qiao H, He X, Schymura D, Ban L, Field L, Dani FR, Michelucci E, Caputo B, Della Torre A, Iatrou K, Zhou JJ, Krieger J, Pelosi P. Cooperative interactions between odorant-binding proteins of *Anopheles gambiae*. *Cell Mol Life Sci*. 2011;68(10):1799-1813. doi:10.1007/s00018-010-0539-8.

221. Thailayil J, Gabrieli P, Caputo B, Bascuñán P, South A, Diabate A, Dabire R, della Torre A, Catteruccia F. Analysis of natural female post-mating responses of *Anopheles gambiae* and *Anopheles coluzzii* unravels similarities and differences in their reproductive ecology. *Sci Rep*. 2018;8(1):6594. doi:10.1038/s41598-018-24923-w.

222. Huff RM, Jason Pitts R. An odorant receptor from *Anopheles gambiae* that demonstrates enantioselectivity to the plant volatile, linalool. *PLoS One*. 2019;14(11):1-15. doi:10.1371/journal.pone.0225637.

223. Carey AF, Wang G, Su C-Y, Zwiebel LJ, Carlson JR. Odorant reception in the malaria mosquito *Anopheles gambiae*. *Nature*. 2010;464(7285):66-71. doi:10.1038/nature08834.

224. Xia Y, Wang G, Buscariollo D, Pitts RJ, Wenger H, Zwiebel LJ. The molecular

and cellular basis of olfactory-driven behavior in *Anopheles gambiae* larvae. *Proc Natl Acad Sci U S A*. 2008;105(17):6433-6438. doi:10.1073/pnas.0801007105.

225. Kreher SA, Mathew D, Kim J, Carlson JR. Translation of Sensory Input into Behavioral Output via an Olfactory System. *Neuron*. 2008;59(1):110-124. doi:10.1016/j.neuron.2008.06.010.

226. van Schooten B, Jiggins CD, Briscoe AD, Papa R. Genome-wide analysis of ionotropic receptors provides insight into their evolution in *Heliconius* butterflies. *BMC Genomics*. 2016;17(1):254. doi:10.1186/s12864-016-2572-y.

227. Missbach C, Vogel H, Hansson BS, Große-Wilde E, Vilcinskas A, Kaiser TS. Developmental and sexual divergence in the olfactory system of the marine insect *Clunio marinus*. *Sci Rep*. 2020;10(1):1-17. doi:10.1038/s41598-020-59063-7.

228. Biessmann H, Andronopoulou E, Biessmann MR, Douris V, Dimitratos SD, Eliopoulos E, Guerin PM, Iatrou K, Justice RW, Kröber T, Marinotti O, Tsitoura P, Woods DF, Walter MF. The *Anopheles gambiae* odorant binding protein 1 (AgamOBP1) mediates indole recognition in the antennae of female mosquitoes. *PLoS One*. 2010;5(3). doi:10.1371/journal.pone.0009471.

229. Schultze A, Schymura D, Forstner M, Krieger J. Expression pattern of a "Plus-C" class odorant binding protein in the antenna of the malaria vector *Anopheles gambiae*. *Insect Mol Biol*. 2012;21(2):187-195. doi:10.1111/j.1365-2583.2011.01125.x.

230. Bohbot JD, Jones PL, Wang G, Pitts RJ, Pask GM, Zwiebel LJ. Conservation of Indole Responsive Odorant Receptors in Mosquitoes Reveals an Ancient Olfactory Trait. *Chem Senses*. 2011;36(2):149-160. doi:10.1093/chemse/bjq105.

231. Andronopoulou E, Labropoulou V, Douris V, Woods DF, Biessmann H, Iatrou K. Specific interactions among odorant-binding proteins of the African malaria vector *Anopheles gambiae*. *Insect Mol Biol*. 2006;15(6):797-811. doi:10.1111/j.1365-2583.2006.00685.x.

232. Bentzur A, Shmueli A, Omesi L, Ryvkin J, Knapp JM, Parnas M, Davis FP, Shohat-Ophir G. *Odorant binding protein 69a* connects social interaction to modulation of social responsiveness in *Drosophila*. *PLoS Genet*. 2018;14(4):1-23. doi:10.1371/journal.pgen.1007328.

233. Takken W, Verhulst NO. Host Preferences of Blood-Feeding Mosquitoes. *Annu Rev Entomol*. 2013:433-453. doi:10.1146/annurev-ento-120811-153618.

234. Carey AF, Carlson JR. Insect olfaction from model systems to disease control. *Proc Natl Acad Sci U S A*. 2011;108(32):12987-12995.

doi:10.1073/pnas.1103472108.

235. Syed Z. Chemical ecology and olfaction in arthropod vectors of diseases. *Curr Opin Insect Sci*. 2015;10:83-89. doi:10.1016/j.cois.2015.04.011.

236. Pitts RJ, Fox a N, Zwiebel LJ. A highly conserved candidate chemoreceptor expressed in both olfactory and gustatory tissues in the malaria vector *Anopheles gambiae*. *Proc Natl Acad Sci*. 2004;101(14):5058-5063. doi:10.1073/pnas.0308146101.

237. Davis EE, Sokolove PG. Temperature Responses of Antennal Receptors of the Mosquito, *Aedes aegypti*. *J Comp Physiol A*. 1975;96(3):223-236. doi:10.1007/BF00612696.

238. Besansky NJ, Hill CA, Costantini C. No accounting for taste: Host preference in malaria vectors. *Trends Parasitol*. 2004;20(6):249-251. doi:10.1016/j.pt.2004.03.007.

239. Iovinella I, Caputo B, Calzetta M, Zwiebel LJ, Dani FR, Torre A della. Profiles of soluble proteins in chemosensory organs of three members of the afro-tropical *Anopheles gambiae* complex. *Comp Biochem Physiol Part D Genomics Proteomics*. 2017;24(May):41-50. doi:https://doi.org/10.1016/j.cbd.2017.07.005.

240. Syed Z, Leal WS. Maxillary Palps Are Broad Spectrum Odorant Detectors in *Culex quinquefasciatus*. *Chem Senses*. 2007;32(8):727-738. doi:10.1093/chemse/bjm040.

241. Knols BGJ, de Jong R, Takken W. Trapping system for testing olfactory responses of the malaria mosquito *Anopheles gambiae* in a wind tunnel. *Med Vet Entomol*. 1994;8(4):386-388. http://www.ncbi.nlm.nih.gov/pubmed/7841494.

242. R Core Team. R: A Language and Environment for Statistical Computing. 2020. https://www.r-project.org/.

243. Fox J, Weisberg S. *An {R} Companion to Applied Regression*. Third. Thousand Oaks {CA}: Sage; 2019. https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

244. Venables WN, Ripley BD. *Modern Applied Statistics with S*. Fourth. New York: Springer; 2002. http://www.stats.ox.ac.uk/pub/MASS4.

245. Brooks ME, Kristensen K, van Benthem KJ, Magnusson A, Berg CW, Nielsen A, Skaug HJ, Mächler M, Bolker BM. glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *R J*. 2017;9(2):378-400. doi:10.32614/rj-2017-066.

246. Lenth R. emmeans: Estimated Marginal Means, aka Least-Squares Means. 2020. https://cran.r-project.org/package=emmeans.

247. Mangiafico S. rcompanion: Functions to Support Extension Education Program Evaluation. 2020. https://cran.r-project.org/package=rcompanion.

248. Meijerink J, van Loon JJA. Sensitivities of antennal olfactory neurons of the malaria mosquito, *Anopheles gambiae*, to carboxylic acids. *J Insect Physiol*. 1999;45:365-373.

249. DeGennaro M, McBride CS, Seeholzer L, Nakagawa T, Dennis EJ, Goldman C, Jasinskiene N, James AA, Vosshall LB. *orco* mutant mosquitoes lose strong preference for humans and are not repelled by volatile DEET. *Nature*. 2013;498(7455):487-491. doi:10.1038/nature12206.

250. Liu H, Liu T, Xie L, Wang X, Deng Y, Chen C, James AA, Chen X. Functional analysis of *Orco* and odorant receptors in odor recognition in *Aedes albopictus*. *Parasit Vectors*. 2016:1-10. doi:10.1186/s13071-016-1644-9.

251. McBride CS. Genes and Odors Underlying the Recent Evolution of Mosquito Preference for Humans. *Curr Biol*. 2016;26:R41-R46. doi:10.1016/j.cub.2015.11.032.

252. Omondi BA, Majeed S, Ignell R. Functional development of carbon dioxide detection in the maxillary palp of *Anopheles gambiae*. *J Exp Biol*. 2015;218(15):2482-2488. doi:10.1242/jeb.116798.

253. McMeniman CJ, Corfas RA, Matthews BJ, Ritchie SA, Vosshall LB. Multimodal integration of carbon dioxide and other sensory cues drives mosquito attraction to humans. *Cell*. 2014;156(5):1060-1071. doi:10.1016/j.cell.2013.12.044.

254. Swarup S, Williams TI, Anholt RRH. Functional dissection of *Odorant binding protein* genes in *Drosophila melanogaster*. *Genes, Brain Behav*. 2011;10(6):648-657. doi:10.1111/j.1601-183X.2011.00704.x.

255. Robertson HM. Molecular Evolution of the Major Arthropod Chemoreceptor Gene Families. *Annu Rev Entomol*. 2019;64(1):227-242. doi:10.1146/annurev-ento-020117-043322.

256. Kirkness EF, Haas BJ, Sun W, Braig HR, Perotti MA, Clark JM, Lee SH, Robertson HM, Kennedy RC, Elhaik E, Gerlach D, Kriventseva E V., Elsik CG, Graur D, Hill CA, Veenstra JA, Walenz B, Tubío JMC, Ribeiro JMC, Rozas J, Johnston JS, Reese JT, Popadic A, Tojo M, Raoult D, Reed DL, Tomoyasu Y, Krause E, Mittapalli O, Margam VM, Li HM, Meyer JM, Johnson RM, Romero-Severson J, VanZee JP, Alvarez-Ponce D, Vieira FG, Aguadé M, Guirao-Rico S,

Anzola JM, Yoon KS, Strycharz JP, Unger MF, Christley S, Lobo NF, Seufferheld MJ, Wang NK, Dasch GA, Struchiner CJ, Madey G, Hannick LI, Bidwell S, Joardar V, Caler E, Shao R, Barker SC, Cameron S, Bruggner R V., Regier A, Johnson J, Viswanathan L, Utterback TR, Sutton GG, Lawson D, Waterhouse RM, Venter JC, Strausberg RL, Berenbaum MR, Collins FH, Zdobnov EM, Pittendrigh BR. Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proc Natl Acad Sci U S A*. 2010;107(27):12168-12173. doi:10.1073/pnas.1003379107.

257. Robertson HM, Baits RL, Walden KKO, Wada-Katsumata A, Schal C. Enormous expansion of the chemosensory gene repertoire in the omnivorous German cockroach *Blattella germanica*. *J Exp Zool Part B Mol Dev Evol*. 2018;330(5):265-278. doi:10.1002/jez.b.22797.

258. Macharia R, Mireji P, Murungi E, Murilla G, Christoffels A, Aksoy S, Masiga D. Genome-Wide Comparative Analysis of Chemosensory Gene Families in Five Tsetse Fly Species. *PLoS Negl Trop Dis*. 2016;10(2):1-30. doi:10.1371/journal.pntd.0004421.

259. Obiero GFO, Mireji PO, Nyanjom SRG, Christoffels A, Robertson HM, Masiga DK. Odorant and Gustatory Receptors in the Tsetse Fly *Glossina morsitans morsitans*. Ribeiro JMC, ed. *PLoS Negl Trop Dis*. 2014;8(4):e2663. doi:10.1371/journal.pntd.0002663.

260. Scott JG, Warren WC, Beukeboom LW, Bopp D, Clark AG, Giers SD, Hediger M, Jones AK, Kasai S, Leichter CA, Li M, Meisel RP, Minx P, Murphy TD, Nelson DR, Reid WR, Rinkevich FD, Robertson HM, Sackton TB, Sattelle DB, Thibaud-Nissen F, Tomlinson C, van de Zande L, Walden KKO, Wilson RK, Liu N. Genome of the house fly, *Musca domestica* L., a global vector of diseases with adaptations to a septic environment. *Genome Biol*. 2014;15(10):466. doi:10.1186/s13059-014-0466-3.

261. Sparks JT, Bohbot JD, Dickens JC. *Olfactory Disruption: Toward Controlling Important Insect Vectors of Disease*. Vol 130. 1st ed. Elsevier Inc.; 2015. doi:10.1016/bs.pmbts.2014.11.004.

262. Robertson HM, Kent LB. Evolution of the Gene Lineage Encoding the Carbon Dioxide Receptor in Insects. *J Insect Sci*. 2009;9(19):1-14. doi:10.1673/031.009.1901.

263. Dunipace L, Meister S, Mcnealy C, Amrein H. Spatially restricted expression of candidate taste receptors in the *Drosophila* gustatory system. *Curr Biol*. 2001;(11):822-835.

264. Scott K, Brady R, Cravchik A, Morozov P, Rzhetsky A, Zuker C, Axel R. A Chemosensory Gene Family Encoding Candidate Gustatory and Olfactory Receptors in *Drosophila*. *Cell*. 2001;104(5):661-673. doi:10.1016/S0092-8674(01)00263-X.

265. Kwon JY, Dahanukar A, Weiss LA, Carlson JR. The molecular basis of CO2 reception in *Drosophila*. *Proc Natl Acad Sci*. 2007;104(9):3574-3578. doi:10.1073/pnas.0700079104.

266. Park J-H, Kwon JY. Heterogeneous Expression of *Drosophila* Gustatory Receptors in Enteroendocrine Cells. Marion-Poll F, ed. *PLoS One*. 2011;6(12):e29022. doi:10.1371/journal.pone.0029022.

267. Thorne N, Amrein H. Atypical Expression of *Drosophila Gustatory Receptor* Genes in Sensory and Central Neurons. *J Comp Neurol*. 2008;506(4):548-568. doi:10.1002/cne.21547.

268. Wolff GH, Riffell JA. Olfaction, experience and neural mechanisms underlying mosquito host preference. *J Exp Biol*. 2018;221. doi:10.1242/jeb.157131.

269. Gardiner A, Butlin RK, Jordan WC, Ritchie MG. Sites of evolutionary divergence differ between olfactory and gustatory receptors of *Drosophila*. *Biol Lett*. 2009;5(2):244-247. doi:10.1098/rsbl.2008.0723.

270. Diaz F, Allan CW, Matzkin LM. Positive selection at sites of chemosensory genes is associated with the recent divergence and local ecological adaptation in cactophilic *Drosophila*. *BMC Evol Biol*. 2018;18(1):144. doi:10.1186/s12862-018-1250-x.

271. Briscoe AD, Macias-Muñoz A, Kozak KM, Walters JR, Yuan F, Jamie GA, Martin SH, Dasmahapatra KK, Ferguson LC, Mallet J, Jacquin-Joly E, Jiggins CD. Female Behaviour Drives Expression and Evolution of Gustatory Receptors in Butterflies. *PLoS Genet*. 2013;9(7). doi:10.1371/journal.pgen.1003620.

272. Smadja C, Shi P, Butlin RK, Robertson HM. Large Gene Family Expansions and Adaptive Evolution for Odorant and Gustatory Receptors in the Pea Aphid, *Acyrthosiphon pisum*. *Mol Biol Evol*. 2009;26(9):2073-2086. doi:10.1093/molbev/msp116.

273. Suzuki HC, Ozaki K, Makino T, Uchiyama H, Yajima S, Kawata M. Evolution of Gustatory Receptor Gene Family Provides Insights into Adaptation to Diverse Host Plants in Nymphalid Butterflies. *Genome Biol Evol*. 2018;10(6):1351-1362. doi:10.1093/gbe/evy093.

274. Wada-Katsumata A, Silverman J, Schal C. Changes in Taste Neurons Support the

Emergence of an Adaptive Behavior in Cockroaches. *Science (80-).* 2013;340(6135):972-975. doi:10.1126/science.1234854.

275. Wisotsky Z, Medina A, Freeman E, Dahanukar A. Evolutionary differences in food preference rely on Gr64e, a receptor for glycerol. *Nat Neurosci*. 2011;14(12):1534-1541. doi:10.1038/nn.2944.

276. Ozaki K, Ryuda M, Yamada A, Utoguchi A, Ishimoto H, Calas D, Marion-Poll F, Tanimura T, Yoshikawa H. A gustatory receptor involved in host plant recognition for oviposition of a swallowtail butterfly. *Nat Commun*. 2011;2(1). doi:10.1038/ncomms1548.

277. Wanner KW, Robertson HM. The gustatory receptor family in the silkworm moth *Bombyx mori* is characterized by a large expansion of a single lineage of putative bitter receptors. *Insect Mol Biol*. 2008;17(6):621-629. doi:10.1111/j.1365-2583.2008.00836.x.

278. Nourmohammad A, Rambeau J, Held T, Kovacova V, Berg J, Lässig M. Adaptive Evolution of Gene Expression in *Drosophila*. *Cell Rep*. 2017;20(6):1385-1395. doi:10.1016/j.celrep.2017.07.033.

279. Githeko AK, Adungo NI, Karanja DM, Hawley WA, Vulule JM, Seroney IK, Ofulla AVO, Atieli FK, Ondijo SO, Genga IO, Odada PK, Situbi PA, Oloo JA. Some Observations on the Biting Behavior of *Anopheles gambiae s.s., Anopheles arabiensis*, and *Anopheles funestus* and Their Implications for Malaria Control. *Exp Parasitol*. 1996;82(3):306-315. doi:10.1006/expr.1996.0038.

280. Fontenille D, Lochouarn L, Diatta M, Sokhna C, Dia I, Diagne N, Lemasson J-J, Ba K, Tall A, Rogier C, Trape J-F. Four years' entomological study of the transmission of seasonal malaria in Senegal and the bionomics of *Anopheles gambiae* and *A. arabiensis*. *Trans R Soc Trop Med Hyg*. 1997;91(6):647-652. doi:10.1016/S0035-9203(97)90506-X.

281. Lemasson J-J, Fontenille D, Lochouarn L, Dia I, Simard F, Ba K, Diop A, Diatta M, Molez J-F. Comparison of Behavior and Vector Efficiency of *Anopheles gambiae* and *An. arabiensis* (Diptera: Culicidae) in Barkedji, a Sahelian Area of Senegal. *J Med Entomol*. 1997;34(4):396-403. doi:10.1093/jmedent/34.4.396.

282. Githeko AK, Service MW, Mbogo CM, Atieli FK, Juma FO. Origin of blood meals in indoor and outdoor resting malaria vectors in western Kenya. *Acta Trop*. 1994;58(3-4):307-316. doi:10.1016/0001-706X(94)90024-8.

283. Petrarca V, Beier JC. Intraspecific Chromosomal Polymorphism in the *Anopheles gambiae* Complex as a Factor Affecting Malaria Transmission in the Kisumu Area of Kenya. *Am J Trop Med Hyg*. 1992;46(2):229-237.

doi:10.4269/ajtmh.1992.46.229.

284. Main BJ, Lee Y, Ferguson H, Kihonda A, Govella N, Kreppel KS, Collier TC, Cornel AJ, Eskin E, Kang EY, Nieman CC, Weakley AM, Lanzaro G. The genetic basis of host choice and resting behavior in the major African malaria vector, *Anopheles arabiensis*. *bioRxiv*. 2016:044701. doi:10.1101/044701.

285. Lulu M, Hadis M, Makonnen Y, Asfaw T. Chromosomal inversion polymorphisms of Anopheles arabiensis from some localities in Ethiopia in relation to host Chromosomal inversion polymorphisms of Anopheles arabiensis from some localities in Ethiopia in relation to host feeding choice. *Ethiop J Health Sci*. 1998;12(January):23-28.

286. Besansky NJ, Powell JR, Caccone A, Hamm DM, Scott JA, Collins FH. Molecular phylogeny of the *Anopheles gambiae* complex suggests genetic introgression between principal malaria vectors. *Proc Natl Acad Sci U S A*. 1994;91(15):6885-6888. doi:10.1073/pnas.91.15.6885.

287. Slotman MA, Della Torre A, Calzetta M, Powell JR. Differential introgression of chromsomal regions between *Anopheles gambiae* and *An. arabiensis*. *Am J Trop Med Hyg*. 2005;73(2):326-335. http://www.ncbi.nlm.nih.gov/pubmed/16103599.

288. Fontaine MC, Pease JB, Steele A, Waterhouse RM, Neafsey DE, Sharakhov I V., Jiang X, Hall AB, Catteruccia F, Kakani E, Mitchell SN, Wu Y-C, Smith HA, Love RR, Lawniczak MKN, Slotman MA, Emrich SJ, Hahn MW, Besansky NJ. Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science (80-)*. 2015;347(6217):1258524. doi:10.1126/science.1258524.

289. Cheng C, White BJ, Kamdem C, Mockaitis K, Costantini C, Hahn MW, Besansky NJ. Ecological Genomics of *Anopheles gambiae* Along a Latitudinal Cline: A Population-Resequencing Approach. *Genetics*. 2012;190(4):1417-1432. doi:10.1534/genetics.111.137794.

290. White BJ, Hahn MW, Pombi M, Cassone BJ, Lobo NF, Simard F, Besansky NJ. Localization of Candidate Regions Maintaining a Common Polymorphic Inversion (2La) in *Anopheles gambiae*. Malik HS, ed. *PLoS Genet*. 2007;3(12):e217. doi:10.1371/journal.pgen.0030217.

291. Bryan JH. *Anopheles gambiae* and *A. melas* at Brefet, The Gambia, and their role in malaria transmission. *Ann Trop Med Parasitol*. 1983;77(1):1-12. doi:10.1080/00034983.1983.11811667.

292. Bryan JH, Petrarca V, Di Deco MA, Coluzzi M. Adult behaviour of members of the *Anopheles gambiae* complex in the Gambia with special reference to *An.*

melas and its chromosomal variants. *Parassitologia*. 1987;29(2-3):221-249. http://www.ncbi.nlm.nih.gov/pubmed/3508262. Accessed May 4, 2020.

293. Temu EA, Minjas JN, Coetzee M, Hunt RH, Shiff CJ. The role of four anopheline species (Diptera: Culicidae) in malaria transmission in coastal Tanzania. *Trans R Soc Trop Med Hyg*. 1998;92(2):152-158. doi:10.1016/S0035-9203(98)90724-6.

294. Cuamba N, Mendis C. The role of *Anopheles merus* in malaria transmission in an area of southern Mozambique. *J Vector Borne Dis*. 2009;46(2):157-159. http://www.ncbi.nlm.nih.gov/pubmed/19502697.

295. Caputo B, Nwakanma D, Jawara M, Adiamoh M, Dia I, Konate L, Petrarca V, Conway DJ, della Torre A. *Anopheles gambiae* complex along The Gambia river, with particular reference to the molecular forms of *An. gambiae* s.s. *Malar J*. 2008;7(1):182. doi:10.1186/1475-2875-7-182.

296. Mutero CM, Mosha FW, Subra R. Biting activity and resting behaviour of *Anopheles merus* Donitz (Diptera: Gulicidae) on the Kenya Coast. *Ann Trop Med Parasitol*. 1984;78(1):43-47. doi:10.1080/00034983.1984.11811771.

297. Kipyab PC, Khaemba BM, Mwangangi JM, Mbogo CM. The bionomics of *Anopheles merus* (Diptera: Culicidae) along the Kenyan coast. *Parasit Vectors*. 2013;6(1):37. doi:10.1186/1756-3305-6-37.

298. White GB. *Anopheles gambiae* complex and disease transmission in Africa. *Trans R Soc Trop Med Hyg*. 1974;68(4):278-298. doi:10.1016/0035-9203(74)90035-2.

299. Coluzzi M. Heterogeneities of the malaria vectorial system in tropical Africa and their significance in malaria epidemiology and control. *Bull World Health Organ*. 1984;62(SUPPL.):107-113.

300. Prior A, Torr SJ. Host selection by *Anopheles arabiensis* and *An. quadriannulatus* feeding on cattle in Zimbabwe. *Med Vet Entomol*. 2002;16(2):207-213. doi:10.1046/j.1365-2915.2002.00367.x.

301. Majeed S, Hill SR, Birgersson G, Ignell R. Detection and perception of generic host volatiles by mosquitoes modulate host preference: Context dependence of (R)-1-octen-3-ol. *R Soc Open Sci*. 2016;3(11). doi:10.1098/rsos.160467.

302. Kent LB, Walden KKO, Robertson HM. The Gr family of candidate gustatory and olfactory receptors in the yellow-fever mosquito *Aedes aegypti*. *Chem Senses*. 2008;33(1):79-93. doi:10.1093/chemse/bjm067.

303. Freeman EG, Wisotsky Z, Dahanukar A. Detection of sweet tastants by a conserved group of insect gustatory receptors. *Proc Natl Acad Sci*.

2014;111(4):1598-1603. doi:10.1073/pnas.1311724111.

304. Miles A, Harding NJ, Bottà G, Clarkson CS, Antão T, Kozak K, Schrider DR, Kern AD, Redmond S, Sharakhov I, Pearson RD, Bergey C, Fontaine MC, Donnelly MJ, Lawniczak MKN, Ayala D, Besansky NJ, Burt A, Caputo B, Torre A Della, Godfray HCJ, Hahn MW, Midega J, Neafsey DE, O'Loughlin S, Pinto J, Riehle MM, Vernick KD, Weetman D, Wilding CS, White BJ, Troco AD, Diabaté A, Costantini C, Rohatgi KR, Elissa N, Coulibaly B, Dinis J, Mbogo C, Bejon P, Mawejje HD, Stalker J, Rockett K, Drury E, Mead D, Jeffreys A, Hubbart C, Rowlands K, Isaacs AT, Jyothi D, Malangone C, Vauterin P, Jeffery B, Wright I, Hart L, Kluczyński K, Cornelius V, Macinnis B, Henrichs C, Giacomantonio R, Kwiatkowski DP. Genetic diversity of the African malaria vector *Anopheles gambiae*. *Nature*. 2017;552(7683):96-100. doi:10.1038/nature24995.

305. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754-1760. doi:10.1093/bioinformatics/btp324.

306. Giraldo-Calderón GI, Emrich SJ, MacCallum RM, Maslen G, Emrich S, Collins F, Dialynas E, Topalis P, Ho N, Gesing S, Madey G, Collins FH, Lawson D, Kersey P, Allen J, Christensen M, Hughes D, Koscielny G, Langridge N, Gallego EL, Megy K, Wilson D, Gelbart B, Emmert D, Russo S, Zhou P, Christophides G, Brockman A, Kirmitzoglou I, MacCallum B, Tiirikka T, Louis K, Dritsou V, Mitraka E, Werner-Washburn M, Baker P, Platero H, Aguilar A, Bogol S, Campbell D, Carmichael R, Cieslak D, Davis G, Konopinski N, Nabrzyski J, Reinking C, Sheehan A, Szakonyi S, Wieck R. VectorBase: An updated Bioinformatics Resource for invertebrate vectors and other organisms related with human diseases. *Nucleic Acids Res*. 2015;43(D1):D707-D713. doi:10.1093/nar/gku1117.

307. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215(3):403-410. doi:10.1016/S0022-2836(05)80360-2.

308. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 2011;27(21):2987-2993. doi:10.1093/bioinformatics/btr509.

309. Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841-842. doi:10.1093/bioinformatics/btq033.

310. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. DnaSP 6: DNA Sequence Polymorphism

Analysis of Large Data Sets. *Mol Biol Evol*. 2017;34(12):3299-3302. doi:10.1093/molbev/msx248.

311. McDonald JH, Kreitman M. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature*. 1991;351(6328):652-654. doi:10.1038/351652a0.

312. Fu YX, Li WH. Statistical tests of neutrality of mutations. *Genetics*. 1993;133(3):693-709. http://www.ncbi.nlm.nih.gov/pubmed/8454210.

313. Tajima F. Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. *Genetics*. 1989;123(3):585-595. doi:PMC1203831.

314. Fay JC, Wu CI. Hitchhiking Under Positive Darwinian Selection. *Genetics*. 2000;155(3):1405-1413. http://www.ncbi.nlm.nih.gov/pubmed/10880498.

315. Zeng K, Fu YX, Shi S, Wu CI. Statistical tests for detecting positive selection by utilizing high-frequency variants. *Genetics*. 2006;174(3):1431-1439. doi:10.1534/genetics.106.061432.

316. Leigh JW, Bryant D. POPART: Full-feature software for haplotype network construction. *Methods Ecol Evol*. 2015;6(9):1110-1116. doi:10.1111/2041-210X.12410.

317. Booker TR, Jackson BC, Keightley PD. Detecting positive selection in the genome. *BMC Biol*. 2017;15(1):98. doi:10.1186/s12915-017-0434-y.

318. Veillard F, Troxler L, Reichhart J-M. *Drosophila melanogaster* clip-domain serine proteases: Structure, function and regulation. *Biochimie*. 2016;122:255-269. doi:10.1016/j.biochi.2015.10.007.

319. Kwon H, Smith RC. Chemical depletion of phagocytic immune cells in *Anopheles gambiae* reveals dual roles of mosquito hemocytes in anti- *Plasmodium* immunity. *Proc Natl Acad Sci*. 2019;116(28):14119-14128. doi:10.1073/pnas.1900147116.

320. He J, Sun H, Zhang D, Sun Y, Ma L, Chen L, Liu Z, Xiong C, Yan G, Zhu C. Cloning and characterization of 60S ribosomal protein L22 (*RPL22*) from *Culex pipiens pallens*. *Comp Biochem Physiol Part B Biochem Mol Biol*. 2009;153(2):216-222. doi:10.1016/j.cbpb.2009.03.003.

321. Agnihotri AR, Roy AA, Joshi RS. Gustatory receptors in Lepidoptera: chemosensation and beyond. *Insect Mol Biol*. 2016;25(5):519-529. doi:10.1111/imb.12246.

322. Zhang Y V., Ni J, Montell C. The Molecular Basis for Attractive Salt-Taste Coding in Drosophila. *Science (80-)*. 2013;340(6138):1334-1338.

doi:10.1126/science.1234133.

323. Charlwood JD, Jones MDR. Mating in the mosquito, *Anopheles gambiae* s.l. II. Swarming behaviour. *Physiol Entomol*. 1980;5(4):315-320. doi:10.1111/j.1365-3032.1980.tb00241.x.

324. Charlwood JD, Jones MDR. Mating behaviour in the mosquito, *Anopheles gambiae* s.l. I. Close range and contact behaviour. *Physiol Entomol*. 1979;4(2):111-120. doi:10.1111/j.1365-3032.1979.tb00185.x.

325. Stocker RF, Gendre N. Courtship behavior of *Drosophila* genetically or surgically deprived of basiconic sensilla. *Behav Genet*. 1989;19(3):371-385. doi:10.1007/BF01066165.

326. Moon SJ, Lee Y, Jiao Y, Montell C. A *Drosophila* Gustatory Receptor Essential for Aversive Taste and Inhibiting Male-to-Male Courtship. *Curr Biol*. 2009;19(19):1623-1627. doi:10.1016/j.cub.2009.07.061.

327. Dabire KR, Sawadodgo S, Diabate A, Toe KH, Kengne P, Ouari A, Costantini C, Gouagna C, Simard F, Baldet T, Lehmann T, Gibson G. Assortative mating in mixed swarms of the mosquito *Anopheles gambiae s.s.* M and S molecular forms, in Burkina Faso, West Africa. *Med Vet Entomol*. 2013;27(3):298-312. doi:10.1111/j.1365-2915.2012.01049.x.

328. Charlesworth J, Eyre-Walker A. The McDonald-Kreitman test and slightly deleterious mutations. *Mol Biol Evol*. 2008;25(6):1007-1015. doi:10.1093/molbev/msn005.

329. Eyre-Walker A. Changing Effective Population Size and the McDonald-Kreitman Test. *Genetics*. 2002;162(4):2017-2024.

330. Impoinvil DE, Kongere JO, Foster WA, Njiru BN, Killeen GF, Githure JI, Beier JC, Hassanali A, Knols BGJ. Feeding and survival of the malaria vector *Anopheles gambiae* on plants growing in Kenya. *Med Vet Entomol*. 2004;18(2):108-115.

331. Yu J, Chen B, Zheng B, Qiao C, Chen X, Yan Y, Luan X, Xie B, Liu J, Shen C, He Z, Hu X, Liu M, Li H, Shao Q, Fang J. ATP synthase is required for male fertility and germ cell maturation in *Drosophila* testes. *Mol Med Rep*. 2019;19(3):1561-1570. doi:10.3892/mmr.2019.9834.

332. Tarr PT, Tarling EJ, Bojanic DD, Edwards PA, Baldán Á. Emerging new paradigms for ABCG transporters. *Biochim Biophys Acta - Mol Cell Biol Lipids*. 2009;1791(7):584-593. doi:10.1016/j.bbalip.2009.01.007.

333. Hill CA, Fox AN, Pitts RJ, Kent LB, Tan PL, Chrystal MA, Cravchik A, Collins FH, Robertson HM, Zwiebel LJ. G Protein-Coupled Receptors in *Anopheles gambiae*. *Science (80-)*. 2002;298(5591):176-178. doi:10.1126/science.1076196.

334. Pel J, Leung A, Choi WWY, Despotovic M, Ung WL, Shibahara G, Gelinas L, Marziali A. Rapid and highly-specific generation of targeted DNA sequencing libraries enabled by linking capture probes with universal primers. Kalendar R, ed. *PLoS One*. 2018;13(12):e0208283. doi:10.1371/journal.pone.0208283.

335. Kent LB, Robertson HM. Evolution of the sugar receptors in insects. *BMC Evol Biol*. 2009;9(1):41. doi:10.1186/1471-2148-9-41.

336. Lutz EK, Grewal TS, Riffell JA. Computational and experimental insights into the chemosensory navigation of *Aedes aegypti* mosquito larvae. *Proc R Soc B Biol Sci*. 2019;286(1915):20191495. doi:10.1098/rspb.2019.1495.

337. Tinbergen N. On aims and methods of Ethology. *Z Tierpsychol*. 1963;20(4):410-433. doi:10.1111/j.1439-0310.1963.tb01161.x.

APPENDIX A

Table A1 – Average Expression Values of Chemosensory Genes in Labella

Gene ID	Gene Name	Female *An. coluzzii* (CPM)	Female *An. quadriannulatus* (CPM)	Male *An. coluzzii* (CPM)	Male *An. quadriannulatus* (CPM)
AGAP004114	Gr1	0.101	0.000	0.200	0.368
AGAP002275	Gr2	6.643	0.177	0.000	0.000
AGAP009858	Gr3	0.923	0.000	0.183	0.235
AGAP009857	Gr4	0.323	0.375	0.599	9.995
AGAP009853	Gr5	0.824	1.778	0.000	0.327
AGAP009854	Gr6	0.173	0.000	0.237	0.069
AGAP009855	Gr7	0.698	0.236	0.973	0.250
AGAP009856	Gr8	1.681	3.123	1.251	1.167
AGAP009805	Gr9	0.314	0.985	0.291	0.955
AGAP009804	Gr10	0.787	0.588	0.618	1.075
AGAP009803	Gr11	0.260	0.116	0.279	0.302
AGAP009802	Gr12	0.341	0.339	0.184	0.168
AGAP002635	Gr13	0.546	0.060	0.987	0.082
AGAP006399	Gr14	2.449	2.636	2.533	1.632
AGAP003253	Gr15	0.411	0.060	0.302	0.144
AGAP003254	Gr16	0.000	0.000	0.128	0.000
AGAP003255	Gr17	16.547	10.903	11.325	12.035
AGAP003256	Gr18	0.040	0.236	0.093	0.446
AGAP003258	Gr19	0.503	0.049	0.361	0.000
AGAP003259	Gr20	0.000	0.116	0.195	0.000
AGAP003260	Gr21	1.556	3.936	1.150	6.306
AGAP009999	Gr22	0.000	0.060	0.000	0.000
AGAP003098	Gr23	0.235	0.545	0.000	0.434
AGAP001915	Gr24	0.162	1.005	0.054	0.245
AGAP004727	Gr25	0.395	0.120	0.587	0.000
AGAP006717	Gr26	0.000	0.000	0.064	0.082

Table A1 Continued

Gene ID	Gene Name	Female *An. coluzzii* (CPM)	Female *An. quadriannulatus* (CPM)	Male *An. coluzzii* (CPM)	Male *An. quadriannulatus* (CPM)
AGAP006716	Gr27	0.108	0.000	0.000	0.082
AGAP006713	Gr28	0.081	0.346	0.000	0.000
AGAP006874	Gr29	2.018	2.200	2.679	2.747
AGAP006875	Gr30	0.209	1.080	0.000	0.441
AGAP006876	Gr31	1.921	0.165	1.672	0.307
AGAP006877	Gr32	0.270	0.851	0.285	1.788
AGAP010195	Gr33	6.675	5.244	8.211	5.629
AGAP006450	Gr34	0.065	0.116	0.000	0.000
AGAP011915	Gr35	0.000	0.000	0.000	0.000
AGAP001123	Gr36	0.000	0.000	0.000	0.000
AGAP001117	Gr37	0.215	0.178	0.145	0.716
AGAP001114	Gr38	0.000	0.236	0.109	0.139
AGAP001119	Gr39	0.161	0.452	0.183	0.372
AGAP001120	Gr40	0.000	0.000	0.000	0.000
AGAP001122	Gr41	0.040	0.000	0.000	0.000
AGAP001115	Gr42	0.125	0.885	0.000	0.072
AGAP005047	Gr43	2.027	1.836	3.404	1.251
AGAP009256	Gr44	0.119	2.935	3.080	0.245
AGAP007757	Gr45	0.000	0.000	0.000	0.000
AGAP007756	Gr46	0.108	0.127	0.400	0.301
AGAP005514	Gr47	0.000	0.219	0.054	0.163
AGAP001170	Gr48	0.578	0.885	0.398	4.697
AGAP001169	Gr49	2.223	0.412	2.481	0.278
AGAP001171	Gr50	0.000	0.000	0.366	0.399
AGAP001172	Gr51	0.779	1.130	0.474	1.043
AGAP001173	Gr52	0.422	0.219	0.213	0.163
AGAP002633	Gr53	0.065	0.000	0.000	0.000
AGAP004313	Gr54	0.000	0.127	0.000	0.168

Table A1 Continued

Gene ID	Gene Name	Female *An. coluzzii* (CPM)	Female *An. quadriannulatus* (CPM)	Male *An. coluzzii* (CPM)	Male *An. quadriannulatus* (CPM)
AGAP006917	Gr55	1.093	0.841	0.934	0.845
AGAP006143	Gr56	0.000	0.699	0.000	0.151
AGAP004716	Gr57	0.393	0.000	0.107	0.000
AGAP001125	Gr58	0.000	0.233	0.000	0.163
AGAP001137	Gr59	0.276	0.000	0.000	0.561
AGAP001121	Gr60	0.000	0.000	0.000	0.000
AGAP029199	Gr62	0.856	1.229	1.774	0.830
AGAP013154	IR7h.1	0.132	0.000	0.064	0.072
AGAP013363	IR7i	0.242	0.000	0.000	0.000
AGAP000714	IR7n	0.430	1.306	0.429	1.018
AGAP013409	IR7s	7.711	5.404	2.742	1.899
AGAP002763	IR7t	0.503	0.000	0.000	0.000
AGAP013285	IR7u	0.323	0.000	0.128	0.245
AGAP013416	IR7w	4.657	0.408	4.035	1.284
AGAP013520	IR7x	1.414	7.485	0.425	4.294
AGAP013172	IR7y	0.000	0.049	0.000	0.245
AGAP010411	IR8a	0.471	0.243	0.456	0.450
AGAP008511	IR21a	0.192	0.405	0.500	0.322
AGAP010272	IR25a	65.547	59.535	52.720	76.420
AGAP009014	IR31a	0.132	0.000	0.000	0.000
AGAP004021	IR40a	0.054	0.064	0.000	0.215
AGAP002904	IR41a	23.118	11.283	20.857	13.240
AGAP008759	IR41b	0.054	0.000	0.156	0.245
AGAP012951	IR41c	0.054	0.292	0.182	0.332
AGAP003531	IR41n	0.094	0.000	0.099	0.247
AGAP004432	IR41t.1	0.094	0.000	0.000	0.478
AGAP012969	IR41t.2	0.000	0.000	0.000	0.082
AGAP011943	IR60a	0.000	0.000	0.000	0.000

Table A1 Continued

Gene ID	Gene Name	Female *An. coluzzii* (CPM)	Female *An. quadriannulatus* (CPM)	Male *An. coluzzii* (CPM)	Male *An. quadriannulatus* (CPM)
AGAP004923	IR64a	1.317	1.563	1.019	1.397
AGAP007951	IR68a	0.000	0.000	0.104	0.000
AGAP004969	IR75d	0.047	0.000	0.339	0.168
AGAP013085	IR75g	0.310	0.622	0.909	0.998
AGAP001811	IR75h.1	0.000	0.000	0.000	0.000
AGAP001812	IR75h.2	0.054	0.060	0.586	0.082
AGAP007498	IR75k	0.421	0.296	0.052	0.000
AGAP005466	IR75l	0.943	0.601	0.809	0.678
AGAP011968	IR76b	164.836	170.147	143.545	174.797
AGAP000256	IR93a	0.248	2.141	0.674	0.133
AGAP000140	IR100a	1.621	1.189	4.309	1.534
AGAP000293	IR100h	0.709	0.116	0.159	0.000
AGAP004475	IR100i	0.476	0.098	0.433	0.409
AGAP013425	IR101	0.000	0.000	0.000	0.000
AGAP005677	IR133	0.000	0.000	0.000	0.000
AGAP005678	IR134	0.000	0.060	0.000	0.327
AGAP005679	IR135	0.108	0.000	0.091	0.000
AGAP006440	IR136	0.000	0.000	0.000	0.000
AGAP006691	IR139	0.040	0.000	0.000	0.082
AGAP013242	IR140.1	0.000	0.116	0.000	0.082
AGAP013436	Ir140.2	0.000	0.000	0.143	0.000
AGAP029188	Ir141	0.000	0.000	0.000	0.082
AGAP006407	IR142	0.188	0.064	0.000	0.082
AGAP029212	Ir144	0.293	0.191	0.193	0.924
AGAP029208	Ir146	0.000	0.116	0.175	0.329
AGAP029217	Ir149	0.000	0.049	0.183	0.000
AGAP029215	Ir150	0.000	0.480	0.266	0.405
AGAP029230	Ir152	0.108	0.000	1.702	0.764

Table A1 Continued

Gene ID	Gene Name	Female *An. coluzzii* (CPM)	Female *An. quadriannulatus* (CPM)	Male *An. coluzzii* (CPM)	Male *An. quadriannulatus* (CPM)
AGAP029222	Ir153	0.000	0.000	0.000	0.000
AGAP029225	Ir154	0.219	0.000	0.600	0.278
AGAP029216	Ir155	0.569	0.889	1.149	0.422
AGAP029220	Ir156	0.601	0.339	0.348	0.168
AGAP029209	Ir162	0.094	0.000	0.000	0.000
AGAP029210	Ir168	1.687	4.170	6.133	1.392
AGAP029223	Ir170	0.215	0.000	0.091	0.408
AGAP029233	Ir175	0.040	0.242	0.611	0.000
AGAP029227	Ir176	0.226	0.497	0.385	1.469
AGAP029232	Ir179	4.020	1.450	0.193	15.015
AGAP029229	Ir180	0.108	0.802	0.390	0.816
AGAP029226	Ir195	0.000	0.000	0.000	0.000
AGAP029228	Ir196	0.000	0.000	0.000	0.000
AGAP029201	Ir199	16.059	3.771	13.669	5.317
AGAP029202	Ir203	0.000	0.000	0.362	0.082
AGAP029200	Ir206	0.000	0.000	0.181	0.000
AGAP002560	Orco	90.649	80.615	103.667	113.985
AGAP009640	Or1	0.132	0.000	0.000	0.245
AGAP009519	Or2	0.212	0.107	0.189	0.084
AGAP011469	OR3	2.825	0.794	1.532	0.969
AGAP011468	Or4	4.355	3.055	2.680	4.282
AGAP011467	Or5	7.274	6.108	9.226	11.056
AGAP006167	Or6	0.879	0.064	1.537	0.466
AGAP001912	Or8	0.000	0.181	0.000	0.000
AGAP008333	Or9	0.107	0.000	0.000	0.082
AGAP009520	Or10	0.577	0.412	0.636	0.215
AGAP011631	Or11	0.047	0.000	0.000	0.082
AGAP009396	Or13	0.047	0.000	0.091	0.072

Table A1 Continued

Gene ID	Gene Name	Female *An. coluzzii* (CPM)	Female *An. quadriannulatus* (CPM)	Male *An. coluzzii* (CPM)	Male *An. quadriannulatus* (CPM)
AGAP009408	Or14	0.000	0.455	0.000	0.000
AGAP009398	Or15	1.305	1.121	1.367	1.122
AGAP009394	Or16	0.347	0.622	0.341	1.501
AGAP009395	Or17	0.065	0.064	0.392	0.072
AGAP009410	Or18	0.612	0.234	0.757	0.000
AGAP009411	Or20	1.016	0.631	1.124	1.895
AGAP029499	Or21	4.806	1.242	4.519	1.734
AGAP008114	Or22	0.000	0.000	0.000	0.000
AGAP007797	Or23	0.000	0.060	0.052	0.000
AGAP010507	Or24	0.435	0.064	0.211	0.371
AGAP003054	Or25	0.000	0.000	0.000	0.000
AGAP004354	Or26	0.000	0.000	0.181	0.632
AGAP004355	Or27	0.415	0.864	0.903	1.088
AGAP002722	Or28	0.000	0.174	0.000	0.000
AGAP009111	Or29	3.205	7.554	3.388	5.779
AGAP009391	Or30	0.000	0.000	0.000	0.000
AGAP004974	Or31	0.000	0.000	0.000	0.082
AGAP004951	Or32	0.033	0.000	0.195	0.327
AGAP005760	Or33	0.087	0.121	0.245	0.302
AGAP002125	Or34	0.000	0.000	0.000	0.000
AGAP004971	Or35	0.000	0.000	0.000	0.000
AGAP001012	Or36	0.317	0.000	0.000	0.000
AGAP002126	Or37	0.000	0.000	0.000	0.000
AGAP002640	Or38	0.054	0.263	0.209	0.069
AGAP002639	Or39	0.033	0.000	0.000	0.000
AGAP002558	Or40	0.054	1.521	16.431	0.572
AGAP000226	Or41	0.000	0.000	0.000	0.215
AGAP004278	Or42	0.000	0.000	0.000	0.000

Table A1 Continued

Gene ID	Gene Name	Female *An. coluzzii* (CPM)	Female *An. quadriannulatus* (CPM)	Male *An. coluzzii* (CPM)	Male *An. quadriannulatus* (CPM)
AGAP010504	Or43	0.121	0.000	0.183	0.000
AGAP010505	Or44	0.108	0.000	0.000	0.226
AGAP003053	Or45	0.000	0.000	0.000	0.000
AGAP009392	Or46	0.000	0.216	0.091	0.000
AGAP009393	Or47	0.040	0.000	0.000	0.163
AGAP006666	Or48	0.047	0.000	0.000	0.000
AGAP006667	Or49	0.040	0.000	0.000	0.163
AGAP029705	Or50	0.397	0.098	1.231	0.213
AGAP009409	Or51	0.424	0.060	0.692	0.163
AGAP000230	Or52	0.293	0.265	0.000	0.215
AGAP009390	Or53	1.416	1.752	0.956	1.842
AGAP011813	Or54	0.214	0.826	0.306	0.886
AGAP009397	Or55	0.146	0.000	0.000	0.000
AGAP004356	Or56	0.222	0.176	0.140	0.000
AGAP004357	Or57	0.000	0.064	0.000	0.000
AGAP004067	Or58	0.000	0.000	0.000	0.000
AGAP002995	Or59	0.000	0.000	0.000	0.000
AGAP011979	Or60	0.033	0.000	0.000	0.082
AGAP011991	Or61	0.000	0.000	0.000	0.000
AGAP011978	Or62	0.000	0.000	0.000	0.000
AGAP011989	Or63	0.000	0.000	0.000	0.000
AGAP011990	Or64	0.000	0.000	0.000	0.000
AGAP008894	Or65	0.000	0.000	0.000	0.000
AGAP003310	Or66	0.087	0.000	0.000	0.000
AGAP009704	Or68	0.300	0.275	0.970	0.000
AGAP009705	Or69	0.378	0.174	0.207	0.305
AGAP009706	Or70	0.356	0.177	0.464	0.678
AGAP009707	Or71	0.422	0.127	0.457	0.409

Table A1 Continued

Gene ID	Gene Name	Female *An. coluzzii* (CPM)	Female *An. quadriannulatus* (CPM)	Male *An. coluzzii* (CPM)	Male *An. quadriannulatus* (CPM)
AGAP009718	Or72	0.304	0.000	0.184	0.327
AGAP009719	Or73	0.344	0.000	0.155	0.240
AGAP009720	Or74	0.000	0.000	0.091	0.302
AGAP002045	Or75	0.000	0.289	0.143	0.205
AGAP002046	Or76	0.000	1.405	1.141	0.000
AGAP002044	Or77	0.112	0.124	0.091	0.726
AGAP005495	Or80	0.132	0.000	0.000	0.000
AGAP013512	Or81	0.047	0.000	0.091	0.000
AGAP029062	OBP1	273.688	188.973	368.997	235.485
AGAP003306	OBP2	3.279	1.996	3.768	2.665
AGAP001409	OBP3	78.720	58.307	117.773	66.464
AGAP010489	OBP4	0.000	0.000	0.000	0.000
AGAP009629	OBP5	37.979	25.311	41.159	30.061
AGAP003530	OBP6	25.232	55.863	26.853	56.746
AGAP001556	OBP7	120.859	98.412	147.953	117.139
AGAP000279	Obp8	0.141	0.000	0.689	0.000
AGAP000278	OBP9	2.081	2.486	1.161	2.607
AGAP001189	Obp10	1,353.581	1,319.310	1,698.480	1,176.919
AGAP002025	OBP11	0.000	0.000	0.401	0.069
AGAP002188	OBP12	0.000	0.049	0.000	0.000
AGAP002905	OBP13	2,490.079	2,695.098	2,619.672	3,334.414
AGAP002189	OBP14	0.000	0.000	0.000	0.163
AGAP003307	OBP15	0.000	0.000	0.000	0.000
AGAP004433	Obp19	2.721	5.002	2.015	7.400
AGAP005208	OBP20	85.880	69.650	99.328	79.231
AGAP008398	OBP21	0.000	0.064	0.000	0.000
AGAP010409	OBP22	61.809	49.496	88.265	65.022
AGAP012318	OBP23	55.525	1.891	80.484	3.458

Table A1 Continued

Gene ID	Gene Name	Female *An. coluzzii* (CPM)	Female *An. quadriannulatus* (CPM)	Male *An. coluzzii* (CPM)	Male *An. quadriannulatus* (CPM)
AGAP012319	OBP24	0.214	0.124	1.153	0.144
AGAP012320	OBP25	571.571	382.905	618.469	554.428
AGAP012321	OBP26	1,472.918	129.103	81.465	24.486
AGAP012323	OBP27	0.000	0.000	0.000	0.000
AGAP012325	OBP28	2.030	1.036	2.033	2.159
AGAP012331	OBP29	0.444	0.301	0.290	0.252
AGAP010649	Obp31	0.000	0.000	0.000	0.000
AGAP000638	OBP32	0.000	0.060	0.000	0.000
AGAP000640	OBP33	0.000	0.000	0.000	0.000
AGAP000641	OBP34	0.000	0.000	0.000	0.000
AGAP000642	OBP35	0.000	0.000	0.000	0.000
AGAP000643	OBP36	0.000	0.000	0.000	0.000
AGAP000644	OBP37	0.000	0.000	0.000	0.000
AGAP029660	Obp38	0.192	0.191	0.272	13.144
AGAP002190	OBP39	0.000	0.000	0.000	0.000
AGAP002191	OBP40	0.000	0.000	0.000	0.000
AGAP005182	OBP41	0.000	0.000	0.000	0.133
AGAP009065	OBP42	0.000	0.000	0.000	0.000
AGAP009402	OBP43	0.000	0.000	0.000	0.000
AGAP010648	OBP44	0.000	0.000	0.000	0.084
AGAP010650	OBP45	0.000	0.000	0.000	0.000
AGAP007289	OBP46	0.310	1.376	0.301	2.578
AGAP007287	OBP47	49.424	39.226	57.438	47.949
AGAP007286	Obp48	1,603.747	3,445.310	1,693.067	4,280.548
AGAP006075	OBP49	3.316	0.366	1.541	0.228
AGAP006076	OBP50	338.533	646.601	338.889	497.528
AGAP006077	Obp51	394.574	186.576	274.379	172.626
AGAP006078	Obp52	1.349	1.150	1.228	1.209

Table A1 Continued

Gene ID	Gene Name	Female *An. coluzzii* (CPM)	Female *An. quadriannulatus* (CPM)	Male *An. coluzzii* (CPM)	Male *An. quadriannulatus* (CPM)
AGAP006079	Obp53	0.725	0.000	0.000	0.307
AGAP006080	OBP54	1,380.586	266.436	1,057.866	276.815
AGAP006081	OBP55	1.380	1.457	1.002	1.966
AGAP011367	OBP56	14.870	4.164	12.505	6.597
AGAP011368	OBP57	4,346.634	11,705.754	3,134.650	5,679.227
AGAP006760	Obp62	0.197	0.598	0.235	0.139
AGAP007281	Obp63	0.000	0.000	0.104	0.082
AGAP007282	Obp64	0.000	0.000	0.000	0.000
AGAP002556	Obp66	71.921	56.047	96.064	65.687
AGAP012322	Obp67	0.168	0.000	0.000	0.163
AGAP012659	Obp67	0.000	0.000	0.000	0.000
AGAP012324	Obp68	0.000	0.000	0.000	0.000
AGAP012658	Obp68	0.000	0.000	0.091	0.000
AGAP013182	Obp69	0.000	0.000	0.000	0.139
AGAP006368	Obp69	1.871	0.277	0.195	0.072
AGAP007283	Obp70	0.033	0.000	0.000	0.000
AGAP006074	Obp71	42.176	19.611	26.688	24.966
AGAP012714	Obp72	1.009	1.412	1.217	3.208

APPENDIX B

Table A2 – Raw Results of Dual-Port Olfactometer Experiments

Date	Condition	Left Port	Right Port	Cow Port	Human Port	Release Cage	Wind Tunnel	All Non-Responders	Starting #	Alive	Active
2/5/2019	intact	cow	sock	0	15	4	29	33	52	48	44
2/9/2019	intact	sock	cow	3	32	4	24	28	64	63	59
2/10/2019	intact	sock	cow	0	10	5	14	19	44	29	24
2/17/2019	leg	sock	cow	0	3	13	23	36	45	39	26
2/17/2019	intact	sock	cow	0	8	16	15	31	45	39	23
2/17/2019	labella	sock	cow	0	0	20	12	32	44	32	12
2/18/2019	intact	cow	sock	0	5	11	30	41	60	46	35
3/24/2019	intact	sock	cow	0	14	7	21	28	49	42	35
3/24/2019	antenna	sock	cow	0	0	30	13	43	49	43	13
3/25/2019	antenna	cow	sock	0	0	21	25	46	63	46	25
3/25/2019	intact	cow	sock	1	16	4	42	46	64	63	59
3/26/2019	antenna	sock	cow	0	0	25	7	32	47	32	7
3/26/2019	intact	sock	cow	0	14	8	22	30	47	44	36
4/1/2019	intact	sock	cow	0	20	3	29	32	53	52	49
4/1/2019	MP	sock	cow	0	18	11	23	34	53	52	41
4/3/2019	labella	cow	sock	0	4	15	17	32	52	36	21
4/3/2019	intact	cow	sock	0	20	7	24	31	52	51	44
4/4/2019	labella	sock	cow	0	4	10	17	27	55	31	21
4/4/2019	intact	sock	cow	0	6	6	41	47	55	53	47
4/4/2019	MP	sock	cow	0	1	12	37	49	55	50	38
4/5/2019	intact	cow	sock	0	12	8	49	57	73	69	61
4/5/2019	MP	cow	sock	0	0	20	38	58	73	58	38
4/8/2019	MP	cow	sock	0	1	4	23	27	30	28	24
4/8/2019	intact	cow	sock	0	2	1	27	28	30	30	29
4/10/2019	leg	cow	sock	0	8	3	45	48	56	56	53
7/30/2019	antenna	sock	cow	0	0	24	14	38	68	38	14
7/30/2019	intact	sock	cow	1	5	2	52	54	68	60	58

Table A2 Continued

Date	Condition	Left Port	Right Port	Cow Port	Human Port	Release Cage	Wind Tunnel	All Non-Responders	Starting #	Alive	Active
9/21/2019	no CO2	sock	cow	0	0	25	40	65	5	70	65
9/21/2019	intact	sock	cow	0	6	13	40	53	11	70	59
1/23/2020	no CO2	sock	cow	0	1	12	53	65	2	68	66
1/23/2020	intact	sock	cow	0	15	3	51	54	0	69	69
1/27/2020	no CO2	sock	cow	0	0	8	41	49	1	50	49
1/27/2020	labella	sock	cow	0	0	10	28	38	15	53	38
1/27/2020	intact	sock	cow	0	5	5	36	41	4	50	46
1/28/2020	intact	sock	cow	1	3	19	22	41	1	46	45
1/28/2020	leg	sock	cow	0	2	11	27	38	7	47	40
2/4/2020	intact	cow	sock	0	11	6	26	32	12	55	43
2/8/2020	no CO2	cow	sock	0	6	4	40	44	1	51	50
2/8/2020	leg	cow	sock	0	2	9	37	46	3	51	48
2/8/2020	intact	cow	sock	0	10	3	37	40	1	51	50
2/10/2020	antenna	cow	sock	0	0	35	18	53	0	53	53
2/10/2020	intact	cow	sock	2	26	5	20	25	0	53	53
2/11/2020	no CO2	cow	sock	0	0	25	19	44	0	44	44
2/11/2020	intact	cow	sock	0	23	4	17	21	0	44	44
2/11/2020	MP	cow	sock	0	1	7	31	38	5	44	39
2/11/2020	leg	cow	sock	0	1	30	8	38	4	43	39
2/14/2020	no CO2	sock	cow	0	0	34	16	50	0	50	50
2/14/2020	labella	sock	cow	0	5	32	7	39	6	50	44
2/14/2020	leg	sock	cow	0	4	13	33	46	0	50	50
2/14/2020	antenna	sock	cow	0	0	24	23	47	3	50	47
2/14/2020	MP	sock	cow	0	2	10	36	46	1	49	48
2/14/2020	intact	sock	cow	0	21	7	22	29	0	50	50
2/16/2020	no CO2	cow	sock	0	0	43	5	48	2	50	48
2/16/2020	intact	cow	sock	1	7	25	17	42	0	50	50
2/16/2020	leg	cow	sock	0	12	19	16	35	3	50	47
2/16/2020	MP	cow	sock	0	0	16	30	46	4	50	46

Table A2 Continued

Date	Condition	Left Port	Right Port	Cow Port	Human Port	Release Cage	Wind Tunnel	All Non-Responders	Starting #	Alive	Active
2/17/2020	antenna	sock	cow	0	0	21	2	23	22	45	23
2/17/2020	labella	sock	cow	0	4	15	15	30	12	46	34
2/17/2020	intact	sock	cow	1	3	14	26	40	1	45	44
2/24/2020	no CO2	cow	sock	0	8	27	60	87	5	100	95
2/24/2020	MP	cow	sock	0	5	17	75	92	3	100	97
2/24/2020	antenna	cow	sock	0	0	48	30	78	22	100	78
2/24/2020	leg	cow	sock	0	14	22	57	79	7	100	93
2/24/2020	intact	cow	sock	0	14	14	63	77	9	100	91
3/2/2020	labella	sock	cow	0	9	29	19	48	91	148	57
3/2/2020	intact	sock	cow	0	4	40	46	86	54	144	90
3/6/2020	labella	cow	sock	0	0	6	8	14	50	64	14
3/6/2020	intact	cow	sock	0	12	8	16	24	4	40	36
3/7/2020	labella	cow	sock	0	1	2	4	6	31	38	7
3/7/2020	intact	cow	sock	0	7	9	17	26	4	37	33
3/14/2020	labella	sock	cow	0	5	31	22	53	42	100	58
3/14/2020	intact	sock	cow	0	10	3	79	82	8	100	92

APPENDIX C

Table A3 – McDonald-Kreitman Test Results for *An. arabiensis – An. coluzzii*

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
2R	50,165,767	*Gr1*	1	0	56	21	1	NA
2R	18,508,124	*Gr2*	0	0	88	29	NA	NA
3R	44,359,442	*Gr3*	0	0	39	35	NA	NA
3R	44,357,873	*Gr4*	0	0	57	56	NA	NA
3R	44,353,861	*Gr7*	2	1	41	53	0.582657	2.585
3R	44,356,455	*Gr8*	1	2	41	36	0.601777	0.439
3R	43,668,907	*Gr11*	0	0	50	30	NA	NA
3R	43,667,301	*Gr12*	0	0	46	55	NA	NA
2R	24,811,173	*Gr13-RA*	2	0	52	59	0.226138	NA
2R	24,811,174	*Gr13-RB*	2	0	52	53	0.495327	NA
2L	31,086,199	*Gr14*	0	0	72	25	NA	NA
2R	34,483,756	*Gr15*	0	0	61	15	NA	NA
2R	34,486,991	*Gr16*	0	0	55	32	NA	NA
2R	34,490,489	*Gr17*	0	0	42	22	NA	NA
2R	34,492,401	*Gr18*	0	1	41	17	0.305085	0
2R	34,498,848	*Gr19*	0	0	46	32	NA	NA
2R	34,500,539	*Gr20-RA*	0	0	37	9	NA	NA
2R	34,500,539	*Gr20-RB*	0	0	37	8	NA	NA
2R	34,504,196	*Gr21*	3	1	64	16	1	0.75
3R	47,434,756	*Gr22*	2	0	37	8	1	NA
2R	32,427,076	*Gr23*	0	0	65	18	NA	NA
2R	12,028,579	*Gr24*	0	0	50	19	NA	NA
2L	2,812,195	*Gr25*	2	1	15	7	1	0.933
2L	37,150,360	*Gr26*	0	0	91	38	NA	NA
2L	37,148,677	*Gr27*	0	0	66	65	NA	NA
2L	37,132,998	*Gr28-RA/RB*	0	0	60	28	NA	NA
2L	39,308,977	*Gr29*	0	0	91	55	NA	NA
2L	39,310,745	*Gr30*	0	0	55	67	NA	NA
2L	39,312,570	*Gr31*	0	0	67	44	NA	NA
2L	39,314,564	*Gr32-RA*	0	0	58	37	NA	NA
2L	39,314,564	*Gr32-RB*	0	0	59	28	NA	NA

Table A3 Continued

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
3R	50,417,385	*Gr33*	1	0	23	4	1	NA
2L	32,255,763	*Gr34*	0	0	71	17	NA	NA
3L	34,950,515	*Gr35*	0	0	75	20	NA	NA
2R	449,457	*Gr36*	0	0	36	29	NA	NA
2R	374,651	*Gr37-RA*	0	0	34	27	NA	NA
2R	374,651	*Gr37-RB*	0	0	35	23	NA	NA
2R	374,651	*Gr37-RC*	0	0	41	16	NA	NA
2R	374,651	*Gr37-RD*	0	0	40	20	NA	NA
2R	374,651	*Gr37-RE*	0	0	33	27	NA	NA
2R	374,651	*Gr37-RF*	0	0	20	23	NA	NA
2R	374,651	*Gr37-RG/RH*	0	0	32	23	NA	NA
2R	346,604	*Gr38*	0	0	29	28	NA	NA
2R	439,782	*Gr39*	0	0	38	23	NA	NA
2R	443,778	*Gr40*	0	0	34	33	NA	NA
2R	446,998	*Gr41*	0	0	13	43	NA	NA
2R	355,689	*Gr42*	0	0	38	22	NA	NA
2L	8,794,945	*Gr43*	1	0	83	29	1	NA
3R	29,972,930	*Gr44-RA*	2	7	100	80	0.082906	0.229
3R	29,972,930	*Gr44-RB*	1	0	90	62	1	NA
3R	29,972,930	*Gr44-RC*	0	0	42	19	NA	NA
3R	29,972,930	*Gr44-RD*	3	5	76	61	0.469029	0.482
3R	29,972,930	*Gr44-RE*	0	0	48	52	NA	NA
3R	315,137	*Gr45*	1	0	36	38	0.493333	NA
3R	307,779	*Gr46*	2	2	31	26	1	0.839
2L	16,537,983	*Gr47*	0	0	88	24	NA	NA
2R	688,812	*Gr48*	0	0	61	53	NA	NA
2R	685,937	*Gr49-RA*	0	0	49	29	NA	NA
2R	685,937	*Gr49-RB*	4	0	55	55	0.119316	NA
2R	690,556	*Gr50*	3	7	51	40	0.181513	0.336
2R	691,958	*Gr51*	0	0	62	56	NA	N
2R	693,555	*Gr52*	0	0	63	54	NA	NA
2R	24,694,665	*Gr53*	0	0	57	4	NA	NA
2R	54,384,164	*Gr54*	0	0	32	33	NA	NA
2L	39,994,248	*Gr55*	0	0	68	66	NA	NA
2L	27,137,739	*Gr56-RA*	1	1	62	66	1	1.065
2L	27,137,739	*Gr56-RB*	0	2	53	61	0.4994	0

Table A3 Continued

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
2L	27,137,739	*Gr56-RC*	0	0	50	43	NA	NA
2L	27,137,739	*Gr56-RD*	0	0	66	24	NA	NA
2L	27,137,739	*Gr56-RE*	1	1	46	57	1	1.239
2L	27,137,739	*Gr56-RF*	0	0	64	38	NA	NA
2L	2,624,121	*Gr57*	0	1	16	25	1	0
2R	453,144	*Gr58*	0	0	33	26	NA	NA
2R	567,440	*Gr59*	0	0	52	26	NA	NA
2R	445,479	*Gr60*	0	0	37	18	NA	NA

Table A4 – McDonald-Kreitman Test Results for *An. arabiensis* – *An. gambiae* s.s.

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
2R	50,165,767	*Gr1*	0	0	66	29	NA	NA
2R	18,508,124	*Gr2*	0	0	118	47	NA	NA
3R	44,359,442	*Gr3*	0	0	51	62	NA	NA
3R	44,357,873	*Gr4*	0	0	74	65	NA	NA
3R	44,353,861	*Gr7*	2	0	53	69	0.194729	NA
3R	44,356,455	*Gr8*	1	2	60	51	0.596655	0.425
3R	43,668,907	*Gr11*	0	0	55	38	NA	NA
3R	43,667,301	*Gr12*	0	0	45	53	NA	NA
2R	24,811,173	*Gr13-RA*	0	0	62	64	NA	NA
2R	24,811,174	*Gr13-RB*	0	0	60	57	NA	NA
2L	31,086,199	*Gr14*	0	0	107	42	NA	NA
2R	34,483,756	*Gr15*	0	0	98	21	NA	NA
2R	34,486,991	*Gr16*	0	0	77	43	NA	NA
2R	34,490,489	*Gr17*	0	0	72	35	NA	NA
2R	34,492,401	*Gr18*	0	0	63	48	NA	NA
2R	34,498,848	*Gr19*	0	0	66	44	NA	NA
2R	34,500,539	*Gr20-RA*	0	0	37	9	NA	NA
2R	34,500,539	*Gr20-RB*	0	0	37	8	NA	NA
2R	34,504,196	*Gr21*	1	0	93	30	1	NA

Table A4 Continued

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
3R	47,434,756	*Gr22*	2	0	47	9	1	NA
2R	32,427,076	*Gr23*	0	0	90	23	NA	NA
2R	12,028,579	*Gr24*	0	0	84	33	NA	NA
2L	2,812,195	*Gr25*	2	1	15	8	1	1.067
2L	37,150,360	*Gr26*	0	0	93	34	NA	NA
2L	37,148,677	*Gr27*	0	0	77	74	NA	NA
2L	37,132,998	*Gr28-RA/RB*	0	0	70	43	NA	NA
2L	39,308,977	*Gr29*	0	0	110	76	NA	NA
2L	39,310,745	*Gr30*	0	0	77	100	NA	NA
2L	39,312,570	*Gr31*	0	0	74	57	NA	NA
2L	39,314,564	*Gr32-RA*	0	0	88	49	NA	NA
2L	39,314,564	*Gr32-RB*	0	0	75	48	NA	NA
3R	50,417,385	*Gr33*	1	0	60	10	1	NA
2L	32,255,763	*Gr34*	0	0	103	19	NA	NA
3L	34,950,515	*Gr35*	0	0	109	23	NA	NA
2R	449,457	*Gr36*	0	0	45	37	NA	NA
2R	374,651	*Gr37-RA*	0	0	38	25	NA	NA
2R	374,651	*Gr37-RB*	0	0	34	17	NA	NA
2R	374,651	*Gr37-RC*	0	0	45	20	NA	NA
2R	374,651	*Gr37-RD*	0	0	41	19	NA	NA
2R	374,651	*Gr37-RE*	0	0	37	27	NA	NA
2R	374,651	*Gr37-RF*	0	0	30	25	NA	NA
2R	374,651	*Gr37-RG/RH*	0	0	36	22	NA	NA
2R	346,604	*Gr38*	0	0	39	34	NA	NA
2R	439,782	*Gr39*	0	0	45	33	NA	NA
2R	443,778	*Gr40*	0	0	41	38	NA	NA
2R	446,998	*Gr41*	0	0	10	60	NA	NA
2R	355,689	*Gr42*	0	0	40	36	NA	NA
2L	8,794,945	*Gr43*	0	0	108	29	NA	NA
3R	29,972,930	*Gr44-RA*	1	3	100	95	0.363724	0.317
3R	29,972,930	*Gr44-RB*	1	0	105	69	1	NA
3R	29,972,930	*Gr44-RC*	0	0	62	23	NA	NA
3R	29,972,930	*Gr44-RD*	2	4	91	73	0.411823	0.401
3R	29,972,930	*Gr44-RE*	0	0	65	74	NA	NA
3R	315,137	*Gr45*	0	0	43	41	NA	NA
3R	307,779	*Gr46*	0	0	33	37	NA	NA

Table A4 Continued

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
2L	16,537,983	*Gr47*	0	0	103	40	NA	NA
2R	688,812	*Gr48*	0	0	88	52	NA	NA
2R	685,937	*Gr49-RA*	0	0	58	47	NA	NA
2R	685,937	*Gr49-RB*	4	1	63	65	0.365494	4.127
2R	690,556	*Gr50*	3	7	73	41	0.044982	0.241
2R	691,958	*Gr51*	0	0	80	80	NA	NA
2R	693,555	*Gr52*	0	0	69	70	NA	NA
2R	24,694,665	*Gr53*	0	0	69	3	NA	NA
2R	54,384,164	*Gr54*	0	0	42	46	NA	NA
2L	39,994,248	*Gr55*	0	0	80	68	NA	NA
2L	27,137,739	*Gr56-RA*	0	1	82	81	1	0
2L	27,137,739	*Gr56-RB*	0	2	74	67	0.231065	0
2L	27,137,739	*Gr56-RC*	0	0	74	60	NA	NA
2L	27,137,739	*Gr56-RD*	0	0	87	30	NA	NA
2L	27,137,739	*Gr56-RE*	1	1	64	59	1	0.922
2L	27,137,739	*Gr56-RF*	0	0	78	48	NA	NA
2L	2,624,121	*Gr57*	0	1	9	27	1	0
2R	453,144	*Gr58*	0	0	39	34	NA	NA
2R	567,440	*Gr59*	0	0	56	25	NA	NA
2R	445,479	*Gr60*	0	0	38	26	NA	NA

Table A5 – McDonald-Kreitman Test Results for *An. arabiensis* – *An. melas*

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
2R	50,165,767	*Gr1*	13	6	30	21	0.584289	1.517
3R	44,359,442	*Gr3*	9	4	16	21	0.196326	2.953
3R	44,357,873	*Gr4*	7	10	34	27	0.410839	0.556
3R	44,353,861	*Gr7*	9	5	24	20	0.555114	1.5
3R	44,356,455	*Gr8*	20	13	16	18	0.330183	1.731
3R	43,668,907	*Gr11*	11	16	33	17	0.052678	0.354
3R	43,667,301	*Gr12*	9	13	20	35	0.796511	1.212

Table A5 Continued

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
2R	24,811,173	*Gr13-RA*	7	2	42	37	0.2883	3.083
2R	24,811,174	*Gr13-RB*	6	2	41	32	0.457443	2.341
2L	31,086,199	*Gr14*	10	7	41	16	0.373928	0.557
2R	34,483,756	*Gr15*	12	4	37	11	1	0.892
2R	34,486,991	*Gr16*	11	6	44	21	1	0.875
2R	34,490,489	*Gr17*	8	6	23	17	1	0.986
2R	34,492,401	*Gr18*	20	11	32	18	1	1.023
2R	34,498,848	*Gr19*	9	11	35	25	0.314261	0.584
2R	34,500,539	*Gr20-RA*	10	1	47	12	0.675705	2.553
2R	34,500,539	*Gr20-RB*	10	1	47	11	0.674436	2.34
2R	34,504,196	*Gr21*	21	7	33	9	0.776479	0.818
3R	47,434,756	*Gr22*	12	1	16	8	0.119346	6
2R	32,427,076	*Gr23*	15	2	32	12	0.311173	2.813
2R	12,028,579	*Gr24*	7	0	43	14	0.196566	NA
2L	2,812,195	*Gr25*	11	3	13	7	0.467408	1.974
2L	37,150,360	*Gr26*	19	5	26	18	0.11408	2.631
2L	37,148,677	*Gr27*	13	13	29	27	1	0.931
2L	37,132,998	*Gr28-RA/RB*	10	7	27	19	1	1.005
2L	39,308,977	*Gr29*	15	11	44	31	1	0.961
2L	39,310,745	*Gr30*	11	15	33	45	1	1
2L	39,312,570	*Gr31*	22	13	24	17	0.81816	1.113
2L	39,314,564	*Gr32-RA*	17	10	28	19	0.809544	1.154
2L	39,314,564	*Gr32-RB*	19	7	23	15	0.422419	1.77
3R	50,417,385	*Gr33*	12	0	16	6	0.069121	NA
2L	32,255,763	*Gr34*	9	2	51	11	1	0.971
3L	34,950,515	*Gr35*	21	7	54	18	1	1
2R	449,457	*Gr36*	20	13	19	14	1	1.134
2R	374,651	*Gr37-RA*	11	13	27	14	0.127552	0.439
2R	374,651	*Gr37-RB*	10	1	20	12	0.129055	6
2R	374,651	*Gr37-RC*	12	2	24	8	0.476917	2
2R	374,651	*Gr37-RD*	11	3	22	10	0.724105	1.667
2R	374,651	*Gr37-RE*	14	8	17	16	0.417456	1.647
2R	374,651	*Gr37-RF*	10	3	18	15	0.196833	2.778
2R	374,651	*Gr37-RG/RH*	11	13	25	13	0.186193	0.44
2R	346,604	*Gr38*	16	19	14	21	0.809403	1.263
2R	439,782	*Gr39*	12	12	18	14	0.787537	0.778

Table A5 Continued

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
2R	443,778	*Gr40*	13	10	23	22	0.798622	1.243
2R	446,998	*Gr41*	7	19	8	22	1	1.013
2R	355,689	*Gr42*	14	17	17	18	0.809541	0.872
2L	8,794,945	*Gr43*	16	7	36	18	1	1.143
3R	315,137	*Gr45*	10	10	26	18	0.590206	0.692
3R	307,779	*Gr46*	11	10	13	18	0.573456	1.523
2L	16,537,983	*Gr47*	4	5	56	19	0.110756	0.271
2R	688,812	*Gr48*	21	14	25	18	1	1.08
2R	685,937	*Gr49-RA*	19	17	26	16	0.493085	0.688
2R	685,937	*Gr49-RB*	18	15	28	17	0.641667	0.729
2R	690,556	*Gr50*	16	13	29	23	1	0.976
2R	691,958	*Gr51*	17	12	23	30	0.248882	1.848
2R	693,555	*Gr52*	12	15	22	51	0.235001	1.855
2R	24,694,665	*Gr53*	8	1	40	3	1	0.6
2R	54,384,164	*Gr54*	12	12	15	19	0.790441	1.267
2L	39,994,248	*Gr55*	11	11	32	36	1	1.125
2R	453,144	*Gr58*	18	20	25	21	0.661345	0.756
2R	567,440	*Gr59*	9	6	23	17	1	1.109
2R	445,479	*Gr60*	15	14	11	17	0.428656	1.656

Table A6 – McDonald-Kreitman Test Results for *An. arabiensis – An. merus*

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
2R	50,165,767	*Gr1*	16	7	36	17	1	1.079
2R	18,508,124	*Gr2*	8	6	64	28	0.369637	0.583
3R	44,359,442	*Gr3*	9	6	23	23	0.562468	1.5
3R	44,357,873	*Gr4*	20	20	36	27	0.544872	0.75
3R	44,350,967	*Gr6*	NA	NA	NA	NA	NA	NA
3R	44,353,861	*Gr7*	10	13	21	26	1	0.952
3R	44,356,455	*Gr8*	18	13	20	16	1	1.108
3R	43,670,895	*Gr10*	NA	NA	NA	NA	NA	NA
3R	43,668,907	*Gr11*	12	7	49	19	0.782203	0.814

Table A6 Continued

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
3R	43,667,301	*Gr12*	9	10	28	37	0.796439	1.189
2R	24,811,173	*Gr13-RA*	15	8	46	49	0.169156	1.997
2R	24,811,174	*Gr13-RB*	13	8	46	41	0.477235	1.448
2L	31,086,199	*Gr14*	13	9	47	17	0.281771	0.522
2R	34,483,756	*Gr15*	14	2	48	14	0.500885	2.042
2R	34,486,991	*Gr16*	10	6	48	21	0.766382	0.729
2R	34,490,489	*Gr17*	14	4	24	18	0.154534	2.625
2R	34,492,401	*Gr18*	17	10	31	21	0.812564	1.152
2R	34,498,848	*Gr19*	12	8	39	29	1	1.115
2R	34,500,539	*Gr20-RA*	10	4	47	17	1	0.904
2R	34,500,539	*Gr20-RB*	10	4	47	17	1	0.904
2R	34,504,196	*Gr21*	17	5	46	11	0.760101	0.813
3R	47,434,756	*Gr22*	17	0	17	7	0.029471	NA
2R	32,427,076	*Gr23*	15	2	39	13	0.326145	2.5
2R	12,028,579	*Gr24*	7	1	57	21	0.673597	2.579
2L	2,812,195	*Gr25*	18	5	10	5	0.472559	1.8
2L	37,150,360	*Gr26*	12	12	30	20	0.45975	0.667
2L	37,148,677	*Gr27*	14	21	31	35	0.534702	0.753
2L	37,132,998	*Gr28-RA/RB*	11	9	43	32	1	0.91
2L	39,308,977	*Gr29*	12	7	61	34	1	0.956
2L	39,310,745	*Gr30*	12	17	33	42	0.829403	0.898
2L	39,312,570	*Gr31*	20	14	27	24	0.659457	1.27
2L	39,314,564	*Gr32-RA*	17	10	21	21	0.329765	1.7
2L	39,314,564	*Gr32-RB*	16	16	26	21	0.654289	0.808
3R	50,417,385	*Gr33*	18	0	13	3	0.093583	NA
2L	32,255,763	*Gr34*	7	3	57	14	0.679535	0.573
3L	34,950,515	*Gr35*	10	5	62	14	0.293753	0.452
2R	449,457	*Gr36*	24	20	25	20	1	0.96
2R	374,651	*Gr37-RA*	11	8	23	14	0.779778	0.837
2R	374,651	*Gr37-RB*	4	0	20	13	0.275645	NA
2R	374,651	*Gr37-RC*	9	1	26	10	0.410253	3.462
2R	374,651	*Gr37-RD*	8	1	20	11	0.233181	4.4
2R	374,651	*Gr37-RE*	13	5	16	19	0.085399	3.088
2R	374,651	*Gr37-RF*	7	3	20	19	0.477811	2.217
2R	374,651	*Gr37-RG/RH*	11	8	22	13	0.775344	0.813
2R	346,604	*Gr38*	21	4	17	17	0.012366	5.25

Table A6 Continued

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
2R	439,782	*Gr39*	15	5	15	14	0.139272	2.8
2R	443,778	*Gr40*	9	8	24	21	1	0.984
2R	446,998	*Gr41*	7	20	8	28	0.771825	1.225
2R	355,689	*Gr42*	12	8	20	16	0.785263	1.2
2L	8,794,945	*Gr43*	14	6	37	17	1	1.072
3R	315,137	*Gr45*	18	8	23	17	0.438209	1.663
3R	307,779	*Gr46*	12	6	13	17	0.144671	2.615
2L	16,537,983	*Gr47*	10	4	62	23	1	0.927
2R	688,812	*Gr48*	14	13	30	20	0.629948	0.718
2R	690,556	*Gr50*	17	20	31	24	0.396357	0.658
2R	691,958	*Gr51*	7	19	36	33	0.037337	0.338
2R	693,555	*Gr52*	13	14	28	39	0.648401	1.293
2R	54,384,164	*Gr54*	22	11	13	23	0.016033	3.538
2L	39,994,248	*Gr55*	8	10	29	30	0.791962	0.828
2L	27,137,739	*Gr56-RA*	13	14	34	46	0.657832	1.256
2L	27,137,739	*Gr56-RB*	11	6	40	39	0.422456	1.788
2L	27,137,739	*Gr56-RC*	10	17	35	39	0.377377	0.655
2L	27,137,739	*Gr56-RD*	5	10	38	17	0.016953	0.224
2L	27,137,739	*Gr56-RE*	11	10	23	46	0.130247	2.2
2L	27,137,739	*Gr56-RF*	7	6	40	29	1	0.846
2R	453,144	*Gr58*	16	16	26	19	0.642774	0.731
2R	567,440	*Gr59*	13	12	24	15	0.60447	0.677
2R	445,479	*Gr60*	23	12	13	16	0.129917	2.359

Table A7 – McDonald-Kreitman Test Results for *An. arabiensis – An. quadriannulatus*

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
2R	50,165,767	*Gr1*	2	0	57	27	0.564159	NA
2R	18,508,124	*Gr2*	0	1	76	40	0.350427	0
3R	44,359,442	*Gr3*	1	1	29	31	1	1.069
3R	44,357,873	*Gr4*	0	5	67	57	0.023511	0
3R	44,353,861	*Gr7*	3	2	39	39	1	1.5
3R	44,356,455	*Gr8*	5	3	40	26	1	1.083
3R	43,668,907	*Gr11*	0	1	57	21	0.278481	0
3R	43,667,301	*Gr12*	1	1	44	46	1	NA
2R	24,811,173	*Gr13-RA*	5	4	58	46	1	0.991

Table A7 Continued

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
2L	31,086,199	*Gr14*	2	2	87	22	0.197489	0.253
2R	34,483,756	*Gr15*	3	2	54	13	0.27726	0.361
2R	34,486,991	*Gr16*	5	4	60	30	0.714464	0.625
2R	34,490,489	*Gr17*	4	1	41	21	0.660591	2.049
2R	34,492,401	*Gr18*	2	6	51	20	0.013557	0.131
2R	34,498,848	*Gr19*	7	0	46	28	0.050412	NA
2R	34,500,539	*Gr20-RA*	5	2	57	15	0.639477	0.658
2R	34,500,539	*Gr20-RB*	5	2	57	15	0.639477	0.658
2R	34,504,196	*Gr21*	6	4	64	13	0.198994	0.305
3R	47,434,756	*Gr22*	9	0	29	10	0.172248	NA
2R	32,427,076	*Gr23*	2	0	48	19	1	NA
2R	12,028,579	*Gr24*	0	0	62	20	NA	NA
2L	2,812,195	*Gr25*	6	2	13	6	1	1.385
2L	37,150,360	*Gr26*	7	0	69	31	0.104657	NA
2L	37,148,677	*Gr27*	15	8	44	58	0.066505	2.472
2L	37,132,998	*Gr28-RA/RB*	5	3	49	24	1	0.816
2L	39,308,977	*Gr29*	10	8	75	43	0.603318	0.717
2L	39,310,745	*Gr30*	4	7	60	48	0.342001	0.457
2L	39,312,570	*Gr31*	13	10	45	36	1	1.04
2L	39,314,564	*Gr32-RA*	5	5	58	30	0.487435	0.517
2L	39,314,564	*Gr32-RB*	11	8	54	28	0.597514	0.713
3R	50,417,385	*Gr33*	5	0	27	6	0.57015	NA
2L	32,255,763	*Gr34*	0	0	65	11	NA	NA
3L	34,950,515	*Gr35*	0	0	78	23	NA	NA
2R	449,457	*Gr36*	11	7	21	13	1	0.973
2R	374,651	*Gr37-RA*	7	7	27	16	0.532435	
2R	374,651	*Gr37-RB*	6	1	20	16	0.214871	4.8
2R	374,651	*Gr37-RC*	5	0	31	15	0.304711	NA
2R	374,651	*Gr37-RD*	4	0	26	14	0.289692	NA
2R	374,651	*Gr37-RE*	9	1	21	21	0.031572	9
2R	374,651	*Gr37-RF*	4	2	23	20	0.677878	1.739
2R	374,651	*Gr37-RG/RH*	7	7	26	15	0.52876	0.577
2R	346,604	*Gr38*	14	4	15	12	0.204147	2.8
2R	439,782	*Gr39*	14	0	25	16	0.005139	NA
2R	443,778	*Gr40*	8	3	26	20	0.49655	2.051
2R	446,998	*Gr41*	2	13	7	31	1	0.681
2R	355,689	*Gr42*	10	4	22	19	0.349779	2.159

Table A7 Continued

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
2L	8,794,945	*Gr43*	8	2	46	18	0.718969	1.565
3R	29,972,930	*Gr44-RA*	4	4	93	86	1	0.925
3R	29,972,930	*Gr44-RB*	5	4	89	71	1	0.997
3R	29,972,930	*Gr44-RC*	2	1	58	21	1	0.724
3R	29,972,930	*Gr44-RD*	6	6	69	55	0.767184	0.797
3R	29,972,930	*Gr44-RE*	12	2	52	56	0.009646	6.462
3R	315,137	*Gr45*	3	0	24	24	0.237647	NA
3R	307,779	*Gr46*	5	3	17	23	0.441736	2.255
2L	16,537,983	*Gr47*	2	1	77	23	1	0.597
2R	688,812	*Gr48*	6	12	40	26	0.06037	0.325
2R	685,937	*Gr49-RA*	6	6	33	29	1	0.879
2R	685,937	*Gr49-RB*	9	6	38	32	0.779336	1.263
2R	690,556	*Gr50*	4	2	43	30	1	1.395
2R	691,958	*Gr51*	4	6	51	40	0.505573	0.523
2R	693,555	*Gr52*	3	6	44	48	0.497591	0.545
2R	24,694,665	*Gr53*	0	0	52	4	NA	NA
2R	54,384,164	*Gr54*	11	9	26	40	0.302887	1.88
2L	39,994,248	*Gr55*	1	3	47	42	0.350955	0.298
2L	27,137,739	*Gr56-RA*	5	3	68	70	0.718776	1.716
2L	27,137,739	*Gr56-RB*	5	6	63	51	0.752848	0.675
2L	27,137,739	*Gr56-RC*	2	3	56	49	0.665704	0.583
2L	27,137,739	*Gr56-RD*	1	0	61	30	1	NA
2L	27,137,739	*Gr56-RE*	7	6	30	46	0.372668	1.789
2L	27,137,739	*Gr56-RF*	3	0	63	45	0.27027	NA
2L	2,624,121	*Gr57*	0	6	5	23	0.5585	0
2R	453,144	*Gr58*	10	14	23	26	0.803376	0.807
2R	567,440	*Gr59*	3	1	29	23	0.627281	2.379
2R	445,479	*Gr60*	12	9	19	22	0.592092	1.544

Table A8 – McDonald-Kreitman Test Results for *An. coluzzii* – *An. gambiae* s.s.

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
2R	50,165,767	*Gr1*	0	0	60	27	NA	NA
2R	18,508,124	*Gr2*	0	0	99	32	NA	NA

Table A8 Continued

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
3R	44,359,442	*Gr3*	0	0	50	60	NA	NA
3R	44,357,873	*Gr4*	0	0	64	65	NA	NA
3R	44,350,967	*Gr6*	0	0	72	54	NA	NA
3R	44,353,861	*Gr7*	0	0	49	81	NA	NA
3R	44,356,455	*Gr8*	0	0	56	48	NA	NA
3R	43,668,907	*Gr11*	0	0	44	30	NA	NA
3R	43,667,301	*Gr12*	0	0	53	50	NA	NA
2R	24,811,173	*Gr13-RA*	0	0	43	53	NA	NA
2R	24,811,174	*Gr13-RB*	0	0	41	49	NA	NA
2L	31,086,199	*Gr14*	0	0	113	40	NA	NA
2R	34,483,756	*Gr15*	0	0	85	22	NA	NA
2R	50,165,767	*Gr1*	0	0	60	27	NA	NA
2R	18,508,124	*Gr2*	0	0	99	32	NA	NA
3R	44,359,442	*Gr3*	0	0	50	60	NA	NA
3R	44,357,873	*Gr4*	0	0	64	65	NA	NA
3R	44,350,967	*Gr6*	0	0	72	54	NA	NA
3R	44,353,861	*Gr7*	0	0	49	81	NA	NA
3R	44,356,455	*Gr8*	0	0	56	48	NA	NA
3R	43,670,895	*Gr10*	NA	NA	NA	NA	NA	NA
3R	43,668,907	*Gr11*	0	0	44	30	NA	NA
3R	43,667,301	*Gr12*	0	0	53	50	NA	NA
2R	24,811,173	*Gr13-RA*	0	0	43	53	NA	NA
2R	24,811,174	*Gr13-RB*	0	0	41	49	NA	NA
2L	31,086,199	*Gr14*	0	0	113	40	NA	NA
2R	34,483,756	*Gr15*	0	0	85	22	NA	NA
2R	34,486,991	*Gr16*	0	0	58	36	NA	NA
2R	34,490,489	*Gr17*	0	0	70	33	NA	NA
2R	34,492,401	*Gr18*	0	0	56	42	NA	NA
2R	34,498,848	*Gr19*	0	0	50	36	NA	NA
2R	34,500,539	*Gr20-RA*	0	0	0	0	NA	NA
2R	34,500,539	*Gr20-RB*	0	0	0	0	NA	NA
2R	34,504,196	*Gr21*	0	0	90	34	NA	NA
3R	47,434,756	*Gr22*	0	0	50	7	NA	NA
2R	32,427,076	*Gr23*	0	0	97	18	NA	NA

Table A8 Continued

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
2R	12,028,579	*Gr24*	0	0	74	26	NA	NA
2L	2,812,195	*Gr25*	0	0	11	8	NA	NA
2L	37,150,360	*Gr26*	0	0	106	34	NA	NA
2L	37,148,677	*Gr27*	0	0	83	71	NA	NA
2L	37,132,998	*Gr28-RA/RB*	0	0	76	41	NA	NA
2L	39,308,977	*Gr29*	0	0	105	74	NA	NA
2L	39,310,745	*Gr30*	0	0	79	103	NA	NA
2L	39,312,570	*Gr31*	0	0	78	57	NA	NA
2L	39,314,564	*Gr32-RA*	0	0	90	49	NA	NA
2L	39,314,564	*Gr32-RB*	0	0	81	47	NA	NA
3R	50,417,385	*Gr33*	0	0	59	7	NA	NA
2L	32,255,763	*Gr34*	0	0	98	19	NA	NA
3L	34,950,515	*Gr35*	0	0	112	22	NA	NA
2R	449,457	*Gr36*	0	0	44	38	NA	NA
2R	374,651	*Gr37-RA*	0	0	31	31	NA	NA
2R	374,651	*Gr37-RB*	0	0	31	22	NA	NA
2R	374,651	*Gr37-RC*	0	0	42	21	NA	NA
2R	374,651	*Gr37-RD*	0	0	39	20	NA	NA
2R	374,651	*Gr37-RE*	0	0	32	25	NA	NA
2R	374,651	*Gr37-RF*	0	0	24	24	NA	NA
2R	374,651	*Gr37-RG/RH*	0	0	30	26	NA	NA
2R	346,604	*Gr38*	0	0	34	35	NA	NA
2R	439,782	*Gr39*	0	0	44	33	NA	NA
2R	443,778	*Gr40*	0	0	41	36	NA	NA
2R	446,998	*Gr41*	0	0	12	60	NA	NA
2R	355,689	*Gr42*	0	0	37	35	NA	NA
2L	8,794,945	*Gr43*	0	0	109	34	NA	NA
3R	29,972,930	*Gr44-RA*	0	0	85	83	NA	NA
3R	29,972,930	*Gr44-RB*	0	0	97	50	NA	NA
3R	29,972,930	*Gr44-RC*	0	0	48	23	NA	NA
3R	29,972,930	*Gr44-RD*	0	0	83	61	NA	NA
3R	29,972,930	*Gr44-RE*	0	0	50	46	NA	NA
3R	315,137	*Gr45*	0	0	34	38	NA	NA

Table A8 Continued

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
3R	307,779	*Gr46*	0	0	34	35	NA	NA
2L	16,537,983	*Gr47*	0	0	93	33	NA	NA
2R	688,812	*Gr48*	0	0	91	57	NA	NA
2R	685,937	*Gr49-RA*	0	0	61	48	NA	NA
2R	685,937	*Gr49-RB*	0	0	72	77	NA	NA
2R	690,556	*Gr50*	0	0	86	43	NA	NA
2R	691,958	*Gr51*	0	0	90	88	NA	NA
2R	693,555	*Gr52*	0	0	83	70	NA	NA
2R	24,694,665	*Gr53*	0	0	58	4	NA	NA
2R	54,384,164	*Gr54*	0	0	38	38	NA	NA
2L	39,994,248	*Gr55*	0	0	88	77	NA	NA
2L	27,137,739	*Gr56-RA*	0	0	83	81	NA	NA
2L	27,137,739	*Gr56-RB*	0	0	68	80	NA	NA
2L	27,137,739	*Gr56-RC*	0	0	70	64	NA	NA
2L	27,137,739	*Gr56-RD*	0	0	94	33	NA	NA
2L	27,137,739	*Gr56-RE*	0	0	78	59	NA	NA
2L	27,137,739	*Gr56-RF*	0	0	83	40	NA	NA
2L	2,624,121	*Gr57*	0	0	15	23	NA	NA
2R	453,144	*Gr58*	0	0	25	27	NA	NA
2R	567,440	*Gr59*	0	0	60	27	NA	NA
2R	445,479	*Gr60*	0	0	42	22	NA	NA

Table A9 – McDonald-Kreitman Test Results for *An. coluzzii – An. melas*

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
2R	50,165,767	*Gr1*	10	7	42	20	0.567821	0.68
3R	44,359,442	*Gr3*	7	5	35	32	0.761855	1.28
3R	44,357,873	*Gr4*	6	5	47	46	1	1.174
3R	44,350,967	*Gr6*	14	6	44	32	0.442357	1.697
3R	44,353,861	*Gr7*	8	5	39	47	0.37421	1.928
3R	44,356,455	*Gr8*	12	7	34	34	0.43636	1.714
3R	43,668,907	*Gr11*	10	13	36	20	0.131403	0.427
3R	43,667,301	*Gr12*	8	14	40	36	0.228389	0.514

Table A9 Continued

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
2R	24,811,173	*Gr13-RA*	7	1	27	32	0.0544	8.296
2R	24,811,174	*Gr13-RB*	6	1	26	30	0.104196	6.923
2L	31,086,199	*Gr14*	5	3	59	19	0.416027	0.537
2R	34,483,756	*Gr15*	11	6	31	11	0.534566	0.651
2R	34,486,991	*Gr16*	11	5	28	18	0.765216	1.414
2R	34,490,489	*Gr17*	4	6	35	18	0.160817	0.343
2R	34,492,401	*Gr18*	19	10	33	13	0.613474	0.748
2R	34,498,848	*Gr19*	10	13	25	19	0.317543	0.585
2R	34,500,539	*Gr20-RA*	22	3	10	3	0.642588	2.2
2R	34,500,539	*Gr20-RB*	22	3	10	3	0.642588	2.2
2R	34,504,196	*Gr21*	15	6	45	15	0.776743	0.833
3R	47,434,756	*Gr22*	7	1	34	8	1	1.647
2R	32,427,076	*Gr23*	14	1	53	11	0.444721	2.906
2R	12,028,579	*Gr24*	6	1	28	10	0.662754	2.143
2L	2,812,195	*Gr25*	12	3	12	8	0.281424	2.667
2L	37,150,360	*Gr26*	9	5	81	27	0.518116	0.6
2L	37,148,677	*Gr27*	6	10	60	48	0.192309	0.48
2L	37,132,998	*Gr28-RA/RB*	6	6	48	18	0.172207	0.375
2L	39,308,977	*Gr29*	6	5	77	49	0.751975	0.764
2L	39,310,745	*Gr30*	8	10	52	69	1	1.062
2L	39,312,570	*Gr31*	6	7	64	38	0.36576	0.509
2L	39,314,564	*Gr32-RA*	6	4	59	30	0.732593	0.763
2L	39,314,564	*Gr32-RB*	10	5	56	24	1	0.857
3R	50,417,385	*Gr33*	9	0	20	4	0.312024	NA
2L	32,255,763	*Gr34*	8	1	57	16	0.676911	2.246
3L	34,950,515	*Gr35*	19	9	52	18	0.617862	0.731
2R	449,457	*Gr36*	17	8	32	25	0.340356	1.66
2R	374,651	*Gr37-RA*	9	12	27	26	0.610517	0.722
2R	374,651	*Gr37-RB*	9	1	25	18	0.075637	6.48
2R	374,651	*Gr37-RC*	9	1	30	11	0.417394	3.3
2R	374,651	*Gr37-RD*	8	3	34	15	1	1.176
2R	374,651	*Gr37-RE*	11	8	25	18	1	0.99
2R	374,651	*Gr37-RF*	6	3	21	17	0.712959	1.619
2R	374,651	*Gr37-RG/RH*	9	12	25	22	0.600245	0.66
2R	346,604	*Gr38*	10	18	25	29	0.480655	0.644

Table A9 Continued

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
2R	439,782	*Gr39*	10	10	34	24	0.603483	0.706
2R	443,778	*Gr40*	9	10	32	28	0.793227	0.793227
2R	446,998	*Gr41*	6	16	12	35	1	1.094
2R	355,689	*Gr42*	10	14	30	20	0.212499	0.476
2L	8,794,945	*Gr43*	11	6	74	25	0.553883	0.619
3R	315,137	*Gr45*	10	9	30	28	1	1.037
3R	307,779	*Gr46*	7	8	22	17	0.556651	0.676
2L	16,537,983	*Gr47*	3	5	62	17	0.022536	0.165
2R	688,812	*Gr48*	13	7	62	52	0.467297	1.558
2R	685,937	*Gr49-RA*	15	14	41	22	0.255351	0.575
2R	685,937	*Gr49-RB*	11	7	55	56	0.449271	1.6
2R	690,556	*Gr50*	15	15	49	39	0.59	0.796
2R	691,958	*Gr51*	11	6	55	52	0.433592	1.733
2R	693,555	*Gr52*	7	9	58	61	0.793434	0.818
2R	24,694,665	*Gr53*	10	1	27	3	1	1.111
2R	54,384,164	*Gr54*	9	11	25	20	0.591259	0.655
2L	39,994,248	*Gr55*	8	5	58	57	0.562798	1.572
2L	27,137,739	*Gr56-RA*	NA	NA	NA	NA	NA	NA
2L	27,137,739	*Gr56-RB*	NA	NA	NA	NA	NA	NA
2L	27,137,739	*Gr56-RC*	NA	NA	NA	NA	NA	NA
2L	27,137,739	*Gr56-RD*	NA	NA	NA	NA	NA	NA
2L	27,137,739	*Gr56-RE*	NA	NA	NA	NA	NA	NA
2L	27,137,739	*Gr56-RF*	NA	NA	NA	NA	NA	NA
2L	2,624,121	*Gr57*	NA	NA	NA	NA	NA	NA
2R	453,144	*Gr58*	23	21	20	19	1	1.04
2R	567,440	*Gr59*	6	5	45	24	0.515493	0.64
2R	445,479	*Gr60*	7	13	34	16	0.015958	0.253

Table A10 – McDonald-Kreitman Test Results for *An. coluzzii – An. merus*

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
2R	50,165,767	*Gr1*	10	7	48	17	0.243333	0.506
2R	18,508,124	*Gr2*	7	8	60	18	0.02655	0.263
3R	44,359,442	*Gr3*	6	5	41	34	1	0.995
3R	44,357,873	*Gr4*	19	14	49	46	0.685978	1.274
3R	44,350,967	*Gr6*	16	4	49	36	0.076662	2.939
3R	44,353,861	*Gr7*	8	10	38	53	1	1.116
3R	44,356,455	*Gr8*	14	8	37	31	0.471505	1.466
3R	43,668,907	*Gr11*	12	5	43	22	0.781423	1.228
3R	43,667,301	*Gr12*	10	10	46	38	0.804488	0.826
2R	24,811,173	*Gr13-RA*	13	6	32	42	0.071309	2.844
2R	24,811,174	*Gr13-RB*	11	6	32	37	0.278675	2.12
2L	31,086,199	*Gr14*	10	5	64	19	0.513824	0.594
2R	34,483,756	*Gr15*	14	4	42	15	0.770163	1.25
2R	34,486,991	*Gr16*	11	5	32	18	0.773738	1.238
2R	34,490,489	*Gr17*	9	4	36	19	1	1.188
2R	34,492,401	*Gr18*	15	9	30	16	1	0.889
2R	34,498,848	*Gr19*	12	11	28	22	0.804088	0.857
2R	34,500,539	*Gr20-RA*	22	6	13	8	0.221874	2.256
2R	34,500,539	*Gr20-RB*	22	6	13	8	0.221874	2.256
2R	34,504,196	*Gr21*	11	4	57	16	0.738379	0.772
3R	47,434,756	*Gr22*	10	0	35	7	0.321761	NA
2R	32,427,076	*Gr23*	15	1	59	12	0.447028	3.051
2R	12,028,579	*Gr24*	8	2	42	18	0.712965	1.714
2L	2,812,195	*Gr25*	19	4	9	6	0.149902	3.167
2L	37,150,360	*Gr26*	6	10	84	31	0.007758	0.221
2L	37,148,677	*Gr27*	5	15	58	56	0.050239	0.322
2L	37,132,998	*Gr28-RA/RB*	9	7	60	31	0.572367	0.664
2L	39,308,977	*Gr29*	7	3	91	51	0.751704	1.308
2L	39,310,745	*Gr30*	8	11	55	66	0.810101	0.873
2L	39,312,570	*Gr31*	5	6	69	44	0.348154	0.531
2R	32,427,076	*Gr23*	15	1	59	12	0.447028	3.051
2L	39,314,564	*Gr32-RA*	10	6	53	33	1	1.038
2L	39,314,564	*Gr32-RB*	10	13	56	31	0.093812	0.426
3R	50,417,385	*Gr33*	15	0	18	1	1	NA
2L	32,255,763	*Gr34*	8	3	64	19	1	0.792

Table A10 Continued

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
3L	34,950,515	*Gr35*	9	6	60	14	0.093544	0.35
2R	449,457	*Gr36*	23	14	39	31	0.544369	1.306
2R	374,651	*Gr37-RA*	11	5	22	25	0.156277	2.5
2R	374,651	*Gr37-RB*	4	1	24	18	0.634799	3
2R	374,651	*Gr37-RC*	7	1	31	12	0.661832	2.71
2R	374,651	*Gr37-RD*	6	0	32	15	0.167239	NA
2R	374,651	*Gr37-RE*	11	5	23	20	0.379531	1.913
2R	374,651	*Gr37-RF*	5	3	22	20	0.710539	1.515
2R	374,651	*Gr37-RG/RH*	11	5	21	21	0.24647	2.2
2R	346,604	*Gr38*	17	2	28	25	0.005321	7.589
2R	439,782	*Gr39*	12	5	31	22	0.408823	1.703
2R	443,778	*Gr40*	7	7	32	27	1	0.844
2R	446,998	*Gr41*	6	17	12	42	0.77195	1.235
2R	355,689	*Gr42*	10	5	32	18	1	1.125
2L	8,794,945	*Gr43*	9	5	75	23	0.333671	0.552
3R	315,137	*Gr45*	17	7	27	27	0.136992	2.429
3R	307,779	*Gr46*	7	4	22	16	1	1.273
2L	16,537,983	*Gr47*	9	2	67	21	1	1.41
2R	688,812	*Gr48*	7	6	66	54	1	0.955
2R	690,556	*Gr50*	10	7	52	38	1	1.044
2R	691,958	*Gr51*	3	16	66	54	0.002289	0.153
2R	693,555	*Gr52*	8	10	64	50	0.446989	0.625
2R	54,384,164	*Gr54*	18	9	25	23	0.236817	1.84
2L	39,994,248	*Gr55*	8	5	56	51	0.571194	1.457
2L	27,137,739	*Gr56-RA*	11	6	62	71	0.200918	2.099
2L	27,137,739	*Gr56-RB*	9	6	41	57	0.264754	2.085
2L	27,137,739	*Gr56-RC*	7	16	43	48	0.16564	0.488
2L	27,137,739	*Gr56-RD*	4	10	57	23	0.004502	0.161
2L	27,137,739	*Gr56-RE*	6	6	41	51	0.765387	1.244
2L	27,137,739	*Gr56-RF*	6	3	61	33	1	1.082
2R	54,384,164	*Gr54*	18	9	25	23	0.236817	1.84
2R	453,144	*Gr58*	20	16	21	18	1	1.071
2R	567,440	*Gr59*	11	13	44	23	0.143425	0.442
2R	445,479	*Gr60*	16	12	36	15	0.321392	0.556

Table A11 – McDonald-Kreitman Test Results for *An. coluzzii – An. quadriannulatus*

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
2R	50,165,767	*Gr1*	2	0	64	26	1	NA
2R	18,508,124	*Gr2*	0	0	71	32	NA	NA
3R	44,359,442	*Gr3*	0	2	45	36	0.206582	0
3R	44,357,873	*Gr4*	1	0	66	68	0.496296	NA
3R	44,350,967	*Gr6*	5	2	68	46	0.701847	1.691
3R	44,353,861	*Gr7*	2	0	55	63	0.223529	NA
3R	44,356,455	*Gr8*	3	2	53	38	1	1.075
3R	43,668,907	*Gr11*	0	0	58	23	NA	NA
3R	43,667,301	*Gr12*	0	1	61	45	0.429907	0
2R	24,811,173	*Gr13-RA*	5	2	46	42	0.444648	2.283
2R	24,811,174	*Gr13-RB*	5	2	45	38	0.455708	2.111
2L	31,086,199	*Gr14*	0	0	98	24	NA	NA
2R	34,483,756	*Gr15*	4	3	48	13	0.342292	0.361
2R	34,486,991	*Gr16*	8	4	40	27	0.755703	1.35
2R	34,490,489	*Gr17*	2	1	50	21	1	0.84
2R	34,492,401	*Gr18*	1	5	52	15	0.005055	0.058
2R	34,498,848	*Gr19*	8	1	39	21	0.253873	4.308
2R	34,500,539	*Gr20-RA*	16	3	22	6	0.720447	1.455
2R	34,500,539	*Gr20-RB*	16	3	22	6	0.720447	1.455
2R	34,504,196	*Gr21*	2	2	71	19	0.214775	0.268
3R	47,434,756	*Gr22*	3	0	44	10	1	NA
2R	32,427,076	*Gr23*	1	0	67	16	1	NA
2R	12,028,579	*Gr24*	0	0	48	16	NA	NA
2L	2,812,195	*Gr25*	7	1	12	7	0.364497	4.083
2L	37,150,360	*Gr26*	0	0	104	40	NA	NA
2L	37,148,677	*Gr27*	0	0	66	73	NA	NA
2L	37,132,998	*Gr28-RA/RB*	0	0	59	21	NA	NA
2L	39,308,977	*Gr29*	2	0	100	59	0.532764	NA
2L	39,310,745	*Gr30*	1	2	71	73	1	0.514
2L	39,312,570	*Gr31*	0	1	76	53	0.415385	0
2L	39,314,564	*Gr32-RA*	0	1	80	42	0.349594	0
2L	39,314,564	*Gr32-RB*	5	5	80	37	0.29682	0.463
3R	50,417,385	*Gr33*	2	0	32	3	1	NA
2L	32,255,763	*Gr34*	1	0	70	15	1	NA
3L	34,950,515	*Gr35*	0	0	69	22	NA	NA
2R	449,457	*Gr36*	9	3	34	23	0.354383	2.029

Table A11 Continued

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
2R	374,651	*Gr37-RA*	5	6	26	26	1	0.833
2R	374,651	*Gr37-RB*	4	1	25	20	0.382979	3.2
2R	374,651	*Gr37-RC*	3	0	36	16	0.548085	NA
2R	374,651	*Gr37-RD*	1	0	37	17	1	NA
2R	374,651	*Gr37-RE*	7	1	28	21	0.134445	5.25
2R	374,651	*Gr37-RF*	2	2	25	20	1	0.8
2R	374,651	*Gr37-RG/RH*	5	6	25	22	0.743706	0.733
2R	346,604	*Gr38*	9	2	26	20	0.174019	3.462
2R	439,782	*Gr39*	10	0	39	25	0.013353	NA
2R	443,778	*Gr40*	4	3	35	27	1	1.029
2R	446,998	*Gr41*	2	9	12	43	1	0.796
2R	355,689	*Gr42*	7	1	35	21	0.245267	4.2
2L	8,794,945	*Gr43*	4	1	84	26	1	1.238
3R	29,972,930	*Gr44-RA*	4	1	89	88	0.368653	3.955
3R	29,972,930	*Gr44-RB*	6	3	88	61	0.739889	1.386
3R	29,972,930	*Gr44-RC*	3	0	51	24	0.548557	NA
3R	29,972,930	*Gr44-RD*	4	2	76	59	0.698231	1.553
3R	29,972,930	*Gr44-RE*	14	2	45	36	0.023232	5.6
3R	315,137	*Gr45*	4	0	28	34	0.049895	NA
3R	307,779	*Gr46*	0	1	26	24	0.490196	0
2L	16,537,983	*Gr47*	2	1	83	19	0.473114	0.458
2R	688,812	*Gr48*	1	3	70	56	0.328809	0.267
2R	685,937	*Gr49-RA*	5	3	47	33	1	1.17
2R	685,937	*Gr49-RB*	1	1	61	67	1	1.098
2R	690,556	*Gr50*	3	7	63	41	0.092326	0.279
2R	691,958	*Gr51*	1	4	77	59	0.172572	0.192
2R	693,555	*Gr52*	0	2	73	58	0.201641	0
2R	24,694,665	*Gr53*	0	0	37	4	NA	NA
2R	54,384,164	*Gr54*	8	7	35	40	0.778732	1.306
2L	39,994,248	*Gr55*	0	0	71	60	NA	NA
2L	27,137,739	*Gr56-RA*	2	1	86	88	0.620726	2.047
2L	27,137,739	*Gr56-RB*	3	3	61	64	1	1.049
2L	27,137,739	*Gr56-RC*	0	1	62	60	0.495935	0
2L	27,137,739	*Gr56-RD*	0	0	79	35	NA	NA
2L	27,137,739	*Gr56-RE*	3	3	45	50	1	1.111

Table A11 Continued

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
2L	27,137,739	*Gr56-RF*	0	0	78	46	NA	NA
2L	2,624,121	*Gr57*	0	5	13	21	0.148773	0
2R	453,144	*Gr58*	11	13	18	24	1	1.128
2L	27,137,739	*Gr56-RF*	0	0	78	46	NA	NA
2R	567,440	*Gr59*	1	1	47	30	1	0.638
2R	445,479	*Gr60*	4	8	42	22	0.053031	0.262

Table A12 – McDonald-Kreitman Test Results for *An. gambiae* s.s. – *An. melas*

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
2R	50,165,767	*Gr1*	10	7	55	27	0.578857	0.701
3R	44,359,442	*Gr3*	6	5	47	60	0.539214	1.532
3R	44,357,873	*Gr4*	6	6	63	54	1	0.857
3R	44,350,967	*Gr6*	12	6	79	52	0.797375	1.316
3R	44,353,861	*Gr7*	7	3	51	63	0.186774	2.882
3R	44,356,455	*Gr8*	12	7	52	47	0.457567	1.549
3R	43,668,907	*Gr11*	11	12	40	28	0.466893	0.642
3R	43,667,301	*Gr12*	8	15	40	37	0.162759	0.493
2R	24,811,173	*Gr13-RA*	5	1	39	36	0.211825	4.615
2R	24,811,174	*Gr13-RB*	4	1	36	33	0.366171	3.667
2L	31,086,199	*Gr14*	4	2	100	38	1	0.76
2R	34,483,756	*Gr15*	8	5	79	18	0.140419	0.365
2R	34,486,991	*Gr16*	10	5	57	29	1	1.018
2R	34,490,489	*Gr17*	4	6	65	33	0.163631	0.338
2R	34,492,401	*Gr18*	19	10	56	43	0.520706	1.459
2R	34,498,848	*Gr19*	8	13	49	34	0.139737	0.427
2R	34,500,539	*Gr20-RA*	22	3	10	3	0.642588	2.2
2R	34,500,539	*Gr20-RB*	22	3	10	3	0.642588	2.2
2R	34,504,196	*Gr21*	13	6	73	28	0.783545	0.831
3R	47,434,756	*Gr22*	6	1	43	9	1	1.256
2R	32,427,076	*Gr23*	12	1	84	17	0.688637	2.429
2R	12,028,579	*Gr24*	6	1	71	25	0.675914	2.113
2L	2,812,195	*Gr25*	12	3	12	9	0.177512	3

Table A12 Continued

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
2L	37,150,360	*Gr26*	10	6	84	24	0.213585	0.476
2L	37,148,677	*Gr27*	6	10	75	61	0.197264	0.488
2L	37,132,998	*Gr28-RA/RB*	5	5	65	39	0.505196	0.6
2L	39,308,977	*Gr29*	6	4	101	73	1	1.084
2L	39,310,745	*Gr30*	4	10	72	103	0.410095	0.572
2L	39,312,570	*Gr31*	7	5	69	41	1	1.035
2L	39,314,564	*Gr32-RA*	5	4	86	43	0.717964	0.625
2L	39,314,564	*Gr32-RB*	8	5	71	43	1	0.969
3R	50,417,385	*Gr33*	8	0	57	9	0.584267	NA
2L	32,255,763	*Gr34*	5	1	95	18	1	0.947
3L	34,950,515	*Gr35*	15	9	98	21	0.050561	0.357
2R	449,457	*Gr36*	17	8	40	33	0.347856	1.753
2R	374,651	*Gr37-RA*	9	12	27	24	0.604715	0.667
2R	374,651	*Gr37-RB*	9	1	23	12	0.237979	4.696
2R	374,651	*Gr37-RC*	9	1	32	15	0.253431	4.219
2R	374,651	*Gr37-RD*	8	3	30	14	1	1.244
2R	374,651	*Gr37-RE*	8	7	29	18	0.763232	0.709
2R	374,651	*Gr37-RF*	7	3	20	18	0.477979	2.1
2R	374,651	*Gr37-RG/RH*	9	12	25	21	0.437224	0.63
2R	346,604	*Gr38*	8	18	35	33	0.104694	0.419
2R	439,782	*Gr39*	9	10	42	34	0.611453	0.729
2R	443,778	*Gr40*	10	10	39	32	0.801176	0.821
2R	446,998	*Gr41*	6	16	9	51	0.333091	2.125
2R	355,689	*Gr42*	10	14	32	35	0.641105	0.781
2L	8,794,945	*Gr43*	9	6	101	25	0.097598	0.371
3R	315,137	*Gr45*	8	9	37	33	0.788573	0.793
3R	307,779	*Gr46*	6	8	26	31	1	0.894
2L	16,537,983	*Gr47*	3	5	85	32	0.049038	0.226
2R	688,812	*Gr48*	10	7	88	52	0.793911	0.844
2R	685,937	*Gr49-RA*	13	13	55	43	0.659559	0.782
2R	685,937	*Gr49-RB*	9	7	67	66	0.792928	1.267
2R	690,556	*Gr50*	15	14	73	40	0.283476	0.587
2R	691,958	*Gr51*	11	7	76	79	0.456093	1.633
2R	693,555	*Gr52*	7	10	68	79	0.799581	0.813
2R	24,694,665	*Gr53*	8	1	45	3	1	0.533

Table A12 Continued

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
2R	54,384,164	*Gr54*	9	12	35	32	0.61766	0.686
2L	39,994,248	*Gr55*	9	5	75	60	0.58408	1.44
2R	453,144	*Gr58*	24	20	27	27	0.688498	1.2
2R	567,440	*Gr59*	6	5	48	24	0.503887	0.6
2R	445,479	*Gr60*	8	13	32	23	0.132012	0.442

Table A13 – McDonald-Kreitman Test Results for *An. gambiae* s.s. – *An. merus*

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
2R	50,165,767	*Gr1*	8	8	61	24	0.140286	0.393
2R	18,508,124	*Gr2*	7	8	97	38	0.072623	0.343
3R	44,359,442	*Gr3*	6	5	53	61	0.754691	1.381
3R	44,357,873	*Gr4*	19	15	63	54	0.847907	1.086
3R	44,350,967	*Gr6*	13	4	84	56	0.290232	2.167
3R	44,353,861	*Gr7*	8	11	50	68	1	0.989
3R	44,356,455	*Gr8*	13	8	55	46	0.632233	1.359
3R	43,668,907	*Gr11*	13	5	47	30	0.428581	1.66
3R	43,667,301	*Gr12*	10	10	46	39	0.806117	0.848
2R	24,811,173	*Gr13-RA*	11	6	44	48	0.291363	2
2R	24,811,174	*Gr13-RB*	9	6	42	42	0.579396	1.5
2L	31,086,199	*Gr14*	9	5	104	38	0.533349	0.658
2R	34,483,756	*Gr15*	9	3	88	22	0.709333	0.75
2R	34,486,991	*Gr16*	11	5	63	30	1	1.048
2R	34,490,489	*Gr17*	8	4	67	34	1	1.015
2R	34,492,401	*Gr18*	17	8	54	45	0.263324	1.771
2R	34,498,848	*Gr19*	10	10	51	38	0.622107	0.745
2R	34,500,539	*Gr20-RA*	22	6	13	8	0.221874	2.256
2R	34,500,539	*Gr20-RB*	22	6	13	8	0.221874	2.256
2R	34,504,196	*Gr21*	8	4	85	30	0.732318	0.706
3R	47,434,756	*Gr22*	10	0	44	8	0.333359	NA
2R	32,427,076	*Gr23*	13	1	90	18	0.468205	2.6
2R	12,028,579	*Gr24*	8	2	82	33	0.7243	1.61

Table A13 Continued

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
2L	2,812,195	*Gr25*	19	4	9	7	0.146115	3.694
2L	37,150,360	*Gr26*	6	9	88	28	0.011325	0.212
2L	37,148,677	*Gr27*	6	14	75	69	0.093493	0.394
2L	37,132,998	*Gr28-RA/RB*	9	7	79	53	0.793178	0.863
2L	39,308,977	*Gr29*	7	2	116	76	0.486823	2.293
2L	39,310,745	*Gr30*	5	9	76	101	0.78039	0.738
2L	39,312,570	*Gr31*	7	6	73	57	1	0.911
2L	39,314,564	*Gr32-RA*	11	6	79	45	1	1.044
2L	39,314,564	*Gr32-RB*	10	12	71	50	0.349782	0.587
3R	50,417,385	*Gr33*	13	0	55	7	0.343064	NA
2L	32,255,763	*Gr34*	7	3	100	21	0.38973	0.49
3L	34,950,515	*Gr35*	8	5	106	17	0.037348	0.257
2R	449,457	*Gr36*	21	13	46	39	0.540749	1.37
2R	374,651	*Gr37-RA*	11	5	22	24	0.244389	2.4
2R	374,651	*Gr37-RB*	4	1	22	13	0.64043	2.364
2R	374,651	*Gr37-RC*	7	1	33	17	0.413497	3.606
2R	374,651	*Gr37-RD*	6	0	28	15	0.158596	NA
2R	374,651	*Gr37-RE*	10	4	27	21	0.367376	1.944
2R	374,651	*Gr37-RF*	5	3	21	22	0.703002	1.746
2R	374,651	*Gr37-RG/RH*	11	5	21	21	0.24647	2.2
2R	346,604	*Gr38*	16	2	38	29	0.01306	6.105
2R	439,782	*Gr39*	10	5	39	33	0.408716	1.692
2R	443,778	*Gr40*	7	7	39	31	0.773027	0.795
2R	446,998	*Gr41*	6	17	9	58	0.197224	2.275
2R	355,689	*Gr42*	10	4	34	33	0.238676	2.426
2L	8,794,945	*Gr43*	8	5	101	23	0.139846	0.364
3R	315,137	*Gr45*	15	7	34	31	0.222598	1.954
3R	307,779	*Gr46*	7	4	26	30	0.340487	2.019
2L	16,537,983	*Gr47*	8	2	88	36	0.724249	1.636
2R	688,812	*Gr48*	7	6	93	54	0.556539	0.677
2R	690,556	*Gr50*	10	7	78	40	0.592066	0.733
2R	691,958	*Gr51*	4	11	87	81	0.103378	0.339
2R	693,555	*Gr52*	8	10	74	67	0.619379	0.724
2R	54,384,164	*Gr54*	18	10	34	35	0.261188	1.853
2L	39,994,248	*Gr55*	8	6	72	54	1	1

Table A13 Continued

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
2L	27,137,739	*Gr56-RA*	10	6	86	84	0.437926	1.628
2L	27,137,739	*Gr56-RB*	10	7	63	66	0.606977	1.497
2L	27,137,739	*Gr56-RC*	6	12	74	73	0.215218	0.493
2L	27,137,739	*Gr56-RD*	3	11	83	30	0.000228	0.099
2L	27,137,739	*Gr56-RE*	7	5	69	54	1	1.096
2L	27,137,739	*Gr56-RF*	5	2	78	44	1	1.41
2R	453,144	*Gr58*	21	15	28	26	0.666363	1.3
2R	567,440	*Gr59*	11	13	49	22	0.05222	0.38
2R	445,479	*Gr60*	18	11	35	24	0.82177	1.122

Table A14 – McDonald-Kreitman Test Results for *An. gambiae* s.s. – *An. quadriannulatus*

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
2R	50,165,767	*Gr1*	2	0	79	31	1	NA
2R	18,508,124	*Gr2*	0	0	107	51	NA	NA
3R	44,359,442	*Gr3*	0	2	57	61	0.497059	0
3R	44,357,873	*Gr4*	1	0	81	76	1	NA
3R	44,350,967	*Gr6*	5	2	100	66	0.70561	1.65
3R	44,353,861	*Gr7*	2	0	67	79	0.215665	NA
3R	44,356,455	*Gr8*	3	2	68	53	1	1.169
3R	43,668,907	*Gr11*	0	0	60	31	NA	NA
3R	43,667,301	*Gr12*	1	1	61	46	1	0.754
2R	24,811,173	*Gr13-RA*	5	2	57	44	0.696302	1.93
2R	24,811,174	*Gr13-RB*	5	2	54	39	0.696916	1.806
2L	31,086,199	*Gr14*	1	0	136	43	NA	NA
2R	34,483,756	*Gr15*	1	1	93	20	0.333181	0.215
2R	34,486,991	*Gr16*	7	3	71	39	1	1.282
2R	34,490,489	*Gr17*	1	1	80	36	1	0.45
2R	34,492,401	*Gr18*	2	5	76	45	0.109429	0.237
2R	34,498,848	*Gr19*	5	1	60	37	0.409414	3.083
2R	34,500,539	*Gr20-RA*	16	3	22	6	0.720447	1.455
2R	34,500,539	*Gr20-RB*	16	3	22	6	0.720447	1.455
2R	34,504,196	*Gr21*	0	2	98	32	0.064886	0

Table A14 Continued

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
3R	47,434,756	*Gr22*	3	0	53	11	1	NA
2R	32,427,076	*Gr23*	0	0	93	22	NA	NA
2R	12,028,579	*Gr24*	0	0	88	31	NA	NA
2L	2,812,195	*Gr25*	7	1	11	8	0.201119	5.091
2L	37,150,360	*Gr26*	1	0	112	38	1	NA
2L	37,148,677	*Gr27*	0	0	86	87	NA	NA
2L	37,132,998	*Gr28-RA/RB*	0	0	77	42	NA	NA
2L	39,308,977	*Gr29*	2	0	123	83	0.51807	NA
2L	39,310,745	*Gr30*	0	1	88	107	1	0
2L	39,312,570	*Gr31*	1	1	80	65	1	0.813
2L	39,314,564	*Gr32-RA*	0	1	104	54	0.345912	0
2L	39,314,564	*Gr32-RB*	4	3	92	55	1	0.797
3R	50,417,385	*Gr33*	2	0	68	9	1	NA
2L	32,255,763	*Gr34*	0	0	105	18	NA	NA
3L	34,950,515	*Gr35*	0	0	113	24	NA	NA
2R	449,457	*Gr36*	8	2	42	31	0.301874	2.952
2R	374,651	*Gr37-RA*	5	6	25	26	1	0.867
2R	374,651	*Gr37-RB*	4	1	23	16	0.633704	2.783
2R	374,651	*Gr37-RC*	3	0	38	12	0.30722	NA
2R	374,651	*Gr37-RD*	1	0	33	17	1	NA
2R	374,651	*Gr37-RE*	4	1	32	23	0.639177	2.875
2R	374,651	*Gr37-RF*	2	2	24	23	1	0.958
2R	374,651	*Gr37-RG/RH*	5	6	24	23	1	0.799
2R	346,604	*Gr38*	7	2	36	24	0.466375	2.333
2R	439,782	*Gr39*	8	0	47	35	0.020782	NA
2R	443,778	*Gr40*	5	3	42	31	1	1.23
2R	446,998	*Gr41*	2	9	9	59	0.646091	1.457
2R	355,689	*Gr42*	7	1	38	36	0.067239	6.632
2L	8,794,945	*Gr43*	3	1	109	25	1	0.688
3R	29,972,930	*Gr44-RA*	4	1	98	105	0.205461	4.286
3R	29,972,930	*Gr44-RB*	5	3	114	70	1	1.023
3R	29,972,930	*Gr44-RC*	3	0	72	25	0.570965	NA
3R	29,972,930	*Gr44-RD*	4	2	94	69	1	1.468
3R	29,972,930	*Gr44-RE*	12	2	66	56	0.024906	5.091
3R	315,137	*Gr45*	1	0	35	39	0.48	NA
3R	307,779	*Gr46*	0	1	29	37	1	0

Table A14 Continued

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
2L	16,537,983	*Gr47*	1	1	106	34	0.433523	0.321
2R	688,812	*Gr48*	0	3	95	56	0.054463	0
2R	685,937	*Gr49-RA*	4	4	61	55	1	0.902
2R	685,937	*Gr49-RB*	1	1	71	79	1	1.113
2R	690,556	*Gr50*	4	7	88	42	0.048929	0.273
2R	691,958	*Gr51*	0	2	100	83	0.209753	0
2R	693,555	*Gr52*	0	2	82	74	0.229783	0
2R	24,694,665	*Gr53*	0	0	51	4	NA	NA
2R	54,384,164	*Gr54*	7	6	44	52	0.768409	1.379
2L	39,994,248	*Gr55*	1	0	87	62	1	NA
2L	27,137,739	*Gr56-RA*	1	1	111	99	1	0.892
2L	27,137,739	*Gr56-RB*	3	3	85	78	1	0.918
2L	27,137,739	*Gr56-RC*	0	1	89	79	0.473373	0
2L	27,137,739	*Gr56-RD*	0	0	103	42	NA	NA
2L	27,137,739	*Gr56-RE*	1	4	73	54	0.168392	0.185
2L	27,137,739	*Gr56-RF*	0	0	95	56	NA	NA
2L	2,624,121	*Gr57*	0	5	6	24	0.561023	0
2R	453,144	*Gr58*	11	13	25	32	1	1.083
2R	567,440	*Gr59*	1	1	52	29	1	0.558
2R	445,479	*Gr60*	5	8	41	30	0.235751	0.457

Table A15 – McDonald-Kreitman Test Results for *An. melas – An. merus*

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
2R	50,165,767	*Gr1*	24	11	12	15	0.072229	2.727
3R	44,359,442	*Gr3*	13	10	17	19	0.59591	1.453
3R	44,357,873	*Gr4*	23	20	12	10	1	0.958
3R	44,350,967	*Gr6*	19	7	28	20	0.311798	1.939
3R	44,353,861	*Gr7*	17	16	21	20	1	1.012
3R	44,356,455	*Gr8*	29	8	10	14	0.005941	5.075
3R	43,668,907	*Gr11*	24	15	13	6	0.772612	0.738
3R	43,667,301	*Gr12*	15	25	16	12	0.140461	0.45
2R	24,811,173	*Gr13-RA*	15	8	17	18	0.283031	1.985
2R	24,811,174	*Gr13-RB*	14	8	16	15	0.414777	1.641

Table A15 Continued

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
2L	31,086,199	*Gr14*	21	10	19	11	0.791071	1.216
2R	34,483,756	*Gr15*	21	6	15	10	0.231663	2.333
2R	34,486,991	*Gr16*	22	8	18	6	1	0.917
2R	34,490,489	*Gr17*	14	6	12	13	0.224439	2.528
2R	34,492,401	*Gr18*	25	14	19	15	0.631989	1.41
2R	34,498,848	*Gr19*	19	20	12	15	0.804757	1.188
2R	34,500,539	*Gr20-RA*	13	5	23	11	0.765105	1.243
2R	34,500,539	*Gr20-RB*	13	5	23	11	0.765105	1.243
2R	34,504,196	*Gr21*	25	10	20	8	1	1
3R	47,434,756	*Gr22*	17	1	11	7	0.040752	10.818
2R	32,427,076	*Gr23*	34	2	19	5	0.104392	4.474
2R	12,028,579	*Gr24*	13	1	28	12	0.145806	5.571
2L	2,812,195	*Gr25*	22	7	7	6	0.27844	2.694
2L	37,150,360	*Gr26*	23	16	11	9	0.787772	1.176
2L	37,148,677	*Gr27*	16	16	21	16	0.633234	0.762
2L	37,132,998	*Gr28-RA/RB*	18	11	19	19	0.457374	1.636
2L	39,308,977	*Gr29*	17	12	33	21	1	0.902
2L	39,310,745	*Gr30*	13	15	26	32	1	1.067
2L	39,312,570	*Gr31*	13	11	18	14	1	0.081
2L	39,314,564	*Gr32-RA*	19	11	12	13	0.286399	1.871
2L	39,314,564	*Gr32-RB*	22	16	17	14	0.812308	1.132
3R	50,417,385	*Gr33*	19	0	9	3	0.048943	NA
2L	32,255,763	*Gr34*	17	3	30	11	0.352949	2.078
3L	34,950,515	*Gr35*	38	16	17	8	1	1.118
2R	449,457	*Gr36*	30	21	15	16	0.371573	1.524
2R	374,651	*Gr37-RA*	14	14	9	9	1	1
2R	374,651	*Gr37-RB*	12	2	7	6	0.103188	5.143
2R	374,651	*Gr37-RC*	15	3	7	3	0.634496	2.143
2R	374,651	*Gr37-RD*	14	4	8	4	0.677946	1.75
2R	374,651	*Gr37-RE*	13	7	6	8	0.296004	2.476
2R	374,651	*Gr37-RF*	10	6	7	9	0.479488	2.143
2R	374,651	*Gr37-RG/RH*	14	14	8	9	1	1.125
2R	346,604	*Gr38*	21	21	9	16	0.315838	1.778
2R	439,782	*Gr39*	17	18	8	13	0.580418	1.535
2R	443,778	*Gr40*	16	12	12	11	0.782328	1.222
2R	446,998	*Gr41*	10	19	6	16	0.761935	1.404

Table A15 Continued

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
2R	355,689	*Gr42*	19	19	11	14	0.797142	1.273
2L	8,794,945	*Gr43*	17	9	10	11	0.250282	2.078
3R	315,137	*Gr45*	22	13	13	5	0.554787	0.651
3R	307,779	*Gr46*	8	10	2	8	0.2474	3.2
2L	16,537,983	*Gr47*	13	9	21	12	0.782256	0.825
2R	688,812	*Gr48*	25	14	17	16	0.34073	1.681
2R	690,556	*Gr50*	29	32	12	13	1	0.982
2R	691,958	*Gr51*	19	25	19	19	0.657632	0.76
2R	693,555	*Gr52*	17	24	11	36	0.10755	2.318
2R	54,384,164	*Gr54*	24	17	6	9	0.243177	2.118
2L	39,994,248	*Gr55*	22	13	11	12	0.289747	1.846
2R	453,144	*Gr58*	27	22	9	7	1	0.955
2R	567,440	*Gr59*	18	14	10	12	0.580397	1.543
2R	445,479	*Gr60*	22	16	6	15	0.055471	3.438

Table A16 – McDonald-Kreitman Test Results for *An. melas – An. quadriannulatus*

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
2R	50,165,767	*Gr1*	13	6	35	24	0.591696	1.486
3R	44,359,442	*Gr3*	8	6	23	28	0.548893	1.623
3R	44,357,873	*Gr4*	6	6	49	46	1	0.939
3R	44,350,967	*Gr6*	13	8	54	32	1	0.963
3R	44,353,861	*Gr7*	9	5	39	33	0.56571	1.523
3R	44,356,455	*Gr8*	12	10	32	23	0.803191	0.863
3R	43,668,907	*Gr11*	12	15	33	9	0.004906	0.218
3R	43,667,301	*Gr12*	7	18	35	22	0.007919	0.244
2R	24,811,173	*Gr13-RA*	9	4	31	17	1	1.234
2R	24,811,174	*Gr13-RB*	9	4	29	15	1	1.164
2L	31,086,199	*Gr14*	11	4	61	16	0.732675	0.721
2R	34,483,756	*Gr15*	11	4	22	8	1	1
2R	34,486,991	*Gr16*	22	8	18	6	1	0.917
2R	34,490,489	*Gr17*	6	5	28	17	0.73572	0.729
2R	34,492,401	*Gr18*	19	11	42	14	0.320792	0.576
2R	34,498,848	*Gr19*	19	20	12	15	0.804757	1.188
2R	34,500,539	*Gr20-RA*	8	4	32	9	0.459032	0.563

Table A16 Continued

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
2R	34,500,539	*Gr20-RB*	8	4	32	9	0.459032	0.563
2R	34,504,196	*Gr21*	15	5	36	10	1	0.833
3R	47,434,756	*Gr22*	9	1	25	10	0.408741	3.6
2R	32,427,076	*Gr23*	16	1	31	11	0.092354	5.677
2R	12,028,579	*Gr24*	7	1	37	9	1	1.703
2L	2,812,195	*Gr25*	16	3	10	7	0.139368	3.733
2L	37,150,360	*Gr26*	9	6	54	21	0.368281	0.583
2L	37,148,677	*Gr27*	16	15	35	40	0.674166	1.219
2L	37,132,998	*Gr28-RA/RB*	10	8	26	11	0.367685	0.529
2L	39,308,977	*Gr29*	10	7	49	31	1	0.904
2L	39,310,745	*Gr30*	5	9	49	37	0.159051	0.42
2L	39,312,570	*Gr31*	6	8	35	25	0.374961	0.536
2L	39,314,564	*Gr32-RA*	8	6	51	23	0.535995	0.601
2L	39,314,564	*Gr32-RB*	13	11	47	22	0.321856	0.553
3R	50,417,385	*Gr33*	12	0	23	6	0.155654	NA
2L	32,255,763	*Gr34*	11	2	41	8	1	1.073
3L	34,950,515	*Gr35*	20	11	40	18	0.812748	0.818
2R	449,457	*Gr36*	16	8	11	9	0.538514	1.636
2R	374,651	*Gr37-RA*	8	11	12	11	0.551202	0.667
2R	374,651	*Gr37-RB*	13	3	7	9	0.065893	5.571
2R	374,651	*Gr37-RC*	11	3	13	8	0.460645	2.256
2R	374,651	*Gr37-RD*	9	4	12	7	1	1.313
2R	374,651	*Gr37-RE*	14	9	10	10	0.547169	1.556
2R	374,651	*Gr37-RF*	10	6	9	10	0.500103	1.852
2R	374,651	*Gr37-RG/RH*	8	11	11	11	0.755844	0.727
2R	346,604	*Gr38*	17	21	7	11	0.77654	1.272
2R	439,782	*Gr39*	16	13	18	15	1	1.026
2R	443,778	*Gr40*	14	10	14	11	1	1.1
2R	446,998	*Gr41*	7	18	6	18	1	1.167
2R	355,689	*Gr42*	17	15	13	16	0.611129	1.395
2L	8,794,945	*Gr43*	14	7	21	13	0.779057	1.238
3R	315,137	*Gr45*	13	11	14	12	1	1.013
3R	307,779	*Gr46*	7	9	6	16	0.322658	2.074
2L	16,537,983	*Gr47*	7	7	42	12	0.050779	0.286
2R	688,812	*Gr48*	16	14	30	23	0.82086	0.876
2R	685,937	*Gr49-RA*	17	18	22	15	0.478195	0.644

Table A16 Continued

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
2R	685,937	*Gr49-RB*	18	15	30	21	0.821933	0.84
2R	690,556	*Gr50*	19	16	27	21	1	0.924
2R	691,958	*Gr51*	12	13	37	28	0.485598	0.699
2R	693,555	*Gr52*	9	16	26	49	1	1.06
2R	24,694,665	*Gr53*	9	1	20	3	1	1.35
2R	54,384,164	*Gr54*	16	20	17	26	0.819087	1.224
2L	39,994,248	*Gr55*	15	10	32	26	0.810321	1.219
2R	453,144	*Gr58*	31	30	6	13	0.189965	2.239
2R	567,440	*Gr59*	9	7	15	21	0.378305	1.8
2R	445,479	*Gr60*	9	9	12	21	0.384988	1.75

Table A17 – McDonald-Kreitman Test Results for *An. merus – An. quadriannulatus*

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
2R	50,165,767	*Gr1*	13	8	42	21	0.792326	0.813
2R	18,508,124	*Gr2*	6	7	48	30	0.365519	0.536
3R	44,359,442	*Gr3*	6	7	29	29	1	0.857
3R	44,357,873	*Gr4*	17	14	51	46	0.839654	1.095
3R	44,350,967	*Gr6*	13	6	60	37	0.617736	1.336
3R	44,353,861	*Gr7*	9	11	38	39	0.804686	0.84
3R	44,356,455	*Gr8*	11	10	36	21	0.440222	0.642
3R	43,668,907	*Gr11*	12	6	40	11	0.350782	0.55
3R	43,667,301	*Gr12*	9	12	43	24	0.126014	0.419
2R	24,811,173	*Gr13-RA*	16	8	36	29	0.467872	1.611
2R	24,811,174	*Gr13-RB*	15	8	35	24	0.801602	1.286
2L	31,086,199	*Gr14*	8	5	66	16	0.153072	0.388
2R	34,483,756	*Gr15*	9	4	33	12	1	0.818
2R	34,486,991	*Gr16*	16	6	34	15	1	1.176
2R	34,490,489	*Gr17*	11	3	30	18	0.34579	2.2
2R	34,492,401	*Gr18*	14	6	39	17	1	1.017
2R	34,498,848	*Gr19*	14	11	27	17	0.799421	0.801
2R	34,500,539	*Gr20-RA*	9	4	34	14	1	0.926
2R	34,500,539	*Gr20-RB*	9	4	34	14	1	0.926

Table A17 Continued

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
2R	34,504,196	*Gr21*	8	7	49	12	0.045437	0.28
3R	47,434,756	*Gr22*	10	0	26	9	0.096201	NA
2R	32,427,076	*Gr23*	17	1	39	11	0.160295	4.795
2R	12,028,579	*Gr24*	8	2	51	18	1	1.412
2L	2,812,195	*Gr25*	21	5	7	5	0.234686	3
2L	37,150,360	*Gr26*	8	7	56	24	0.237381	0.49
2L	37,148,677	*Gr27*	10	17	36	48	0.657847	0.784
2L	37,132,998	*Gr28-RA/RB*	10	8	41	24	0.592926	0.732
2L	39,308,977	*Gr29*	11	7	65	35	0.792432	0.846
2L	39,310,745	*Gr30*	5	11	48	37	0.100035	0.35
2L	39,312,570	*Gr31*	6	6	39	32	0.764815	0.821
2L	39,314,564	*Gr32-RA*	12	10	44	25	0.460311	0.682
2L	39,314,564	*Gr32-RB*	6	9	47	26	0.090847	0.369
3R	50,417,385	*Gr33*	19	0	20	3	0.238676	NA
2L	32,255,763	*Gr34*	9	3	47	10	0.684932	0.638
3L	34,950,515	*Gr35*	9	4	48	14	0.721665	0.656
2R	449,457	*Gr36*	22	14	17	15	0.624502	1.387
2R	374,651	*Gr37-RA*	8	9	8	10	1	1.111
2R	374,651	*Gr37-RB*	7	3	7	9	0.247517	3
2R	374,651	*Gr37-RC*	9	2	15	9	0.435397	2.7
2R	374,651	*Gr37-RD*	6	1	11	7	0.362297	3.818
2R	374,651	*Gr37-RE*	14	7	9	12	0.214603	2.667
2R	374,651	*Gr37-RF*	7	5	11	12	0.724672	1.527
2R	374,651	*Gr37-RG/RH*	8	9	8	10	1	1.111
2R	346,604	*Gr38*	21	4	10	7	0.086378	3.675
2R	439,782	*Gr39*	12	5	14	14	0.222393	2.4
2R	443,778	*Gr40*	14	5	15	10	0.521867	1.867
2R	446,998	*Gr41*	7	17	6	25	0.525055	1.716
2R	355,689	*Gr42*	13	6	16	15	0.376369	2.031
2L	8,794,945	*Gr43*	14	6	22	12	0.771184	1.273
3R	315,137	*Gr45*	21	10	10	11	0.164239	2.31
3R	307,779	*Gr46*	7	5	6	15	0.142135	3.5
2L	16,537,983	*Gr47*	11	1	49	16	0.281737	3.592
2R	688,812	*Gr48*	9	9	33	25	0.786766	0.758
2R	690,556	*Gr50*	17	19	29	20	0.378464	0.617
2R	691,958	*Gr51*	4	11	49	30	0.020777	0.223

Table A17 Continued

Chromosome	Starting Position	Gene	dS	dN	pS	pN	p-value	NI
2R	693,555	*Gr52*	6	4	32	35	0.517019	1.641
2R	54,384,164	*Gr54*	21	16	17	29	0.081197	0.081197
2L	39,994,248	*Gr55*	10	9	29	20	0.785363	0.766
2L	27,137,739	*Gr56-RA*	6	1	58	61	0.114837	6.31
2L	27,137,739	*Gr56-RB*	7	4	38	41	0.521626	1.888
2L	27,137,739	*Gr56-RC*	4	4	43	47	1	1.093
2L	27,137,739	*Gr56-RD*	7	9	40	28	0.401805	0.544
2L	27,137,739	*Gr56-RE*	11	12	23	25	1	0.996
2L	27,137,739	*Gr56-RF*	7	4	48	36	0.755705	1.313
2R	453,144	*Gr58*	28	24	7	12	0.284773	2
2R	567,440	*Gr59*	13	14	15	19	0.800165	1.176
2R	445,479	*Gr60*	16	12	14	20	0.307293	1.905

APPENDIX D

The following tables present results from the DH suite of tests for each species. *Grs* are only included if there were ten or more sequences available to analyze from a given species. D refers to Tajima's D, H refers to normalized Fay and Wu's H, DH refers to the joint DH test and is only presented as a p-value, and E refers to the E test.

Table A18 – DH Suite of Tests for *An. arabiensis*

Chromosome	Starting Position	Gene	D	p(D)	H	p(H)	DH	E	p(E)
3R	50,417,385	Gr33	0.100	0.598	-0.656	0.157	0.339	0.656	0.806
2R	54,384,164	Gr54	-0.032	0.537	0.663	0.715	0.494	-0.649	0.347
2R	445,479	Gr60	-1.089	0.144	0.397	0.527	0.289	-1.206	0.135

Table A19 – DH Suite of Tests for *An. coluzzii*

Chromosome	Starting Position	Gene	D	p(D)	H	p(H)	DH	E	p(E)
2R	50,165,767	Gr1	0.428	0.726	-0.337	0.222	0.526	0.743	0.844
2R	18,508,124	Gr2	-0.608	0.294	-0.301	0.235	0.165	-0.229	0.520
3R	44,359,442	Gr3	-0.897	0.189	0.479	0.573	0.365	-1.320	0.104
3R	44,357,873	Gr4	-0.252	0.449	-0.042	0.311	0.278	-0.192	0.536
3R	44,350,967	Gr6	-0.509	0.335	-0.232	0.256	0.183	-0.227	0.528
3R	44,353,861	Gr7	-0.983	0.161	0.238	0.425	0.252	-1.184	0.151
3R	44,356,455	Gr8	-0.803	0.222	-0.051	0.307	0.157	-0.683	0.331
3R	43,668,907	Gr11	0.466	0.744	0.126	0.377	0.548	0.297	0.715
3R	43,667,301	Gr12	-1.001	0.155	-1.163	0.097	0.074	0.275	0.711
2R	24,811,173	Gr13	-1.125	0.124	-0.821	0.135	0.068	-0.186	0.536

Table A19 Continued

Chromosome	Starting Position	Gene	D	p(D)	H	p(H)	DH	E	p(E)
2R	34,483,756	Gr15	-0.935	0.179	-0.746	0.148	0.084	-0.050	0.593
2R	34,490,489	Gr17	-1.236	0.095	-1.573	0.058	0.047	0.481	0.775
2R	34,492,401	Gr18	-0.815	0.218	0.483	0.573	0.354	-1.223	0.135
2R	34,498,848	Gr19	-0.521	0.337	-0.904	0.122	0.185	0.457	0.764
2R	34,500,539	Gr20*	NA	NA	NA	NA	NA	NA	NA
2R	34,504,196	Gr21	0.520	0.761	-0.577	0.175	0.579	1.096	0.906
3R	47,434,756	Gr22	-0.646	0.279	-0.261	0.246	0.150	-0.321	0.487
2R	32,427,076	Gr23	-0.826	0.213	0.018	0.333	0.179	-0.772	0.292
2R	12,028,579	Gr24	0.438	0.715	-0.491	0.191	0.503	0.861	0.858
2L	2,812,195	Gr25	-0.704	0.262	0.575	0.653	0.430	-1.214	0.142
2L	37,148,677	Gr27	0.524	0.759	0.170	0.395	0.562	0.312	0.717
2L	39,308,977	Gr29	-0.014	0.536	0.493	0.573	0.368	-0.574	0.392
2L	39,310,745	Gr30	-0.233	0.459	0.149	0.385	0.275	-0.375	0.462
2L	39,312,570	Gr31	0.635	0.791	-0.453	0.202	0.608	1.069	0.899
2L	39,314,564	Gr32	0.303	0.617	0.136	0.393	0.380	0.114	0.631
3R	50,417,385	Gr33	-0.473	0.361	-0.354	0.221	0.184	-0.050	0.582
2L	32,255,763	Gr34	-0.564	0.314	-1.115	0.099	0.175	0.642	0.812
3L	34,950,515	Gr35	0.077	0.590	-0.202	0.263	0.383	0.283	0.700
2R	449,457	Gr36	-0.903	0.189	-0.528	0.182	0.085	-0.280	0.503
2R	374,651	Gr37	-0.558	0.312	0.169	0.394	0.235	-0.733	0.322
2R	346,604	Gr38	-0.217	0.466	-0.311	0.235	0.276	0.127	0.650
2R	439,782	Gr39	-0.266	0.448	0.565	0.640	0.419	-0.825	0.276
2R	443,778	Gr40	-0.516	0.332	-0.074	0.297	0.184	-0.399	0.458
2R	446,998	Gr41	-0.157	0.489	0.110	0.364	0.298	-0.256	0.513
2R	355,689	Gr42	-0.538	0.324	-0.184	0.270	0.177	-0.308	0.489
2L	8,794,945	Gr43	-0.446	0.358	0.212	0.413	0.245	-0.661	0.344
3R	315,137	Gr45	-0.595	0.302	-0.207	0.261	0.155	-0.329	0.477
2L	16,537,983	Gr47	-0.290	0.425	-0.478	0.196	0.250	0.248	0.698

*No polymorphisms were identified in this *Gr*

Table A19 Continued

Chromosome	Starting Position	Gene	D	p(D)	H	p(H)	DH	E	p(E)
2R	688,812	Gr48	-0.140	0.498	-0.212	0.262	0.309	0.092	0.642
2R	685,937	Gr49	-0.390	0.387	-0.063	0.306	0.227	-0.307	0.503
2R	690,556	Gr50	-0.273	0.442	-0.176	0.271	0.262	-0.069	0.591
2R	691,958	Gr51	-0.280	0.426	-0.149	0.282	0.257	-0.101	0.579
2R	693,555	Gr52	-0.128	0.506	-0.223	0.261	0.314	0.115	0.660
2R	24,694,665	Gr53	-0.588	0.302	-0.400	0.218	0.176	-0.119	0.569
2R	54,384,164	Gr54	0.383	0.714	0.328	0.476	0.504	0.013	0.617
2L	39,994,248	Gr55	-0.047	0.534	0.245	0.434	0.337	-0.310	0.489
2L	27,137,739	Gr56	-0.219	0.451	-0.274	0.249	0.277	0.101	0.642
2L	2,624,121	Gr57	-0.810	0.220	0.795	0.827	0.669	-1.589	0.052
2R	453,144	Gr58	0.458	0.738	0.081	0.354	0.536	0.327	0.723
2R	567440	Gr59	-0.680	0.265	-0.012	0.319	0.175	-0.612	0.361
2R	445479	Gr60	-0.608	0.298	0.400	0.519	0.319	-0.958	0.225

Table A20 – DH Suite of Tests for *An. gambiae* s.s.

Chromosome	Starting Position	Gene	D	p(D)	H	p(H)	DH	E	p(E)
2R	50,165,767	Gr1	-0.761	0.242	-0.279	0.242	0.111	-0.424	0.438
2R	18,508,124	Gr2	-1.357	0.063	-0.039	0.312	0.161	-1.222	0.122
3R	44,359,442	Gr3	-1.522	0.038	0.116	0.373	0.197	-1.506	0.047
3R	44,357,873	Gr4	-0.808	0.225	-0.166	0.271	0.136	-0.585	0.361
3R	44,350,967	Gr6	-1.114	0.120	-0.586	0.176	0.083	-0.461	0.420
3R	44,353,861	Gr7	-1.495	0.041	-0.034	0.316	0.169	-1.359	0.079
3R	44,356,455	Gr8	-1.588	0.030	-0.530	0.185	0.083	-0.943	0.212
3R	43,668,907	Gr11	-0.609	0.306	-0.753	0.147	0.157	0.163	0.678
3R	43,667,301	Gr12	-0.586	0.316	-0.687	0.154	0.164	0.121	0.663
2R	24,811,173	Gr13	-1.047	0.141	-0.576	0.177	0.083	-0.404	0.456

Table A20 Continued

Chromosome	Starting Position	Gene	D	p(D)	H	p(H)	DH	E	p(E)
2R	34,483,756	Gr15	-1.533	0.039	-0.129	0.282	0.148	-1.282	0.101
2R	34,486,991	Gr16	-0.424	0.374	0.237	0.425	0.250	-0.652	0.348
2R	34,490,489	Gr17	-0.862	0.207	-0.705	0.154	0.092	-0.114	0.574
2R	34,492,401	Gr18	-1.448	0.048	0.511	0.609	0.390	-1.816	0.010
2R	34,498,848	Gr19	-0.658	0.285	0.103	0.368	0.202	-0.706	0.314
2R	34,500,539	Gr20	*	*	*	*	*	*	*
2R	34,504,196	Gr21	-0.632	0.296	-0.332	0.230	0.155	-0.259	0.510
3R	47,434,756	Gr22	-1.301	0.078	-0.528	0.184	0.089	-0.673	0.327
2R	32,427,076	Gr23	-1.152	0.112	0.286	0.454	0.259	-1.335	0.092
2R	12,028,579	Gr24	-1.390	0.058	0.414	0.531	0.327	-1.675	0.025
2L	2,812,195	Gr25	-0.364	0.412	-0.622	0.167	0.229	0.262	0.703
2L	37,150,360	Gr26	-0.496	0.330	-0.798	0.145	0.197	0.429	0.745
2L	37,148,677	Gr27	0.015	0.585	0.404	0.528	0.371	-0.377	0.456
2L	39,308,977	Gr29	-0.301	0.439	0.540	0.628	0.628	-0.830	0.265
2L	39,310,745	Gr30	-0.491	0.358	0.321	0.475	0.284	-0.770	0.293
2L	39,312,570	Gr31	-0.059	0.546	-0.655	0.164	0.335	0.613	0.803
2L	39,314,564	Gr32	-0.332	0.417	-0.230	0.254	0.253	-0.078	0.594
3R	50,417,385	Gr33	-1.483	0.043	-0.017	0.321	0.160	-1.315	0.095
2L	32,255,763	Gr34	-1.016	0.149	-0.286	0.242	0.125	-0.660	0.341
3L	34,950,515	Gr35	-1.187	0.102	-0.589	0.172	0.081	-0.519	0.399
2R	449,457	Gr36	-1.275	0.081	-0.776	0.141	0.056	-0.415	0.439
2R	374,651	Gr37	-0.811	0.221	-0.453	0.202	0.109	-0.314	0.487
2R	346,604	Gr38	-0.944	0.171	-0.538	0.183	0.090	-0.348	0.467
2R	439,782	Gr39	-1.096	0.127	-0.162	0.270	0.134	-0.839	0.255
2R	443,778	Gr40	-0.705	0.265	-0.231	0.252	0.132	-0.419	0.429
2R	446,998	Gr41	-0.451	0.374	-0.017	0.322	0.209	-0.391	0.458
2R	355,689	Gr42	-1.350	0.066	-0.415	0.209	0.101	-0.839	0.255
2L	8,794,945	Gr43	-0.916	0.183	0.167	0.395	0.219	-1.023	0.196

*No polymorphisms were identified in this *Gr*

Table A19 Continued

Chromosome	Starting Position	Gene	D	p(D)	H	p(H)	DH	E	p(E)
3R	315,137	Gr45	-0.133	0.519	-0.171	0.267	0.305	0.043	0.630
3R	307,779	Gr46	-1.417	0.060	-0.815	0.141	0.071	-0.487	0.413
2L	16,537,983	Gr47	-0.589	0.317	-0.151	0.278	0.170	-0.400	0.450
2R	688,812	Gr48	-0.973	0.165	-0.446	0.202	0.098	-0.468	0.426
2R	685,937	Gr49	-1.084	0.129	-0.672	0.161	0.075	-0.353	0.467
2R	690,556	Gr50	-0.649	0.285	-0.095	0.292	0.151	-0.504	0.401
2R	691,958	Gr51	-0.922	0.179	-0.452	0.202	0.107	-0.402	0.450
2R	693,555	Gr52	-0.543	0.331	-0.016	0.318	0.184	-0.489	0.407
2R	24,694,665	Gr53	-0.840	0.210	0.269	0.444	0.261	-1.066	0.178
2R	54,384,164	Gr54	-0.484	0.358	0.284	0.452	0.268	-0.731	0.313
2L	39,994,248	Gr55	-1.013	0.151	-0.107	0.291	0.149	-0.832	0.259
2L	27,137,739	Gr56	-0.204	0.454	-0.347	0.230	0.287	0.209	0.670
2L	2,624,121	Gr57	-1.201	0.100	-0.047	0.310	0.147	-1.026	0.189
2R	453,144	Gr58	-0.760	0.246	-0.087	0.298	0.151	-0.597	0.356
2R	567,440	Gr59	-1.383	0.059	-0.881	0.127	0.054	-0.413	0.441
2R	445,479	Gr60	-1.207	0.098	-0.217	0.256	0.128	-0.872	0.245

Table A21 – DH Suite of Tests for *An. melas*

Chromosome	Starting Position	Gene	D	p(D)	H	p(H)	DH	E	p(E)
3R	44,359,442	Gr3	-1.072	0.151	-2.099	0.036	0.044	1.085	0.894
3R	43,668,907	Gr11	-1.095	0.157	-0.989	0.126	0.046	0.125	0.651
3R	43,667,301	Gr12	-1.273	0.103	-0.800	0.139	0.052	-0.163	0.551
2R	34,504,196	Gr21	0.177	0.614	0.437	0.550	0.360	-0.300	0.493
3R	47,434,756	Gr22	-0.658	0.277	-3.071	0.010	0.131	2.566	0.995
2R	449,457	Gr36	-1.095	0.143	0.525	0.614	0.349	-1.245	0.121
2L	39,994,248	Gr55	0.896	0.844	0.191	0.404	0.707	0.482	0.765

Table A22 – DH Suite of Tests for *An. merus*

Chromosome	Starting Position	Gene	D	p(D)	H	p(H)	DH	E	p(E)
2L	2,812,195	Gr25	0.818	0.804	-1.377	0.091	0.656	1.656	0.952
3R	50,417,385	Gr33	0.096	0.560	-2.459	0.040	0.353	1.928	0.971
3R	315,137	Gr45	-1.245	0.136	0.738	0.751	0.564	-1.461	0.089
2L	39994248	Gr55	0.878	0.839	0.200	0.436	0.802	0.312	0.720
2R	453144	Gr58	-0.537	0.343	0.743	0.859	0.731	-1.000	0.213
2R	567440	Gr59	1.789	0.968	0.257	0.405	0.946	0.900	0.870

Table A23 – DH Suite of Tests for *An. quadriannulatus*

Chromosome	Starting Position	Gene	D	p(D)	H	p(H)	DH	E	p(E)
2R	50,165,767	Gr1	-0.034	0.565	-1.767	0.048	0.358	1.567	0.954
3R	44,359,442	Gr3	0.913	0.863	0.527	0.632	0.741	0.322	0.723
3R	44,357,873	Gr4	0.873	0.857	-0.226	0.253	0.730	1.043	0.898
3R	44,350,967	Gr6	0.195	0.657	-0.379	0.218	0.447	0.529	0.790
3R	44,356,455	Gr8	1.390	0.937	-1.347	0.078	0.877	2.352	0.991
3R	43,668,907	Gr11	-0.072	0.551	-1.139	0.095	0.342	0.926	0.877
3R	43,667,301	Gr12	0.949	0.873	-0.798	0.140	0.754	1.595	0.956
2R	34,483,756	Gr15	-0.203	0.489	-2.850	0.014	0.272	2.455	0.992
2R	34,490,489	Gr17	0.085	0.583	-0.468	0.202	0.367	0.588	0.794
2R	34,492,401	Gr18	-0.405	0.391	-1.106	0.101	0.217	0.728	0.837
2R	34,498,848	Gr19	-1.690	0.023	-3.481	0.006	0.007	1.766	0.969
3R	47,434,756	Gr22	0.369	0.692	-0.418	0.209	0.457	0.773	0.850
2R	32,427,076	Gr23	0.810	0.815	-0.110	0.339	0.638	0.581	0.801
2R	12,028,579	Gr24	0.133	0.622	-0.229	0.253	0.410	0.331	0.726
2L	37,148,677	Gr27	0.160	0.635	-0.354	0.220	0.416	0.490	0.770
2L	39,308,977	Gr29	-0.171	0.490	-1.527	0.063	0.287	1.412	0.946
2L	39,310,745	Gr30	0.630	0.798	-0.668	0.161	0.624	1.232	0.923
2L	39,312,570	Gr31	-0.013	0.561	-1.243	0.086	0.345	1.253	0.924
3R	50,417,385	Gr33	-0.931	0.184	0.442	0.569	0.351	-1.154	0.126
2L	32,255,763	Gr34	0.389	0.723	-0.893	0.126	0.529	1.173	0.912
3L	34,950,515	Gr35	0.926	0.866	-0.455	0.199	0.739	1.264	0.923
2R	449,457	Gr36	1.139	0.896	-1.529	0.065	0.806	2.140	0.979
2R	439,782	Gr39	-0.465	0.376	-0.389	0.210	0.190	-0.038	0.587
2R	443,778	Gr40	-0.801	0.235	0.230	0.423	0.207	-0.839	0.248
2R	446,998	Gr41	-1.358	0.069	-1.560	0.059	0.019	0.267	0.701
2L	8,794,945	Gr43	-0.703	0.260	-1.614	0.058	0.135	1.028	0.893
3R	315,137	Gr45	-0.267	0.463	-0.014	0.317	0.234	-0.209	0.520
2R	688,812	Gr48	0.054	0.594	-1.045	0.104	0.370	1.048	0.895

Table A23 Continued

Chromosome	Starting Position	Gene	D	p(D)	H	p(H)	DH	E	p(E)
2R	690,556	Gr50	0.414	0.737	-1.359	0.075	0.544	1.551	0.953
2R	691,958	Gr51	-0.788	0.234	-3.022	0.009	0.108	2.189	0.990
2R	693,555	Gr52	0.047	0.588	-0.863	0.129	0.381	0.892	0.870
2R	54,384,164	Gr54	-1.138	0.120	-0.264	0.241	0.108	-0.698	0.318
2L	39,994,248	Gr55	0.258	0.681	-0.316	0.230	0.468	0.510	0.782
2R	453,144	Gr58	-0.496	0.362	0.744	0.851	0.729	-1.003	0.175
2R	567,440	Gr59	-0.124	0.529	-1.785	0.047	0.308	1.443	0.939
2R	445,479	Gr60	-0.690	0.277	0.421	0.549	0.320	-0.946	0.202

APPENDIX E

The following tables present basic genetic variation statistics calculated by DNAsp. Only *Grs* where there were more ten or more full sequences available for a species are included.

Table A24 – Genetic Variation Statistics for *An. arabiensis*

Chromosome	Starting Point	Gene	N	π	Haplotype Diversity	Haplotype Number
2L	39,314,564	Gr33	26	0.002	0.655	8
2R	54,384,164	Gr54	14	0.002	0.945	10
2R	445,479	Gr60	16	0.002	0.858	9

Table A25 – Genetic Variation Statistics for *An. coluzzii*

Chromosome	Starting Point	Gene	N	π	Haplotype Diversity	Haplotype Number
2R	50,165,767	Gr1	22	0.015	0.987	19
2R	18,508,124	Gr2	14	0.008	1	14
3R	44,359,442	Gr3	24	0.011	0.978	21
3R	44,357,873	Gr4	22	0.018	1	22
3R	44,350,967	Gr6	20	0.011	1	20
3R	44,353,861	Gr7	22	0.014	1	22
3R	44,356,455	Gr8	24	0.012	1	24
3R	43,668,907	Gr11	22	0.014	0.996	21
3R	43,667,301	Gr12	24	0.011	0.989	21
2R	24,811,173	Gr13	22	0.010	1	22
2R	34,483,756	Gr15	16	0.006	0.992	15
2R	34,490,489	Gr17	24	0.008	0.938	17
2R	34,492,401	Gr18	20	0.005	0.984	18
2R	34,498,848	Gr19	22	0.009	0.97	18
2R	34,500,539	Gr20	24	0.000	No polymorphisms	No polymorphisms
2R	34,500,539	Gr21	20	0.014	0.984	18
2R	34,504,196	Gr22	22	0.007	0.991	20
2R	32,427,076	Gr24	12	0.004	0.879	7
2R	12,028,579	Gr25	20	0.004	0.979	16

Table A25 Continued

Chromosome	Starting Point	Gene	N	π	Haplotype Diversity	Haplotype Number
2L	37,150,360	Gr27	18	0.024	1	18
2L	37,132,998	Gr29	12	0.026	1	12
2L	39,308,977	Gr30	24	0.022	0.996	23
2L	39,310,745	Gr31	16	0.012	1	16
2L	39,314,564	Gr33	22	0.003	0.987	19
3R	50,417,385	Gr34	22	0.014	1	22
2L	32,255,763	Gr35	18	0.012	1	18
3L	34,950,515	Gr36	24	0.009	1	24
2R	374,651	Gr37	14	0.008	1	14
2R	346,604	Gr38	20	0.010	0.995	19
2R	439,782	Gr39	22	0.010	0.987	20
2R	443,778	Gr40	22	0.012	1	22
2R	446,998	Gr41	20	0.011	0.995	19
2R	355,689	Gr42	22	0.009	0.996	21
2L	8,794,945	Gr43	16	0.022	1	16
3R	315,137	Gr45	22	0.009	0.991	20
2L	16,537,983	Gr47	16	0.013	1	16
2R	688,812	Gr48	22	0.022	1	22
2R	685,937	Gr49	20	0.019	1	20
2R	690,556	Gr50	20	0.018	1	20
2R	691,958	Gr51	14	0.023	1	14
2R	24,694,665	Gr53	14	0.014	1	14
2R	54,384,164	Gr54	22	0.014	0.991	20
2L	39,994,248	Gr55	18	0.026	0.993	17
2L	27,137,739	Gr56	12	0.022	1	12
2L	2,624,121	Gr57	20	0.004	0.979	16
2R	453,144	Gr58	22	0.009	0.991	20
2R	567,440	Gr59	18	0.011	0.993	17
2R	445,479	Gr60	24	0.009	0.986	20

Table A26 – Genetic Variation Statistics for *An. gambiae* s.s.

Chromosome	Starting Point	Gene	N	π	Haplotype Diversity	Haplotype Number
2R	50,165,767	Gr1	48	0.011	0.97	34
2R	18,508,124	Gr2	42	0.013	0.999	41
3R	44,359,442	Gr3	50	0.011	0.974	40
3R	44,357,873	Gr4	50	0.017	0.999	49
3R	44,350,967	Gr6	50	0.014	0.999	49
3R	44,353,861	Gr7	44	0.013	1	44
3R	44,356,455	Gr8	52	0.010	0.995	47

Table A26 Continued

Chromosome	Starting Point	Gene	N	π	Haplotype Diversity	Haplotype Number
3R	43,668,907	Gr11	52	0.009	0.977	34
3R	43,667,301	Gr12	52	0.012	0.981	38
2R	24,811,173	Gr13	52	0.011	0.971	42
2R	34,483,756	Gr15	44	0.010	0.998	42
2R	34,486,991	Gr16	20	0.016	0.995	19
2R	34,490,489	Gr17	50	0.015	1	50
2R	34,492,401	Gr18	52	0.008	0.995	46
2R	34,498,848	Gr19	52	0.015	0.999	51
2R	34,500,539	Gr20	52	0.000	No polymorphisms	No polymorphisms
2R	34,500,539	Gr21	39	0.017	1	40
2R	34,504,196	Gr22	42	0.007	0.991	35
2R	32,427,076	Gr24	48	0.012	0.998	46
2R	12,028,579	Gr25	46	0.003	0.957	28
2L	2,812,195	Gr26	12	0.013	0.985	11
2L	37,150,360	Gr27	50	0.024	0.992	44
2L	37,132,998	Gr29	34	0.026	0.996	32
2L	39,308,977	Gr30	46	0.026	0.999	45
2L	39,310,745	Gr31	34	0.028	0.998	33
2L	39,312,570	Gr32	28	0.022	0.997	27
2L	39,314,564	Gr33	50	0.006	0.988	40
3R	50,417,385	Gr34	48	0.016	0.996	44
2L	32,255,763	Gr35	44	0.014	1	44
3L	34,950,515	Gr36	52	0.008	0.998	49
2R	374,651	Gr37	44	0.008	1	44
2R	346,604	Gr38	44	0.009	0.998	42
2R	439,782	Gr39	46	0.008	0.970	37
2R	443,778	Gr40	44	0.011	0.996	40
2R	446,998	Gr41	46	0.011	0.998	42
2R	355,689	Gr42	48	0.008	0.999	47
2L	8,794,945	Gr43	40	0.021	1	40
3R	315137	Gr45	48	0.010	0.992	40
3R	307779	Gr46	22	0.006	1	22
2L	16537983	Gr47	46	0.017	1	46
2R	688812	Gr48	48	0.020	0.999	47
2R	685937	Gr49	44	0.016	1	44
2R	690556	Gr50	44	0.020	1	44
2R	691958	Gr51	28	0.021	1	28
2R	24694665	Gr53	38	0.016	1	38
2R	54384164	Gr54	34	0.014	0.996	32
2L	39994248	Gr55	52	0.022	0.998	50
2L	27137739	Gr56	10	0.022	1	10
2L	2624121	Gr57	40	0.002	0.94	22

Table A26 Continued

Chromosome	Starting Point	Gene	N	π	Haplotype Diversity	Haplotype Number
2R	453144	Gr58	52	0.007	0.984	42
2R	567440	Gr59	50	0.008	0.988	48
2R	445479	Gr60	50	0.007	0.988	40

Table A27 – Genetic Variation Statistics for *An. melas*

Chromosome	Starting Point	Gene	N	π	Haplotype Diversity	Haplotype Number
3R	44,359,442	Gr3	22	0.001	0.632	6
3R	43,668,907	Gr11	18	0.001	0.556	5
3R	43,667,301	Gr12	12	0.002	0.667	5
2R	34,500,539	Gr21	10	0.001	0.844	6
2R	34,504,196	Gr22	12	0.002	0.803	5
3L	34,950,515	Gr36	28	0.001	0.661	6
2L	39,994,248	Gr55	10	0.004	0.800	5

Table A28 – Genetic Variation Statistics for *An. merus*

Chromosome	Starting Point	Gene	N	π	Haplotype Diversity	Haplotype Number
2R	12,028,579	Gr25	44	0.0004	0.555	4
2L	39,314,564	Gr33	66	0.000	0.4	4
3R	315,137	Gr45	10	0.001	0.667	5
2R	54,384,164	Gr54	22	0.000	No polymorphisms	No polymorphisms
2R	453,144	Gr58	22	0.001	0.671	5
2R	567,440	Gr59	14	0.001	0.747	4

Table A29 – Genetic Variation Statistics for *An. quadriannulatus*

Chromosome	Starting Point	Gene	N	π	Haplotype Diversity	Haplotype Number
2R	50,165,767	Gr1	74	0.008	0.996	65
3R	44,359,442	Gr3	74	0.009	0.996	62
3R	44,357,873	Gr4	28	0.016	0.989	24
3R	44,350,967	Gr6	62	0.013	0.996	55
3R	44,356,455	Gr8	144	0.011	0.999	129
3R	43,668,907	Gr11	104	0.006	0.993	73
3R	43,667,301	Gr12	46	0.010	0.99	36
2R	34,483,756	Gr15	46	0.005	0.995	41
2R	34,490,489	Gr17	10	0.007	0.978	9
2R	34,492,401	Gr18	30	0.006	0.993	27
2R	34,498,848	Gr19	44	0.002	0.943	27
2R	34,500,539	Gr21	144	0.008	0.999	131
2R	34,504,196	Gr22	12	0.006	0.985	11
2R	32,427,076	Gr24	38	0.005	0.99	31
2L	37,150,360	Gr27	34	0.010	0.995	31
2L	37,132,998	Gr29	22	0.009	0.991	20
2L	39,308,977	Gr30	34	0.013	1	34
2L	39,310,745	Gr31	26	0.010	0.997	25
2L	39,314,564	Gr33	90	0.002	0.959	43
3R	50,417,385	Gr34	50	0.008	0.994	43
2L	32,255,763	Gr35	20	0.008	0.974	15
3L	34,950,515	Gr36	106	0.003	0.931	36
2R	439,782	Gr39	50	0.002	0.956	29
2R	443,778	Gr40	62	0.002	0.857	16
2R	446,998	Gr41	48	0.002	0.94	25
2L	8,794,945	Gr43	20	0.006	0.995	19
3R	315,137	Gr45	80	0.003	0.971	41
2R	688,812	Gr48	36	0.007	0.994	32
2R	690,556	Gr50	92	0.007	0.991	67
2R	691,958	Gr51	44	0.008	0.978	34
2R	54,384,164	Gr54	24	0.003	0.96	17
2L	39,994,248	Gr55	80	0.009	0.997	72
2R	453,144	Gr58	88	0.001	0.872	24
2R	567,440	Gr59	84	0.004	0.99	60
2R	445,479	Gr60	74	0.003	0.973	43