

**Analysis of *Anopheles gambiae* ovary-specific  
Odorant Binding Proteins as potential targets  
for reducing female fertility**

Ανάλυση οσμοδεσμευόμενων πρωτεϊνών που εκφράζονται ειδικά  
στις ωοθήκες του κουνουπιού της ελονοσίας *Anopheles gambiae* ως  
πιθανών στόχων για τη μείωση της γονιμότητας θηλυκών εντόμων

Diploma Project  
Of

**Kyrou Kyros**

Department of Biochemistry & Biotechnology  
Faculty of Life Sciences  
University of Thessaly

And

Division of Cell & Molecular Biology  
Department of Life Sciences  
Imperial College London

London 2014



| <b>Three-Member Board</b>   | <b>Signature</b> |
|---|------------------|
| Crisanti Andrea - Professor<br>Molecular Parasitology<br>Division of Cell & Molecular Biology<br>Department of Life Sciences<br>Imperial College London |                  |
| Mathiopoulos Konstantinos - Professor<br>Molecular Biology<br>Department of Biochemistry and Biotechnology<br>University of Thessaly                    |                  |
| Sarafidou Theologia - Assistant Professor<br>Animal Molecular Genetics<br>Department of Biochemistry and Biotechnology<br>University of Thessaly        |                  |

## ABSTRACT

Malaria is a devastating disease that since antiquity continues to cause severe problems and millions of casualties upon the human race. One of the main vectors that can transmit Malaria disease is *Anopheles gambiae*, a mosquito species that thrives in vast numbers across equatorial regions in Africa. Because of the nearly “pandemic” spread of the disease, many studies have been focused in developing methods that could reduce or even eliminate the spread of the vector, which cause the death of hundreds of thousands of people per year, including a high number in children. Understanding of the insect’s biology is crucial in order to reveal the molecular components that are used in cell pathways which have a primary role in mosquito fertility. Uncovering these functions and processes, new methods could be developed in which crucial components could be targeted, achieving in final the suppression of the mosquito population. The current thesis focuses on a subgroup of the odorant binding proteins (OBPs) which are expressed highly in the ovaries of the blood-fed female mosquitoes, which is the main vector for the transmission of the *Plasmodium falciparum* that cause the disease. Due to their high expression in the ovaries, especially after blood meal, it is believed that these odorant-binding proteins have a crucial role in the development of the eggs. If so, they should be promising candidates for creating transgenic mosquito lines that produce sterile offspring that could drive the population of the mosquitoes to critically low numbers. The aim is to uncover, through in situ hybridization, in which part of the ovary each ovary-specific OBP is expressed. Revealing the within-tissue location of the OBP mRNA, and thus the potential role in the developmental process of the eggs is important. The information can be accessed by further studies which attempt the creation of transgenic mosquito lines that promise population suppression. One such study uses a novel method for editing the genome of the mosquito, known as the CRISPR/Cas9 system, in order to target the candidate genes that seem to have the highest impact on the mosquito fertility.

## ΠΕΡΙΛΗΨΗ

Η Ελονοσία είναι μια καταστροφική ασθένεια η οποία από την αρχαιότητα συνεχίζει να προκαλεί σοβαρά προβλήματα και να προσβάλλει εκατομμύρια ανθρώπους ανά το παγκόσμιο. Ένας από τους κύριους φορείς που μεταδίδουν την ασθένεια της ελονοσίας είναι κουνούπια του είδους *Anopheles gambiae*, τα οποία ευδοκιμούν σε τεράστιους αριθμούς στις περιοχές της τροπικής Αφρικής. Λόγω της σχεδόν «πανδημικής» εξάπλωσης της νόσου, πολλές μελέτες έχουν επικεντρωθεί στην ανάπτυξη μεθόδων που θα μπορούσαν να μειώσουν την εξάπλωση του φορέα ή ακόμη και να εξαλείψουν το φορέα, που προκαλεί το θάνατο εκατοντάδων χιλιάδων ανθρώπων ανά έτος, συμπεριλαμβανομένου ενός μεγάλου αριθμού παιδιών. Η κατανόηση της βιολογίας του εντόμου είναι θέμα ζωτικής σημασίας ούτως ώστε να αποκαλυφθούν οι μοριακοί παράγοντες που συμβάλουν στα μοριακά μονοπάτια που έχουν πρωταρχικό ρόλο στη γονιμότητα των κουνουπιών. Κατανοώντας τις λειτουργίες αυτές και αποκαλύπτοντας καίρια μόρια που συμβάλουν στις επί μέρους διαδικασίες της γονιμότητας, θα μπορούσαν να αναπτυχθούν νέες μέθοδοι οι οποίες θα έχουν στο στόχαστρο τους τα εν λόγω μοριακά συστατικά. Η απαλοιφή των γονιδίων των μοριακών αυτών συστατικών, μπορεί να οδηγήσει εν τέλει στον έλεγχο του πληθυσμού των κουνουπιών. Η τρέχουσα διατριβή εστιάζει σε μια υποομάδα οσμοδεσμευόμενων πρωτεϊνών (OBPs) οι οποίες εκφράζονται ειδικά στις ωθήκες των θηλυκών κουνουπιών που μόλις τράφηκαν με αίμα, τα οποία είναι και ο κύριος φορέας μετάδοσης του *Plasmodium falciparum* που προκαλεί την ασθένεια. Λόγω της υψηλής έκφρασής τους στις ωθήκες, ειδικά μετά από “γεύμα-αίματος”, εικάζεται ότι αυτές οι πρωτεΐνες έχουν έναν κρίσιμο ρόλο στην ωογένεση. Αν αυτό ισχύει, θα μπορούσαν να αποτελέσουν σημαντικούς υποψήφιους παράγοντες για τη δημιουργία διαγονιδιακών κουνουπιών, των οποίων οι απόγονοι θα είναι στείροι και οι οποίοι σταδιακά θα οδηγούσαν τον πληθυσμό των κουνουπιών σε κρίσιμα μικρούς αριθμούς. Ο στόχος είναι να αποκαλυφθεί, μέσω in-situ υβριδοποίησης, σε ποιο μέρος της ωθήκης εκφράζονται μέλη της ομάδας των πρωτεϊνών αυτών. Αποκαλύπτοντας τη θέση των μεταγράφων των πρωτεϊνών εντός του ιστού είναι σημαντικό, αφού έτσι έμμεσα αποκαλύπτεται και ο δυνητικός τους ρόλος στην αναπτυξιακή διαδικασία των ωαρίων. Οι πληροφορίες αυτές μπορεί να προσεγγιστούν και να αξιοποιηθούν από περαιτέρω μελέτες που έχουν σαν σκοπό τη δημιουργία διαγονιδιακών κουνουπιών και υπόσχονται καταστολή των πληθυσμού τους. Μια τέτοια μελέτη αξιοποιεί μια νέα μέθοδο για την επεξεργασία του γονιδιώματος των κουνουπιών, γνωστό ως το σύστημα CRISPR/Cas9, και σκοπό έχει την απαλοιφή υποψήφιων γονιδίων, τα οποία φαίνεται να έχουν την υψηλότερη επίδραση στη γονιμότητα του κουνουπιού.

## Table of Contents

|  |    |
|--|----|
| ABSTRACT .....   | 4  |
| ΠΕΡΙΛΗΨΗ.....  | 5  |
| CHAPTER 1 INTRODUCTION .....   | 9  |
| 1.1 Malaria: A Devastating Disease.....  | 10 |
| 1.1.1 Defining Malaria .....   | 10 |
| 1.1.2 An ancient disease - from past to present .....  | 11 |
| 1.1.3 Global impact .....  | 12 |
| 1.2 The Biology of Malaria.....  | 13 |
| 1.2.1 The <i>Plasmodium</i> genus.....   | 13 |
| 1.2.2 Malaria life cycle - an infrequent “dual host” system .....  | 15 |
| 1.3 Mosquito Vectors .....   | 16 |
| 1.3.1 The <i>Anophelinae</i> subfamily.....  | 16 |
| 1.3.2 Mosquito anatomy .....   | 18 |
| 1.3.3 Mosquito life cycle .....  | 19 |
| 1.4 Deep Sight in the Female Mosquito Reproductive System: Gonadogenesis, Oogenesis, Egg Development and Fertilisation ..... | 23 |
| 1.4.1 Structure of the mosquito ovaries, ovarioles and follicles .....   | 23 |
| 1.4.2 Mosquito gonadogenesis and germ line determination .....   | 25 |
| 1.4.3 Mosquito oogenesis .....   | 25 |
| 1.4.4 Mosquito egg development.....  | 27 |
| 1.4.5 Mosquito egg fertilisation.....  | 28 |
| 1.5 The Homing Endonuclease System .....   | 28 |
| 1.5.1 HEGs: selfish genetic elements .....   | 28 |
| 1.5.2 HEGs and mosquito control .....  | 30 |
| 1.5.3 Targeting of female germline- and soma- specific genes .....   | 31 |
| 1.6 Insect Olfactory System .....  | 32 |
| 1.6.1 An olfactory system is essential .....   | 32 |
| 1.6.2 Odorant binding proteins .....   | 33 |
| CHAPTER 2 AIM OF THE PROJECT.....  | 37 |
| CHAPTER 3 MATERIALS AND METHODS .....  | 39 |
| 3.1 Mosquito Strain Maintenance And Tissue Dissection.....   | 40 |
| 3.1.1 Mosquito strains and husbandry.....  | 40 |
| 3.1.2 Ovary dissection .....   | 40 |

|   |    |
|---|----|
| 3.2 Nucleic Acid Manipulation.....  | 40 |
| 3.2.1 RNA extraction .....  | 40 |
| 3.2.2 DNase treatment .....   | 41 |
| 3.2.3 Phenol/Chloroform extraction .....  | 41 |
| 3.2.4 First strand cDNA synthesis.....  | 42 |
| 3.2.5 PCR.....  | 42 |
| 3.2.6 PCR purification.....   | 43 |
| 3.2.7 DNA quantification .....  | 43 |
| 3.2.8 <i>In vitro</i> transcription.....  | 44 |
| 3.2.9 LiCl precipitation.....   | 45 |
| 3.2.10 Gel electrophoresis.....   | 45 |
| 3.3 Tissue Manipulation .....   | 46 |
| 3.3.1 Fixation .....  | 46 |
| 3.3.2 Proteinase K treatment.....   | 47 |
| 3.4 In-situ Hybridization .....   | 47 |
| 3.4.1 Hybridization .....   | 47 |
| 3.4.2 RNase A treatment .....   | 48 |
| 3.4.3 Antibody incubation .....   | 48 |
| 3.4.4 Antibody treatment.....   | 49 |
| 3.4.5 Glycerol mounting.....  | 50 |
| 3.5 Bioinformatics.....   | 51 |
| 3.5.1 Primer designing for PCR .....  | 51 |
| 3.5.2 Phylogenetic Analysis of the <i>Anopheles gambiae</i> OBPs.....                     | 52 |
| 3.5.3 <i>In silico</i> analysis of OBP genes expression in <i>Anopheles gambiae</i> ..... | 53 |
| 3.5.4 Candidate gene selection.....   | 54 |
| APPENDIX A: Buffers & Solutions.....  | 56 |
| APPENDIX B: Primer Sets used for PCRs.....  | 57 |
| CHAPTER 4 RESULTS .....   | 59 |
| 4.1 Project Strategy.....   | 60 |
| 4.2 In-Silico Analysis of the OBP Family .....  | 61 |
| 4.2.1 Listing and grouping of the OBPs .....  | 61 |
| 4.2.2 Clustering the OBPs .....   | 63 |
| 4.2.3 Selection of <i>Anopheles gambiae</i> candidate genes.....                          | 64 |
| 4.2.4 Genome clusters.....  | 67 |

|   |    |
|---|----|
| 4.3 In-Situ Hybridisation Experiment .....  | 69 |
| 4.3.1 PCR amplification of DNA template.....                                      | 69 |
| 4.3.2 In vitro transcription .....  | 71 |
| 4.3.3 RNA probe validation - positive & negative controls .....                   | 71 |
| 4.3.4 In situ hybridisation .....   | 74 |
| APPENDIX C: List of <i>Anopheles gambiae</i> OBP Genes.....                       | 79 |
| APPENDIX D: Expression charts for the candidate OBPs.....                         | 81 |
| CHAPTER 5 DISCUSSION.....   | 85 |
| 5.1 OBPs may be functionally analogous to the <i>Drosophila</i> “gap genes” ..... | 87 |
| 5.2 OBPs may be under the regulation of “maternal effect” genes .....             | 88 |
| 5.3 OBPs may be components of the eggshell .....                                  | 88 |
| 5.4 OBPs may be chemo-attractants for sperm .....                                 | 89 |
| ACKNOWLEDGMENTS .....   | 91 |
| ABBREVIATIONS .....   | 92 |
| REFERENCES .....  | 94 |



# **CHAPTER 1**

## **INTRODUCTION**

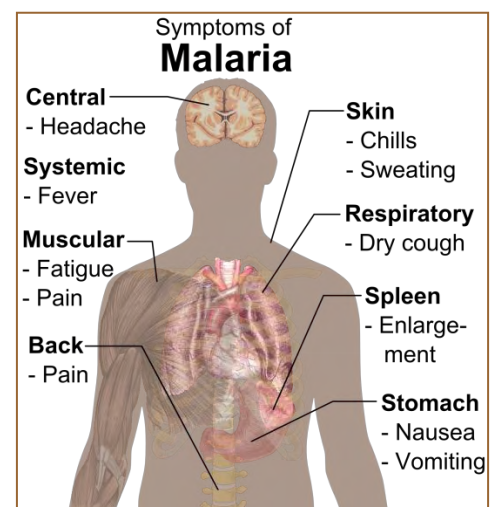
## CHAPTER 1: INTRODUCTION

### 1.1 Malaria: A Devastating Disease

#### 1.1.1 Defining Malaria

“Mal-aria”, which means “bad air” in Latin (Sallares, 2002), is a mosquito-borne serious and sometimes fatal disease, caused by a parasitic protozoon of the Plasmodium type. The protozoon infects certain types of mosquitoes (commonly of the *Anophelinae* subfamily) and thus the disease is transmitted with the bite of an infected female mosquito since they feed on vertebrates. The bite introduces the parasites from the mosquito's saliva into the person's blood, which then they travel through the blood to reach the liver, where they mature and reproduce. From there the parasites invade the red blood cells which they destroy in order to be released, since within the red blood cells they grow and multiply. These series of destruction of the liver cells and the red blood cells are the factors that cause the pathophysiology of the disease (Bledsoe, 2005; Vaughan, 2008).

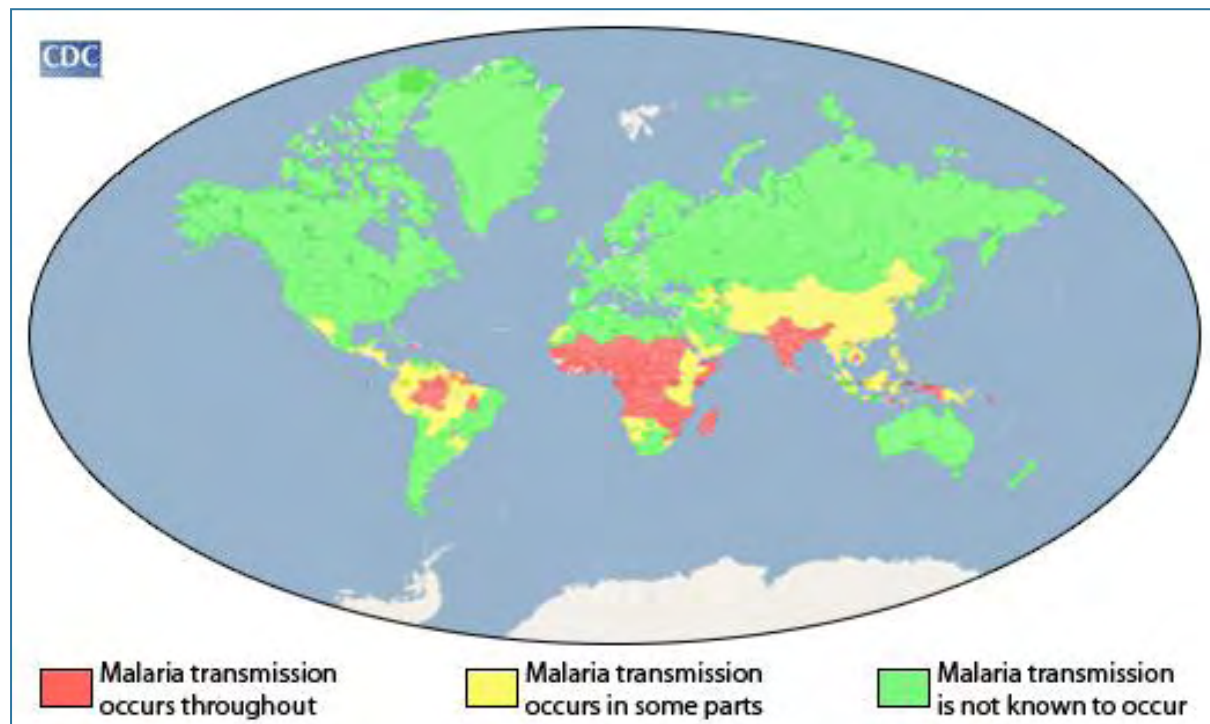
The classic symptom of malaria is “paroxysm” which describes a cyclical occurrence of sudden coldness followed by shivering, fever and sweating. People who get malaria are typically very sick with high fevers, shaking chills, and flu-like illness. Other symptoms include fatigue, vomiting and headaches. In its severe cases it can cause yellow skin, seizures, coma and finally, death. The symptoms of the disease (Figure 1.1) appear usually ten to fifteen days after the infectious bite. Malaria is typically diagnosed either by microscopic examination of the blood, using specific blood films, or by molecular methods such as antigen-based rapid diagnostic tests (RDTs), and even polymerase chain reactions (Caraballo, 2014). The latter methods are not widely used in areas where malaria is common due to their cost and complexity. It is interesting to state that despite the fact that malaria can be a deadly disease, illness and death from malaria can usually be prevented!



**Figure 1.1.** The typical symptoms of malaria. The symptoms usually appear ten to fifteen days after the infection.

Malaria is a disease which is common in the areas where the mosquito-vectors thrive. It is widespread in tropical and subtropical regions which are present in a broad band around the equator. This includes much of Sub-Saharan Africa, Asia and Latin America. Here, the wet climate,

the high humidity and the heat help the population of mosquitoes to increase and established (CDC, 2014) (Figure 1.2).



**Figure 1.2.** The frequency of malaria transmission on the globe. Note the equatorial spread of the disease (CDC, Centers for Disease Control and Prevention, 2014).

### 1.1.2 An ancient disease - from past to present

Humanity has suffered from malaria for over four thousand of years, and the disease had back then, and continues to have, strong influence on the course of humanity. The first reference of what seemed to be an accurate description of the disease dates back to 2.700 BC in the Province of China. Nei Ching, a medical doctor, had recorded malaria symptoms in his medical writings “The Canon of Medicine”. Two thousand years later Hippocrates, the “Father of Medicine” was able to show the connection between stagnant bodies of water and incidents of fevers in nearby local populations within Egypt, outlining thus the main symptoms of the disease. Hippocrates’s findings were backed up by the Romans which they had observed similar patterns of malaria symptoms near marshes, ponds and swamps. The spread of the newly called “roman fever” was under restriction for the Romans were trying to drain these stagnant bodies of water. (CDC, 2014)

It wasn't until 1880 that the factor which is responsible for the disease was discovered. The discovery was accomplished by a French physician, called Charles L.A. Laveran, who had the idea to observe under a microscope the blood of a patient who had passed away from the disease in Algeria (Bruce-Chwatt, 1987). Laveran though, which was awarded with the Nobel Prize for Physiology or Medicine in 1907 for this discovery, did not determine how these protozoa had been inserted inside

the body of the patient. This discovery was made seventeen years later in 1897 in India, when Ronald Ross a British military doctor, demonstrated the mode of the transmission of malaria. Ross detected the malaria parasite inside the gastrointestinal tract of mosquitoes, and this led him to hypothesize that the parasite transmission between its vertebrate hosts is achieved by mosquitoes. For this discovery Ross received the Nobel Prize for Physiology or Medicine in 1902. Finally, following Ross, one year later some other Italian scientists, G.B Grassi, A. Bignami and G. Bastianelli, will be the first to document the complete transmission cycle of the parasite inside the mosquitoes (CDC, 2014).



**Figure 1.3.** (Left) Illustration drawn by Laveran of various stages of malaria parasites as seen on fresh blood. Dark pigment granules are present in most stages. The bottom row shows an exflagellating male gametocyte, which produces "filiform elements which move with great vivacity." (Right) Page from notebook where Sir Ronald Ross records his discovery of the mosquito transmission of malaria, 20 August 1897 (Courtesy: London School of Hygiene and Tropical Medicine).

### 1.1.3 Global impact

With modern medicine malaria is preventable and treatable; even though, it still remains a worldwide deadly disease due to environmental disturbance, malnutrition and poverty. In many of the countries affected by malaria it is a leading cause of illness and death. The two most vulnerable groups, in areas with high transmission, are young children and pregnant women. This is because young children have not developed immunity to malaria yet, and the pregnant women because their immunity has been decreased by pregnancy.

In 2013, 97 countries had ongoing malaria transmission. In 2012 there were around 207 million cases of malaria, and an estimated 627.000 people died of the disease. The 482.000 of that number were children under five years of age. That is 1.300 children every day or one child almost every minute! An estimated 3.4 billion people are at risk of malaria, of which 1.2 billion are at high risk. In high-risk areas, more than one malaria case occurs per 1.000 of population! Within the last decade though, increasing numbers of partners and resources have rapidly increased malaria control efforts. This scale-up of interventions has saved 3.3 million lives globally and cut malaria mortality by 45%, leading to hopes and plans for elimination and ultimately eradication. (WHO, 2013)

Malaria is commonly associated with poverty and has a major negative effect on economic development, especially in poor tropical and subtropical areas of the world (Worrall, 2005). It has been judged as a disease of poverty, being a cause and a consequence of poverty at the same time (Gollin & Zimmermann, 2007). The costs of malaria to individuals, families, communities and nations are enormous; in Africa it is estimated to result in losses of \$12 billion USD a year due to increased healthcare costs, lost ability to work and effects on tourism (WHO, 2013).

## 1.2 The Biology of Malaria

### 1.2.1 The *Plasmodium* genus

Malaria is caused by five species of parasite that affect humans, and all of these species belong to the genus *Plasmodium*: *P.falciparum*, *P.vivax*, *P.ovale*, *P.malariae* and *P.knowlesi*. Of these five species, *P.falciparum* and *P.vivax* are the most important since they cause the most human deaths. *P.falciparum* is dominated in central Africa and *P.vivax*, although it's less dangerous it's more widespread, and that makes it as important. *P.ovale* and *P.malariae* generally cause a milder form of malaria in contrast to the species *P.knowlesi*, which rarely causes the disease in humans (WHO, 2013).

As mentioned above, *Plasmodium* is a protist; it is a single cellular eukaryote which displays unusual structural characteristics, in contrast with general eukaryotes (Figure 1.5). Because of its pathogenesis it was amongst the first organisms that the genome was sequenced. The sequencing revealed that the average genome size of the *Plasmodium* genus is approximately 25 megabases and that is organised into 14 chromosomes. These chromosomes' length varies from 500 kilobases to 3.5 megabases (Hall, 2005).

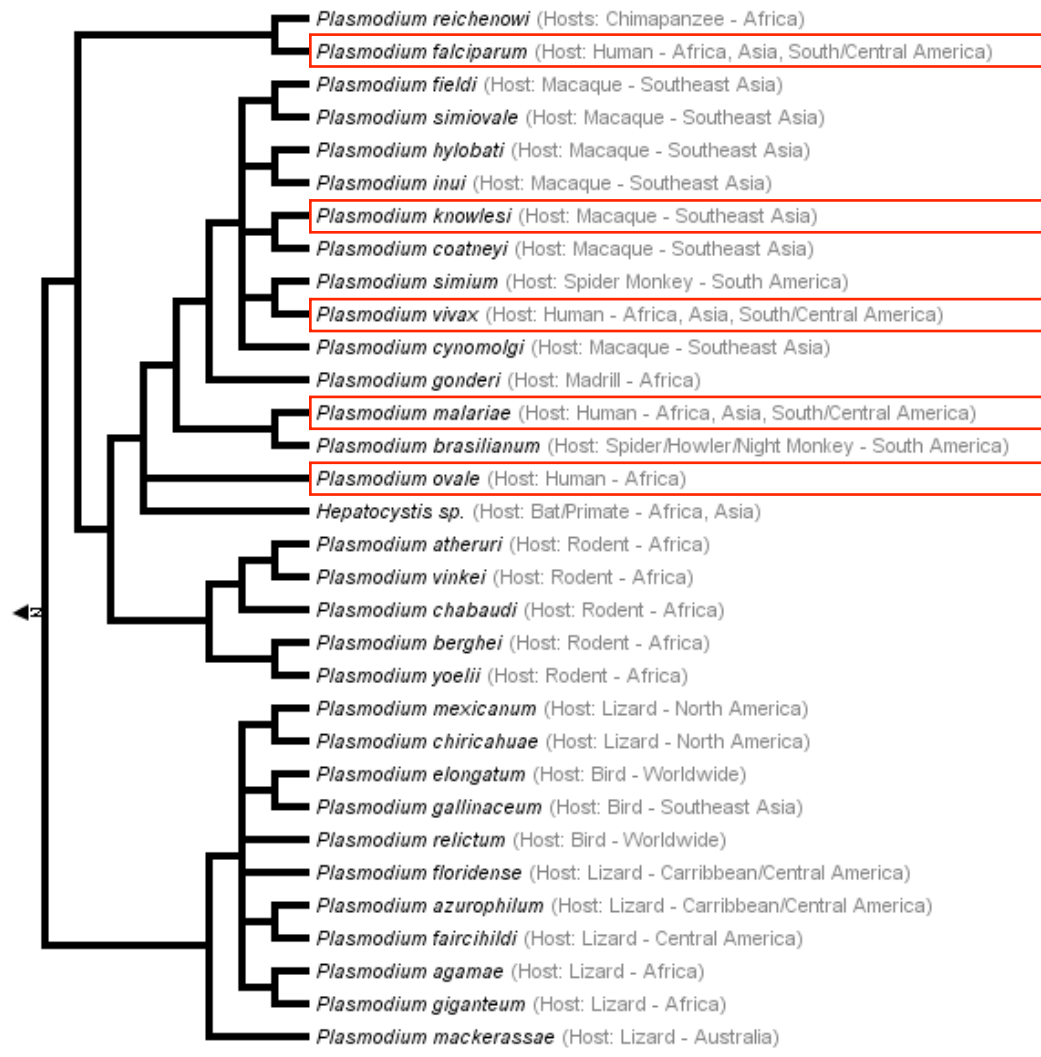


Figure 1.4. Phylogeny of the *Plasmodium* genus. The five species of parasite that affect humans are annotated (Zilversmit & Susan, 2008).

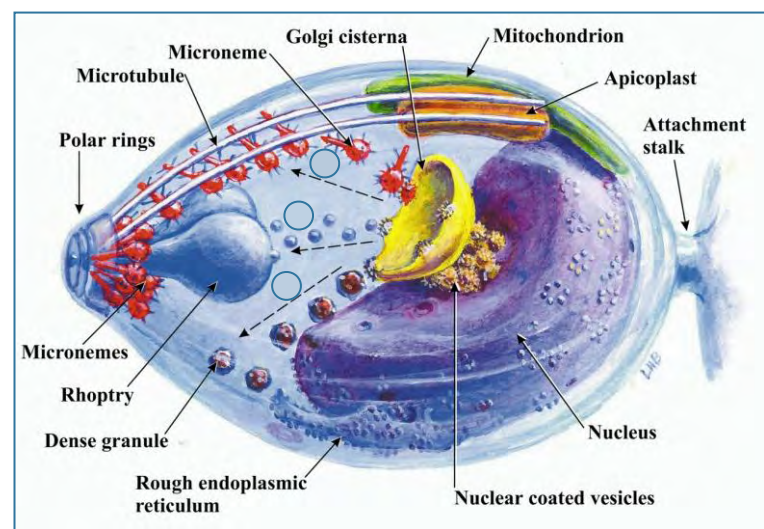
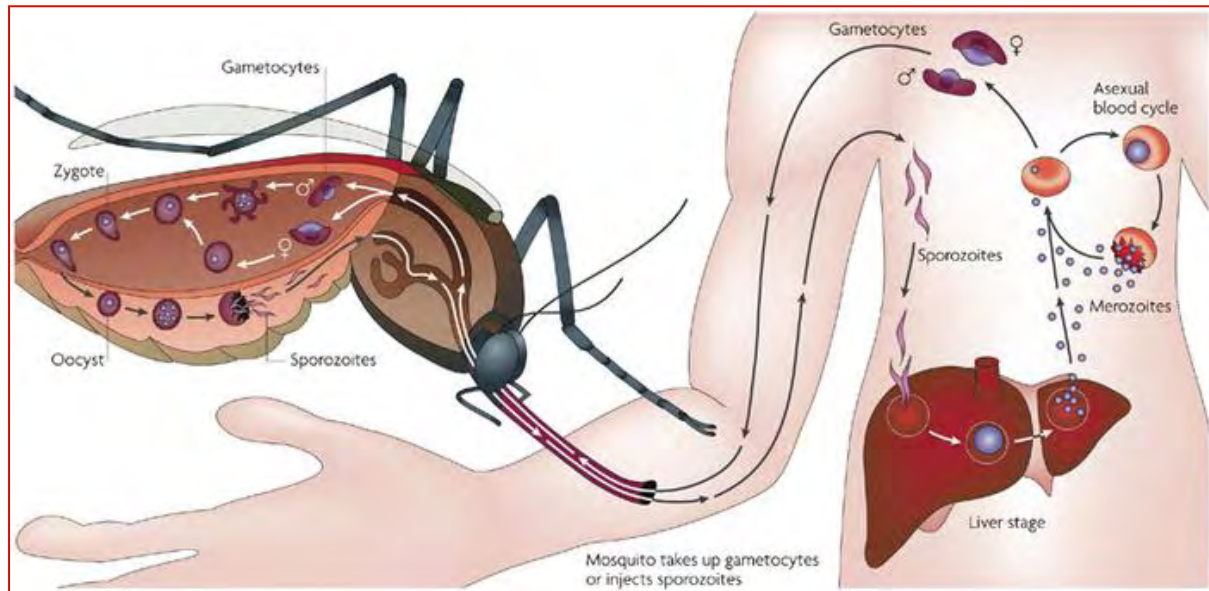


Figure 1.5. Representation of *Plasmodium* significant cellular components. Unusual structures of this organism in comparison to general eukaryotes include the apicoplasts, micronemes, rhoptry, and polar rings near the apical end (Bannister *et al.*, 2003).

### 1.2.2 Malaria life cycle - an infrequent "dual host" system

The natural ecology of malaria involves malaria parasites infecting successively two types of hosts: the human-host and the female *Anopheles* mosquito-vector. This dual host system is very rare in nature and that is why it draws the attention and interest of the scientific community. A synoptic schematic *Plasmodium* life cycle is displayed in Figure 1.6.



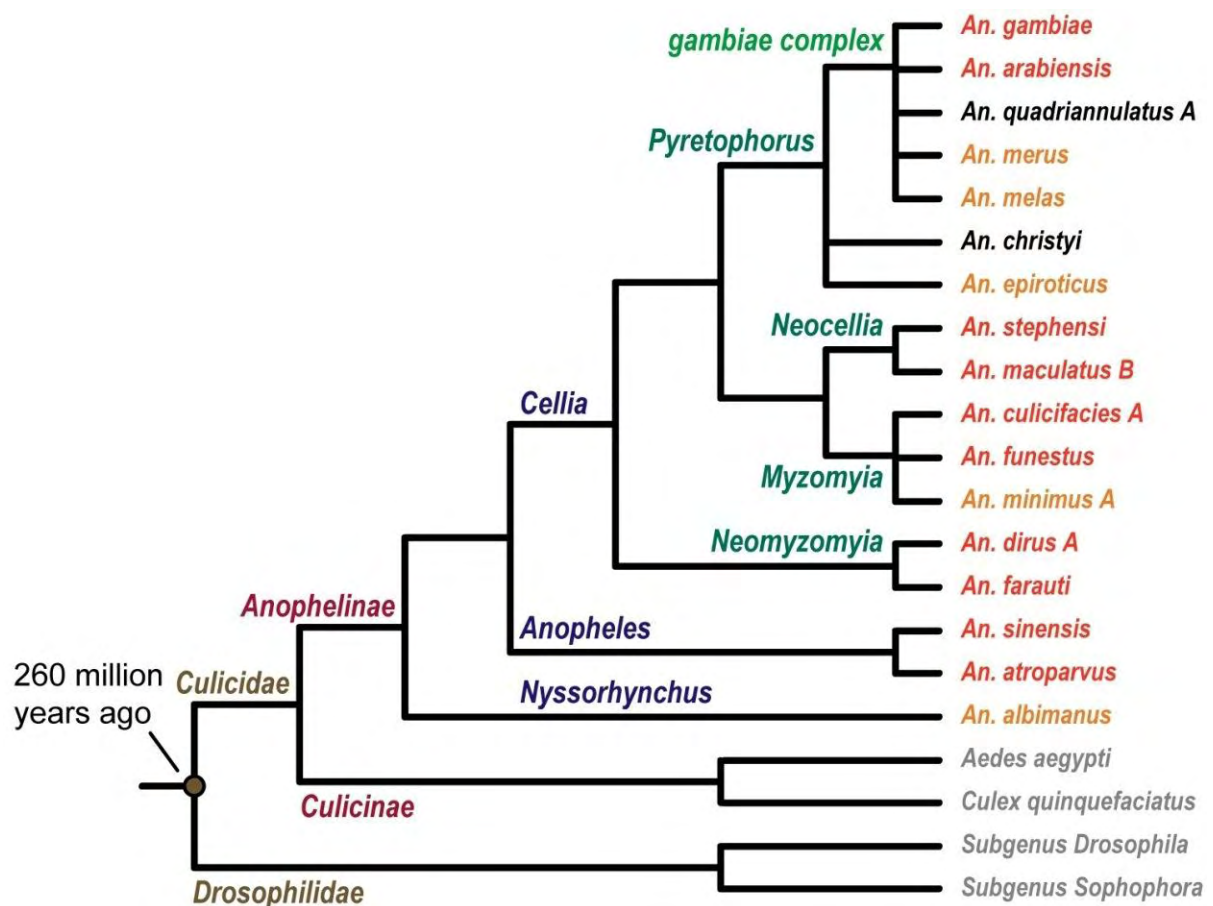
**Figure 1.6.** The life cycle of malaria parasites. A mosquito causes an infection by a bite: First, sporozoites enter the bloodstream and migrate to the liver. They infect liver cells, where they multiply into merozoites, rupture the liver cells and return to the bloodstream. Then, the merozoites infect red blood cells, where they develop into ring forms, trophozoites and schizonts that in turn produce further merozoites. Finally, sexual forms are also produced, which, if taken up by a mosquito, will infect the insect and continue the life cycle (Su *et al.*, 2007).

Briefly, the parasites are found (as "sporozoites") in the mosquito's salivary glands. When the *Anopheles* mosquito takes a blood meal on a human, the sporozoites are injected with the mosquito's saliva into the blood stream. The released sporozoites parasitize the liver, and thus how the infection begins. Inside the human body the parasites grow and multiply; first, the parasites rupture and destroy the liver cells and then the parasites migrate to the blood. In the blood, successive broods of parasites grow inside the red blood cells and destroy them, releasing daughter parasites ("merozoites"). The merozoite, which causes the symptoms of malaria, is the parasite form which carries on the red blood cells destruction by invading other red blood cells. Through this process sexual forms of the parasite ("gametocytes") are also produced. Finally, when gametocytes are picked up by a female *Anopheles* mosquito during a blood meal they start another, different cycle of growth and multiplication, this time inside the mosquito. Thus, the mosquito carries the disease from one human to another (acting as a "vector"). It is surprising and kind of "unfair" that the mosquito vector does not suffer from the presence of the parasites as the human host! (Warhurst, 2008; Cowman, 2012).

## 1.3 Mosquito Vectors

### 1.3.1 The *Anophelinae* subfamily

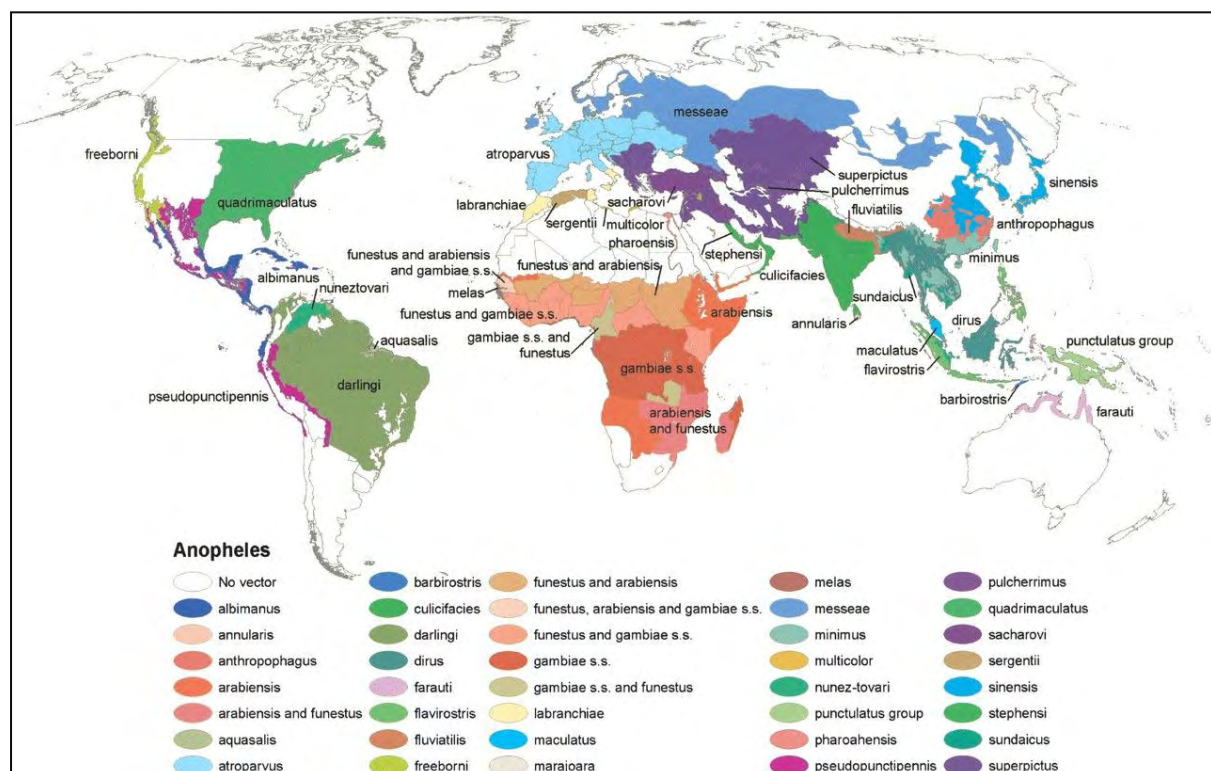
In nature there are approximately 3.500 species of mosquitoes which grouped into 41 genera (CDC, 2014). Human malaria is transmitted only by female mosquitoes of the genus *Anopheles*, which means “without benefit” in Greek. The *Anopheles* genus belongs to a wider family of mosquitoes called *Culicidae* which is divided into two subfamilies: the *Anophelinae* and the *Culicinae*. The former subfamily covers the *Anopheles* divisions to lower classifications, and the latter subfamily covers two different genera of mosquitoes: the *Culex* and the *Aedes*. Experiments made by Battista Grassi G. and his team in 1898, showed that members of the *Culicinae* subfamily are not capable of transmitting the disease to humans (Cox, 2010). Nowadays, it is well known that malaria is spread only by members of the *Anophelinae* subfamily. Specifically, of the approximately 430 *Anopheles* species that had been discovered so far, only 30 to 40 are capable to transmit malaria in nature, i.e., are “vectors” (CDC, 2014; Neafsey *et al.*, 2013).



**Figure 1.7.** Dendrogrammatic illustration of the 16 *Anophelinae* species and their putative evolutionary relationships. The two sequenced *Culicinae*: *Aedes aegypti* and *Culex quinquefasciatus*, and the sequenced *Drosophila* species are also demonstrated. *Anopheles* species that comprise the major human malaria vectors are labeled in red; minor vectors are labeled in orange, and species that are not human malaria vectors are labeled in black (Neafsey *et al.*, 2013).



The *Anophelinae* are found worldwide except of the subcontinent of Antarctica (Figure 1.8). Malaria is transmitted by different *Anopheles* species, depending on the region and the environment. Members of the genus that can transmit the disease are found not only in malaria-endemic areas, but also in areas where malaria has been eliminated; the latter areas are thus constantly at risk of re-introduction of the disease. Only female mosquitoes are capable of transmitting the disease, since they feed on blood in order to stimulate and complete the process of oogenesis; male mosquitoes on the other hand feed on plant nectar, and thus do not transmit the disease. The females of the *Anopheles* genus prefer to feed at night. They usually start searching for a meal at dusk, and continue throughout the night until taking a meal (Arrow *et al.*, 2004).



**Figure 1.8.** Global distribution of dominant or potentially important malaria vectors. Note the *A.gambiae s.s.* distribution in central Africa. (Kiszewski *et al.*, 2004).

The most important and thus most characterised species of the *Anophelinae* subfamily is the *Anopheles gambiae* complex. The complex consists of seven morphologically indistinguishable species: *A.arabiensis*, *A.melas*, *A.quadriannulatus*, *A.bwambae*, *A.merus* and *A.gambiae sensu stricto (s.s)* (Besansky *et al.*, 1994; Wilkins *et al.*, 2006). Members of this complex have a predominant role in the transmission of the most dangerous of the *Plasmodiums*, the *P.falciparum*. Also, the antropophilic nature (preference of human blood) of some of the species in this complex determines their efficiency as human malaria vectors, making *Anopheles gambiae s.s.* - a highly antropophilic species - the dominant malaria vector in Africa.

It has been suggested that *Anopheles gambiae* s.s. is currently in a state of diverging into two different species: the Mopti (M) and Savannah (S) strains; since 2007 though, the two strains are still considered to be a single species (Lawniczak *et al.*, 2010). Recently a new “cryptic” subgroup - the Goundry subgroup - of *Anopheles gambiae sensu stricto* has been described (Yakob, 2011).

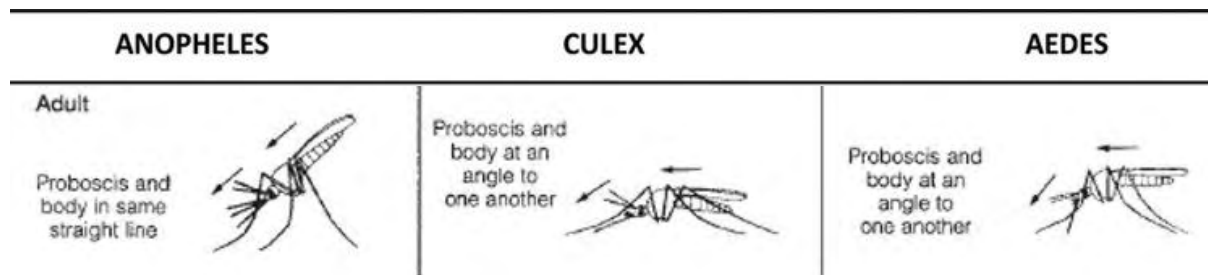
### 1.3.2 Mosquito anatomy

Adult *Anopheles*, like all mosquitoes, have sylphlike bodies consist of 3 sections: a head, a thorax and an abdomen. The head as expected contains the eyes, a pair of long, many-segmented antennae, two sensory palps and an elongate forward-projecting proboscis. The eyes, the antennae and the papls, like in all insects, are specialized for acquiring sensory information and for detecting host odors, as well as odors for breeding and oviposition sites. The head also has the purpose to feed the insect; the proboscis is a complex structure that contains six parts: two pairs of sharp and flexible cutters, and a pair of fine tubes. The cutters surround the tubes which are used for sucking up blood and for dripping a chemical into the wound at the same time, that keeps the blood flowing. The head also is used to distinguish between male and female individuals. For example, the males have bigger and bushier antennae in contrast to females, which possess smaller and slenderer antennae (Figure 1.9). Also, the different lengths of the palps can be used to distinguish between different types of mosquitoes, since many different species carry different lengths of palps. For example, *Anopheles* adult mosquitoes carry long pulps which are as long as their proboscis, in contrast to *Culex* and *Aedes* which carry smaller palps in relevance to their antennae.



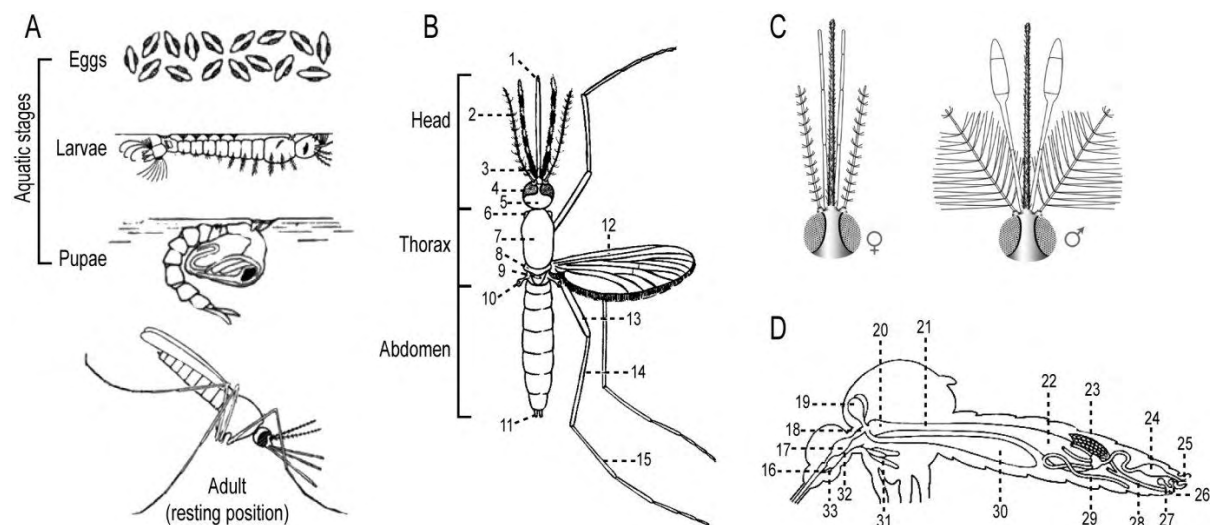
**Figure 1.9.** Difference between a male (left) and a female (right) *Anopheles gambiae* mosquito heads. Note the bushy antennae of the male mosquito.

The thorax is specialized for locomotion and flight; three pairs of legs and one pair of wings are attached to different segments of the thorax. The wings of the different species of mosquitoes have different patterns of discrete scales of black and white, which are also used to distinguish between the different species. It is interesting to state that different species of mosquitoes have different resting positions; for example *Anopheles* species rest with their abdomens sticking up in the air, rather than parallel to the surface - a resting preference of *Culicinae* (Figure 1.10).



**Figure 1.10.** Difference between *Anophelinae* (*Anopheles*) and *Culicinae* (*Culex* and *Aedes*) mosquito resting positions. Anopheles are known for their straight bodies in relevance to their antennae.

Finally, the abdomen is a segmented body part that is specialized in the completion of the key procedures in an organism’s life: nutrient digestion and production of offspring. In the abdomen there is the plethora of organs that compose the mosquito body; midgut, stomach, ovaries, rectum, spermatheca, oviduct, malpighian tubules and ventral diverticulum (Figure 1.11). When a female takes a blood meal, the abdomen expands considerably since it fills with the sucked blood. The blood components are digested over time, serving as a source of protein and nutrients for the production of eggs. At the end of the egg development the majority of the abdomen is occupied by the ovaries that consist of the mature eggs.

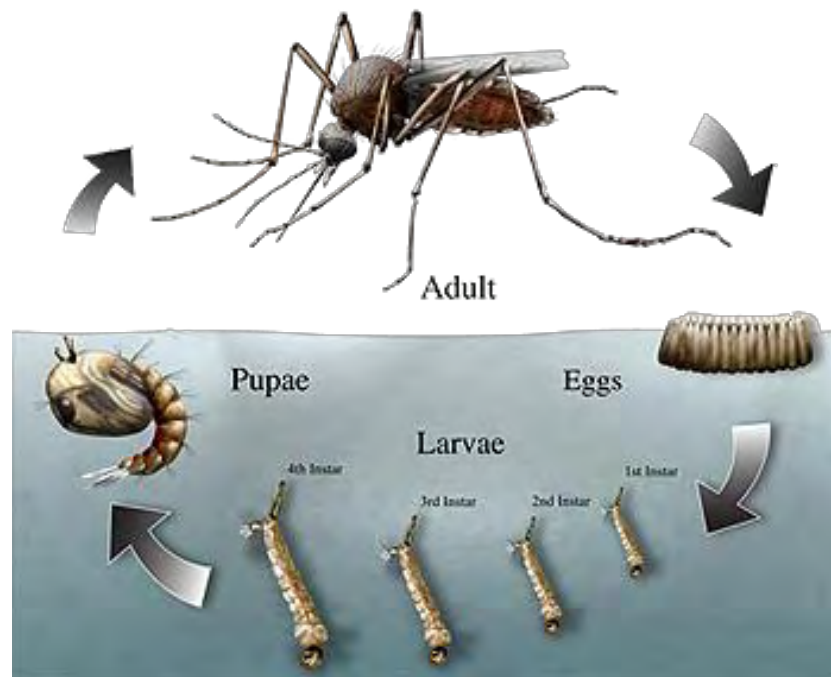


**Figure 1.11.** Schematic representation of *Anopheles* life cycle and anatomy. (A) The four developmental stages of *Anopheles* mosquitoes. (B) External anatomy of a female *Anopheles* mosquito: 1. proboscis; 2. antenna; 3. maxillary palpus; 4. eye; 5. occiput; 6. anterior pronotal lobe; 7. scutum; 8. scutellum; 9. postnotum; 10. halter; 11. cercus; 12. wing; 13. femur; 14. tibia; 15. tarsus. (C) Structure of a male and female *Anopheles* mosquito head (D) Internal anatomy of a female *Anopheles* mosquito: 16. pharynx; 17. pharyngeal pump; 18. oesophagus; 19. dorsal diverticula; 20. proventriculus; 21. midgut; 22. stomach; 23. ovary; 24. rectum; 25. cerci; 26. atrium; 27. spermatheca; 28. oviduct; 29. malpighian tubules; 30. ventral diverticulum; 31. salivary glands; 32. salivary duct; 33. salivary pump (Gilles *et al.*, 1993).

### 1.3.3 Mosquito life cycle

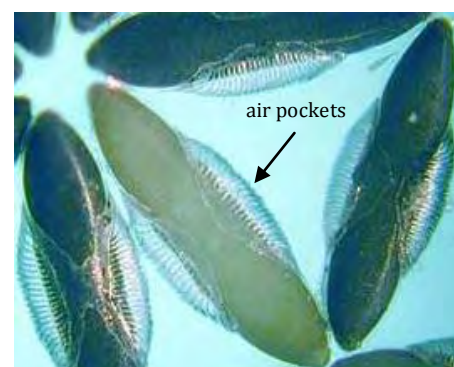
Mosquitoes as “holometabola” insects require four stages to reach adulthood starting from the phase of the egg. The stages compose the insects’ life cycle, and they are very easy to distinguish. The mosquito pass through the stage of the egg, to the stage of the larva, to the stage of the pupa,

to reach the final adult form which is also the mating form of the insect. (Figure 1.12). Because the *Anophelinae* have not lost the connection with water, like butterflies had, the phases of the egg, larva and pupa are aquatic phases and they require water in order to survive; the stages are also depended on the environment and the ambient temperature. Depending on the species, the whole process from oviposition to pupation last for about 5 to 14 days.



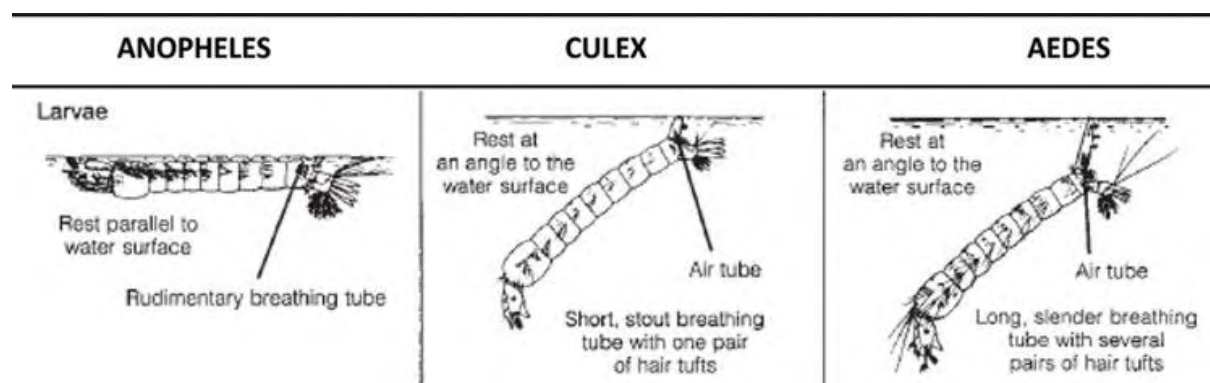
**Figure 1.12.** The life cycle of the mosquito. The mosquito pass through the stage of the egg, to the stage of the larva, to the stage of the pupa, to reach the final adult form which is also the mating form of the insect. Note that the larva stage consists of 4 phases (L1 to L4).

The egg of the mosquito is a compact oval structure which, as it mentioned above, is totally depended on water to survive. The egg is actually an embryo since during oviposition gets fertilised by male sperm, which is stored in the spermatheca of the female mosquito. It is wise to mention that the mating happens before the egg development, since male sperm contains molecules and steroid hormones that trigger and regulate female oogenesis in the female mosquito (Robinson, 2013; Baldini *et al.*, 2013). The adult female mosquitoes lay from 50 to 200 eggs per oviposition. The eggs are laid on the surface of stagnant body of waters, and in order to float they equipped with two air pockets on either side (Figure 1.13). Depending on the temperature, the eggs can hatch within 2 or 3 days, although hatching may take up to 2 to 3 weeks in colder climates.



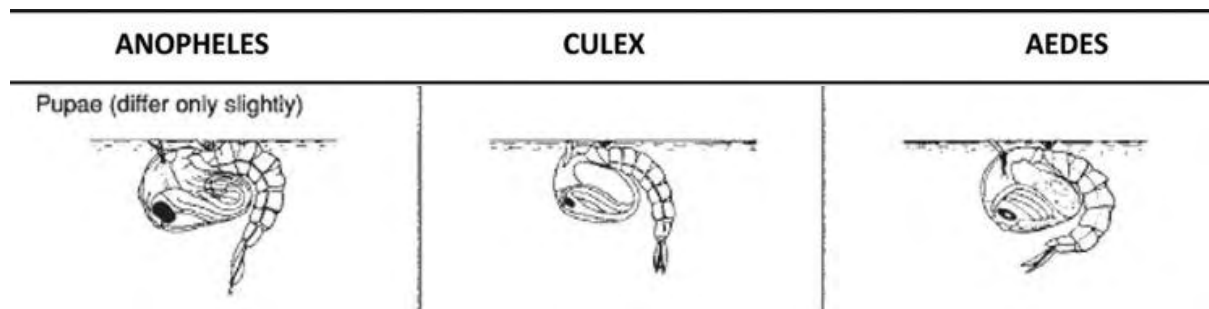
**Figure 1.13.** *Anopheles gambiae* mosquito eggs. The eggs carry two air pockets that enable them to float. Note the middle egg that has yet to melanise.

The larva of the mosquito passes through 4 “instars” (L1-L4) in order to grow in a preferable size and pupate. Each instar is followed by shedding of the skin, process necessary for the larvae to grow in size. The larva’s body consists of: a well-developed head which possess mouth brushes used for feeding, a distinguishable large precursor of thorax, and a segmented abdomen. The larvae lack the legs and the wings, structures that are developed later, from imaginal disks, during the phase of the pupa. The majority of the mosquitoes larvae species possess spiracles located on the 8th abdominal segment and through them are able to breathe by reaching the water’s surface in an angle. *Anopheles* larvae however, lack a respiratory siphon and for this reason they position themselves parallel to the surface of the water in order to breathe through the body (Figure 1.14). The larvae are able to swim by whole body movements in order reach food sources such as algae and small floating particles.



**Figure 1.14.** Differences between *Anophelinae* (*Anopheles*) and *Culicinae* (*Culex* and *Aedes*) mosquito larvae anatomy and position. Note the parallel-to-the-surface position of *Anopheles* larvae.

The larva metamorphoses into pupa within the period of one week. The head and the thorax are merged into a unified structure called “cephalothorax” and the abdomen curves around underneath it. The process gives the pupa a comma-shaped structure which is visible when viewed from the side (Figure 1.15). *Anopheles* pupae are highly active despite the fact that they do not feed; something uncommon amongst other species. Pupae though, have to breathe and thus they possess two trumpet-like structures on the top of the cephalothorax from which they aspirate oxygen. Pupa stage last for at least two days. When they are ready, they melanise; means they acquire a deep black colour, and the adult mosquito emerges by breaking the dorsal surface of the cephalothorax. It is interesting to mention that mosquitoes can be sexed in the stage of pupa. Male pupae acquire a tiny structure at the end of their abdomen which is absent on female pupae (Figure 1.16). Under a microscope the structure is visible and thus the sexing is capable. Sexing pupae is very important in experiments that need virgin females and males.



**Figure 1.15.** Difference between *Anophelinae* (*Anopheles*) and *Culicinae* (*Culex* and *Aedes*) mosquito pupae anatomy and resting positions.



**Figure 1.16.** Difference between a male pupa (right) and a female pupa (left). The arrows point at the tiny structure that there is on the male but it is absent on the female pupa.

Finally the adult stage is when the female *Anopheles* mosquito acts as a malaria vector. The females live for up to one month in contrast to the males which live for about a week. Male and female mosquitoes feed on nectar and other sources of sugar for energy. Adult mosquitoes usually mate within a few days after emerging from the pupal stage. The mosquitoes join their abdomens together (Figure 1.17) and sperm passes from male to female where it is stored inside the spermatheca of the female. Females mate only once since the sperm stored is enough for all the gonotrophic cycles (oogenesis, egg development and oviposition) of the female. Females, as mentioned before, require a full blood meal for every gonotrophic cycle. Mated females obtained full blood meal, will rest for a few days while the blood is digested and the eggs are developed. Between 48 and 72 hours PBM (in tropical conditions) the eggs are fully developed and thus the female lays them and resumes host seeking to start another gonotrophic cycle.

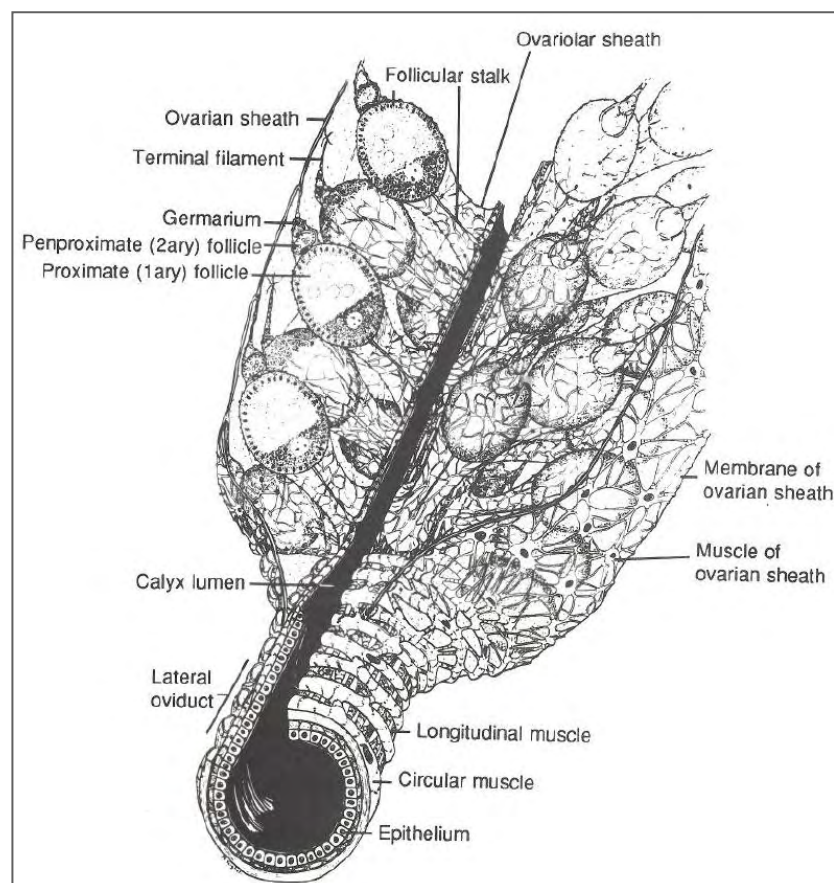


**Figure 1.17.** The mating dance of the mosquitoes. Above, a male (left) and a female (right) join their abdomens to achieve sperm diffusion. The act lasts for about 3 seconds!

## 1.4 Deep Sight in the Female Mosquito Reproductive System: Gonadogenesis, Oogenesis, Egg Development and Fertilisation

### 1.4.1 Structure of the mosquito ovaries, ovarioles and follicles

Female mosquitoes carry one pair of ovaries at the posterior end of their abdomen (Figure 1.11). Each ovary is covered with a fragile veil, the membrane of the “ovarian sheath”, which encloses around 80-100 eggs (ovarioles). Interior, there is one tube-like structure, the “calyx lumen”, which is connected to each ovariole with smaller tubes called the “follicular stalks” (Figure 1.18). The calyx lumen results to the “lateral oviduct” which connects the ovaries with the “common oviduct”, and from there to the vagina and the atrium. The whole of the oviduct structure is covered in circular muscles which are contracted to achieve promotion of the eggs to the common oviduct, and final oviposition.

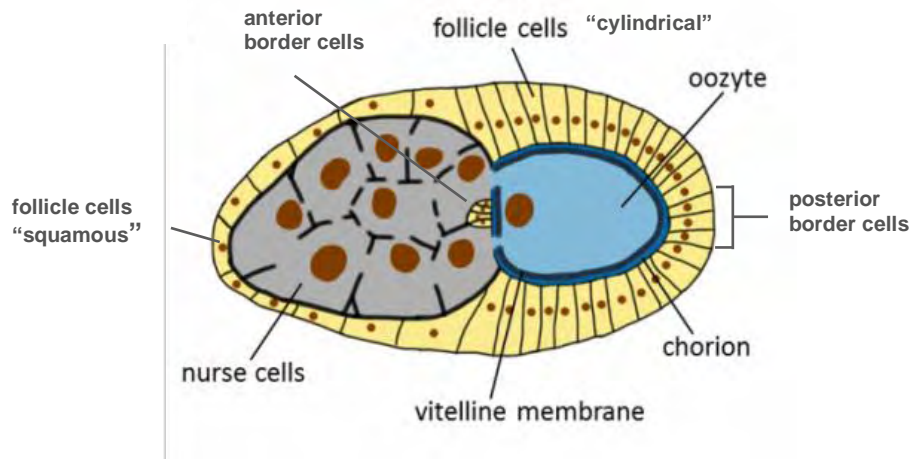


**Figure 1.18.** The ovary of the mosquito adjacent to the lateral oviduct (Clements, 1999). The ovaries consist of hundreds of ovarioles in which the production line of eggs take place. After the eggs maturation they are promoted into the calyx lumen which results to the lateral oviduct. From there, the eggs are migrating inside the common oviduct from which they deployed to the environment through the atrium.

The eggs of the mosquito are produced within the ovarioles, in which precursor germ cells and somatic cells are divided to give the cells which compose the final egg. The ovarioles provide all of the necessary structural, genetic and nutritional requirements for egg development within the ovary and thus they are composed of many and different cell compartment, each one with a distinct role in the process (Figure 1.21). On the anterior of the structure, on the outer region of the ovary - thus in contact with the ovarian sheath, there is a compartment called the "germarium". The germarium compartment is divide to three regions called the "germarial regions" (1, 2a and 2b), in which the cell division of the stem cells takes place. Looking inwards - through the calyx lumen, there is a number of developmentally distinct "follicles" composing different egg chambers which are in different developmental stages. The follicles are named depending on the egg chamber they form, therefore on their maturity, from the most mature one: "proximate/primary" follicle, to the least mature one: "penproximate/secondary" follicle. During oogenesis sometimes a "tertiary" follicle which compose a third egg chamber, is formed near the germarium. The primary follicle is in contact with the follicular stalk, from which is going to pass inside the calyx lumen to be deployed. The ovariole structure is covered by its own veil membrane called the "ovariolar sheath".

With the term "mosquito egg" we usually define the proximate/primary follicle, since it is the first of the follicles which is going to grow and finally, after fertilization, deployed. The egg consists of cells that derive from the germ line, as well as cells come from the somatic line! Specifically, the germ-line cells give birth to the "nurse cells" and the oocyte; the somatic cells that are part of the gonads, give birth to the follicle cells that surround the egg structure. While the egg is growing, the follicle cells are divided into three distinct groups of cells which have different roles in the egg development. The follicle cells that are in touch with the nurse cells are shaping into a "squamous" form in contrast to the cells that cover the oocyte which acquire a more cylindrical shape. The third group is called the "border cells" and they are located in both sides of the oocyte, defining thus the anteroposterior axis, and the right orientation of the embryo. (Figure 1.19). It is interesting, for the aim of the project, to state that the follicle cells are responsible for the formation of the "follicular membrane" that surrounds the oocyte. The membrane consists of two major and structurally distinct layers, the inner vitelline membrane (VM) and the outer chorion (CM) (Trogakos & Margaritis, 2002). The follicular membrane converts to the "eggshell" after oviposition.





**Figure 1.19.** Ovary specific tissues in an insect egg (Alberts *et al.*, 2007). The whole of the structure is called the egg chamber. Note the soatic follicle cells (yellow) which are divided into three groups. When the oocyte is formed, its nucleus migrates to the anterior side of the oocyte.

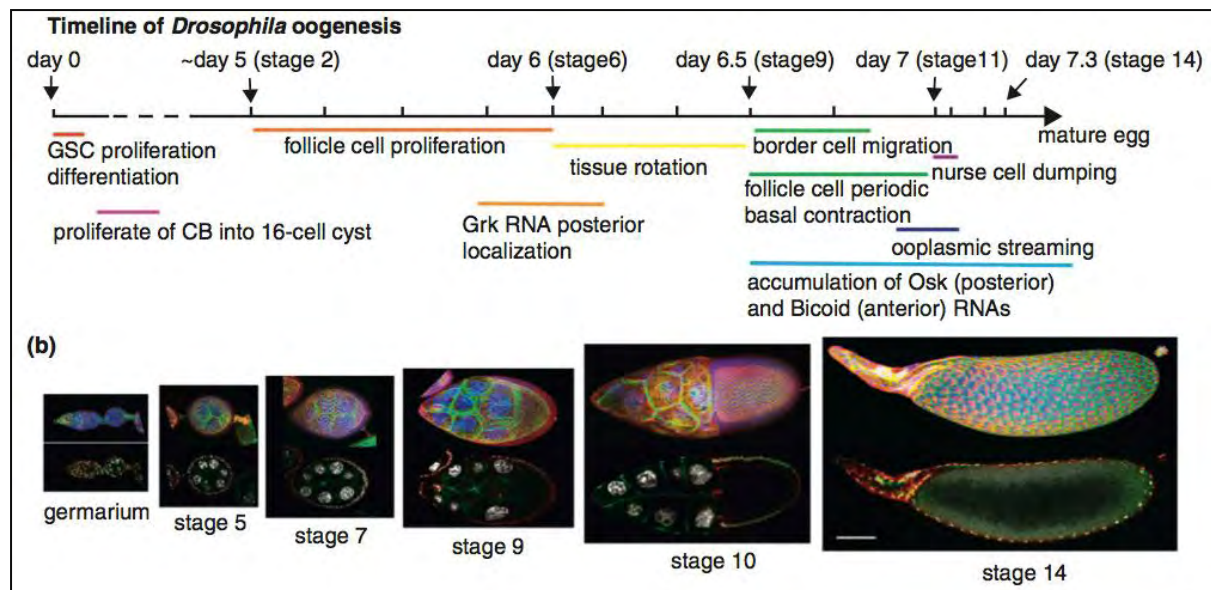
#### 1.4.2 Mosquito gonadogenesis and germ line determination

Pole cells, that create the germ line cells, are defined by a “determinant” early in the embryogenesis. The pole plasm, which consists of the determinant, is located on the posterior of the embryo (near the posterior border cells) and after the eight-first divisions of the embryonic nuclei, with “superficial cleavage”, the pole plasm is engulfed to create the pole cells.

#### 1.4.3 Mosquito oogenesis

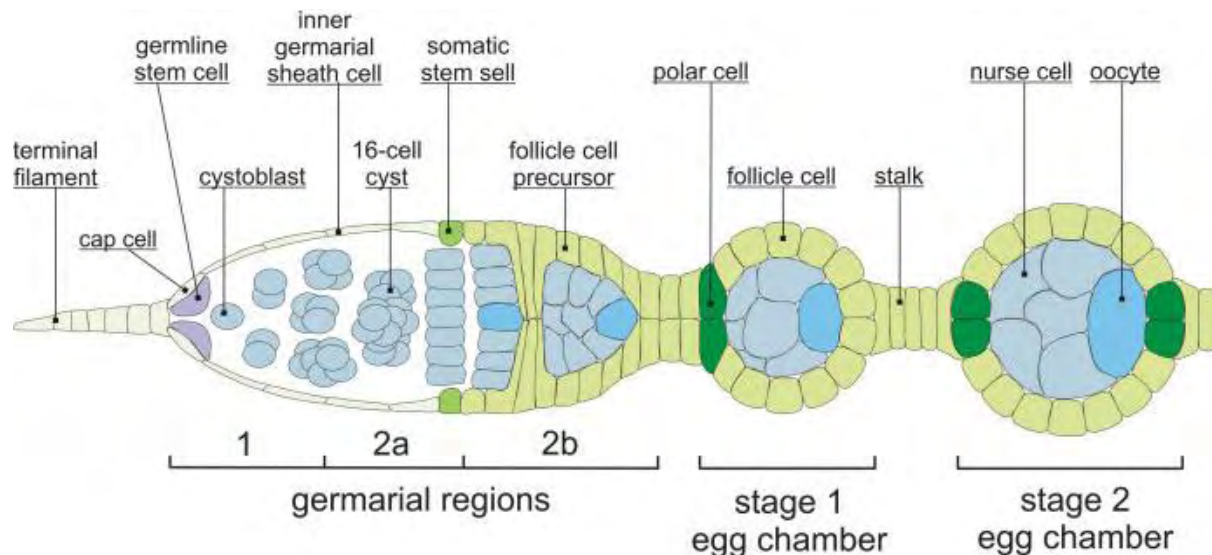
With the term oogenesis we describe the process of the gamete production of the female, in this case a mosquito, which belongs to a wider group of processes that involved in insect fertility. The current knowledge on insects’ fertility comes from studies of *Drosophila melanogaster* in which these events and processes have been studied in detail, since the fruit fly is one great model-organism. Although that *D.melanogaster* and *A.gambiae* have an evolutionary distance of approximately 274.9 million years (Hedges *et al.*, 2006; Kumar & Hedges, 2011) these processes have been kept almost the same and thus, since the publication of the *Anopheles* genome in 2002, scientist have been trying to identify orthologs amongst the two organisms.

Due to their different nature from fruit flies, mosquitoes have changed some of the genes to correspond to the special life cycle the insect has. Processes like odor detection for host seeking, biting to obtain blood and digest blood to nourish the developing eggs, are some of the functions that require different gene in order to be accomplished. For example, the final stages of the ovary maturation and the egg development require a blood meal to be completed. However, the key events in the process of oogenesis in both the organisms have been kept the same (Figure 1.20).



**Figure 1.20.** Timeline of *Drosophila* oogenesis with major developmental events labeled below. The beginning of each developmental stage is indicated by a mark on the line. The interval between stages was drawn in proportion to estimated development time. (b) Micrographs of major developmental stages of *Drosophila* oogenesis. Top panels show three-dimensional projections of z-stacks of confocal images with nuclei labeled in blue, E-cadherin labeled in green, and myosin labeled in red. Bottom panels show single-plane confocal images through the middle of the tissue with nuclei labeled in white, E-cadherin in green, and myosin in red. Scale bar is 50  $\mu$ m (He *et al.*, 2011).

As it shown in Figure 1.21 the oocytes are created through a production line starting from the germarium of the ovariole. In the anterior side of the ovariole (germarial region-1), the “germ-line stem cell” is divided to create the “cytoblast” which undergoes four mitotic divisions, within the germarial region-2a, in order to produce the “16-cell cyst”. From these 16 cells, one is “picked” to produce the oocyte and the other 15 are transformed into the nurse cells. It is necessary to mention that the cell which will be turned to an oocyte is not picked randomly. The factors that turn one cell into the oocyte lay in the posterior-end of the germarium. There, molecules of mRNA placed by the border cells, and which are responsible for the establishment of the anteroposterior axis, interact with the closest cell and promote it to an oocyte; that is the reason why the oocyte is always located at the posterior-end of the follicle. Within the germarial region-2b, the 15 nurse cells and the oocyte are engulfed within the “egg chamber” in order to form the follicle. The egg chamber is formed by somatic follicle cells, which are produced from a “somatic stem cell” at the border of the region. It is interesting to note that distinct ovarioles are formed during the late larva and early pupa stages, when the stem cells are placed within tissue structure in order to form the ovaries. Finally, the follicles mature inwards, towards the follicular stalk, from which they will be dispensed into the calyx lumen. The follicles maturation is described as “egg development” (see below).



**Figure 1.21.** Construction of an egg chamber. The drawing shows events early in oogenesis that contribute to the basic structure of the egg chamber, including the encapsulation of the germline cyst by follicle cell precursors, the differentiation of polar and stalk cells, and the posterior positioning of the oocyte. The large variety of cell types found within the germarium and early egg chambers are indicated (Horne-Badovinac & Bilder, 2005)

#### 1.4.4 Mosquito egg development

The egg development process in the mosquito requires a full blood meal in order to be completed. The process covers the events occur during the maturation of the secondary follicle into the primary follicle. In order for the events to be recorded, scientists have introduced time frames which are defined by hours post blood meal (PBM). The blood meal stimulates processes that lead to the oocyte growth and the final form of the egg. During the events, the nurse cells become polyploid and produce large quantities of mRNA, which they pump insight the oocyte. They also produce large amount of proteins which promote the growth of the oocyte.

The blood meal triggers the upregulation of a big percentage of the transcriptome of the mosquito. Nearly a 16% of all the transcripts are under that effect. Many of these genes are responsible for the digestion of the blood and the egg development. A portion is necessary for immunity where other genes are responsible for the fertility of the egg. Important transcripts that also are abundant in the ovary PBM, are those which regulate the eggshell development and the vitellogenesis (Marinotti *et al.*, 2006). The vitellogenesis is one of the processes which have evolved differently in mosquitoes compared to fruit flies; the process is regulated almost exclusively by the fat body of the insect.

In the final stages of the egg maturation the vitelline membrane matures and the eggshell starts to form. The fat body supplies the developing eggs with components in order to finalise these processes. The knowledge we possess on insect eggshell morphology and composition, comes from studies on *Drosophila melanogaster* (Margaritis, 1985; Waring, 2000; Cavaliere *et al.*, 2008): The

eggshell is divided into two major and structurally distinct layers, the inner vitelline membrane (VM) and the outer chorion figure (CM) (Trogakos & Margaritis, 2002). The insects' eggshells provide the embryo protection from physical and biological insults and insure their survival.

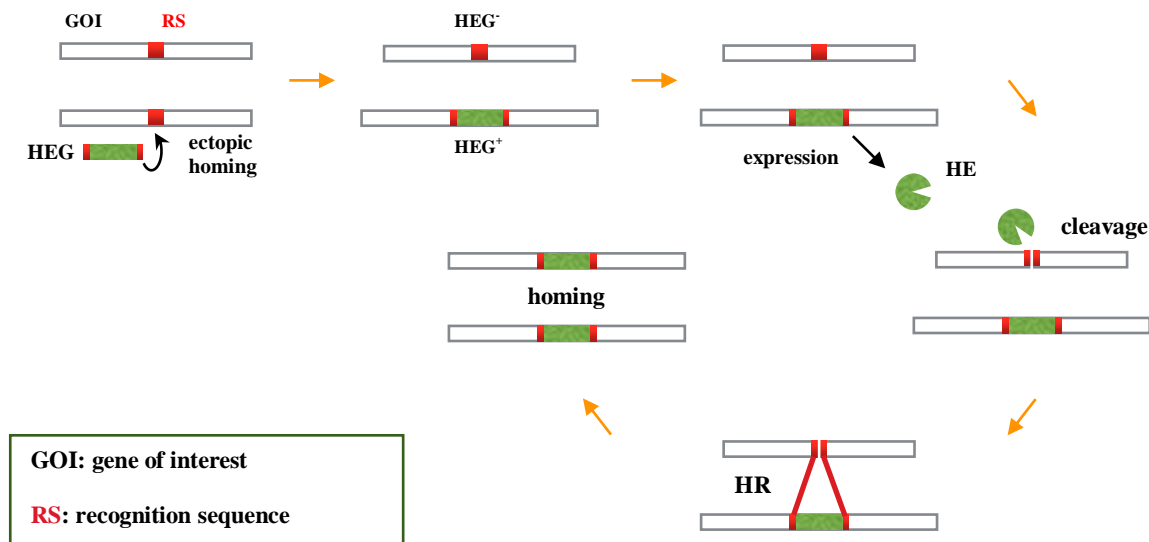
#### 1.4.5 Mosquito egg fertilisation

During the oviposition the eggs pass through the oviduct in order to be deployed. Connected to the common oviduct is the spermatheca which is filled with sperm, since the mating comes usually before the blood meal. The eggs that are placed inside the common oviduct to be deployed, are accepting sperm components from the spermatheca and thus they are fertilised. The sperm enters the egg from the "micropyle", a hole on the chorion, which is located at the anterior side of the egg. As mentioned before, the male sperm contains molecules and steroid hormones that trigger and regulate female oogenesis in the female mosquito (Robinson *et al.*, 2013; Baldini *et al.*, 2013). Recent studies also reveal that the sperm contains components that play a conserved role in fertility (Dottorini *et al.*, 2013). These proteins, called accessory gland proteins (Acps), have been identified in *Anopheles gambiae* in 2007 (Dottirini *et al.*, 2007).

### 1.5 The Homing Endonuclease System

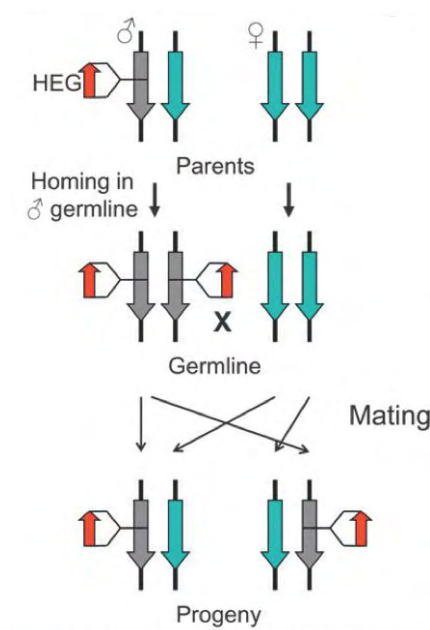
#### 1.5.1 HEGs: selfish genetic elements

Homing Endonucleases (HEs) are modified enzymes that are designed to recognize and cleave a 20-30bp sequence of interest. The gene which encodes the HE (HEG), is inserted in the middle of its own recognition sequence and thus, chromosomes that carry the HE gene (HEG<sup>+</sup>), are protected from cleavage since the recognition site is distorted by the HEG itself (Figure 1.22). The recognisable sequence, which is found on chromosomes that do not contain the HEG (HEG<sup>-</sup>), will be cleaved and the broken HEG<sup>-</sup> chromosome will attempt to be repaired by the cell's homologous recombinational repair system (HR). The system will make use of the homologous chromosome as a template, in this case the HEG<sup>+</sup> chromosome, in order to repair the broken site. Since upstream and downstream of the HE gene, on the HEG<sup>+</sup> chromosome there are homology regions between the two chromosomes, the repair system will insert unintentionally, the HE gene from the HEG<sup>+</sup> to the HEG<sup>-</sup> chromosome, transforming it into a HEG<sup>+</sup>. The process of the insertion of the HE gene into the other chromosome is called "homing". So, after the repair, both chromosomes will contain the HEG, and the former heterozygote will become a homozygote! Therefore, HEGs - the genes which encode the HEs - are described as "selfish" genes that can spread through populations by a non Mendelian inheritance; these genes are spread with a "super-Mendelian" inheritance (Chevalier & Stoddard, 2001; Goddard *et al.*, 2001).



**Figure 1.22.** Schematic illustration of the stages followed for homing of a HEG to occur. The RS, which is found on chromosomes that do not contain the HEG (HEG<sup>-</sup>), will be cleaved by the HE protein and the broken HEG<sup>-</sup> chromosome will attempt to be repaired by the cell's homologous recombination repair system (HR). The system will make use of the homologous chromosome as a template, in this case the HEG<sup>+</sup> chromosome, in order to repair the broken site. Since upstream and downstream of the HE gene, on the HEG<sup>+</sup> chromosome there are homology regions between the two chromosomes (red lines), the repair system will insert the HE gene into the other chromosome. The process of the insertion of the HE gene into the other chromosome is called "homing". So, after the repair, both chromosomes will contain the HEG, and the former heterozygote will become a homozygote!

In order to insert a HEG inside a gene and create a heterozygote, some criteria should be taken into account. First, the target gene has to be chosen so that its knockout mutations have little phenotypic effect on the heterozygote state. Furthermore, the same knockout mutations should be lethal when they meet in the homozygous state. Also, the HEG should be inserted under the control of a meiosis or a germ-line specific promoter so that the heterozygous zygotes develop normally, since the HEG will not be encoded in the somatic cells, but they will transfer (homing) the gene to the gametes, since the HEG will be fully functional in the gonads (Burt, 2003).

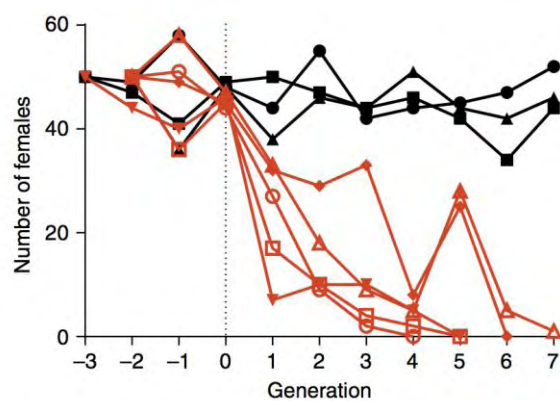


**Figure 1.23.** The population control strategy is illustrated by considering a HEG transgene inserted into and disrupting a target gene that it homes with 100% efficiency in the male germline. It is further assumed that the target gene is recessive lethal because it plays an essential somatic role but is not required for fertility. When a male *trans*-heterozygous for the transgene is mated to a wild-type female, the HEG homes onto its homologous chromosome in the germline. The animal remains viable as somatic tissues remain heterozygous for the HEG insertion; however, all gametes from the male now bear the HEG insertion. As a consequence, all progeny from this mating will be heterozygous for the HEG insertion, thereby increasing the frequency of this insertion within the population (Chan *et al.*, 2011).

### 1.5.2 HEGs and mosquito control

The HEG systems have been introduced as gene spread systems since they can rapidly (in less generations) introduce a gene of interest into a population so that we can control or alter the population. HEGs can be designed to carry additional genetic material or create specific mutations and thus they can act as a gene drive system. In order to use HEG to control an insect population such as mosquitoes, we can choose to spread a lethal allele in the population, or to spread an anti-pathogen effector in the population.

Recent studies have shown that HEGs can be used to drive a sex ratio distorter into the mosquito population (Galizi *et al.*, 2014). The study proposed the introduction of a HE protein (I-PPol a.k.a X-shredder) which specifically cuts a repetitive sequence in the 28S ribosomal gene sides on the X-chromosome. The HE gene is also under the regulation of a germline-specific promoter in the Y-chromosome, and therefore the HE is expressed in the spermatozoa of the male mosquito. The result is that all the spermatozoa that carry the X chromosome are eliminated and only those carrying the Y-chromosome are able to fertilise female eggs. The above resulting into a sex ratio distortion which drives the population into critically low numbers after approximately six generations (Figure 1.24). It is obvious that if this strategy has a high rate of homing the system can be used to suppress populations.



**Figure 1.24.** Reduction of the mosquito population within a cage using the HEG system and the spread of a “selfish” gene that alters population (Galizi *et al.*, 2014).

Other strategies that focuses on the targeting of female fertility genes are currently under develop. For those systems to work the gene knockout mutations have to present little phenotypic effect on the heterozygote state. On the contrary, the same knockout mutations should be lethal when they meet in the homozygous state. This method would allow the system to be used against a number of female fertility genes which are highly specific to the target species and thus negative effects such as the interspecies transmission of the HEG will be avoided. The identification and characterisation of those genes that present mosquito specificity will increase our knowledge on the mosquito vectors and will give insights on the different systems they use in order to thrive. Furthermore, understanding the vectors in depth we will be able to develop novel strategies to control the vectors’ populations and diminish pandemic diseases. The strategies could include from species-specific insecticides or even population suppression using the HEG system (Nolan *et al.*, 2011; Windbichler *et al.*, 2011)

### 1.5.3 Targeting of female germline- and soma- specific genes

In order for a fertility gene to be selected as candidate for the development of a HEG target, some criteria should be met first to ensure high rate of homing and maximum spread of the HEG gene to the population. As mentioned before, the target gene has to be chosen so that its knockout mutations have little phenotypic effect on the heterozygote state and they have a sterile or lethal phenotype while in the homozygous state. Also, the HEG should be inserted under the control of a meiosis or a male germ-line specific promoter so that the heterozygous zygotes develop normally, since the HEG will not be encoded in the somatic cells. The homing of the gene will occur to the gametes, since the HEG will be fully functional in the gonads; the expression of the HEG product in the male germline will ensure a successful gene drive. Another criterion we should think of is the gene not to present any phenotype to male mosquitoes. That means that the gene should be a female specific gene so that we could spread the system with the male population. Genes expressed in the soma of the female mosquito will be proved ineffective since the inheritance will be Mendelian. Also genes that affect female fertility and expressed in the female germline will lead to a dead-end since the female mosquito could not spread the gene. Finally, the candidate genes should be expressed early in the male germline to ensure the maximum homing to the spermatozoan and thus faster gene drive, and they have to be inactive in the fertilised egg since destruction of the female embryos will lead the system to a dead-end!

With the previous in mind, we can find a gene that will be a good target for the development of a HEG system in order to control mosquito populations. Moreover, there are a couple of extra criteria, which when they meet, parallel to the previous, give the perfect gene for the development of a HEG. The first of the two extra criteria, which is the least necessary to accomplish, is the gene to be located in an autosomal chromosome rather than a sex-chromosome. This will ensure faster spread of the gene since the missing X from the males will cause no inefficiencies and homing will occur when the HEG is driven in either sex. The second criterion is the gene to be somatically expressed. The accomplishment of this criterion it will be ideal for the development of a HEG, since the oocyte in the mosquito needs the components donated from the follicle cells, which they are somatic cells, in order to mature. To make things clearer, if the female parent is heterozygous for a mutation on a somatic target gene, the homing will occur in the germline cells and will not affect the fertility of the parent! Furthermore, if the gene is also female specific, the homing will occur in the germline of the male as well increasing the rate of the homing and thus the spread of the HEG, compared to genes that are germline specific. As we observed above, somatic cells have a crucial role in the egg development since they create the egg's membranes and they donate components to the oocyte.

To summarise; we set some criteria for a gene to meet in order to be a good target for a HEG project. The criteria we have set ensure high rate of homing and faster spread of the HE gene. We listed these criteria in order to be accessible and easy to memorise.

The “perfect” gene has to:

- Have NO effects on MALE mosquitoes
- Be an AUTOSOMAL gene
- Interrupt FEMALE fertility
- Be SOMATICALLY expressed
- Be RECESSIVE —> little effects on heterozygotes
- Expressed EARLY in the MALE gametogenesis

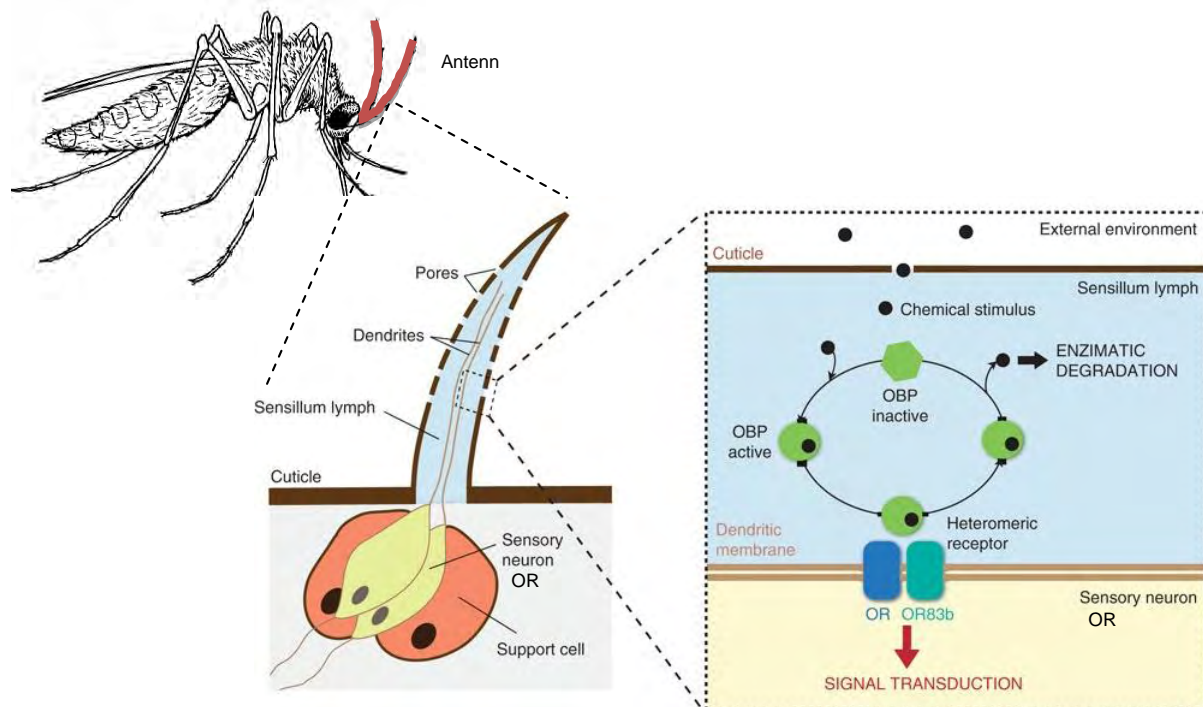
## 1.6 Insect Olfactory System

### 1.6.1 An olfactory system is essential

The olfactory system is a sophisticated, key-physiological element for the survival of insects and their prominence among other animals. The system is orchestrated at many different levels, starting with reception of the chemical molecules at the peripheral sensory organs, processing of the signal the molecules create at the antennal lobes, transfer of the signal to the higher processing centers of the brain, and finally, ultimate translation of the olfactory signals into behaviour. The olfactory system on the mosquito is necessary for key-procedures which are stimulated by sensory information. Mating, communication, detecting hosts and oviposition sites, are some of the processes of the mosquito that are regulated by environmental signs like odors and molecules.

Because it regulates the behaviour of the insect, the olfactory system has to possess the ability to distinguish between different odors, selectively detect, and rapidly inactivate even the smallest amounts of odorants, after they have stimulated the receptors. To select and distinguish amongst the plethora of odors that exist in the environment of the antennae of the mosquito, the system is dependent entirely on the types of odorant receptors (ORs). The receptors are normally expressed and located in the antennae of the insect in olfactory receptor neurons (ORNs). Odorant binding proteins (OBPs) are the link between the odors and the ORs (Figure 1.25). The OBPs seems to uptake hydrophobic odorants that reach the sensillar lymph through the pore tubules. The proteins solubilize and transport these chemical signals throughout the sensillum lymph to the membrane-bound odorant receptors. The ORs after binding these molecules are capable of stimulating the ORNs, which transmit the signal from the antennae towards the brain where the signal is translated into a behavioral response (Leal, 2012).

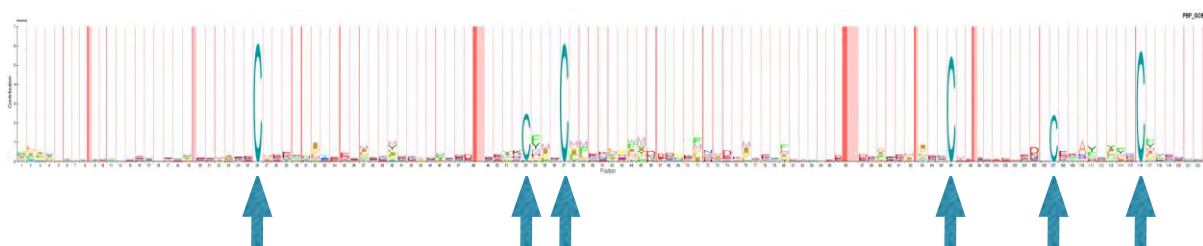




**Figure 1.25.** The olfactory system of the insect and its basic components and features. OBPs uptake hydrophobic odorants that reach the sensillar lymph through the pore tubules. The proteins solubilize and transport these chemical signals throughout the sensillum lymph to the membrane-bound odorant receptors. The ORs after binding these molecules are capable of stimulating the ORNs.

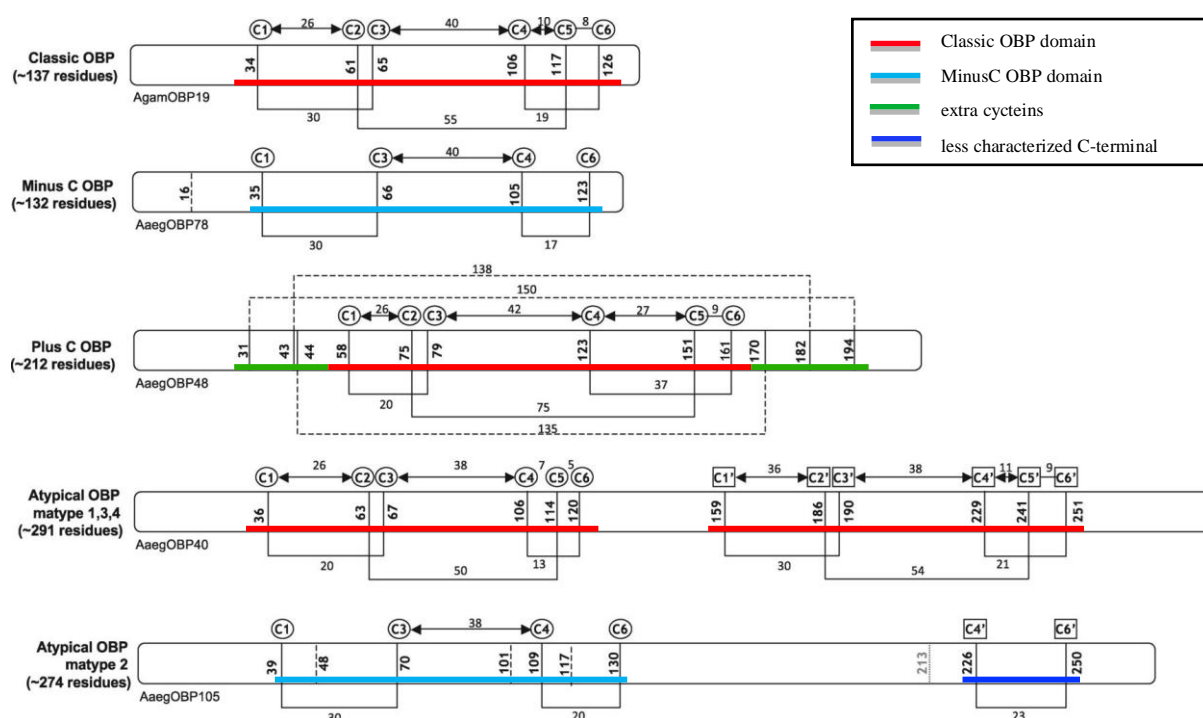
### 1.6.2 Odorant binding proteins

Odorant binding proteins form a large specific multi gene family. They are low molecular weight proteins (10-30 kDa), and are abundant in animals and insects. They are spherical and water-soluble proteins that have pockets for hydrophobic molecules on their structure, ergo they act as molecule binding and carrier proteins. Specifically, as it was mentioned above, they bind and thus solubilize hydrophobic compounds, acting essential for the olfactory system of insects. The family is characterized by a specific six  $\alpha$ -helical domain consisting of six highly conserved cysteines (Cy), that have distinct disulphide connectivities. These structural features and characteristics are now considered the hallmark of this protein family (Figure 1.26). (Calvo *et al.*, 2002, 2006; Valenzuela *et al.*, 2002).

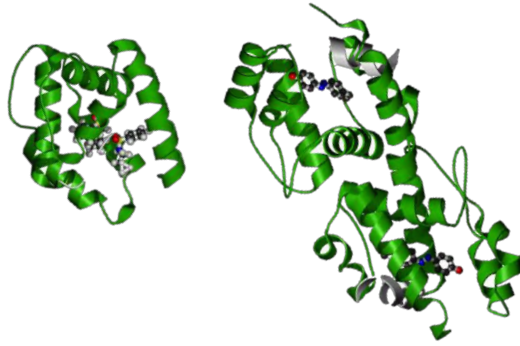


**Figure 1.26.** The Classic OBP domain displayed in weblogo. The frequency of each amino acid on the sequence is relevant to the size of the letter. The hallmark of this protein family are the six highly conserved cysteines (blue arrows), that have distinct disulphide connectivities and compose a specific six  $\alpha$ -helical domain (Calvo *et al.*, 2002, 2006; Valenzuela *et al.*, 2002). Note that the second and fifth cysteines are not so conserved and in some cases, like in MinusC group, they are missing.

The OBP family members are grouped according to their conserved cyctein number and their sequence structure. The hallmark of protein family, the six conserved cycteins, define the Classic OBPs group. One other group of OBPs lacks two of the six conserved cycteins (No 2 & 5) (Figure 1.27), and it is called MinusC OBPs. A different group carries additional conserved cycteins and for that is called PlusC OBPs. (Hekmat-Scafe *et al.*, 2002). The missing cycteins in the MinusC group equals to the lack of some disulphide connectivities, in contrast to PlusC group which the extra cycteins present novel disulphide connectivities. Finally, a fourth group, which is divided further, consists of some of the longest known OBPs. The group, called Atypical OBPs, is divided into two subgroups which have different structures (Figure 1.27). The first subgroup, the matype1,3,4 consists of two domains that are homologous to the Classic OBP domain, and hence they are considered as “dimer OBPs” (Figure 1.28). The second subgroup, the matype2, has initially been described to contain a single MinusC OBP domain in its N-terminal, extended by a less characterized C-terminal extension (Vieira & Rozas, 2011).



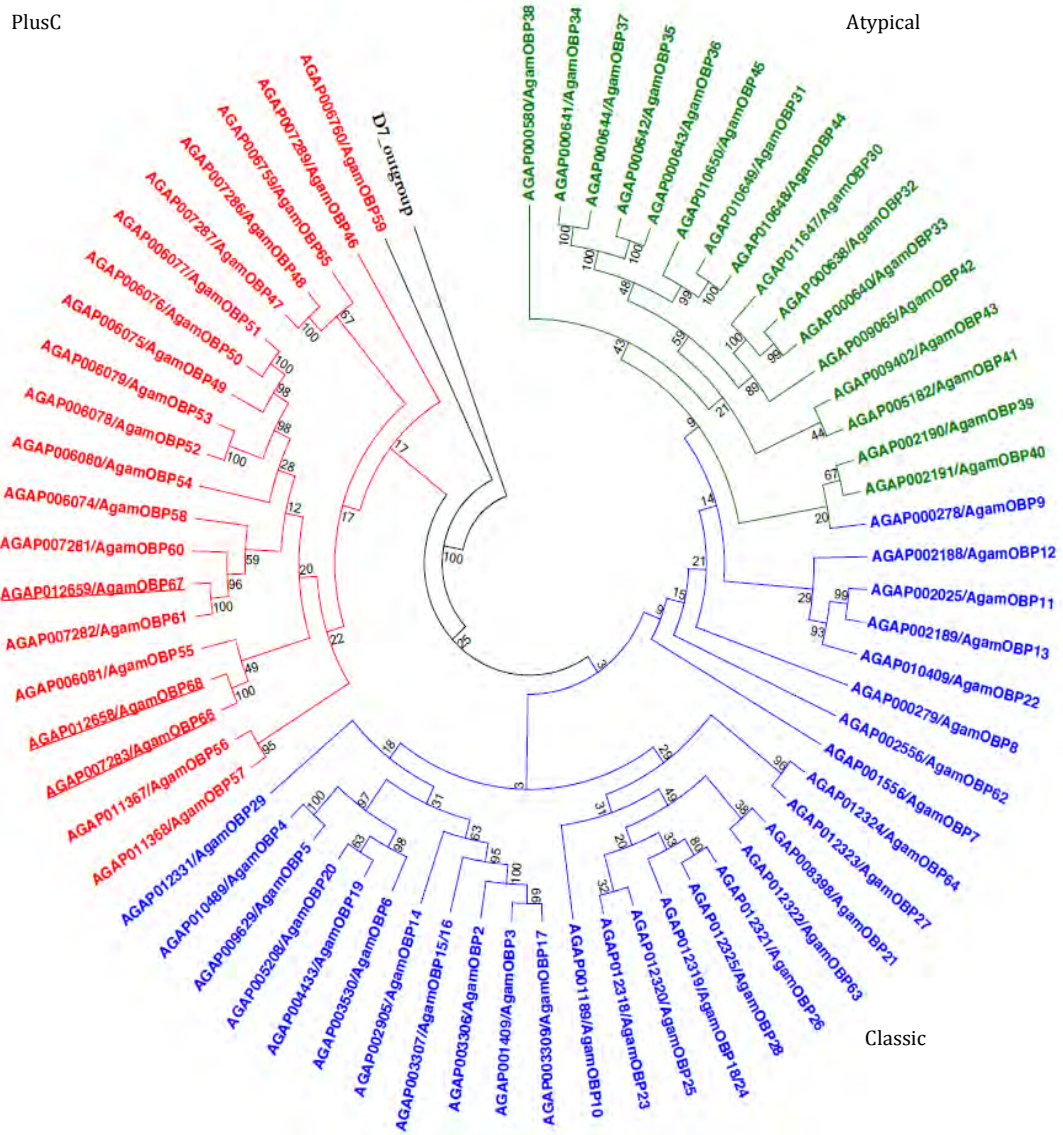
**Figure 1.27.** Cysteine conservation patterns across the different subfamilies and subgroups of OBPs from *Anopheles gambiae*, *Aedes aegypti*, and *Culex quinquefasciatus* genomes. The six conserved cycteins in OBP domain are denoted C1–C6. The six additional cycteins in the C-term of the Atypical OBPs are denoted C1’ –C6’. The lines connecting cycteins represent the disulphide bonds and dotted lines represent the additional disulphide bonds in the PlusC OBPs (Manoharan *et al.*, 2013).



**Figure 1.28.** The first subgroup, the matype1,3,4 (right) consists of two domains that are homologous to the Classic OBP domain (left), and hence they are considered as “dimer OBPs” (Vogt et al., 1991).

To date three groups of OBPs have been characterized in mosquitoes: the Classic OBPs, the PlusC OBPs and the Atypical OBPs. MinusC OBPs have never been described to date in mosquito genomes. In *Anopheles gambiae*, 77 OBPs (including five D7 salivary proteins) have been identified and sorted into the three classes above (Manoharan *et al.*, 2013) (Figure 1.29). The D7 salivary proteins mentioned before, is known to be distantly related to the mosquito OBP superfamily (Calvo *et al.*, 2002, 2006, 2009). The available structures of the D7 proteins indicate that the domains adopt a similar fold to the OBP domains but decorated with additional structural features and a seventh helix. The domains in D7 proteins have been shown to bind to biogenic amines in *A. gambiae* (Calvo *et al.* 2009).

Experiments that were designed to measure the expression levels of the OBPs in different tissues have shown that these proteins are expressed in many parts of the insect’s body and most of them seemed to have an expression peak in specific stages in the development and life-time of the insect. We have listed the OBPs that present both a high expression in the ovaries, and a high expression between 24-48hrs after blood meal. We set these parameters to coincide with the most intense period of egg development because we were interested in the proteins that may play a role in this process.



**Figure 1.29.** Rooted phylogenetic tree of the odorant binding proteins in the *Anopheles gambiae* genome. The bootstrap values of the branches are indicated on the nodes in percentage values (Manoharan *et al.*, 2013).

## **CHAPTER 2**

# **AIM OF THE PROJECT**

## CHAPTER 2: AIM OF THE PROJECT

The current dissertation had as studying object the ovary-specific odorant binding proteins of the *Anopheles gambiae* mosquito. The aim was to analyse this ovary-specific group of the OBP family on this specific species, providing therefore new insights on the subgroups' characteristics. Findings of the project will help future studies uncover the relations exist, between this particular subgroup of OBPs and the processes of oogenesis and egg development on the mosquito.

The first thing the project tried to uncover was the identity of these OBPs. The OBPs' characteristics were revealed through an intensive *in silico* analysis based on previous experiments and information extracted from the literature. The aim was to find the OBPs that are expressed within the ovaries after a blood meal, and provide a list with the most important genes that were useful to study further.

The second part of the project was to determine the exact location of the expression of these OBPs inside the ovary by searching for the mRNAs that encode the proteins. A whole-mount *in situ* hybridisation process was exploited in order for the transcripts to be revealed and thus for different hypothesis to be able to be composed. The location of the transcripts would give new insights on the processes that take place inside the ovary after a blood meal. Finally, it would make clear if the OBPs are expressed from the egg, ergo the germline cells of the mosquito, or from the follicular membrane, ergo the somatic cells of the mosquito.

The ultimate purpose of the current project was to identify if the ovary specific odorant binding proteins are suitable to be considered as candidates for the development of a HEG system, which would be used to insert mutations into the mosquito population and most importantly, spread them using the non-Mendelian inheritance characteristic of the HEG spreading system.

# **CHAPTER 3**

## **MATERIALS AND METHODS**

## CHAPTER 3: MATERIALS AND METHODS

### 3.1 Mosquito Strain Maintenance And Tissue Dissection

#### 3.1.1 Mosquito strains and husbandry

Wild type *Anopheles gambiae* sensu stricto mosquitoes were reared in small cages at 28°C, 80% humidity with 12 hours daylight cycle and fed on 10% glucose solution. Female mosquitoes 5 to 10 days old were blood fed on anaesthetised mice to stimulate the ovary growth and oogenesis. In order to maintain the line, eggs were collected by placing a glass pot, half filled with warm 1% tonic salt solution<sup>1</sup>, and lined with Whatman paper, 48-72hr PBM. The mosquitoes were left to lay overnight. Next day, the glass pot was removed and covered with a petri dish lid. Approximately 24hr later the hatching larvae (L1) were placed in a net covered tray filled with warm 1% tonic salt solution, containing fish-food pellets, and left to pupate for approximately one week. Pupae were collected from the tray and placed in a glass pot half filled with warm 1% tonic salt solution. The glass pot was then placed inside a cage so that the mosquitoes could emerge.

#### 3.1.2 Ovary dissection

48hr PBM female mosquitoes were harvested and anesthetized using CO<sub>2</sub>. Ovary dissection was performed on a microscope using transparent slides covered with 1xPBS buffer. Mosquitoes were placed on the slide by using sterile forceps. Ovaries were removed by holding the thorax with one pair of forceps while another pair of forceps pulls off the penultimate abdominal segment to release the ovaries. Dissected ovaries were placed in a nylon mesh cell strainer (40µm pore size) that was immersed in 4ml 1xPBS solution, inside a small petri dish. The petri dish was maintained on ice until fixation of the ovaries.

### 3.2 Nucleic Acid Manipulation

#### 3.2.1 RNA extraction

RNA molecules were extracted using TRIzol and small scale extraction of total RNA method, optimized for *Drosophila melanogaster* (Auburn, 2006). Mosquitoes from 8 different time frames: 24, 36, 40, 44, 48, 52, 56 and 72 hours PBM, were homogenised in 400µl TRIzol solution using an RNase free pestle. Samples were cryo-centrifuged (4°C) at 13.000rpm for 10 minutes to pellet debris and thus remove unnecessary tissue. Supernatants, containing the RNA were transferred to fresh RNase free tubes and vortexed for 60 seconds with 0.2 volumes of chloroform. The mixtures were

---

<sup>1</sup> All buffers and solutions are described on Appendix A, page 55



centrifuged again at 13.000rpm at 4°C for 15 minutes to separate the phases. Upper, RNA-containing aqueous phases were removed by aspiration and transferred to new RNase free tubes, before the addition of 0.8 volumes of isopropanol and 1µl of glycogen (glycogen is used to maximize precipitation). The tubes were inverted to mix the components and left for 30 minutes to incubate at -20°C before centrifuged again at 13.000rpm at 4°C for 15 minutes to pellet the RNA. The supernatants were discarded and the pellets were washed twice with 500µl of 70% ethanol, which was prepared with RNase free distilled water. Finally, the ethanol was removed and the pellets left to air dry before resuspending in 20µl of warm RNase free water. The RNA concentration of each tube was quantified by Nanodrop® and the RNA containing tubes stored at -80°C.

### 3.2.2 DNase treatment

DNase I (E.C.3.1.21.1) is an endonuclease that hydrolyzes phosphodiester linkages yielding oligonucleotides with a 5' phosphate and a free 3' hydroxyl group (Kunitz, 1950). DNase I has been shown to act on single- and double-stranded DNA, chromatin and RNA:DNA hybrids, although the specific activity of DNase for RNA:DNA hybrids and single-stranded DNA is at least 2 orders of magnitude below that for double- stranded DNA. TURBO™ DNase is a genetically engineered form of bovine DNase I with greater catalytic efficiency, than conventional DNase I at higher salt concentrations and lower DNA concentrations (Clark & Eichhorn, 1974; Price, 1975). TURBO™ DNase is used to degrade DNA in the presence of RNA when the absence of RNase is critical to maintain the integrity of the RNA.

Genomic DNA contamination in RNA samples was reduced by treatment with TURBO™ DNase. According to manufacturers' instructions 1µl of TURBO™ DNase (2U/µl) was added to the samples and incubated at 37°C for 15 minutes. With these reaction conditions, TURBO™ DNase is able to remove up to 2µg of genomic DNA. After the incubation, the DNase was inactivated and removed performing a phenol/chloroform extraction.

### 3.2.3 Phenol/Chloroform extraction

The phenol/chloroform extraction method was used to inactivate and totally remove the TURBO™ DNase from the RNA samples. This step was necessary, since DNase would have degraded any cDNA would had been made from the RNA in later steps.

One volume of tris saturated phenol and one volume of chloroform were added to the RNA samples which were treated with DNase. The components were mixed well and centrifuged at 10.000rpm at room temperature for 5 minutes. Then, the upper aqueous phases containing the RNA, were transferred to new tubes, before an equal volume of chloroform was added. After mixed thoroughly, the samples were centrifuged at 10.000rpm at room temperature for 5 minutes. The

aqueous phases were transferred to new tubes before the addition of 0.1 volumes of 3M sodium acetate. The mixtures were vortexed before the addition of 2.5 volumes of 100% ethanol, and the addition of glycogen (5µg/µl) to a final concentration of 0.2µg/µl to increase the precipitation of the RNA. After a quick vortex the tubes were left to incubate for 15 minutes at -20°C. Next, the tubes were cryo-centrifuged (4°C) at 10.000rpm for 10 minutes to precipitate the RNA; the supernatants were removed and the RNA pellets were rinsed with cold 70% ethanol. After one extra cryo-centrifuge at 10.000rpm for 10 minutes, the supernatants were removed and the pellets were left to air dry, before resuspending in 20µl of warm RNase free water. The RNA concentration of each tube was quantified by Nanodrop® and the RNA containing tubes stored at -80°C.

### 3.2.4 First strand cDNA synthesis

For first strand cDNA synthesis the Superscript™ III First-Strand Synthesis system was used in which a reverse transcriptase and random hexamers are combined to create double-stranded cDNA. RT-PCR is a variant of the standard PCR which is commonly used in molecular biology, since it provides the possibility to assess the transcriptome in vivo or in vitro. The process makes use of reverse transcriptases in order to clone genes of interest by reverse transcribing its mRNA to its DNA complement (cDNA) which can be accessed and amplified using traditional PCR procedure.

The RNA which was extracted from mosquitoes from 8 different time frames (24, 36, 40, 44, 48, 52, 56 and 72 hours PBM) was used to generate complementary double-stranded cDNA through this process. Following manufacturers' instructions, 1µg of total RNA, 1µl of random hexamers (50ng/µl), 1µl of dNTP Mix (10mM) and dH<sub>2</sub>O up to 14µl were placed inside a 0.2ml PCR tube. The components were mixed by aspiration and the tube was placed inside a thermocycler in order to incubate at 65°C for 5 minutes; the thermocycler was used to avoid the time consuming changes of the temperature in the heat block. Then, 4µl of 5xFirst-Strand buffer, 1µl DTT (0.1M) which helps disulphide-bonding breaking (Cleland, 1964), and 1µl Superscript™ III reverse transcriptase (200U/µl) were added to the tube, and after mixing, the tube was incubated inside the thermocycler, first for 5 minutes at 25°C and then at 50°C for 50 minutes in order for reverse transcription to perform. Finally, the Superscript™ III reverse transcriptase was inactivated by heating at 70°C for 5 minutes. The resulting cDNA which produced from the whole procedure, served as template for PCR reactions.

### 3.2.5 PCR

Polymerase Chain Reaction is a simple, yet elegant, enzymatic assay that enables amplification of a specific DNA fragment from a complex pool of DNA. Kary Mullis, who conceptualized the PCR assay, explained that it "lets you pick the piece of DNA you're interested in and have as much of it as you want" (Mullis, 1990). Each PCR assay requires the presence of template DNA, primers,

nucleotides, and DNA polymerase, as well as a suitable buffer for the enzyme to work. The process relies on thermal cycling, consisting of repeated cycles of heating and cooling. This is necessary in order for the DNA to melt and the enzymatic replication to occur.

A standard PCR reaction consists of 5 steps, each important as equal. First, the reaction starts with a quick heat at 95°C for 5 minutes. Then it enters a cycle of temperature changes that repeats approximately 35 times. The cycle consists of: a “denaturation” step which occurs at 96°C for 5 seconds, an “annealing” step which occurs 5°C below the melting temperature ( $T_m$ ) of the two primers, and an “elongation” step at 68°C for 30 seconds. Finally, the last cycle follows a “final elongation” step at 72°C for 1 minute, in which any unfinished elongation comes at an end. These steep changes of the temperature are assisted by the use of the thermocycler, a machine that was invented for PCR reactions.

The QIAGEN Fast Cycling PCR Kit was used to amplify the exon of the genes of the desire OBPs. cDNA from the first-strand cDNA synthesis reaction was used as template, since these genes are transcribed after a blood meal. Briefly, 10µl of 2xQIAGEN Fast Cycling PCR Master Mix and 2µl of 10xCoralLoad Fast Cyclin Dye were added into 0,2ml PCR tubes. Next, we added to the tubes 1µl of each of the two primers (10mM each), in order to reach a final concentration of 0.5mM; each OBP had its own pair of specific primers. Finally, we added 1µl of template DNA to each of the tubes and we brought up to 20µl total volume with dH<sub>2</sub>O. The tubes were added to the thermocycler and the programme above was executed.

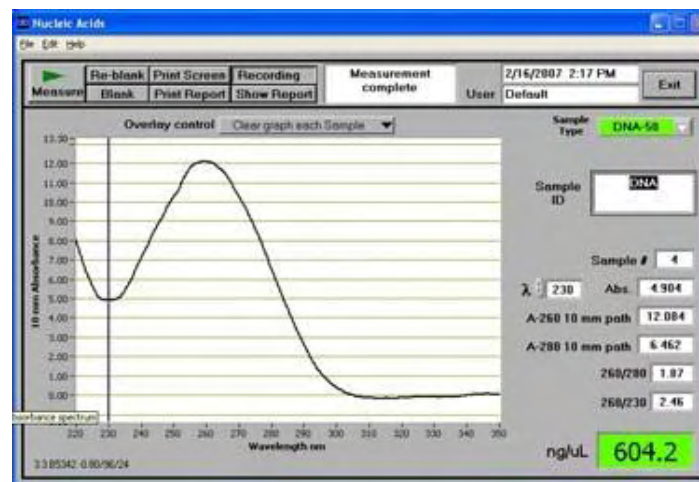
### 3.2.6 PCR purification

The PCR products were purified in order to achieve a clean nucleic acid free of enzymes and buffers. The products were purified using the QiaQuick PCR purification kit which is supplied by Qiagen. The company assists the kit with an easy-to-execute protocol. Briefly, five volumes of Buffer PB are added to one volume of the PCR sample and mixed well. The mixture is then added to a provided QIAquick spin column, which is first inserted inside an also provided 2ml collection tube. The tube is then centrifuged for one minute at 13.000rpm and the flow-through is discarded. Next, the column is washed with 750µl of Buffer PE and then dried at 13.000rpm for one minute. Finally, the DNA is eluted using 30µl of Buffer EB or water, by centrifuging again at 13.000rpm for one minute.

### 3.2.7 DNA quantification

Quantification of the DNA samples was able through NanoDrop<sup>®</sup> ND-1000, a spectrophotometer which had been provided by Thermo Scientific. The machine accepts 1µl of sample volume in which the measurements are taken. The photometric assessment is based on the different

absorbance that different molecules present in different wavelengths, according to their nature. Since nucleic acids absorb at 260nm, they contribute to the total absorbance of the sample in that particular wavelength. It is possible also to measure the purity of the DNA or the RNA, since the programme measures the ratio of the absorbance at 260nm and 280nm. A ratio of  $\sim 1.8$  is generally accepted as “pure” for DNA while a ratio of  $\sim 2.0$  is generally accepted as “pure” for RNA. If the ratio is appreciably lower in either case, it may indicate the presence of protein, phenol or other contaminants that absorb strongly at or near 280nm (NanoDrop® Technical Support Bulletin).



**Figure 3.1.** The typical interface of NanoDrop ND-1000. The right lower corner box indicates the concentration of the nucleic acid in ng/μL.

### 3.2.8 *In vitro* transcription

*In vitro* transcription is the process in which RNA polymerases are used to transcribe genomic material within a tube. Different polymerases are used commonly, such as T7 RNA Pol Plus and T7 MEGAscript®. T7 RNA Polymerase catalyzes the synthesis of RNA in the 5' to 3' direction, in the presence of a DNA template containing a T7 phage promoter. T7 RNA Polymerase is used because it can generate high specific activity labeled RNA probes taking in advantage the run off transcription method (Schenborn & Meirendorf, 1985).

The Ambion T7 PolymerasePlus system was used in order to transcribe the OBPs PCR products that were produced in previous step. It is wise to note that the location of the T7 promoter (5'-end of the reverse primers) will result in the production of a labeled complementary strand of the gene, so that it would be able to bind within the cells. Briefly, 2μL of 10xRNA Transcription Buffer, 1μL of T7 RNAPol Plus (20U/μL), 2μL of 10xDIG RNA labeling mix (Roche) and dH<sub>2</sub>O up to 20μL, were added to the tubes in order to transcribe 200ng of PCR product. The reactions were incubated for two hours at 37°C. It is important to state that the DIG labeling mix contains UTP that carries the DIG molecule,

which is recognised by the anti-DIG antibody. Finally, to terminate the reaction and precipitate the RNA, a lithium chloride (LiCl) precipitation was necessary.

### 3.2.9 LiCl precipitation

LiCl precipitation is useful following *in vitro* transcription! RNA is efficiently precipitated in contrast to protein, carbohydrates and DNA, which are very inefficiently precipitated or are not precipitated at all. LiCl does not precipitate transfer RNA (tRNA) and unincorporated nucleotides. The minimum length required for LiCl precipitation of RNA is unknown, but previous experiments (Applied Biosystems/Austin) have indicated that RNA as small as 100 nucleotides, prepared by *in vitro* transcription, can be efficiently precipitated.

In order to precipitate the products of the *in vitro* transcription, the Ambion® Lithium Chloride Precipitation Solution was used. This solution is provided ready for use and requires no extra preparation. 30µl of the Lithium Chloride Precipitation Solution were added into the tubes and after they were mixed thoroughly they were chilled for 30 minute at -20°C. Then, 1µl of glycogen was added, to increase precipitants, and centrifuged at 14.000rpm for 15 minutes to pellet the RNA. Next, after discarding the supernatant, the pellet washed with 1ml of ice-cold 70% ethanol, and re-centrifuged at the same speed for 15 minutes to succeed removal of residual salt and unincorporated nucleotides. Finally, the ethanol was carefully removed and the tubes left to air dry, for maximum ethanol removal. The pellet was resuspended in 20µl of nuclease free water and the RNA concentration was determined with NanoDrop. The RNAs were stored at -20°C.

### 3.2.10 Gel electrophoresis

Agarose is a polysaccharide polymer material, generally extracted from seaweed and it is a linear polymer made up of the repeating unit of agarobiose. It consists one of the two principal components of agar, and is purified from it by removing agar's other component, agarpectin (Water Structure and Science). The properties of agarose were found very useful in molecular biology for the separation of large molecules, especially DNA, by electrophoresis. The method “gel electrophoresis” proved to be an easy and fast method for the separation and visualisation of nucleic acid fragments, depending on their size and charge. The principle of this method is based on the migration of charged molecules that occurs within a field of an electric current. Thus, the charged molecules are migrating towards the opposite charge of their own. Nucleic acids are negatively charged due to their phosphate groups which compose their backbone; so when DNA is placed on a field with an electric current, these negatively charged molecules migrate towards the positive end of the field. The field in a gel electrophoresis is an agarose slab (0.7 to 2% agarose) immersed in a buffer bath which consists of Tris/Borate/EDTA (TBE) buffer. The agarose gel is a

cross-linked matrix which assists as a three-dimensional mesh, which allows the separation of molecules based on their molecular size. The pores of the agarose gel allow shorter molecules to move faster and migrate farther than longer ones.

DNA is able to be visualised on the gel with the addition of special stains that bind nucleic acids. One commonly used stain for DNA visualisation is the ethidium bromide (EtBr). EtBr binds strongly to DNA by integrating between the bases, and absorbs light from the UV spectrum causing the nucleic acid molecules to fluoresce.

For our experiments, 1% of agarose gel was sufficient and thus it was prepared using 1xTBE buffer and 0.004% EtBr. Briefly, one gramme of agarose was dissolved into 100ml of 1xTBE buffer. The dissolution was achieved with heat in a microwave oven for 2 minutes; in liquid state, agarose is transparent. After the agarose was fully dissolved into the TBE, 4 $\mu$ l of EtBr were added and mixed well. Finally, when the solution cooled down at roughly 60°C, it was poured inside a matrix in order to solidify and take the rectangular shape it is desired, and also for the wells to be formed. In order to identify the molecular weight of the bands, a ladder was included with the samples (1Kb Plus DNA ladder), in a different well. Also, for tracking the molecules during loading and during running of the gel, a dye (CoralLoad<sup>®</sup> PCR Buffer (Qiagen)) was included within the samples. CoralLoad PCR Buffer contains 2 gel-tracking dyes for improved pipetting visibility during PCR setup, enabling immediate gel loading of PCR products for easy visualization of DNA migration.

### 3.3 Tissue Manipulation

#### 3.3.1 Fixation

Fixation in cell biology is called the step in which samples and tissues are treated with solutions and substances that promote the preservation of samples. Fixatives terminate any ongoing biochemical reaction and thus prevent decaying procedures like autolysis and putrefaction. Fixation also increases the mechanical strength or stability of the treated tissues (Carson, 2009).

After tissue dissection, ovaries were fixated based on previous attempts (Juhn & James, 2012). Ovaries were immersed, using the cell strainer, on a small petri dish filled with 4ml of freshly prepared soft tissue fixation solution (STFS) at room temperature. The petri dish was placed on a nutator at minimum level of nutation for 30 minutes to 1 hour. Fixation solution was prepared fresh to avoid oxidation of the formaldehyde fixative. Four extra small petri dishes were prepared in order to rinse out fixative solution and equilibrate tissues for storage. The extra petri dishes were filled with 4ml of PBT/ethanol solutions in different ratios (3:1, 1:1 and 1:3). The last petri dish was filled

with 100% ethanol. Ovaries were incubated for 5 minutes in each solution and then stored at -20°C in 100% ethanol. The tissues were moved by immersing the cell strainer between the petri dishes in order to avoid damaging the ovaries' ultrastructure.

### **3.3.2 Proteinase K treatment**

Proteinase K is a subtilisin-related serine protease that hydrolyses a variety of peptide bonds, leading to degradation most of the proteins in the cell tissue allowing thus the hybridization of the probe. Proteinase K digestion is a critical step for successful in-situ hybridization since insufficient digestion will result in a diminished hybridization signal. On the other hand, if the sample is over digested, tissue morphology will be poor or completely destroyed, making localization of the hybridization signal impossible.

In order to treat the samples with proteinase K, the ovaries should be equilibrated back to PBT by performing stepwise equilibration with PBT/ethanol solutions (1:3, 1:1 and 3:1). After the equilibration the cell strainer containing the ovaries was immersed into 100% PBT. Ovaries were washed three times in fresh BPT for 5 minutes with nutation level at minimum. Proteinase K (23mg/ml) was added to the last wash in a concentration of 0.01mg/ml. The proteinase treatment was occurred for 5 minutes at room temperature with nutation level at minimum. After 5 minutes the cell strainer was immersed into ice-cold PBT in order to terminate the reaction. Then, the ovaries were washed again two times in fresh BPT for 5 minutes at room temperature. Finally, the samples were post-fixated using a small petri dish that contained 4ml of freshly prepared fixation solution (STFS). The fixation process was occurred for 30 minutes at room temperature, keeping nutation level at minimum. Then, the ovaries were washed again for three times in fresh BPT for 5 minutes.

## **3.4 In-situ Hybridization**

### **3.4.1 Hybridization**

Tissues were equilibrated at room temperature into hybridization solution (Hyb) by incubation in a small petri dish which contained 4ml of Hyb/PBT solution in a ratio 1:1. After 30 minutes of incubation, the cell strainer containing the ovaries was placed on a petri dish filled with 100% Hyb solution. At this point the samples were separated in 1.5ml Eppendorf tubes so that the hybridization of different probes could be distinguished. For transferring the ovaries from the cell strainer into the tubes, a p200 pipette tip was used. The end of the tip was cut off and by gentle aspiration the ovaries were collected and transferred into the tubes. The total volume of Hyb in the tubes was brought to 1ml and then the tubes were wrapped in sealed-air protective packaging (bubble wrap) and were placed inside a hybridization bottle. Next, the bottle was placed inside a

hybridization oven and pre-hybridization was performed at 55°C for 30 minute with rotation (7rpm). After pre-hybridization was finished the tubes were removed from the bottle and were let to stand for 5 minutes so that the ovaries could settle at the bottom of the tubes. Using aspiration the pre-hybridization solution was removed before 100ml of Hyb was added to the tubes; tubes were kept at heat block at 55°C. In order to perform the hybridization, 200ng of the different probes had been denatured at 85°C for 8 minutes, before placing on ice for 5 minutes. Finally, the denatured probes were added to the tubes containing the ovaries and mixed by gentle flicks. Hybridization was occurred overnight (16-24hr) fixed on a heat block at 55°C.

### 3.4.2 RNase A treatment

Ribonuclease A (RNase A) is an endoribonuclease that consists of a single chain polypeptide containing 4 disulfide bridges. RNase A attacks at the 3' phosphate of a pyrimidine nucleotide resulting to the degradation of that nucleotide (Burrell, 1993). A major application for Ribonuclease A is the removal of single stranded RNA from preparations of DNA. Degradation of the unhybridized single stranded RNA probe is essential for reliable in-situ hybridization since the presence of any unhybridized probe shall give unspecific signal.

Samples were equilibrated with Hyb solution by adding 1ml into the tubes and inverting 6 times; ovaries were left to settle at 55°C in a heat block. Maintaining the samples at 55°C in heat block is suggested in order to reduce non-specific binding of residual unbound probe. The samples were washed three times in 1ml Hyb solution at 55°C for 30 minutes in a hybridization oven. After the washes, the samples were poured out from the eppendorf tubes into labeled cell strainers on small petri dishes, filled with 4ml of PBT/Hyb at a ratio of 1:1. After 15 minutes equilibration, the cell strainers containing the ovaries were immersed into 100% PBT. Ovaries were washed four times in fresh BPT for 5 minutes with nutation level at minimum. RNase A (24mg/ml) was added to the last wash in a concentration of 0.02mg/ml. The RNase A treatment was occurred for 30 minutes at 37°C in an incubator, without nutation. Finally, the samples were washed five times in fresh BPT for 3 minutes, keeping nutation level at minimum.

### 3.4.3 Antibody incubation

Because for the in situ hybridization a probe that it is labeled with digoxigenin (Dig) is used, it is essential to treat the samples with an anti-Dig antibody. For that an anti-Digoxigenin antibody is used, which is conjugated with the enzyme alkaline phosphatase (AP). The enzyme will help with the detection of the location of the mRNA inside the ovary in a latter step. To treat the samples with the antibody, a freshly prepared blocking solution is necessary. Blocking solution reagents are able to cover any unspecific binding sites that are accessible to the antibody, increasing the reliability of the

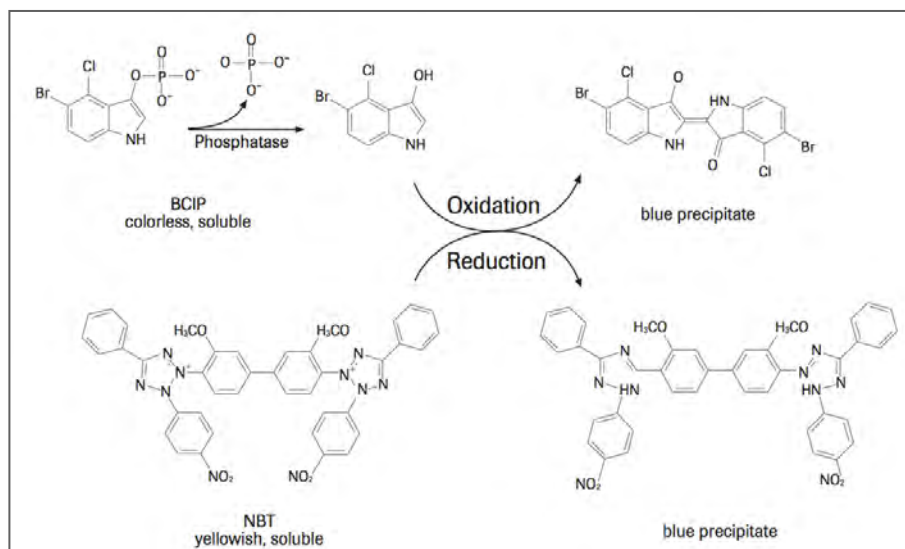


procedure. It is necessary the serum that is used to make the blocking solution, to be from the same organism that was used to make the antibody; in this example, from sheep.

The cell strainers containing the ovaries were immersed into 4ml of fresh blocking solution and incubated at room temperature for 30 minute with minimum nutation. A 1:1000 dilution of anti-DIG conjugated-AP antibody was prepared, according to manufacturer's instructions, and added to the petri dishes containing the cell strainers. The antibody was added at a ratio of 1:5 with the PBT.

### 3.4.4 Antibody treatment

In order to visualize the mRNA inside the ovary and determine its location, a reaction that produces a chromophore substance is necessary to occur. Alkaline phosphatase, which is conjugated upon the anti-Dig antibody, is able to perform a dephosphorylation of the BCIP substrate (5-bromo-4-chloro-3-indolyl phosphate), which is then able to react with NBT (nitro blue tetrazolium chloride). BCIP is the alkaline phosphatase substrate, which after dephosphorylation is oxidized by NBT to yield a dark-blue indigo precipitating dye (Figure 3.2). NBT is thereby reduced to a dark-blue precipitating dye and serves to intensify the color reaction, making the detection more sensitive (Horwitz *et al.*, 1966; Michal *et al.*, 1983). For the experiment we are using NBT/BCIP ready-to-use tablets.



**Figure 3.2.** The BCIP substrate (5-bromo-4-chloro-3-indolyl phosphate) reacts with NBT (nitro blue tetrazolium chloride). BCIP is the alkaline phosphatase substrate, which after dephosphorylation is oxidized by NBT to yield a dark-blue indigo precipitating dye.

Samples were washed five times in PBT solution at room temperature for 10 minutes by transferring the cell strainers into new wells in order to remove the antibody and the blocking solution. Then the ovaries were washed 4 times with AP staining buffer for 5 minutes at room temperature with minimum nutation; AP staining buffer was prepared fresh. After the washes the samples were immersed into AP staining solution containing the NBT/BCIP substrates. AP staining

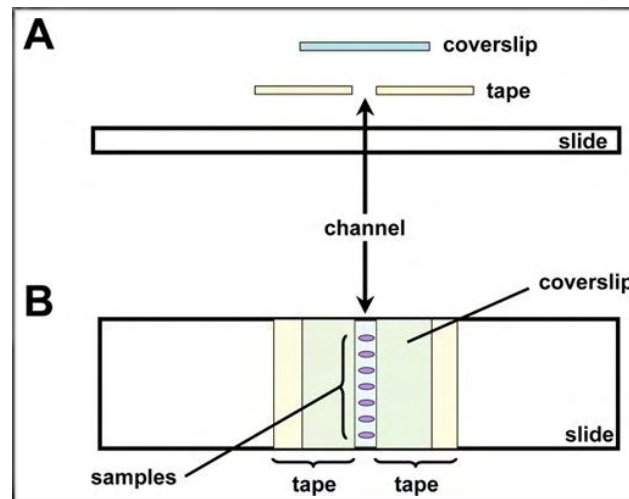
solution was prepared also fresh according to manufacturer's instructions. Each sample was incubated at room temperature with minimum nutation in the dark, since NBT/BCIP substrates are light sensitive. An opaque lid was used to cover the entire wells, ensuring the protection from light. The reaction progression was monitored frequently (every 10 minutes) for the formation of the dark-blue indigo precipitating dye. Depending upon the probe concentration, the concentration of the target mRNA and the presence of non-specific targets, the staining may proceed for several minutes to several hours. According to previous attempts in the lab, three hours were sufficient. Finally, after the formation of the dye, the ovaries were washed five times in PBT for 5 minute at room temperature in order to stop the staining completely.

### 3.4.5 Glycerol mounting

Samples and tissues which have been stained with peroxidase and alkaline phosphatase based systems, as well as with various fluorescent dyes, can be treated with mounting mediums, such as glycerol, for the permanent preservation of the stained tissue sections. Mounting mediums permit the long-term storage of antigens localized using organic insoluble chromogen substrates, including BCIP/NBT, and also prevents desiccation of tissues during slide mounting. Mounting the samples has also one other advantage: Solutions and substances used to achieve mounting have a very high refractive index, which when applied to the stained tissue sections can store the tissue specimens permanently, without the fading of the chromogens (Sigma-Aldrich, 2008).

Using a paintbrush, the ovaries were transferred carefully from the cell strainer to ~100 $\mu$ l/70% glycerol, in the centre of a glass bottom culture plate. The glycerol was allowed to permeate the samples overnight at 4°C. The next day the ovaries were transferred to slides and, using dissecting needles, sections of the ovaries were torn apart so that single follicles and follicle clusters could be released and visualized. The slides were prepared using two pieces of tape, placed parallel to form a narrow channel (Figure 3.3). The samples were placed one-by-one along the length of the channel, positioning each one with the same orientation, with respect to anterior/posterior and dorsal/ventral alignments. After positioning each sample, Kimwipe swab was used to absorb the excess glycerol solution surrounding the tissues to remove excess glycerol. Removal of the excess glycerol is important to prevent the formation of bubbles when sealing the mounted samples onto the slide under a cover slip. The cover slip was centered and placed over the channel containing the aligned samples and fixed semipermanently onto the slide by dabbing the corners of the cover slip with transparent nail polish. After drying, using a 200 $\mu$ l pipettor, glycerol solution 70% was dispensed slowly, into one end of the channel, so that the entire space of the channel and the area under the cover slip were filled by capillary action. Finally, the mounted samples were permanently sealed by applying nail polish along all four sides of the cover slip. Before photographing the

hybridized whole-mount tissues, the nail polish was left to dry sufficiently to avoid misplacing of the slip and thus distortion of the tissues in the channel. Images were taken under phase and contrast was achieved with magnification ranging from 10x to 60x in order to capture whole intact tissue as well as individual follicles/ovarioles.



**Figure 3.3.** Schematic diagram of post-hybridization mounting set-up (Juhn & James, 2012). In A the slide is viewed from the side, and in B the slide is viewed from above.

## 3.5 Bioinformatics

### 3.5.1 Primer designing for PCR

PCR primers are oligonucleotides that usually consist of 15–30 base pairs. The sequences of the primers are designed so that they could hybridize within the region of interest, and thus to be able to select and foremost amplify the region of interest, from a complex pool of DNA. In general, primers should contain 40–60% guanine and cytosine (G+C), since these base pairs create triple hydrogen bonds and thus, they increase the melting temperature ( $T_m$ ) of the PCR reaction. Ideally, both primers should have nearly the same melting temperature so that they both could anneal simultaneously through the annealing step. Also, care should be taken to avoid sequences that might produce internal secondary structure, such as loops and hair-pins. Furthermore, the 3'-end of the primers should not be complementary to each other, in order to avoid the production of primer-dimers. Formation of primer-dimers "occupy" primers from the reaction and results in the formation of a secondary unwanted reaction, that consumes not only primers but dNTPs and the enzyme as well.

Taking into account the previous, the Primer-Blast programme was used in order for primers to be designed, to amplify the genes of the OBPs that we were interested in. The cDNA sequence of each gene was exported from VectorBase® (Megy *et al.*, 2012) by simply searching for the gene and

applying exporting options according to preference. The search was occurred directly on *Anopheles gambiae* species gene set (VectorBase, <http://www.vectorbase.org>, *Anopheles gambiae* PEST AgamP4 AgamP4.2). The exported sequence was derived in FASTA format which is easy to copy and paste to the Primer-Blast programme placeholder. After setting the parameters, which were kept at their default state, except for the length of the PCR product size which was set higher than 600bp, and the “Organism” in which the Blast search to occur which was set to “*Anopheles gambiae* PEST”, the programme was left to run to retrieve the primer sets (Appendix C).

It is important to state that because of the presence of an *in vitro* transcription step, in the experiment process, the 5'-end of the reverse primer of each gene was annotated with the sequence of the T7 promoter. The related sequence (5'-TAATACGACTCACTATAGG) is used with high fidelity by the T7 RNA Pol Plus, and thus is commonly used in the *in vitro* transcription process.

### 3.5.2 Phylogenetic Analysis of the *Anopheles gambiae* OBPs

Phylogenetic analysis is the process in which molecular data, such as genes or peptides sequences, that derive from the same or different organisms, are clustered according to their similarity and thus, according to their between evolutionary distance. Reconstructing a phylogenetic tree from such molecular data involves multiple tasks such as: sequence alignment, selection of sequences and sites to analyse, tree building, rooting, plotting and printing. To execute these tasks, programmers have written different algorithms which help with each one of the steps above specifically. For example, MUSCLE (Edgar, 2004) and T-Coffee (Notredame *et al.*, 2000), are able to perform multiple sequence alignment, in contrast to PhyML (Guindon & Gascuel, 2003) and Neighbor-Joining (Saitu & Nei, 1987; Studier & Keppler, 1988), which are tree reconstruction methods. To bring together all these algorithms into user-friendly interfaces, many different phylogenetic softwares were conceived and invented, in which these algorithms are made easy to execute and manipulate. One basic programme that creates good phylogenetic trees is SeaView<sup>®</sup> (Gouy *et al.*, 2010). SeaView<sup>®</sup> has a basic user interface and a simple structure that enables the user to execute multiple and complex commands with the click of a button. The user sets the parameters he desires by choosing titles, and the only thing he has to consider about is importing the correct sequences, in the correct format.

SeaView<sup>®</sup> version 4 (<http://pbil.univ-lyon1.fr/software/seaview>) was used in order to reconstruct a phylogenetic tree amongst the 77 OBPs that were characterized in *Anopheles gambiae*. The peptide sequence of all the OBPs were exported from VectorBase<sup>®</sup> in FASTA format, and were imported into SeaView's working surface. First, the sequences had to be aligned so that homology regions from the proteins should come together and grouped. For the multiple sequence

alignment we used Clustalw (Larkin *et al.*, 2007). The algorithm was set to run with its default parameter values since they have been chosen by their authors to perform well in most cases (Gouy *et al.*, 2010). When special parameter values were needed, they were specified using SeaView's user interface and the algorithm was reused for subsequent alignment operations. After the Clustalw algorithm was done, the alignment results were used directly through the programme, to reconstruct a phylogenetic tree. The algorithm that was used is maximum-likelihood, which was provided by PhyML (Guindon & Gascuel, 2003), which consists an extension programme to SeaView<sup>®</sup>. The parameters again were left to their default state in order to retrieve the best results.

It is important to state that an additional multi-gene family, known as D7 salivary proteins which they are known to be distantly related to the arthropod OBP superfamily (Calvo *et al.*, 2002, 2006, 2009) was used, as an outgroup in the reconstruction of the OBPs phylogenetic tree. With the insertion of an outgroup in a phylogenetic tree, a root can be established which reveals the evolution process on the molecular data. Finally, SeaView<sup>®</sup> gives the opportunity to manipulate the information that comes out of the algorithms. The user can choose between: a squared tree in which the evolution distances are maintained, a cladogram in which the distances are removed, and a circular tree in which the evolution distances are maintained and displayed in a circular shape.

### **3.5.3 *In silico* analysis of OBP genes expression in *Anopheles gambiae***

In order to be able to perform an *in situ* hybridization, a research had to be done first by which the genes, which qualified to be selected as best candidates for the parameters we had set, had to be investigated. The collection of data stored in VectorBase was used to generate heat maps and visualise the expression data of the OBP genes. Also, experimental microarray data from different past experiments were used in order to collect the relative mean log expression of the OBPs in different stages and tissues. Specifically, for the expression in both sexes the data from Koutsos *et al.* (2007) and Marinotti *et al.* (2006) were used. For the expression of the genes in different PBM time frames, the expression data from Marinotti *et al.* (2006) were used. For the expression numbers in the adult tissues data from Baker *et al.* (2011) were edited, and for the expression in the embryo and during the embryonic development, the data were extracted from Goltsev *et al.* (2009). Finally, for the expression in mated females, data from Roger *et al.* (2009) were consulted, and for the data for two consecutive blood meals, the data within Marinotti *et al.* (2006) were used. Basic statistic values (e.g. median shifted values, mean log expression, standard deviation etc.) were calculated using algorithms from Microsoft Excel, Microsoft.

### 3.5.4 Candidate gene selection

Using the information that was obtained from the in silico analysis, heat maps were able to be created in which the OBPs expression data, in different stages and tissues, would be revealed through the colour changes of the heat map. A heat map is a graphical representation of data, where the individual values contained in a matrix are represented as colors. In molecular biology, heat maps are typically used to represent the level of expression of many genes across a number of comparable samples or situations. For a heat map to be created the data must be inserted and analysed in a specific programme that is suitable to execute multiple and complex algorithms. There are many available free softwares on the web that are able to complete such tasks and present nice and accurate heat maps. One such software is **R**. **R** is a free software programming language and also a software environment for statistical computing and graphics. **R** uses a command line interface; however, several graphical user interfaces are available for use with **R**. One feature of the programme is that it is able to display the results of a cluster analysis by permuting the rows and the columns of a matrix, and placing similar values near each other according to the clustering. Thus, joining cluster to the rows and columns of the data matrix, the user can create trees that referred to the matrix values.

**R** was used in order to create a heat map and visualise the expression profile of the 77 OBPs that were identified in *Anopheles gambiae*. The data were exported from Microsoft excel after they had been modified and edited as described above. Because **R** uses a command line interface, writing commands and executing simulations is a bit tricky, but with the right manual a great heat map with clustering amongst the values can be created. The first that had to be done was to make sure that the data were exported from Excel in the right format; **R** recognises Tab Delimited files with one column and one row as headers. The Excel sheet was adjusted as such and was exported to an easy location (D:\data). Then **R** was launched and the following was written in the command placeholder: `OBP<-read.table("D:/OBPdata.txt",header=T,sep="\t")` and pressed enter (the enter key submits the commands). This command loads the OPB data from the tab delimited ".txt" file and calls it "OBP". Next command was: `library(gplots)` This command loads the gplot package, which contains the heatmap.2 function. The next command was: `OBP=na.omit(OBP)` which throws out missing values from the dataset, and then we proceed with: `View(OBP)` which shows the data as a table in a second window, that is good to be brought to the front and checked. The final command was: `heatmap.2(as.matrix(OBP[3:53]), labRow=OBP$Gene.ID, col=redgreen(75), density.info="none", trace="none", cexRow=0.4, cexCol=0.5, key=T, keysize=0.85, symkey=F, Colv = FALSE, dendrogram = c("row"), scale="row", srtCol=70)` which sets

many parameters for the heat map and presents it as well. On the next paragraph some of the commands shown above are explained.

`heatmap.2(as.matrix(OBP[2:53]))` executes the `heatmap.2` function, tells it to use the "OBP" data as a matrix/table, where columns from 2:55 contain the data; column 1 contains the labels  
`labRow=OBP$Gene.ID` identifies where the row labels are kept; here the row labels are under the column "Gene ID". Note that a "." is needed to separate spaces because R doesn't recognise spaces.

`col=redgreen(75)` uses a range of 75 colours from red to green for the heatmap

`density.info="none"` chooses whether or not to superimpose a histogram; histogram can be removed in the scale bar with "histogram" instead of "none"

`trace="none"` removes lines around the coloured blocks

`cexRow=0.4, cexCol=0.5` chooses the size of row and column labels

`key=T, keysize=0.85` shows the key and makes it relatively small

`symkey=F`, makes the key symmetric around 0

`Colv = FALSE` ensures that the clustering is done by rows only, not the columns

`dendrogram = c("row")` blocks the creation of a dendrogram for the columns

`scale="row"` scale each of the rows rather than scaling the colours by the whole data

`srtCol=70`) puts the column labels at an angle of 70 degrees instead of the default 90

## **APPENDIX A: Buffers & Solutions**

### **PBS (Phosphate-Buffered Saline)**

1xPBS was made from 10X concentrated stock

### **PBT (Phosphate Buffered Tween-20)**

0.1% Tween-20 in PBS buffer

### **STFS (Soft Tissue Fixation Solution)**

4% Formaldehyde, prepared fresh from 37% stock  
Brought to final volume with PBT

### **Hyb buffer (Hybridization buffer)**

64ml autoclaved double distilled water added to one plastic bottle of "DIG Easy Hyb Granules"  
dissolved by stirring immediately for 5 minutes at 37°C

### **Blocking Solution**

5% of sheep serum  
1% of Western Blocking Reagent  
Brought to final volume with PBT

### **AP Staining Buffer (Alkaline Phosphatase staining buffer)**

100mM Tris-HCl, pH 9.5  
50mM MgCl<sub>2</sub>  
100mM NaCl  
0.1% Tween20  
Brought to final volume with double distilled water

### **AP Staining Solution (Alkaline Phosphatase staining solution) - NBT/BCIP staining solution (5-Bromo-4-Chloro-3-Indolyl Phosphate / Nitro Blue Tetrazolium chloride)**

One "NBT/BCIP ready-to-use" tablet was dissolved in 10ml double distilled water



## APPENDIX B: Primer Sets used for PCRs

| Gene ID    | Gene Name | Primer Name | Primer Sequence with <u>T7 promoter</u> (5'-3')   | T <sub>m</sub> (°C) | cDNA amplicon (bp) | gDNA amplicon (bp) |
|------------|-----------|-------------|---|---------------------|--------------------|--------------------|
| AGAP002189 | OBP 14    | pr02189-F   | GCAAGCCAGCTTTCCGATTC                              | 60,18               | 626                | 626                |
|            |           | pr02189-R   | <u>TAATACGACTCACTATAGG</u> CGACAGTGCCTCACCTTAGT   | 60,04               |                    |                    |
| AGAP002025 | OBP 11    | pr02025-F   | CAGCAGGACATTCGCAGGA                               | 60,08               | 570                | 570                |
|            |           | pr02025-R   | <u>TAATACGACTCACTATAGG</u> TCGTGGAAGTGGTAGCCAGA   | 60,2                |                    |                    |
| AGAP002190 | OBP 39    | pr02190-F   | AACTTCCGCGACATTCGGT                               | 60,01               | 719                | 719                |
|            |           | pr02190-R   | <u>TAATACGACTCACTATAGG</u> ACGCCTGAGCATTTTTGTCC   | 59,4                |                    |                    |
| AGAP002191 | OBP 40    | pr02191-F   | GGCAAGCTCTACCAGAAAGC                              | 58,91               | 850                | 850                |
|            |           | pr02191-R   | <u>TAATACGACTCACTATAGG</u> ATACCCATATTGCCCCGAA    | 58,55               |                    |                    |
| AGAP010648 | OBP 44    | pr10648-F   | GCAGCTGGTGTGTATAGTTGTG                            | 59,58               | 929                | 1115               |
|            |           | pr10648-R   | <u>TAATACGACTCACTATAGG</u> AGCCAGGATACGACAGGTTG   | 59,46               |                    |                    |
| AGAP010650 | OBP 45    | pr10650-F   | CTTACGGCGATCATGGCACT                              | 60,53               | 914                | 914                |
|            |           | pr10650-R   | <u>TAATACGACTCACTATAGG</u> TTGAACTCGGTCGTGGAAGG   | 59,97               |                    |                    |
| AGAP002188 | OBP 12    | pr02188-F   | CAGCATGGCACCTGTGAGA                               | 60                  | 845                | 845                |
|            |           | pr02188-R   | <u>TAATACGACTCACTATAGG</u> TGGGTTTTCGGTGATAGCTTTG | 59,18               |                    |                    |
| AGAP005182 | OBP 41    | pr05182-F   | GGTACTGGCCTTGGAAC                                 | 60                  | 808                | 808                |
|            |           | pr05182-R   | <u>TAATACGACTCACTATAGG</u> TAGCCGATATCACGTTGCG    | 60,32               |                    |                    |
| AGAP009065 | OBP 42    | pr09065-F   | TCACTTGGCCTAACGGCTTG                              | 60,61               | 819                | 819                |
|            |           | pr09065-R   | <u>TAATACGACTCACTATAGG</u> AGCTGTCTCCACTTCTGGTTG  | 59,93               |                    |                    |
| AGAP010649 | OBP 31    | pr10649-F   | ATTTGCGTGTGGCGTTGAT                               | 59,69               | 908                | 1301               |
|            |           | pr10649-R   | <u>TAATACGACTCACTATAGG</u> CGTCCAAAGTGTCTGCCTTG   | 60                  |                    |                    |
| AGAP012319 | OBP 18    | pr12319-F   | TTACATTAAGCGCGCCGAC                               | 59                  | 515                | 907                |
|            |           | pr12319-R   | <u>TAATACGACTCACTATAGG</u> TCACTGCTGCATGTTCCCTC   | 58,47               |                    |                    |
| AGAP011647 | OBP 30    | pr11647-F   | ACCACTACGAGGACAACACAC                             | 59,93               | 812                | 812                |
|            |           | pr11647-R   | <u>TAATACGACTCACTATAGG</u> CAGCTGGAAAACATTCGCGT   | 59,76               |                    |                    |
| AGAP012320 | OBP 25    | pr12320-F   | GATCTGCTGGATGCACTCG                               | 59,06               | 700                | 836                |
|            |           | pr12320-R   | <u>TAATACGACTCACTATAGG</u> TGATGAGAACTACCCGCACG   | 59,83               |                    |                    |
| AGAP009402 | OBP 43    | pr09402-F   | GCTATTACTGTTGGCGTGGC                              | 59,62               | 908                | 908                |
|            |           | pr09402-R   | <u>TAATACGACTCACTATAGG</u> CCCTATCGCTCGGTTTCCTT   | 59,54               |                    |                    |

**Table 1.** The primer sets used to amplify the OBP genes.

| Gene ID    | Gene Name | Primer Name | Primer Sequence with <u>T7 promoter</u> (5'-3') | T <sub>m</sub> (°C) | cDNA amplicon (bp) | gDNA amplicon (bp) |
|------------|-----------|-------------|---|---------------------|--------------------|--------------------|
| AGAP010648 | OBP 44    | pr10648-F2  | CTGATTGTCGCCGAGTGTGA                            | 59,58               | 929                | 1115               |
|            |           | pr10648-R2  | <u>TAATACGACTCACTATAGGCGCCGTAATGATACCGGATGT</u> | 59,46               |                    |                    |
| AGAP010650 | OBP 45    | pr10650-F2  | CCGCATCTTACCCTGTCGAA                            | 60,53               | 914                | 914                |
|            |           | pr10650-R2  | <u>TAATACGACTCACTATAGGCCTCGCACACCGTCATCTT</u>   | 59,97               |                    |                    |
| AGAP010649 | OBP31     | pr10649-F2  | CAACGACACCAAGGGACTGA                            | 59,69               | 908                | 1301               |
|            |           | pr10649-R2  | <u>TAATACGACTCACTATAGGGGGACGCCTCAGCAATGATA</u>  | 60,00               |                    |                    |
| AGAP012319 | OBP18     | pr12319-F2  | TACTGAAGGGGGCAGGAATG                            | 59,00               | 515                | 907                |
|            |           | pr12319-R2  | <u>TAATACGACTCACTATAGGCCTTCGGCACCTTGTGATTG</u>  | 58,47               |                    |                    |

**Table 2.** The primer sets used to amplify the OBP genes that did not amplify with the first PCR.

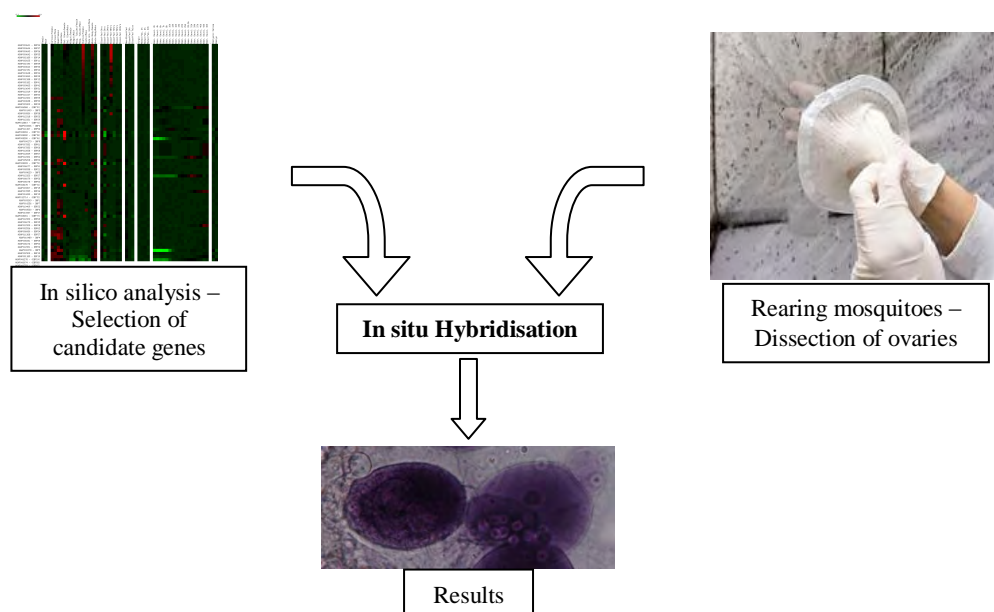
# **CHAPTER 4**

## **RESULTS**

## CHAPTER 4: RESULTS

### 4.1 Project Strategy

The experiment was design having in mind the questions that needed to be answered. The procedure followed a series of experiments and methods outlined in Figure 4.1. As it was mentioned, the project has combined the information derived from existing data within the literature, and new data from in situ Hybridization experiments, in order to create a complete picture of the ovary specific OBPs and their possible role in key processes like the oogenesis and egg development.



**Figure 4.1** Outline of the project strategy. The project starts with the selection of the candidate genes through in silico analysis. Parallel to the analysis mosquito tissues are dissected. Finally, the tissues are used to perform an in situ hybridisation on the genes chosen by the in silico analysis.

The first step of the project strategy included the bioinformatics analysis of the mosquito OBP family, so that the different groups could be studied and compared in order to obtain a wider idea of the family's characteristics. Then, the information extracted from the *in silico* analysis was used in order to list the OBPs according to their expression within the ovaries after the stimulative blood meal. Subsequently, after the revealing of the OBPs that are up-regulated in the ovary after a blood meal, and therefore the selection of the candidate genes, the relevant genomic sequences were used as matrixes to amplify cDNA and create DIG-labeled RNA probes. The probes were used on the second part of the project which was the *in situ* hybridisation experiment. Using the probes it was possible to detect the mRNA of the OBPs within the ovary tissue and even within single ovarioles. The location of the OBPs mRNA provides important and useful information on the expression profile

of those genes. Through the stained tissue it was possible to determine if the OBPs are expressed within the somatic or the germ line of the developing egg. As it was made clear in the introduction, a somatically expressed gene which is necessary for female fertility is the ideal scenario for the development of a HEG against that gene. Through the *in situ* hybridisation though, it is not easy to clear out if the OBP genes are essential for female fertility. This clarification is for other studies to explore.

During this project, fourteen candidate genes were selected through the *in silico* analysis and the expression data exported from VectorBase. For the ten of them, an RNA probe was synthesised and an *in situ* hybridisation was attempted but only for the six of them it was possible to get an idea of the location of the mRNA of the OBP within the ovary.

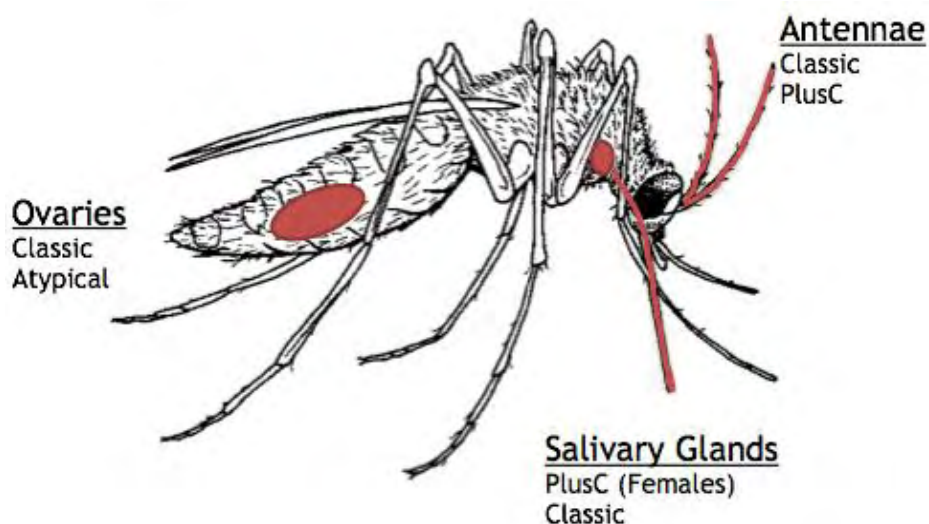
## 4.2 In-Silico Analysis of the OBP Family

### 4.2.1 Listing and grouping of the OBPs

Information on the *Anopheles gambiae* odorant binding proteins were extracted from previous projects that were analysing the entire family and its members in mosquitoes (Manoharan *et al.*, 2013; Xu *et al.*, 2003; Viera & Rozas, 2011). The most recent study listed 69 odorant binding proteins in the *Anopheles gambiae* genome and eight D7 salivary proteins which referred as OBPs but noted with the suffix "D7". The 72 OBPs found in *A.gambiae*, were classified into three of the four families of the OBPs (Classic, PlusC & Atypical) since the MinusC group - the fourth group of OBPs - has never been identified to date in mosquito genomes. The OBPs were listed in the studies, and a specific number-name was given to each one of them within the range of 1 to 77 which follows the prefix "Agam". Some of the OBP names are in agreement with the names within VectorBase database whereas others have not yet been named in VectorBase; furthermore, the different studies suggest different names for a small fraction of the OBPs. Even though there is a small disagreement between the OBP numbers, the references to the gene names of each OBP is always the same amongst the studies and VectorBase, since there is full clarification as the genome annotation concerns. All the genes in the *Anopheles gambiae* genome carry a prefix (AGAP) and a six-digit number, thus the OBPs are named appropriate. Some novel proteins which were recently included in the large family of the OBPs, including some D7 salivary proteins, were given new numbers in the current project according to the list adopted from Manoharan *et al.*, (2013). It is important to note that the OBPs mentioned through this study were adopted from the numbering in Table 3 (Appendix C), so the OBP number is correlated to the gene number mentioned, which to date is in agreement with VectorBase.

In Appendix C, the 77 OBPs are presented in ascending order according to their protein name. The numbering is extending from OBP1 to OBP79 since no protein was given the numbers OBP16 and OBP17 in either study. The type of the OBPs is also referred, except for the D7 OBPs which they consists of OBP-like domains. The gene name correlated to each OBP is the current shown in VectorBase. OBP30, OBP31, OBP48 and OBPs 58-72 were named after suggestion in Manoharan *et al.*, (2013) since VectorBase has yet to label the respective genes. Also, OBP24 and the D7 salivary OBPs numbering, was suggested through this project since there is no numbering neither in VectorBase nor in Manoharan *et al.*, (2013).

The information extracted from the literature help to the separation of the different types of OBPs depending on their expression profile. BioMart, a searching tool within VectorBase, helped the separation of the *Anopheles gambiae* OBPs according to the tissue they are expressed. Correlation between the different sources within the web helped to the creation of a mosquito anatomy map which shows the basic types of OBPs and the most preferred tissue of expression (Figure 4.2).

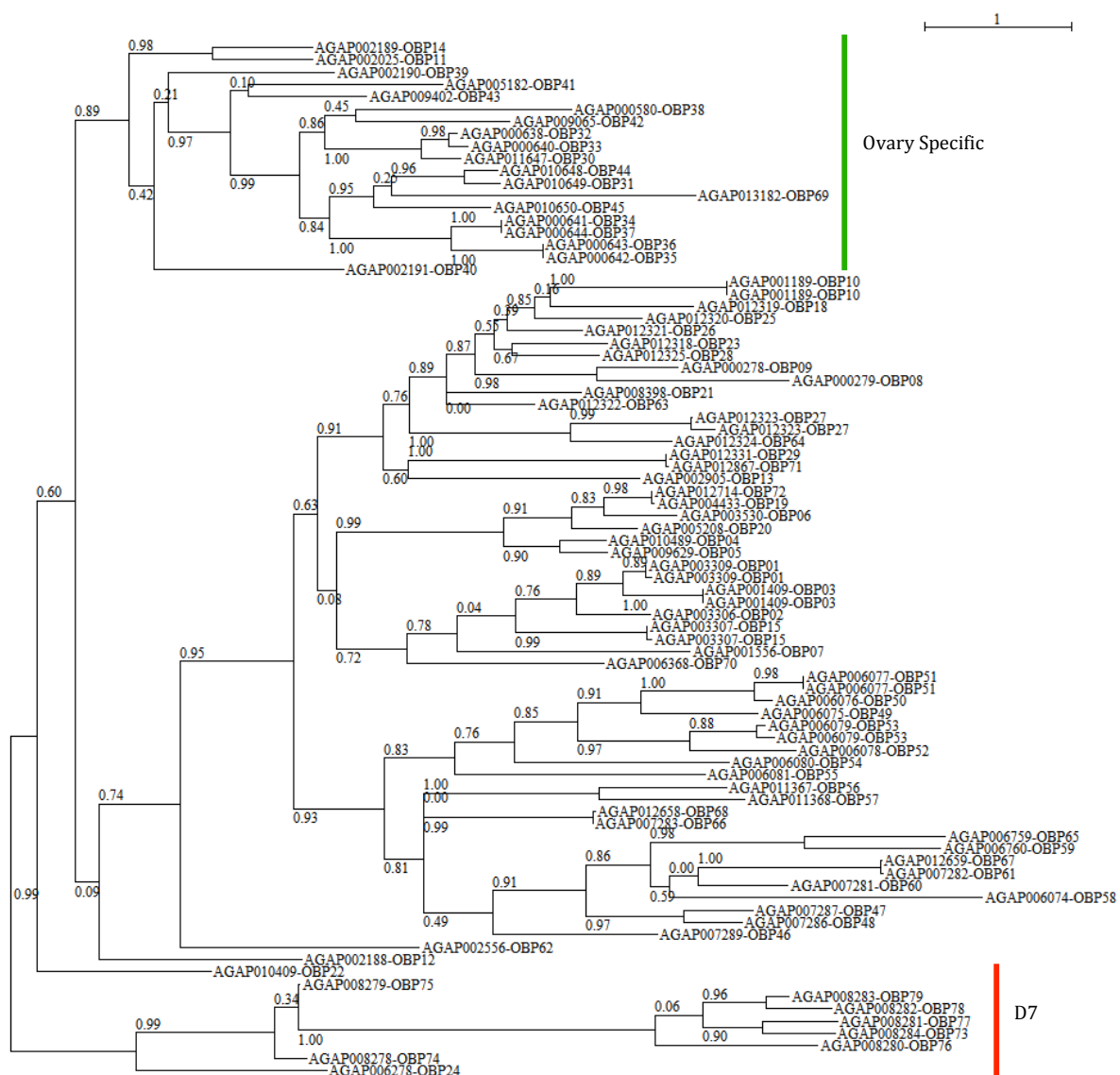


**Figure 4.2.** Tissue localisation of the groups of the OBPs. The Atypical type is a characteristic of the ovary specific OBPs. Note that the salivary glands specific PlusC group is a feature of the female mosquitoes. It is suggested that they have a function in blood feeding.

It was obvious from the different data that Atypical OBPs is a type tightly bound to the ovaries. To remember, Atypical OBPs are considered as “dimer” OBPs due to their double classic OBP domain. Amongst the Atypical OBPs that expressed in the ovaries, there are some Classic OBPs which, as it will proved later, they are amongst the proteins that expressed highly in the ovaries after a blood meal. Also, from the information in the web it was possible to detect some PlusC OBPs that are exclusively expressed in the salivary gland of the female mosquito. Finally, the olfactory system’s OBPs, which usually are called “real OBPs” are members of the Classic and the PlusC groups and they show now different expression amongst the two sexes.

### 4.2.2 Clustering the OBPs

In order to reveal the evolutionary events which led to the separation of the ovary specific OBPs, a phylogenetic tree was created (Figure 4.3). The peptide sequences of the proteins were used to perform a multiple sequence alignment and group the proteins according to their evolutionary distance. The sequences were exported from VectoBase and imported to SeaView programme, where the phylogenesis took place. The D7 salivary OBPs were used as an outgroup, since they are the most diverged within the OBP family. Their cluster used to draw the root of the tree and thus reveal the correct evolutionary events which took place amongst the OBP family. It is important to state that a small number of OBPs present more than one peptide sequences due to alternative splicing that takes place within the cell. The Bootstrap constant shows the reliability of each branch and therefore it is possible to detect branches' division accuracy.



**Figure 4.3.** Phylogenetic analysis of the peptide sequence of the OBPs. The phylogenesis was made using maximum-likelihood approach. The ovary specific OBPs are grouped together and they form a distinct group.

From the peptide sequence clustering, much information was derived according to the OBP family. It was made clear that the ovary specific OBPs are similar as their sequence concerns, making them different from the other OBPs, not only at the expression preference but also at the origin of their sequence. It is obvious that during the evolution, the ovary specific OBPs were divided from the main OBP family, acquiring the functions they have today, which are correlated with the blood meal on the female mosquito. Finally, the D7 branch includes all the eight members discovered so far.

#### 4.2.3 Selection of *Anopheles gambiae* candidate genes

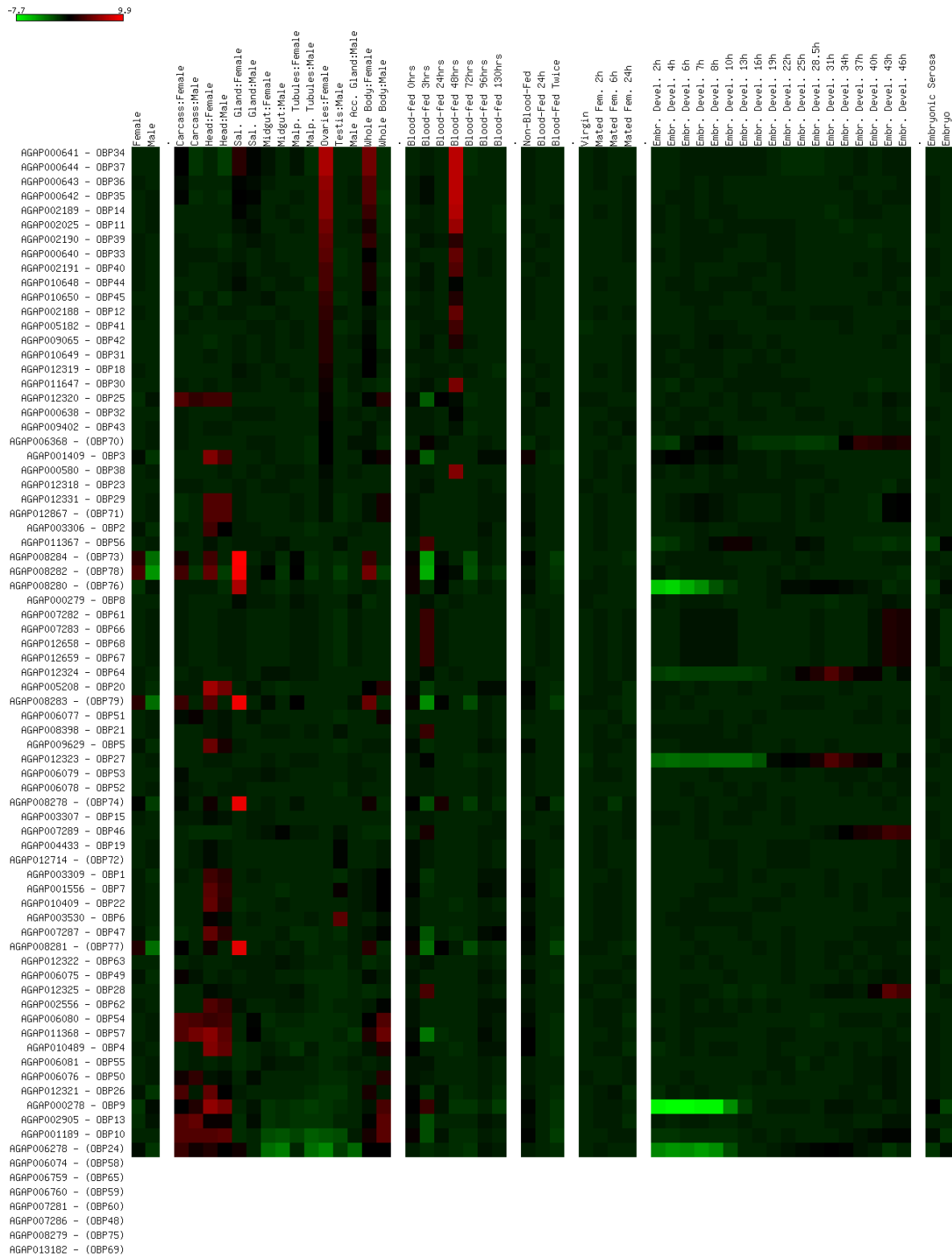
In order to perform a successful *in situ* hybridisation it was necessary to select a group of genes which corresponded to the needs and aims of the project. The appropriate genes to create a labeled RNA probe were those which they are up-regulated after a blood meal. Previous whole genome analysis that was accomplished by the lab using data from MozAtlas (Baker *et al.*, 2011), VectorBase (Megy *et al.*, 2012) and FlyBase (St. Pierre *et al.*, 2014), showed that after a blood meal the ovary specific OBPs are amongst the genes that are highly up-regulated. More specific, it has been shown that 11 OBPs are amongst the 30 highly expressed genes within the whole genome of the mosquito. Also what was striking about this is the fact that these genes are not found in any other tissue rather than the ovaries!

In order to visualise the expression data for each OBP and reveal the genes that are highly expressed in the ovaries, a heat map was created in which each expression value, in this case a median shifted one, is translated into a colour (Figure 4.4). The differences between the values are shown by different exposures of the respective colour. The heat map was generated using a collection of data stored in VectorBase. Because this was a wider analysis, the heat map contained expression values not only for the ovaries but for other stages and tissues as well. Specifically, the heat map contained expression values for the OBPs in both sexes (Koutsos *et al.*, 2007; Marinotti *et al.*, 2006). Also there were included values from genes in different time frames PBM (Marinotti *et al.*, 2006) as well as expression data in the adult tissues (Baker *et al.*, 2011). There were also values from the expression in the embryo and during the embryonic development (Goltsev *et al.*, 2009), and finally there were included expression values for mated females (Roger *et al.*, 2009) and for females that took two consecutive blood meals (Marinotti *et al.*, 2006).

The heat map was created using the median-shifted expression values from these experiments, which they were extracted as a .txt file from VectorBase and imported in Microsoft Excel. The Excel sheet was edited appropriate before the import of the data in the heat map programme. The values of each OBP gene were listed in descending order according to the expression in the ovaries first, and then according to the expression that was observed 48 hours post blood meal, since this was the time

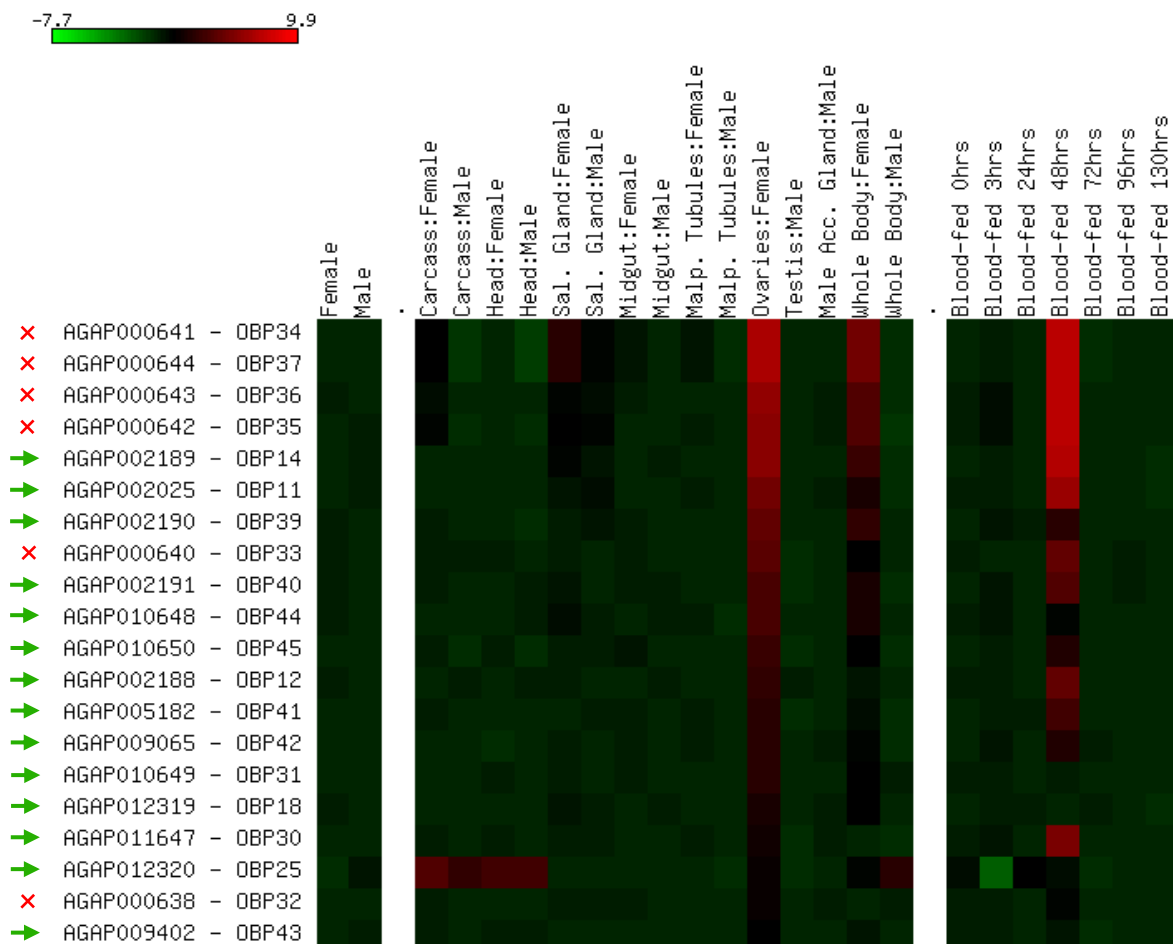


frame that was most important. The colours of the heat map were set so that the up-regulation was signaled with a brighter red whereas the down-regulation was signaled with a brighter green. The baseline of the expression was illustrated with black colour.



**Figure 4.4.** Heat map expression of the OBPs according to their upregulation within the ovaries and at 48 hours PBM. The highly expressed OBPs are at the top of the heat map. The colours of the heat map were set so that the up-regulation was signalled with a brighter red whereas the down-regulation was signalled with a brighter green. The baseline of the expression was illustrated with black colour. For the seven OBPs at the end there are no expression data to date.

From the heat map it was obvious which genes are up-regulated after a blood meal in the ovaries and at 48 hours PBM. The list with the candidate OBPs was looked up closely and the candidate genes were subjected to further investigation about their actual expression and about their location within the genome. It is reminded that according to the list for the HEG parameters (CHAPTER 1) we prefer the gene to be located on an autosomal chromosome.



**Figure 4.5.** A close up of the heat map in Figure 4.3. The OBPs which are located in a sex-chromosome are noted with a red X. The rest OBPs were chosen to perform the *in situ* hybridisation.

From the heat map a first list was available in order to begin the specific research for each of the OBP genes. First of all the location in the genome was the first to be discovered. Sadly, the most up-regulated genes (OBP34, OBP37, OBP36 and OBP35) are located on the X chromosome. That was kind of expectable since these genes are expressed in the ovaries and are female specific. The chromosome location was revealed for all the OBPs and it is mentioned in Appendix C. Furthermore the expression charts of the remnant OBPs were observed to decide which of them were suitable to proceed to the next step of the project, the *in situ* hybridisation. The expression charts used are outlined in Appendix D.

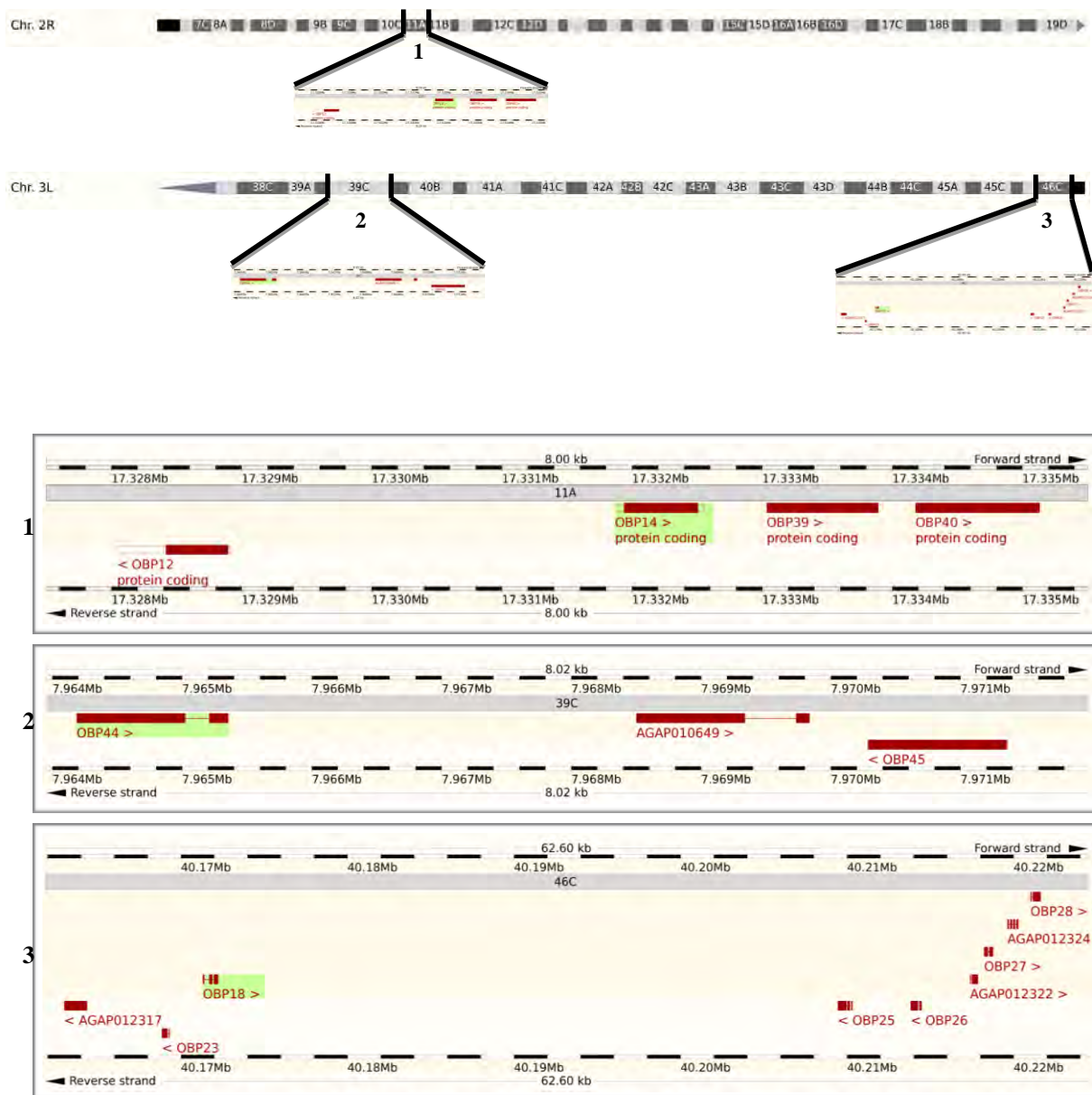
From the charts it was made clear that OBP31 and OBP18 are not the best candidates to include in the *in situ* hybridisation. For those OBPs there was no significant expression in any specific time point. Someone would expect for OBP25 and OBP43 to be excluded but it was decided to keep all of them after all in the list. So, the final list included OBP14, OBP11, OBP39, OBP40, OBP44, OBP45, OBP12, OBP41, OBP42, OBP31, OBP25 and OBP43. The OBPs were listed descending as their importance concerns (Table 4) since the list kept the order displayed in the heat map. Finally, since the candidate genes were sorted the project was ready to move to the next phase, the *in situ* hybridisation.

| Gene ID            | OBP Type | Chromosome |
|--------------------|----------|------------|
| AGAP002189 - OBP14 | Classic  | 2R         |
| AGAP002025 - OBP11 | Classic  | 2R         |
| AGAP002190 - OBP39 | Atypical | 2R         |
| AGAP002191 - OBP40 | Atypical | 2R         |
| AGAP010648 - OBP44 | Atypical | 3L         |
| AGAP010650 - OBP45 | Atypical | 3L         |
| AGAP002188 - OBP12 | Classic  | 2R         |
| AGAP005182 - OBP41 | Atypical | 2L         |
| AGAP009065 - OBP42 | Atypical | 3R         |
| AGAP010649 - OBP31 | Atypical | 3L         |
| AGAP012319 - OBP18 | Classic  | 3L         |
| AGAP011647 - OBP30 | Atypical | 3L         |
| AGAP012320 - OBP25 | Classic  | 3L         |
| AGAP009402 - OBP43 | Atypical | 3R         |

**Table 4.** The candidates OBPs in descending order according to their expression in the ovaries after 48 hours of blood meal.

#### 4.2.4 Genome clusters

While the search for the genome location of the OBPs was undergoing it was interesting that some of the OBPs were found to be grouped together in the genome! (Figure 4.6). This information was very useful since future studies which aim at the production of knock outs of the OBPs with this information would be able to design the experiment so that they could delete the whole clusters. Also, the deletion of whole clusters of the OBPs it will be prove useful if the OBPs appear to be redundant after all, ergo if they present cross functions and thus the deletion of one candidate would give no phenotype.



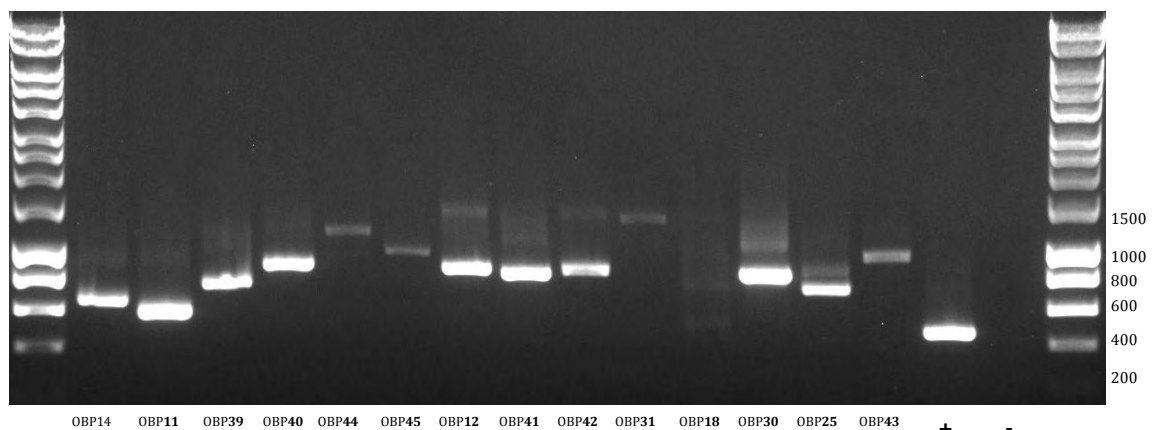
**Figure 4.6.** Some of the candidates OBPs were found to be grouped together in the genome. In case of redundancy of their function, deleting the clusters would bypass the problem.

From the OBPs that were amongst the best candidates it was revealed that some of them were grouped, along with other OBPs, into 3 obvious clusters within the genome. The first cluster was located at chromosome 2R and includes OBP12, OBP14, OBP39 and OBP40. The second cluster was located at chromosome 3L and includes OBP44, OBP31 and OBP45. Finally the third cluster was located also at chromosome 3L and includes OBP23, OBP18, OBP25, OBP26, OBP63, OBP27, OBP64 and OBP28. Some of the OBPs that are included in the clusters above are not up-regulated in the ovaries after blood meal; nevertheless, deletion of the entire clusters should give new insights on the OBPs functions since the redundancy would be at minimum.

## 4.3 In-Situ Hybridisation Experiment

### 4.3.1 PCR amplification of DNA template

For the synthesis of a labeled RNA probe it is necessary first to amplify the desired region of the OBP gene. Since the purpose was to determine the location of the mRNA within the tissue, the amplification had to be performed on the cDNA of the OBPs. The primers (CHAPTER 3) were designed to bind in the exons of each OBP gene so that they could amplify the cDNA. The cDNA was provided by Andrew Hammond (PhD student, Imperial College London) and it was cDNA from blood fed mosquitoes that were harvested 44 and 48 hours PBM. The PCR reaction was set up as described in CHAPTER 3; the Fast Cycling PCR kit was used to amplify the desired region. For this PCR the primers that were used are mentioned in Appendix B. After the PCR was finished, an agarose gel electrophoresis was done in order to verify the PCR reaction. The gel was loaded with 5 $\mu$ l of each product (Figure 4.7), in order to visualise the amplicon of each reaction. The gel was visualised under UV light in a suitable machine.



**Figure 4.7.** PCR amplicons of the PCR which was occurred in order to amplify the template for the *in vitro* transcription.

According to the gel, the PCR was successful for some of the OBPs. OBP14, OBP11, OBP39, OBP40, OBP12, OBP41, OBP42, OBP30, OBP25 and OBP43 gave the expected fragments for amplification from cDNA. On the contrary, OBP44, OBP45, OBP31 and OBP18 didn't present any significant amplicons after the PCR reaction. In fact, only OBP31 and OBP45 gave the expected band for amplification from genomic DNA (gDNA). OBP44 gave irrelevant band size whereas OBP18 didn't give a clear band at all. Also it was observed that the PCR reaction in the tube of OBP12, OBP30 and OBP25 gave a second fainter band that in some cases (OBP25) is the expected band for amplification from gDNA. The band on the other three OBPs was irrelevant and unacceptable; even though, we proceed with RNA *in vitro* transcription since there was a bigger amount of the correct band and because there is already a step of verification after the *in vitro* transcription where it will be clear if

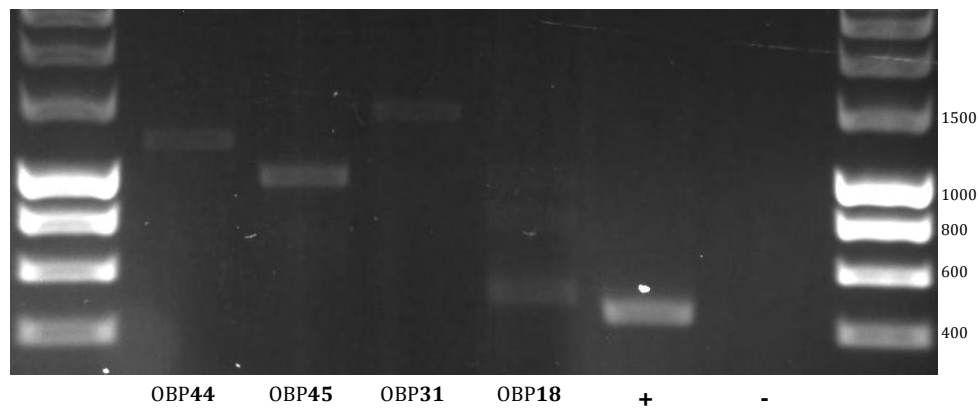
the correct RNA will be created. All the expected bands from the current primers are mentioned in Table 1 (Appendix B).

The positive control (+) that was used was able to reveal if the amplification was done from gDNA or cDNA. The primers that were used for the positive control were called S7, and they amplify an area within the genome which has different sizes according to the nature of the template material. If the template is cDNA the primers give a product that it is 430bp whereas if the template is gDNA the primers give a product that it is 609bp. So, as the current PCR concerns, we can say that there was no genomic DNA in the template used for the PCR, and even if there was, due to the fact that some OBPs gave the expected band for amplification from gDNA, it was in very small quantities.

The negative control (-) that was used did not contain any template material. The primers that were used were the same as the positive control (S7) and thus they are expected to present one of the former two bands if there was any template in the tube. The negative control was included in this PCR to reveal if there are any contaminations in the process of the PCR, since the template is under the same manipulations as the other tubes. A clean negative control means that the PCR process was executed with caution and that all the necessary measures for the avoidance of contaminations were taken.

Finally, after the PCR reaction and the gel validation a PCR purification was done to ensure a clean PCR amplicon, free of salts and enzymes. The PCR purification kit that was used was the QiaGen PCR purification kit which comes always with a protocol. When the PCR product was purified, the concentration in each tube was measured by NanoDrop, which uses a spectrophotometric method to "count" the DNA molecules within 1 $\mu$ l of sample. The concentrations were determined and they were variable from 20 to 100ng/ $\mu$ l. This was more than enough since for the in vitro RNA transcription an amount of 200ng is needed.

It is good to mention that a second PCR was designed in order to try and amplify the OBPs cDNA that did not amplify with the first PCR. To do that, new primers were designed in order to target a different fragment on the cDNA of the OBPs that did not amplify (Appendix B, Table2). The second PCR was an attempt to amplify products from OBP44, OBP45, OBP31 and OBP18. The same procedure was followed as with the first PCR and the results were loaded on a gel (Figure 4.8).



**Figure 4.8.** Repetitive PCR for the OBPs that did not amplify with the first attempt.

According to the gel the PCR was successful since the positive control gave a band around 430bp, which means that the amplification from cDNA was possible. Also, the absence of bands on the negative control suggests that there was no contamination. The amplification reaction in each tube gave faint products that in some cases were the desirable. The amplification for OBP44 gave a band around the gDNA fragment which was 1115bp. The amplification for OBP45 gave a band around 1000bp that was close to the desirable band of 914bp. The amplification for OBP31 gave a band that was close to the gDNA band and finally, the amplification for OBP18 gave the desirable band around 515bp. Although for OBP18 the desirable band was accessible, even faint, was not included in the in vitro transcription process.

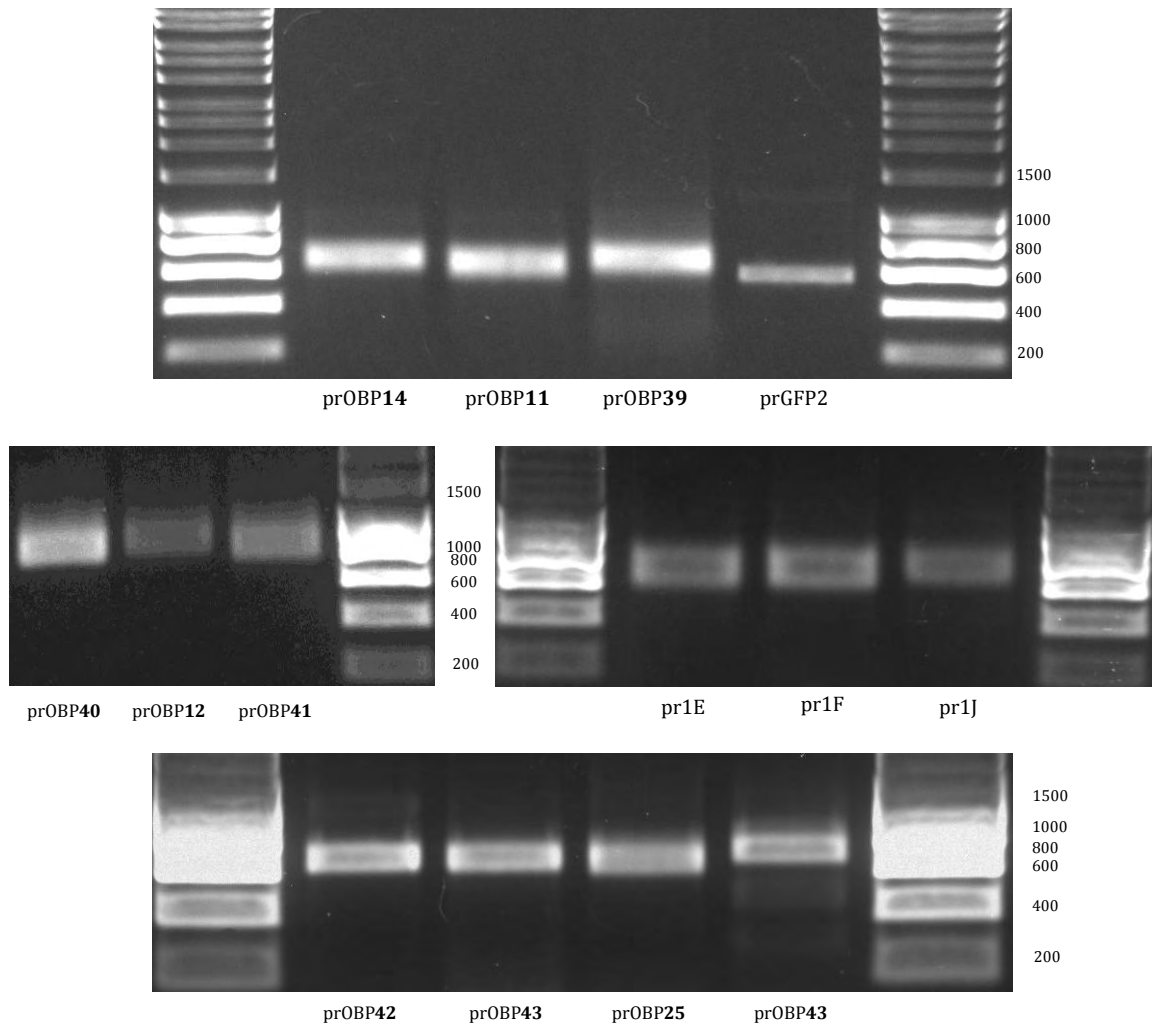
#### 4.3.2 In vitro transcription

The purified amplicons from the PCR reaction were used to create the labeled RNA probes that are essential for the in situ hybridization. For the in vitro transcription was used the template from the OBPs that gave a clear desirable band during the PCR. So, the templates of OBP14, OBP11, OBP39, OBP40, OBP12, OBP41, OBP42, OBP30, OBP25 and OBP43 were used to create labeled RNA probes. The probes, as described in CHAPTER 3 there were created using the T7 Pol Plus system and the DIG-labeled mix of nucleotides. After the in vitro transcription followed a LiCl precipitation in order to precipitate and purify the labeled RNA. After that, the concentration was determined and the RNA was stored at -20°C.

#### 4.3.3 RNA probe validation - positive & negative controls

Prior to the in situ hybridisation the probes “pr” had to be validated using an agarose gel electrophoresis (Figure 4.9). For each probe an amount of 0.5µl of sample was loaded to the gel in order to observe the concentration given by NanoDrop. This method to validate the concentration was not the correct one since it is not possible to correlate each concentration with a different brightness. The correct method was to load on the gel the same concentration of each RNA probe so

that any changes in the brightness of the bands, which they are going to be the same that way, would be visible and detectable.

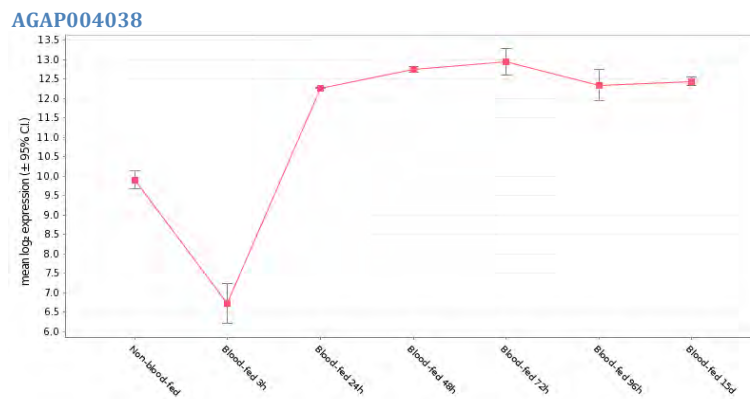


**Figure 4.9.** The gel electrophoreses which have validated the RNA probes.

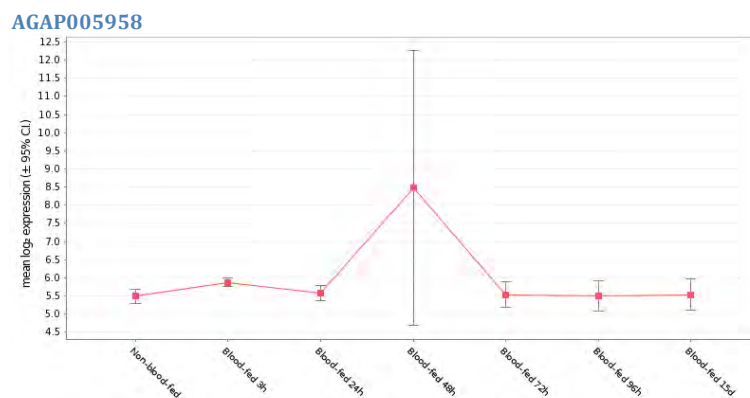
From the gel pictures of the probes it was obvious that there were enough quantities of each probe to proceed to the *in situ* hybridisation. It is worth mentioned that because the molecules that were under detection on the gels were RNA molecules, the bands were expected to be less bright than usual. RNA consists of a single helix and thus the incorporation of the EtBr within the molecule is much less frequent and effective. Also, the RNA molecules are highly sensitive and thus a portion was degraded during the gel electrophoresis accelerating the fading of the band. Finally, it is essential to mention that the ladder in these gels was included in order to give an idea of the size of the RNA probes rather than an accurate measurement. That is because the ladders consists of DNA and thus they couldn't be compared with the RNA molecules that were under detection.



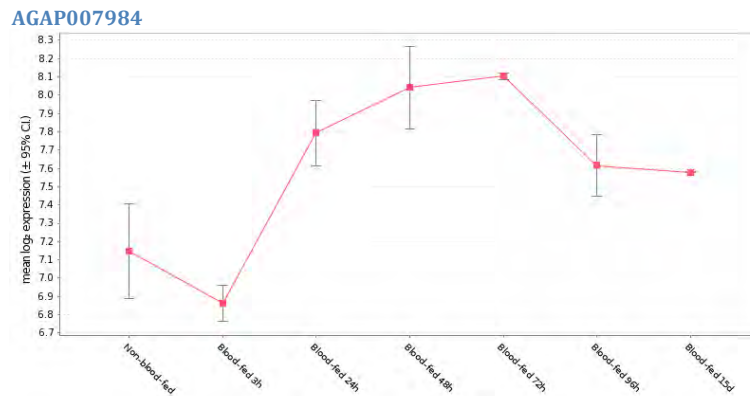
For the *in situ* hybridisation it was necessary to include alongside with the OBP probes, one positive control and one negative control. For the positive control, a choice had to be made alongside three different probes which were used in past *in situ* experiments (pr1E, pr1F and pr1J). The probes were correlated to three different genes and showed to work in the past *in situ* experiment. The probes were checked for their stability and one of them was chosen to participate in the experiment. The probe that was chosen was pr1F because it was a probe of the gene AGA004038 which was expressed also at 48 hours PBM (Figure 4.10). The past *in situ* hybridisation was done on ovaries which were harvested at 24 hours PBM, thus, detecting the gene at 48 hours PBM it would give extra information for this gene, information that was needed from members of the lab. The other two probes pr1E and pr1J were probes of the genes AGAP007984 and AGAP005958 respectively, which were not highly expressed in 48 hours PBM (Figure 4.11 and Figure 4.12).



**Figure 4.10.** The expression of the gene AGAP004038 at time frames PBM.



**Figure 4.11.** The expression of the gene AGAP005958 at time frames PBM.



**Figure 4.12.** The expression of the gene AGAP007984 at time frames PBM.

For the negative control a probe that was also used in past *in situ* experiments within the lab was used. The probe was an RNA probe that was detecting the GFP gene (green fluorescent protein). Since this protein was not expressed in the tissues used for the *in situ* hybridisation, the probe (prGFP2) should not bind to anything within the cells and thus it should not give any signal.

The positive and the negative controls in an *in situ* hybridisation experiment are essential since they validate the method by checking for different things. The positive control is crucial because it shows if the whole of the process was successful; a signal in the positive control means that the *in situ* hybridisation was done according to the protocols. By giving positive results, the positive control shows that first of all there is mRNA within the sample. Furthermore, it shows that there was used a nucleotide that was labeled by DIG (digoxigenin) and that the antibody used to detect the DIG was active and bound to the DIG-labeled nucleotide. Finally, it shows the efficiency of the protocol that was followed and in sometime used as a signal calibrator for the staining amongst other probes. The negative control in the contrary, it is essential to be included in an *in situ* hybridisation experiment since it reveals key aspects of the process. The negative control reveals the sensitivity of the whole process since it reveals the sensitivity of the gene-specific probes and the specificity of those probes relative to the negative control itself. Also, it shows if there is any unspecific binding of the anti-DIG antibody within the ovaries and thus eliminating false results. In an *in situ* experiment where the negative control gives signal the results are ambiguous.

#### 4.3.4 *In situ* hybridisation

After the probes had been validated, the *in situ* hybridisation could occur. The dissected 48 hours PBM ovaries were removed from the -20°C and the experiment was executed following the procedures outlined in CHAPTER 3. The whole of the process took for about 4 days since there are 3 overnight incubations in different steps of the process. Finally, when the experiment was at an end, the pictures were analysed and the results were used to contact different hypotheses on the

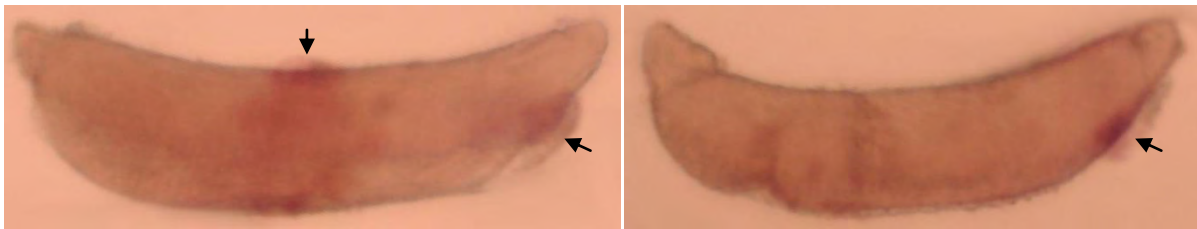
functions of the OBPs. Below, we present some examples of the *in situ* results (all the eggs are displayed with the posterior at the left).

AGAP002189 – OBP14



**Figure 4.13.** The mRNA of the OBP14 was observed at ring-like zones that seem to surround the egg.

AGAP002025 – OBP11



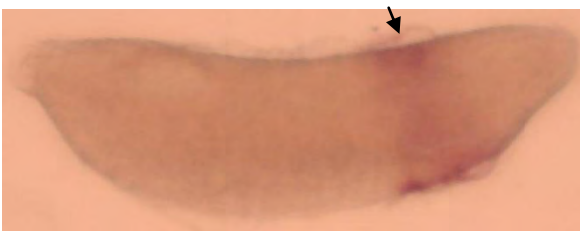
**Figure 4.14.** The mRNA of the OBP11 was observed at a wide ring-like zone and at the anterior ventral of the egg

AGAP002190 – OBP39

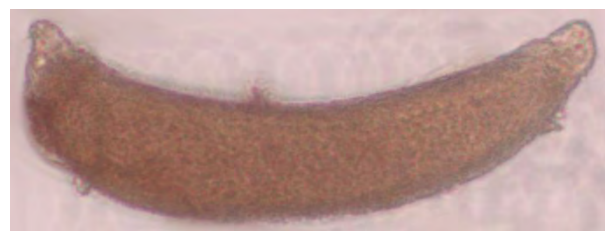


**Figure 4.15.** The mRNA of the OBP39 was observed at the posterior of the egg and at the ventral region. As was mentioned, a lot portion of the eggs were unstained.

AGAP002191 – OBP40



AGAP002188 – OBP12



**Figure 4.16.** (Left) The mRNA of the OBP39 was observed at the anterior of the egg in a wide ring-like zone. (right) the mRNA of OBP12 was difficult to observe in the eggs.

AGAP005182 – OBP41

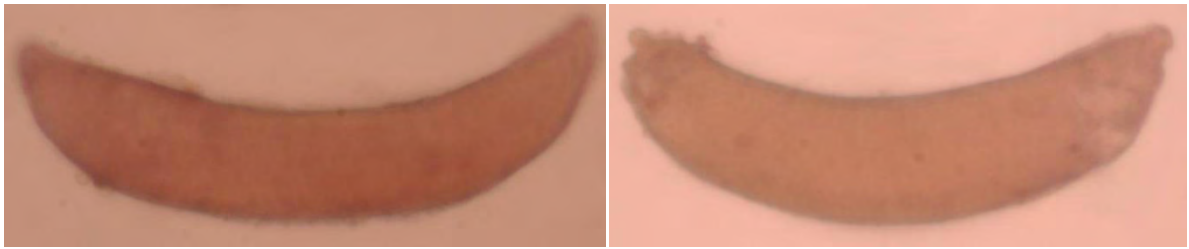


Figure 4.17. The mRNA of the OBP41 was difficult to visualise.

AGAP009065 – OBP42

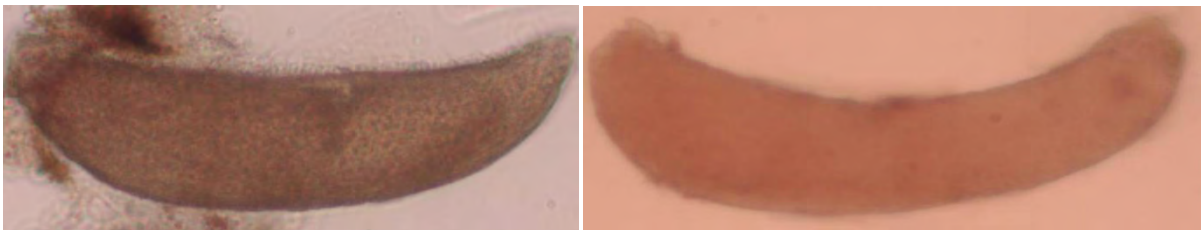


Figure 4.18. The mRNA of the OBP42 was difficult to visualise. The dark shades at the anterior are pigment from the mosquito's cuticle.

AGAP01647– OBP30



AGAP012320 – OBP25



Figure 4.19. The mRNA of the OBP25 (right) was observed at the membrane of the egg in a wide ring-like zone, in contrast with the mRNA of the OBP30 (left) which was difficult to visualise.

AGAP009402 – OBP43

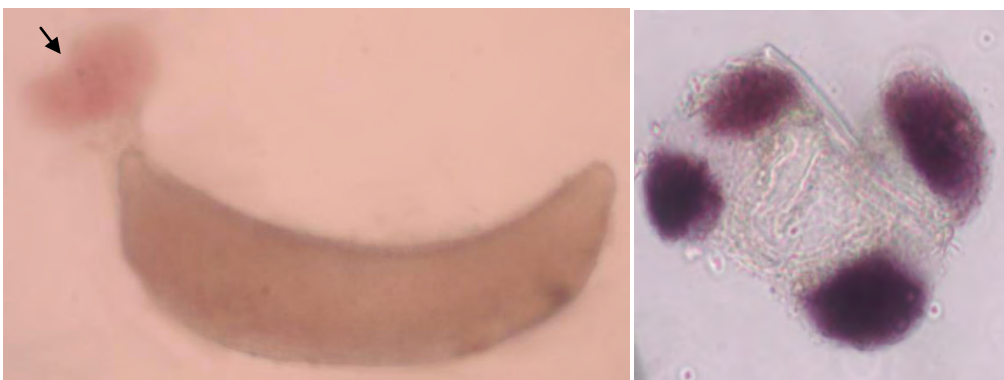
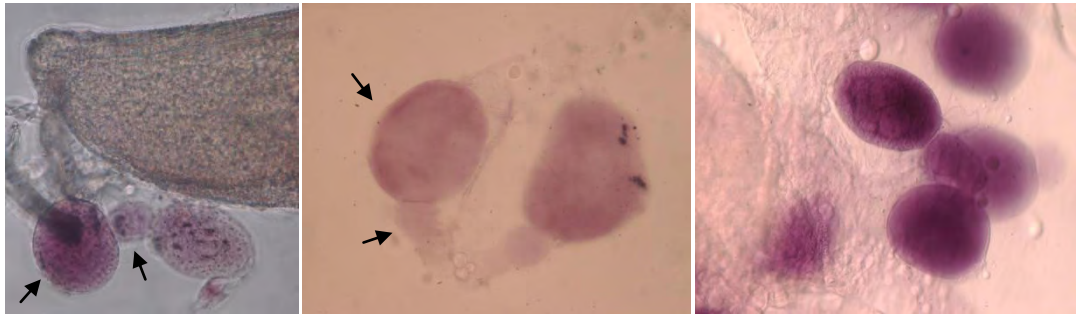


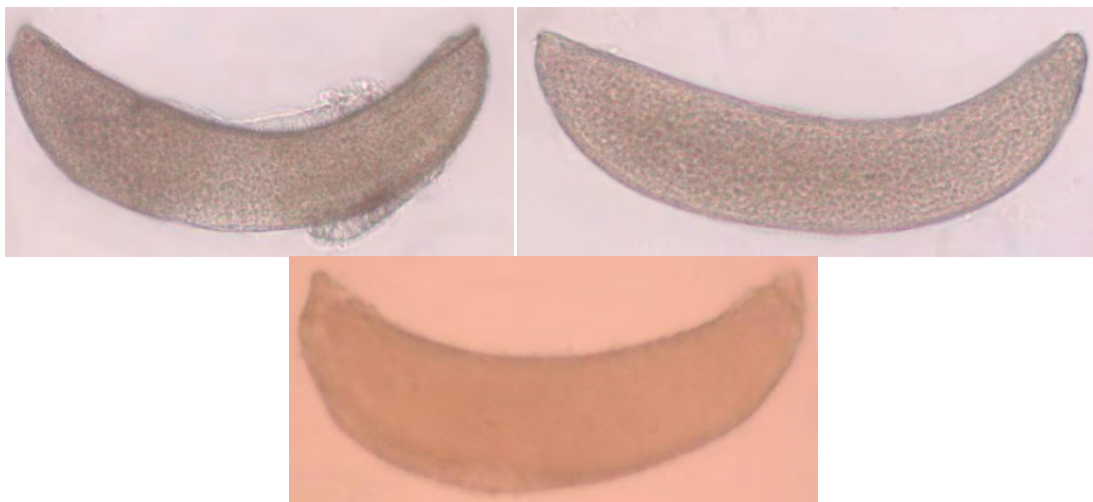
Figure 4.20. The mRNA of the OBP43 was observed in the penultimate/secondary follicle. It is obvious that OBP 43 is needed in the secondary follicle at 48 hours PBM. The pattern was observed at the positive control as well (see below).

#### AGAP004038 – positive control



**Figure 4.21.** The mRNA of the positive control (AGAP004038), was observed at the penproximate/secondary follicle; also the germarium has a smooth staining. The positive control shows that the whole of the process was successful; a signal in the positive control means that the *in situ* hybridisation was done according to the protocols. By giving positive results, the positive control shows that first of all there is mRNA within the sample. Furthermore, it shows that there was used a nucleotide that was labeled by DIG (digoxigenin) and that the antibody used to detect the DIG was active and bound to the DIG-labeled nucleotide. Finally, it shows the efficiency of the protocol that was followed and in sometime used as a signal calibrator for the staining amongst other probes.

#### GFP – negative control



**Figure 4.22.** The mRNA of the negative control (GFP), was not observed on any eggs. The absent of the negative control reveals that there was not unspecific binding of the anti-DIG antibody within the ovaries and thus we do not have false results. The negative control should not present staining; in an *in situ* hybridisation experiment where the negative control gives signal, the results are ambiguous.

The staining on the positive sample was successful. The sample was the first that had to be removed from the staining buffer due to the magnitude of the staining. Actually the sample as removed half an hour after the staining was initialised. Under the microscope it was obvious that the staining was occurred inside the penproximate/secondary follicles. The egg, the proximate/primary follicle was clean of staining. Also, in some majority of the pictures the germarium was stained too. The germarium was visible in this magnification only because of the staining, something that was not

available to see in the negative control. As was expected, the negative control was free of staining. The eggs were transparent without any trace of the dark-blue precipitating dye, product of the staining reaction. It is necessary to mention that the staining that was observed on the sample eggs which were treated with the different probes composed only a portion of the total eggs contained in each reaction. Also many of the eggs were destroyed during the process since under the microscope they looked pretty messy. For the eggs that it was possible to retrieve some good photos, multiple ideas were proposed which might explain the location of the mRNA.

### APPENDIX C: List of *Anopheles gambiae* OBP Genes

| #  | Gene       | OBP              | Length (a.a) | Type     | Genome Location       |
|----|------------|------------------|--------------|----------|-----------------------|
| 1  | AGAP003309 | <b>(OBP01)</b>   | 144          | Classic  | 2R: 35643035-35644609 |
| 2  | AGAP003306 | <b>(OBP02)</b>   | 157          | Classic  | 2R: 35434051-35434604 |
| 3  | AGAP001409 | <b>(OBP03)</b>   | 153          | Classic  | 2R: 4210577-4212097   |
| 4  | AGAP010489 | <b>(OBP04)</b>   | 150          | Classic  | 3L: 4997901-4998953   |
| 5  | AGAP009629 | <b>(OBP05)</b>   | 156          | Classic  | 3R: 37321107-37322204 |
| 6  | AGAP003530 | <b>(OBP06)</b>   | 155          | Classic  | 2R: 39200905-39201673 |
| 7  | AGAP001556 | <b>(OBP07)</b>   | 154          | Classic  | 2R: 6152013-6154470   |
| 8  | AGAP000279 | <b>(OBP08)</b>   | 176          | Classic  | X: 5036744-5037431    |
| 9  | AGAP000278 | <b>(OBP09)</b>   | 139          | Classic  | X: 5035248-5036100    |
| 10 | AGAP001189 | <b>(OBP10)</b>   | 131          | Classic  | 2R: 1034139-1169444   |
| 11 | AGAP002025 | <b>(OBP11)</b>   | 192          | Classic  | 2R: 14069095-14069749 |
| 12 | AGAP002188 | <b>(OBP12)</b>   | 159          | Classic  | 2R: 17328044-17328912 |
| 13 | AGAP002905 | <b>(OBP13)</b>   | 149          | Classic  | 2R: 29134157-29134861 |
| 14 | AGAP002189 | <b>(OBP14)</b>   | 188          | Classic  | 2R: 17331871-17332553 |
| 15 | AGAP003307 | <b>(OBP15)</b>   | 147          | Classic  | 2R: 35436449-35436969 |
| #  | -          | <b>(OBP16)</b>   |              | -        | -                     |
| #  | -          | <b>(OBP17)</b>   |              | -        | -                     |
| 16 | AGAP012319 | <b>(OBP18)</b>   | 176          | Classic  | 3L: 40171315-40172237 |
| 17 | AGAP004433 | <b>(OBP19)</b>   | 137          | Classic  | 2R: 55987079-55987846 |
| 18 | AGAP005208 | <b>(OBP20)</b>   | 142          | Classic  | 2L: 12288238-12289440 |
| 19 | AGAP008398 | <b>(OBP21)</b>   | 131          | Classic  | 3R: 10317255-10317835 |
| 20 | AGAP010409 | <b>(OBP22)</b>   | 144          | Classic  | 3L: 2853087-2853645   |
| 21 | AGAP012318 | <b>(OBP23)</b>   | 131          | Classic  | 3L: 40168852-40169329 |
| 22 | AGAP006278 | <b>(OBP24)**</b> | 320          | D7       | 2L: 29235061-29236536 |
| 23 | AGAP012320 | <b>(OBP25)</b>   | 142          | Classic  | 3L: 40209434-40210326 |
| 24 | AGAP012321 | <b>(OBP26)</b>   | 131          | Classic  | 3L: 40213816-40214477 |
| 25 | AGAP012323 | <b>(OBP27)</b>   | 134          | Classic  | 3L: 40218226-40218764 |
| 26 | AGAP012325 | <b>(OBP28)</b>   | 134          | Classic  | 3L: 40221011-40221620 |
| 27 | AGAP012331 | <b>(OBP29)</b>   | 176          | Classic  | 3L: 40708383-40709402 |
| 28 | AGAP011647 | <b>(OBP30)*</b>  | 289          | Atypical | 3L: 30753689-30754558 |
| 29 | AGAP010649 | <b>(OBP31)*</b>  | 313          | Atypical | 3L: 7968501-7969835   |
| 30 | AGAP000638 | <b>(OBP32)</b>   | 320          | Atypical | X: 11413679-11414777  |
| 31 | AGAP000640 | <b>(OBP33)</b>   | 334          | Atypical | X: 11416806-11417810  |
| 32 | AGAP000641 | <b>(OBP34)</b>   | 311          | Atypical | X: 11419447-11420382  |
| 33 | AGAP000642 | <b>(OBP35)</b>   | 275          | Atypical | X: 11422726-11423559  |
| 34 | AGAP000643 | <b>(OBP36)</b>   | 275          | Atypical | X: 11426182-11427015  |
| 35 | AGAP000644 | <b>(OBP37)</b>   | 311          | Atypical | X: 11428745-11429680  |
| 36 | AGAP000580 | <b>(OBP38)</b>   | 336          | Atypical | X: 10205844-10206981  |
| 37 | AGAP002190 | <b>(OBP39)</b>   | 246          | Atypical | 2R: 17333035-17333892 |
| 38 | AGAP002191 | <b>(OBP40)</b>   | 282          | Atypical | 2R: 17334179-17335132 |
| 39 | AGAP005182 | <b>(OBP41)</b>   | 279          | Atypical | 2L: 11685185-11686024 |

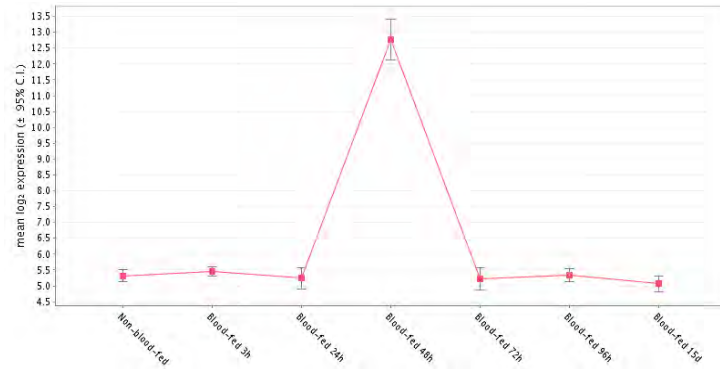
| #  | Gene       | OBP              | Length (a.a) | Type     | Genome Location         |
|----|------------|------------------|--------------|----------|-------------------------|
| 40 | AGAP009065 | <b>(OBP42)</b>   | 288          | Atypical | 3R: 25310798-25311664   |
| 41 | AGAP009402 | <b>(OBP43)</b>   | 333          | Atypical | 3R: 32225167-32226168   |
| 42 | AGAP010648 | <b>(OBP44)</b>   | 327          | Atypical | 3L: 7964187-7965356     |
| 43 | AGAP010650 | <b>(OBP45)</b>   | 356          | Atypical | 3L: 7970287-7971357     |
| 44 | AGAP007289 | <b>(OBP46)</b>   | 202          | Plus C   | 2L: 45014419-45015087   |
| 45 | AGAP007287 | <b>(OBP47)</b>   | 228          | Plus C   | 2L: 45011203-45012023   |
| 46 | AGAP007286 | <b>(OBP48)*</b>  | 200          | Plus C   | 2L: 45008928-45009814   |
| 47 | AGAP006075 | <b>(OBP49)</b>   | 179          | Plus C   | 2L: 26103215-26103894   |
| 48 | AGAP006076 | <b>(OBP50)</b>   | 166          | Plus C   | 2L: 26104148-26104931   |
| 49 | AGAP006077 | <b>(OBP51)</b>   | 203          | Plus C   | 2L: 26129377-26130091   |
| 50 | AGAP006078 | <b>(OBP52)</b>   | 170          | Plus C   | 2L: 26130285-26131425   |
| 51 | AGAP006079 | <b>(OBP53)</b>   | 171          | Plus C   | 2L: 26131618-26132453   |
| 52 | AGAP006080 | <b>(OBP54)</b>   | 181          | Plus C   | 2L: 26133006-26133968   |
| 53 | AGAP006081 | <b>(OBP55)</b>   | 156          | Plus C   | 2L: 26134275-26135119   |
| 54 | AGAP011367 | <b>(OBP56)</b>   | 235          | Plus C   | 3L: 22025830-22026752   |
| 55 | AGAP011368 | <b>(OBP57)</b>   | 204          | Plus C   | 3L: 22028532-22029250   |
| 56 | AGAP006074 | <b>(OBP58)*</b>  | 286          | Plus C   | 2L: 26101764-26135649   |
| 57 | AGAP006760 | <b>(OBP59)*</b>  | 155          | Plus C   | 2L: 37923144-37923831   |
| 58 | AGAP007281 | <b>(OBP60)*</b>  | 198          | Plus C   | 2L: 44998234-44999054   |
| 59 | AGAP007282 | <b>(OBP61)*</b>  | 204          | Plus C   | 2L: 44999450-45000275   |
| 60 | AGAP002556 | <b>(OBP62)*</b>  | 174          | Classic  | 2R: 22787889-22792798   |
| 61 | AGAP012322 | <b>(OBP63)*</b>  | 135          | Classic  | 3L: 40217381-40217853   |
| 62 | AGAP012324 | <b>(OBP64)*</b>  | 142          | Classic  | 3L: 40219631-40220284   |
| 63 | AGAP006759 | <b>(OBP65)*</b>  | 179          | Plus C   | 2L: 37922005-37922885   |
| 64 | AGAP007283 | <b>(OBP66)*</b>  | 212          | Plus C   | 2L: 45000529-45001413   |
| 65 | AGAP012659 | <b>(OBP67)*</b>  | 204          | Plus C   | UNKN: 22699010-22699835 |
| 66 | AGAP012658 | <b>(OBP68)*</b>  | 212          | Plus C   | UNKN: 22697816-22698745 |
| 67 | AGAP013182 | <b>(OBP69)*</b>  | 229          | Plus C   | 2R: 21343359-21344048   |
| 68 | AGAP006368 | <b>(OBP70)*</b>  | 200          | Plus C   | 2L: 30543243-30547869   |
| 69 | AGAP012867 | <b>(OBP71)*</b>  | 228          | Plus C   | UNKN: 35107623-35108585 |
| 70 | AGAP012714 | <b>(OBP72)*</b>  | 121          | Classic  | UNKN: 24728945-24729493 |
| 71 | AGAP008284 | <b>(OBP73)**</b> | 165          | D7       | 3R: 8562770-8563560     |
| 72 | AGAP008278 | <b>(OBP74)**</b> | 311          | long D7  | 3R: 8544002-8545231     |
| 73 | AGAP008279 | <b>(OBP75)**</b> | 315          | long D7  | 3R: 8545899-8555880     |
| 74 | AGAP008280 | <b>(OBP76)**</b> | 166          | D7       | 3R: 8556979-8558021     |
| 75 | AGAP008281 | <b>(OBP77)**</b> | 165          | D7       | 3R: 8558169-8558901     |
| 76 | AGAP008282 | <b>(OBP78)**</b> | 168          | D7       | 3R: 8559675-8560536     |
| 77 | AGAP008283 | <b>(OBP79)**</b> | 179          | D7       | 3R: 8561534-8562256     |

**Table 3.** A list of the OBP genes known to date in *Anopheles gambiae*. \*OBPs not named in VectorBase to date. \*\*OBPs named in this project.

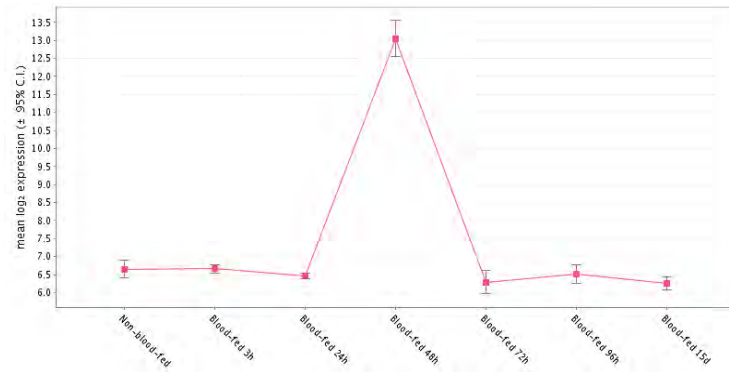


## APPENDIX D: Expression charts for the candidate OBPs

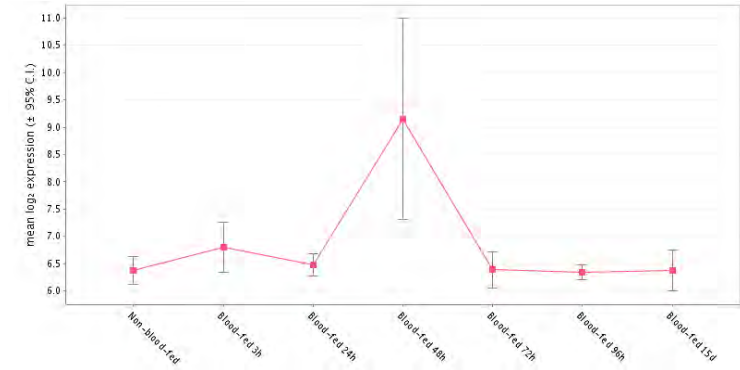
AGAP002189 - OBP14



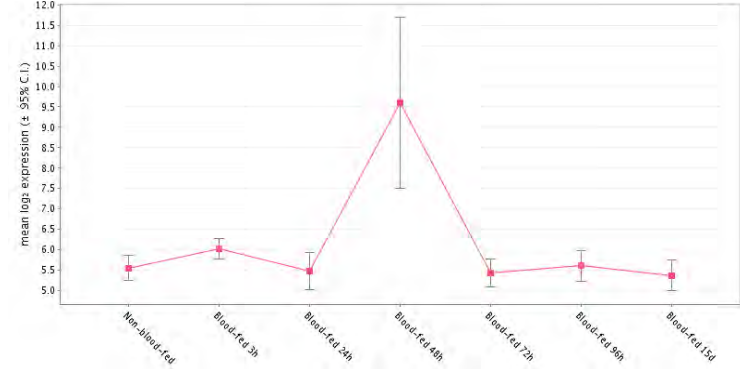
AGAP002025 - OBP11



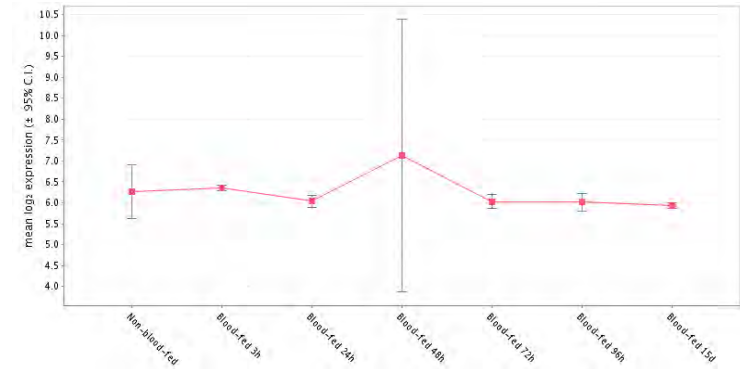
AGAP002190 - OBP39



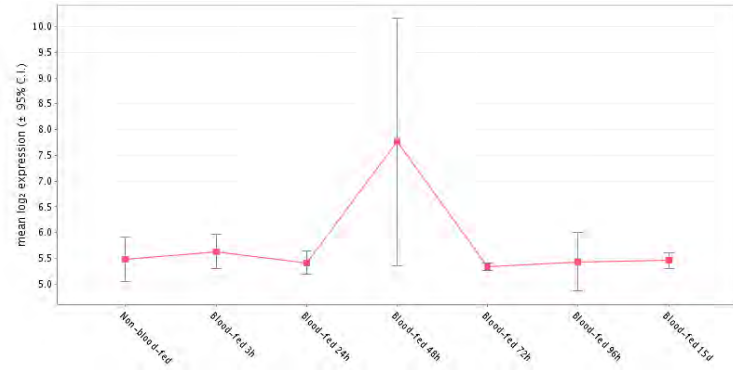
AGAP002191 - OBP40



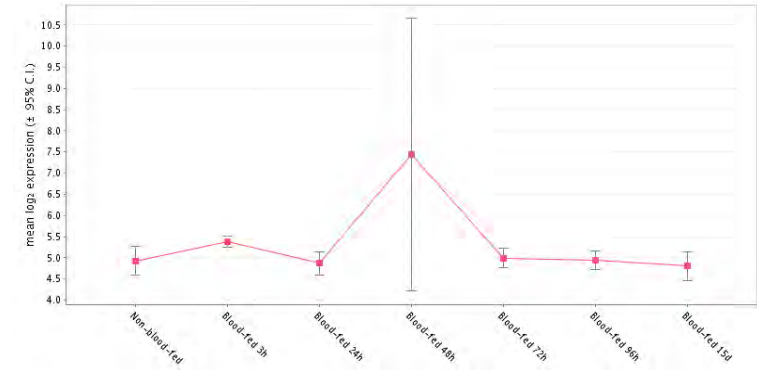
AGAP010648 - OBP44



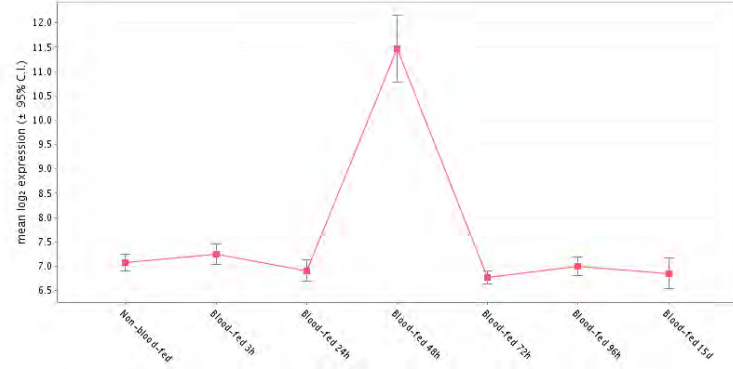
**AGAP010650 - OBP45**



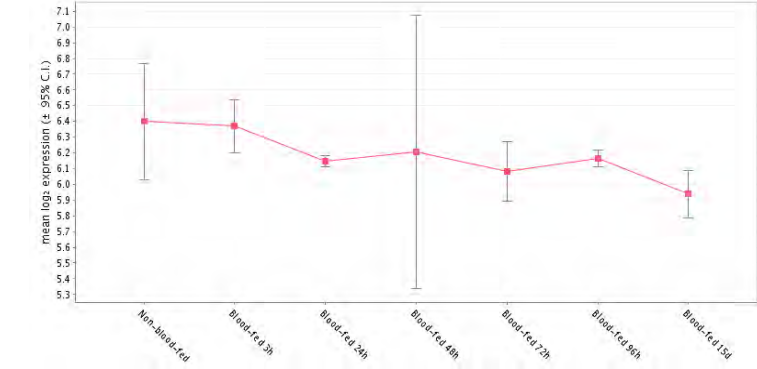
**AGAP009065 - OBP42**



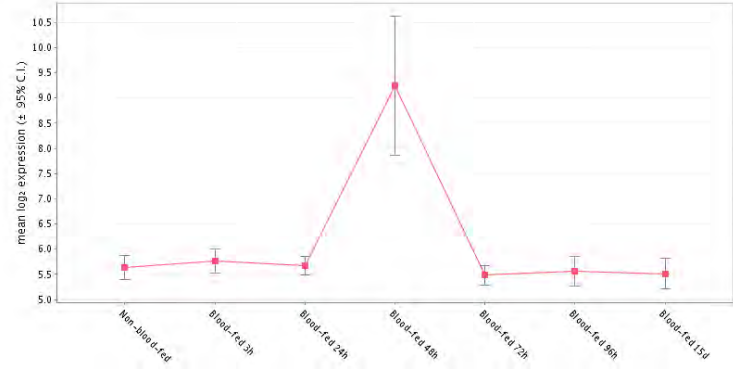
**AGAP002188 - OBP12**



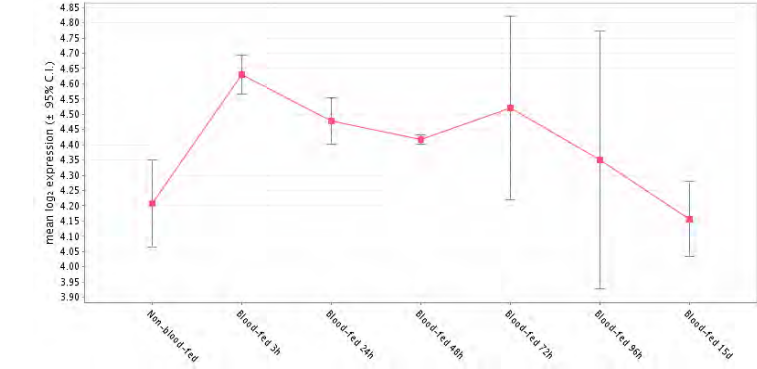
**AGAP010649 - OBP31**



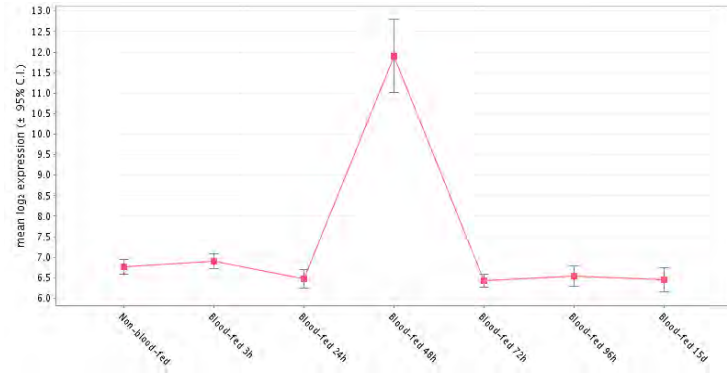
**AGAP005182 - OBP41**



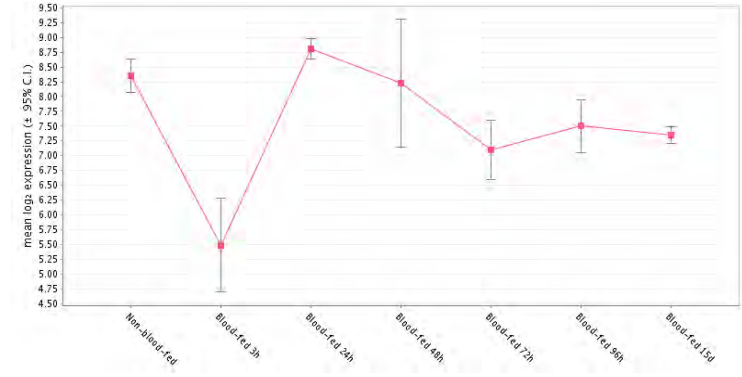
**AGAP012319 - OBP18**



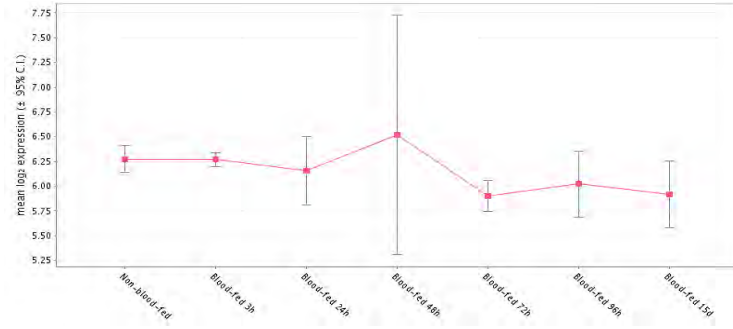
AGAP011647 - OBP30



AGAP009402 - OBP43



AGAP012320 - OBP25





# **CHAPTER 5**

## **DISCUSSION**

## CHAPTER 5: DISCUSSION

The need for an effective control strategy to ensure mosquito population suppression is urgent and essential. Malaria continues to kill hundreds of thousands of kids each year and *Anopheles gambiae*, the main vector for the transmission of the disease, continues to thrive across equatorial region in Africa. Studies which promise mosquito population suppression are in a continuous search for candidate genes, which will ensure through knockouts lines fertility abnormalities within the species, and thus the effective eradication of the disease vectors. Homing endonucleases (HE) and CRISPR/Cas9 system are able to introduce within a population a fertility mutation which spreads by a “super-Mendelian” inheritance ensuring fast and high rate of establishment within a species. In the current project it was attempted to investigate if the ovary-specific odorant binding proteins (OBPs) of the *A.gambiae* mosquito, which are expressed highly after a blood meal, are suitable candidates to be included in one of the studies mentioned. For the genes to be considered as candidates they should obey some basic characteristics which sometimes is difficult to meet in one gene (CHAPTER 1).

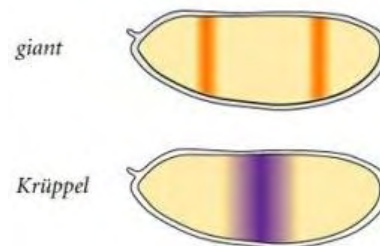
Through an intensive literature research and an *in situ* hybridisation experiment it was attempted to analyse these proteins and investigate their potential role in the egg development within the ovaries. It was obvious from data extracted from previous studies and projects that these ovary-specific OBPs belong to a distinct group of proteins expressed intensively after a blood meal, peaking at the specific time-frame of 48hrs PBM. At this point within the ovaries take place major events resulting to the egg development and maturation (Horne-Badovinac & Bilder, 2005); such events are the completion of vitellogenesis from the follicle cells and the nurse cell dumping of maternal mRNA. Highly expressed proteins from this group were chosen to participate in an *in situ* hybridisation experiment. The *in situ* hybridisation experiment was used to visualise the spatial location of the OBPs transcripts within the whole of the ovary and even within single ovarioles, in order to determine the key features required for the development of a HEG target.

From the fourteen genes candidates that were chosen according to their expression pattern within the ovary (Baker *et al.*, 2011) and after 48 hours PBM (Marinotti *et al.*, 2006), ten were made it through the phase of the probe generation. For those ten genes it was possible to retrieve pictures of proximate/primary follicles and ovarioles stained with the precipitant dye used for the experiment. It was unexpected that the stained ovarioles in all of the attempts were only a low fraction of the total eggs used for the experiments. Many of the eggs were destroyed in the process; something it was attributed to the high temperatures was used or the many wash steps performed. For the six of the OBPs (11, 14, 39, 40, 43 and 25) the staining was successful and an idea of the

location of the OBP was able to be observed with cautious though since as it was mentioned, the staining patterns were a low portion of the total eggs. For the other four OBPs (12, 41, 42 and 30) the staining did not occur at a visible pattern or in other cases the staining did not occur at all. According to the staining pattern that was observed on the six OBPs (11, 14, 39, 40, 43 and 25), many ideas were proposed separately for the functions of that particular OBPs.

### 5.1 OBPs may be functionally analogous to the *Drosophila* “gap genes”

Surprisingly in OBP14 (AGAP002189) and OBP11 (AGAP002025) the pattern of the gene expression was similar to the expression of gap genes in *D.melanogaster* like “Krüppel” and “giant” (Gilbert, 2003) (Figure 5.1), which are two of the many genes that are responsible for the segmentation (development of a section of the organism) of the embryo (Johnston *et al.*, 1992; Monk, 2004; Rivera-Pomar *et al.*, 1996).



**Figure 5.1.** giant and Kruppel expression pattern (Gilbert, 2003)

Gap genes are under the regulation of “maternal effect genes” that are deposited by the mother into the egg through its development within the ovaries. Maternal effect genes like “nanos” and “bicoid” are responsible for the polarity of the egg and the embryo (Russel, 2010; Rivera-Pomar *et al.*, 1996). The polarity of the egg is determined by the differentially localized mRNA molecules of maternal effect genes in contrast of what happens to the fertilized egg, in which the polarity is determined by the expression products of the maternal effect genes, for these genes are translated after the fertilization of the egg (Russel, 2010). Although maternal effect gene transcripts are derived from the mother in the early egg, gap genes on the other hand, are products of the embryo and their present (mRNA and protein) is noticeable in the deployed fertilized egg. In contrast, OBPs’ mRNA is present during oogenesis! The 48hrs PBM timeframe that was examined is relatively a final stage in *Anopheles gambiae* egg development and its related stage in *D.melanogaster* oogenesis seems to be between stages 11 and 13 (Horne-Badovinac & Bilder, 2005) (Figure 1.20). At these stages in *D.melanogaster* there is no evidence of expression of gap genes. These comparisons between *D.melanogaster* and *A.gambiae* help us reject the hypothesis that OBPs act like gap genes. Furthermore, gap genes are in their big majority transcription factors that bind the DNA and regulate

the expression of pair-rule genes and homeotic genes (Ingham *et al.*, 1986), by competing for binding to their enhancer regions. OBPs as far as it is known, have not got regions or domains that could bind DNA or RNA in order to contribute to the regulation of the transcription in the cell. Despite the fact that the previous suggests that they do not act as transcription factors, it should not be rejected due to the fact that other gap genes, like “tailless” and “knirps” are transcription factors as well as steroid receptors. This generates the possibility that OBPs, which have the domains to bind different molecules, could interact with other proteins that bind nucleic acid, and indirectly regulate transcription.

## 5.2 OBPs may be under the regulation of “maternal effect” genes

The basis that OBPs are not gap genes does not mean that they could not be under the regulation of maternal effect genes. Maternal effect genes, through the different gradients they create inside the egg and the embryo, drive the expression of other genes, like gap genes, and localize them in specific parts of the oocyte (Russel, 2010). In *Drosophila*, maternal effect genes are deposited during nurse cell dumping at oogenesis stages 11-13, equivalent to the 48hr blood meal time point in mosquitoes. The gap genes are expressed post deposition, responding directly to cues from the maternal effect genes. As such, it was hypothesised that for OBPs to function like the gap genes of *Drosophila*, maternal effect genes must already have been deposited prior to the OBPs. If something like that is revealed it means that the OBP expression is under the regulation of the gradients created by maternal effect genes. Also, because the egg does not produce any mRNA through the phases of oogenesis (Tadros & Lipdhitz, 2009; He *et al.*, 2011), OBP transcripts (like maternal effect genes) must derive exclusively from the mother (nurse cells).

## 5.3 OBPs may be components of the eggshell

However the patterns of the expression have not been observed inside the oocyte per se but rather around the chorion membrane and even in the layer of the follicle cells surrounding the oocyte. The previous could be observed in OBP39, OBP40, OBP11 and OBP25). That could suggest that OBP transcripts are secreted from the follicle cells into the membrane and so they do not respond to the biological gradient but are instead directly deposited at the functionally active site; something that already happens with other components which are found in the eggshell. One example is the *D.melanogaster* chorion peroxidase (Pxd), which as protein product is actually deposited by the follicle cells in the eggshell and stabilize the final eggshell structure through the formation of cross-links among the eggshell proteins (Konstandi *et al.*, 2005). The previous basis could be supported by the fact that In *Drosophila melanogaster* through the oogenesis, the follicle



cells have already matured and begun the secretion of the vitelline membrane around the oocyte (stage 10), and then add a shell over that membrane (stage 11). Both components are external to the egg and covered by the follicle cells (Trougakos & Margaritis, 2002). Also, it is already known that there is a great amount of these OBPs in the eggshell of *A.gambiae* (Amenya *et al.*, 2010) (Figure 5.2) and that these proteins might have domains which possess the ability to bind odorant hydrophobic molecules (Leal, 2012).

| Gene ID                         | Protein score | Known/putative function/comments |
|---------------------------------|---------------|----------------------------------|
| <b>Odorant binding proteins</b> |               |                                  |
| AGAP000641                      | 1516          | OBP34/37                         |
| AGAP000642                      | 1048          | OBP35                            |
| AGAP000643                      | 1014          | OBP36                            |
| AGAP002025                      | 499           | OBP11                            |
| AGAP011647                      | 152           | OBP1                             |
| AGAP010648                      | 96            | OBP44                            |
| AGAP002189                      | 138           | OBP13                            |

**Figure 5.2.** *Anopheles gambiae* eggshell proteins identified by LC-MSMS (Amenya *et al.*, 2010)

Considering these it could be suggest that the OBPs could participate in the eggshell formation! The hypothesis states that OBPs proteins in the egg are found anchoring to the vitelline membrane and they end up being incorporated into the eggshell as a by-product of its earlier role in chorion localization. To validate the previous, patterns of the expression of OBPs should be observed during earlier stages in the oogenesis (24hrs and 36hrs PBM) as well as stages before the oviposition (72hrs PBM) so that it could be revealed through presence or absence of the transcripts, if indeed the OBPs are there in early or late stages during the egg development.

#### 5.4 OBPs may be chemo-attractants for sperm

A final hypothesis that was suggested base on the patterns observed in OBP39 (AGAP002190) is the idea that OBPs in the shell are one of the connection links between the sperm components and the egg. Having in mind that odorant molecules act as chemo-attractants for sperm (Fukuda *et al.*, 2004), it can be suggested that OBPs' presence in the eggshell may play a role through the fertilization of the egg. OBPs could act like anchor components for binding odorant molecules in turn, like a chain reaction, bind sperm components and trigger the process of the fertilization during oviposition. It has already been proven that male steroid hormones are co-responsible for the

regulation of the oogenesis in the female mosquito (Baldini *et al.*, 2013). That could support the suggestion made before that OBPs involved in the fertilization or even trigger the process. To strongly validate this basis, more evidence must be acquired which show that the OBPs are present in the poles of the egg or even in the anterior in which the micropyle is present (entry of sperm) and furthermore if they are present in the right time-frame.

All the above consist hypothesis that have been proposed in order to explain the expression pattern of the OBPs in the ovaries. To summarise, OBPs are seemed to be transcribed in the follicular membrane; if this is true then they are somatically expressed. They are transcribed in specific parts of the ovaries and they have different patterns which mean that they might not be redundant after all. Finally, there is proving of their existence in the eggshell as proteins. The future research perspectives are to provide enough evidence that will prove or reject the bases explained above. The need for knockout mutants is crucial to reveal OBPs true function within the ovary and also in extent, their redundancy. The need to provide a strongly supported theory about OBPs nature and purpose is essential since these genes could be after all appropriate targets to use with HEGs or CRISPR systems and help control malaria vectors.

## ACKNOWLEDGMENTS

I would like to thank Professor Andrea Crisanti for accepting me to his laboratory and letting me be part of his amazing team. I wish to express my sincere gratitude to Professor Kostas Mathiopoulos, for providing me with the unique and life changing opportunity to study in London. I would also like to thank my supervisor Tony Nolan for introducing me to the interesting world of malaria and providing me with exclusive challenges throughout this demanding project.

Special thanks to the whole lab team, in particular Roberto Galizi, Federica Bernardini, and Giulia Morselli for their support, as well as Christopher Bamikole for his advice on mosquito handling. I would also like to thank my colleague and friend Chrysanthi Taxiarchi for her continuous assistance and support.

I am especially thankful to Andrew Hammond and Danica Fabrigar, who supported me from start to finish as well as lending their skills, advice and insight on science and life.

This final year project thesis was accomplished by the aegis of the LLP/ERASMUS PLACEMENT PROGRAM, powered by the European Union, in collaboration between University of Thessaly (Greece) and Imperial College London (United Kingdom).

## ABBREVIATIONS

|                     |   |
|---------------------|---|
| A                   | adenine   |
| Aa                  | amino acids   |
| Acps                | accessory gland proteins  |
| agarbiose           | disaccharide (D-galactose and 3,6-anhydro-L-galactopyranose)                                |
| anti-Dig-AP         | anti-DIG conjugated-AP antibody (Roche)   |
| BC                  | Before Christ   |
| BCIP                | 5-bromo-4-chloro-3-indolyl phosphate (Roche)  |
| bp                  | base pair   |
| C                   | cytosin   |
| CDC                 | Centers for Disease Control and Prevention ( <a href="http://www.cdc.gov">www.cdc.gov</a> ) |
| cDNA                | complementary DNA   |
| Chloroform          | CHCl <sub>3</sub>   |
| CM                  | outer chorion membrane  |
| CO <sub>2</sub>     | carbon dioxide  |
| Cy                  | cystein   |
| ddH <sub>2</sub> O  | double distilled water  |
| dH <sub>2</sub> O   | distilled water   |
| DNA                 | desoxyribonucleic acids   |
| dNTP                | desoxynucleotide triphosphates (Invitrogen)   |
| DTT                 | dithiothreitol (C <sub>4</sub> H <sub>10</sub> O <sub>2</sub> S <sub>2</sub> )              |
| EtBr                | ethidium bromide  |
| First-Strand Buffer | buffer for the first strand of cDNA synthesis (Invitrogen)                                  |
| G                   | guanine   |
| gDNA                | genomic DNA   |
| GFP                 | green fluorescence protein  |
| Glycogen            | glycogen (Ambion)   |
| H <sub>2</sub> O    | water   |
| HEGs                | homing endonuclease genes   |
| HEs                 | homing endonucleases  |
| HR                  | homologous recombination  |
| Hyb                 | hybridization buffer - DIG easy Hyb granules (Roche)  |
| Isopropanol         | CH <sub>3</sub> CH <sub>2</sub> (OH)CH <sub>3</sub>   |
| kDa                 | kilodalton  |
| LiCl                | lithium chloride precipitation solution (Ambion)  |
| M                   | molarity  |
| mg                  | milligramme   |
| ml                  | millilitre  |

|                     |  |
|---------------------|--|
| mRNA                | messenger RNA  |
| NBT                 | nitro blue tetrazolium chloride (Roche)                    |
| No                  | number   |
| OPB                 | odorant binding protein                                    |
| ORNs                | olfactory receptor neurons                                 |
| ORs                 | odorant receptors  |
| PBM                 | post-blood meal  |
| PBS                 | phosphate-buffered saline (Roche)                          |
| PBT                 | phosphate buffered Tween-20                                |
| PCR                 | polymerase chain reaction                                  |
| Phenol              | tris saturated phenol (Thermo Scientific)                  |
| Proteinase K        | subtilisin-related serine protease (Sigma-Aldrich)         |
| R                   | programming language / software environment                |
| Random Hexamers     | 6 random base paired oligonucleotides (Invitrogen)         |
| RDTs                | antigen-based rapid diagnostic tests                       |
| RNA                 | ribonucleic acid   |
| RNase A             | endoribonuclease A (Sigma-Aldrich)                         |
| RT-PCR              | reverse transcription polymerase chain reaction            |
| s.s                 | sensu stricto  |
| Sodium Acetate      | CH <sub>3</sub> COONa                                      |
| SS                  | sheep serum (Sigma-Aldrich)                                |
| STFS                | soft tissue fixation solution                              |
| Superscript™ III RT | Superscript™ III reversed transcriptase (Invitrogen)       |
| T                   | thymine  |
| T7 RNA Pol Plus     | T7 RNA polymerase plus (Ambion)                            |
| TBE                 | Tris/Borate/EDTA buffer                                    |
| T <sub>m</sub>      | melting temperature  |
| TRIzol              | guanidinium thiocyanate-phenol-chloroform reagent (Ambion) |
| tRNA                | transfer RNA   |
| TURBO™ DNase        | TURBO™ DNase endonuclease (Ambion)                         |
| Tween-20            | polyethylene glycol sorbitan monolaurate (Sigma-Aldrich)   |
| USD                 | United States dollar                                       |
| UTP                 | uridine triphosphate                                       |
| VM                  | inner vitelline membrane                                   |
| WBR                 | western blocking reagent (Roche)                           |
| WHO                 | World Health Organisation                                  |
| µg                  | microgramme  |
| µl                  | microlitre   |
| 3'                  | 3' prime end   |
| 5'                  | 5' prime end   |

## REFERENCES

- Alberts, B., Johnson, A., Lewis, J. *et al.*, *Molecular Biology of the Cell* (2007), USA, Garland Science
- Alkaline phosphatase buffer (NTMT), *Cold Spring Harb Protoc* (2007), doi:10.1101/pdb.rec11088
- Amenya, A.D., Chou, W., Li, J. *et al.*, Proteomics reveal novel components of the *Anopheles gambiae* eggshell, *J Insect Physiol* (2010), 56(10):1414-1419
- Arrow, K.J., Panosian, C. & Gelband, H., Saving Lives, Buying Time: Economics of Malaria Drugs in an Age of Resistance, *National Academies Press* (2004)
- Aurburn, R., Small scale extraction of total RNA from *Drosophila melanogaster*, *FlyChip* (2006)
- Baker, D.A., Baker, D., Nolan, T. *et al.*, A comprehensive gene expression atlas of sex- and -tissue specificity in the malaria vector *Anopheles gambiae*. *BMC Genomics* (2011), 12 (1), 296
- Baldini, F., Gabrieli, P., South, A. *et al.*, The Interaction between a Sexually Transferred Steroid Hormone and a Female Protein Regulates Oogenesis in the Malaria Mosquito *Anopheles gambiae*, *PLoS Biol* (2013), 11(10): e1001695. doi:10.1371/journal.pbio.1001695
- Bannister, H.L., Hopkins, M.J., Dluzewski, R.A. *et al.*, *Plasmodium falciparum* apical membrane antigen 1 (PfAMA-1) is translocated within micronemes along subpellicular microtubules during merozoite development, *J Cell Sci* (2003), 116, 3825-3834, doi: 10.1242/jcs.00665
- Besansky, N.J., Powell, J.R., Caccone A. *et al.*, Molecular phylogeny of the *Anopheles gambiae* complex suggests genetic introgression between principal malaria vectors, *Proc. Natl. Acad. Sci* (1994), 91 (15): 6885-8. doi:10.1073/pnas.91.15.6885
- Bledsoe, G.H., Malaria primer for clinicians in the United States, *Southern Medical Journal* (2005), 98 (12): 1197-204 doi:10.1097/01.smj.0000189904.50838.eb
- Bruce-Chuvatt, L.J., Alphonse Laveran's discovery 100 years ago and today's global fight against malaria, *Journal of the Royal Society of Medicine* (1981), 74.7: 531
- Bruce-Chwatt, L.J., Malaria and its control: present situation and future prospects, *Annual review of public health* (1987), 8.1: 75-110
- Burrell, M.M., *Enzymes of Molecular Biology* (1993), Vol.16, 263-270
- Burt, A., Site-specific selfish genes as tools for the control and genetic engineering of natural populations, *Proc. R. Soc. Lond*, (2003), 270, 921-928 DOI 10.1098/rspb.2002.2319
- Calvo, E., deBianchi, A.G., James, A.A., Marinotti, O., The major acid soluble proteins of adult female *Anopheles darlingi* salivary glands include a member of the D7-related family of proteins, *Insect Biochem Mol Biol* (2002), 32:1419-1427
- Calvo, E., Mans, B.J., Andersen, J.F., Ribeiro, J.M., Function and evolution of a mosquito salivary protein family, *J Biol Chem* (2006), 281:1935-1942
- Calvo, E., Mans, B.J., Ribeiro, J.M., Andersen, J.F., Multifunctionality and mechanism of ligand binding in a mosquito antiinflammatory protein., *Proc Natl Acad Sci USA* (2009), 106:3728-3733
- Caraballo, H., Emergency Department Management Of Mosquito-Borne Illness: Malaria, Dengue, And West Nile Virus, *Emergency Medicine Practice* (2014), 16
- Carson, F.L., Christa, H., Histotechnology: A Self-Instructional Text (3 ed.), *American Society for Clinical Pathology Press* (2009), p.2. ISBN 978-0-89189-581-7
- Cavaliere, V., Bernardi, F., Romani, P., Duchi, S. & Gargiulo, G., Building up the *Drosophila* eggshell: first of all the eggshell genes must be transcribed. *Developmental Dynamics* (2008), 237, 2061-2072
- CDC, Centers for Disease Control and Prevention, 1600 Clifton Rd, Atlanta, GA 30329-4027, USA
- Chan, Y.S., Naujoks, D.A., Huen, D.A. & Russell, S., Insect Population Control by Homing Endonuclease-Based Gene Drive: An Evaluation in *Drosophila melanogaster*, *Genetics* (2011), 188: 33-44
- Chevalier, B.S. & Stoddard, B.L., Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility, *Nucleic Acids Res* (2001), 29, 3757-3774
- Clark, P. & Eichhorn, G.L., Predictable modification of enzyme specificity, Selective alteration of DNA bases by metal ions to promote cleavage specificity by deoxyribonuclease, *Biochemistry* (1974), 13 (25), pp 5098-5102
- Cleland, W.W., Dithiothreitol, a New Protective Reagent for SH Groups, *Biochemistry* (1964), doi: 10.1021/bi00892a002
- Cowman, A.F., Berry, D. & Baum, J., The cellular and molecular basis for malaria parasite invasion of the human red blood cell, *Journal of Cell Biology* (2012), 198 (6): 961-71
- Cox, F.E.G., History of the discovery of the malaria parasites and their vectors, *Parasites & Vectors* (2010), 3 (1): 5. doi:10.1186/1756-3305-3-5
- Darby, A.I. & Hewitson, T.D., In Situ Hybridization Protocols, *Third Edition, Methods in Molecular Biology* (2006)
- Dottorini, T., Nicolaidis, L., Ranson, H. *et al.*, A genome-wide analysis in *Anopheles gambiae* mosquitoes reveals 46 male accessory gland genes, possible modulators of female behavior, *PNAS* (2007), 104, 16215-16220
- Dottorini, T., Persampieri, T., Palladino, P. *et al.*, Regulation of *Anopheles gambiae* male accessory gland genes influences postmating response in female, *FASEB journal* (2013), 27, 86-97
- Edgar, R.C., MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Res* (2004), 32: 1792-1797
- Fairhurst, R.M., Wellem, T.E., *Plasmodium* species (malaria), *Principles and Practice of Infectious Diseases 2* (2010)
- Fukuda, N., Yomogida, K., Okabe, M., Touhara, K., Functional characterization of a mouse testicular olfactory receptor and its role in chemosensing and in regulation of sperm motility, *Journal of Cell Science* (2004), 15, 5835-5845
- Galizi, R., Doyle, L.A., Menichelli, M. *et al.*, A synthetic sex ratio distortion system for the control of the

- human malaria mosquito, *Nat. Commun.* (2014), 5:3977 doi: 10.1038/ncomms4977
- Galtier, N., Gouy, M., Gautier, C., SEAVIEW and PHYLO\_WIN: two graphic tools for sequence alignment and molecular phylogeny, *Comput Appl Biosci* (1996), 12:543–548
- Garibyan, L. & Avashia, N., Polymerase Chain Reaction, *Journal of Investigative Dermatology* (2013), **133**, e6. doi:10.1038/jid.2013.1
- Gaunt, M.W. & Miles, M. A., An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks, *Mol. Biol. Evol* (2002), 19: 748–761
- Genoscope., *Anopheles gambiae*: First genome of a vector for a parasitic disease (2008)
- Gilbert, S., *Developmental Biology* (2003), Sinauer Associates, Inc., Sunderland, Massachusetts
- Gilles, H. M., *Epidemiology of malaria, Bruce-Chwatt's Essential Malariology* (1993), :124-163
- Goddard, M.R., Greig, D., & Burt, A., Outcrossed sex allows a selfish gene to invade yeast populations. *Proc. R. Soc. Lond* (2001), B 268, 2537–2542, doi 10.1098/rspb.2001. 1830
- Gollin, D. & Zimmermann, C., Malaria: Disease impacts and long-run income differences, *Institute for the Study of Labor* (2007)
- Goltsev, Y., Rezende, G., Vranizan, K. *et al.*, Developmental and evolutionary basis for drought tolerance of the *Anopheles gambiae* embryo, *Developmental Biology* (2009), 330 462–470, doi:10.1016/j.ydbio.2009.02.038
- Gouy, M., Guindon, S., Gascuel, O., SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building, *Mol Biol Evol* (2010), 27(2): 221-224
- Guindon, S. & Gascuel, O., A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood, *Syst Biol* (2003), 52:696–704
- Hall, N., Karras, M., Raine, J.D. *et al.*, A comprehensive survey of the *Plasmodium* life cycle by genomic, transcriptomic, and proteomic analyses, *Science* (2005), 307.5706: 82-86
- He, L., Wang, X. & Montell, J.D., Shining light on *Drosophila* oogenesis: live imaging of egg development, *Current Opinion in Genetics & Development* (2011), 21:612–619
- Hedges, S.B., Dudley, J. & Kumar, S., TimeTree: a public knowledge-base of divergence times among organisms, *Bioinformatics* (2006), 22:2971-2972
- Hekmat-Safe, D.S., Safe, C.R., McKinney, A.J., Tanouye, M.A., Genome-wide analysis of the odorant-binding protein gene family in *Drosophila melanogaster*, *Genome Res* (2002), 12:1357–1369
- Horne-Badovinac, S & Bilder, D., Mass Transit: Epithelial Morphogenesis in the *Drosophila* Egg Chamber, *DEVELOPMENTAL DYNAMICS* (2005), 232:559–574, doi 10.1002/dvdy.20286
- Horwitz, J.P., Substrated for Cytochemical Demonstration of Enzyme Activity, Some Dihalo-3-indolyl Phosphates and Sulfates, *J Med Chem* (1996), 9,447
- Ingham, P.W., Ish-Horowicz, D. & Howard, K.R., Correlative changes in homoeotic and segmentation gene expression in Krüppel mutant embryos of *Drosophila*, *The EMBO journal* (1986), 5 (7)
- Johnston, D. & Volhard, C., The Origin of Pattern and Polarity in the *Drosophila* Embryo, *Cell* (1992), 68, 201-219
- Juhn, J. & James, A., Hybridization in situ of salivary glands, ovaries, and embryos of vector mosquitoes, *Journal of visualized experiments : JoVE* (2012), doi:10.3791/3709
- Kiszewski, A., Mellinger, A., Spielman, A. *et al.*, Global Index Representing the Stability of Malaria Transmission, *Am J Trop Med Hyg* (2004), vol. 70 no. 5 486-498
- Konstandi, O.A., Papassideri, I.S., Stravopodis, D.J., *et al.*, The enzymatic component of *Drosophila melanogaster* chorion is the Pxd peroxidase, *Insect Biochemistry and Molecular Biology* (2005), 35, 1043–1057
- Koutsos, C.A., Claudia, B., Stephan, M. *et al.*, Life cycle transcriptome of the malaria mosquito *Anopheles gambiae* and comparison with the fruitfly *Drosophila melanogaster*, *PNAS* (2007), 11304–11309 vol.104 no.27
- Kumar, S. & Hedges, S.B., TimeTree2: species divergence times on the iPhone, *Bioinformatics* (2011), 27:2023-2024
- Kunitz, M., Crystalline desoxyribonuclease isolation and general properties; spectrophotometric method for the measurement of desoxyribonuclease activity, *J Gen. Physiol* (1950), 33,349-62
- Larkin, M.A. *et al.*, Clustal W and Clustal X version 2.0, *Bioinformatics* (2007), 23: 2947–2948
- Lawniczak, M. K., Emrich, S., Holloway, A.K. *et al.*, Widespread divergence between incipient *Anopheles gambiae* species revealed by whole genome sequences, *Science* (2010), 330 (6003): 512–4. doi:10.1126/science.1195755
- Leal, S.W., Odorant Reception in Insects: Roles of Receptors, Binding Proteins, and Degrading Enzymes, *Annu. Rev. Entomol* (2012), 58:373–91
- Manoharan, M., Chong, M., Vartnadapoule, A. *et al.*, Comparative Genomics of Odorant Binding Proteins in *Anopheles gambiae*, *Aedes aegypti*, and *Culex quinquefasciatus*, *Genome Biol Evol* (2013), 163–180 doi:10.1093/gbe/evs131
- Margaritis, L.H., The egg-shell of *Drosophila melanogaster*. III. Covalent cross linking of the chorion proteins involves endogenous hydrogen peroxide. *Tissue Cell* (1985), 17, 553–559
- Marinotti, O., Calvo, E., Nguyen, Q.K. *et al.*, Genome-wide analysis of gene expression in adult *Anopheles gambiae*, *Insect Molecular Biology* (2006), 15(1), 1–12
- Megy, K., Emrich, S.J., Lawson, D. *et al.*, VectorBase: improvements to a bioinformatics resource for invertebrate vector genomics, *Nucleic Acids Research* (2012), 40: D729-734
- Michal, G., Chemical design of indicator reactions for the visible range. *Methods of Enzymatic Analysis* (1983), 3rd edition, Vol I, 197
- Monk, N., Development: Dissecting the Dynamics of Segment Determination, *Curreny Biology* (2004), 14, R705-R707

- Mullis, K.B., The unusual origin of the polymerase chain reaction, *Scientific American* (1990), 262:56–61, 64–5
- Neafsey, D.E., Christophides, G.K., Collins, F.H. *et al.*, The Evolution of the *Anopheles* 16 Genomes Project, *G3-GENES GENOMES GENETICS* (2013), Vol:3, ISSN:2160-1836, Pages:1191-1194
- Nolan, T., Papathanos, P., Windbichler, N. *et al.*, Developing transgenic *Anopheles mosquitoes* for the sterile insect technique, *Genetica* (2011), 139, 33-39
- Notredame, C., Higgins, D.G. & Heringa, J., T-Coffee: a novel method for multiple sequence alignments, *J Mol Biol* (2000), 302: 205–217
- Price, A.P., The essential role of Ca<sup>2+</sup> in the activity of bovine pancreatic deoxyribonucle-ase, *J Biol Chem* (1975) **250**: 1981-1986
- Rivera-Pomar, R. & Jackle, H., From gradients to stripes in *Drosophila* embryogenesis: Filling in the gaps, *Trends Genet* (1996), 12: 478–483. doi:10.1016/0168-9525(96)10044-5
- Robinson, R., His Hormone, Her Oogenesis: How Male Malaria Mosquitoes Trigger Female Egg Development, *PLoS Biol* (2013), 11(10):e1001694, doi:10.1371/journal.pbio.1001694
- Rogers, D.W., Whitten, M., Thailayil, J. *et al.*, Molecular and cellular components of the mating machinery in *Anopheles gambiae* females, *Proc Natl Acad Sci* (2009), 9:105(49):19390-5
- Russel, J.P., iGenetics: a molecular approach. Pearson Education (2010), pp. 564–571, ISBN 978-0-321-56976-9.
- Saitou, N. & Nei, M., The neighbor-joining method: a new method for reconstructing phylogenetic trees, *Mol Biol Evol* (1987), 4:406–425
- Sallares, R., Malaria and Rome: a history of malaria in ancient Italy, *Oxford University Press* (2002), doi:10.1093/acprof:oso/9780199248506.001.0001
- Schenborn, E.T. & Meirendorf, R.C., A novel transcription property of SP6 and T7 RNA polymerases: dependence on template structure, *Nucl Acids Res* (1985), 13, 6223-6236
- St.Pierre, S.E., Ponting, L., Stefancsik, R., McQuilton, P., & the FlyBase Consortium., FlyBase 102-advanced approaches to inter-rogating FlyBase, *Nucleic Acids Res* (2014), 42(D1):D780-D788
- Studier, J.A. & Keppler, K.J., A note on the neighbor-joining algorithm of Saitou and Nei, *Mol Biol Evol* (1988), 5:729–731
- Su, X., Hayton, K. & Wellems, T. E., Genetic linkage and association analyses for trait mapping in *Plasmodium falciparum*, *Nature Reviews Gene-tics* (2007), 8, 497–506
- Tadros, W. & Lipdhit, D.H., The maternal-to-zygotic transition: a play in two acts, *Development* (2009), 136, 3033-3042 (2009) doi:10.1242/dev.033183
- Trougakos, I.P. & Margaritis, L.H., Novel morphological and physiological aspects of insect eggs, *Chemoecology of Insect Eggs and Egg Deposition* (2002), pp. 3–36
- Valenzuela, J.G., Pham, V.M., Garfield, M.K., Francischetti, I.M. & Ribeiro, J.M., Toward a description of the sialome of the adult female mosquito *Aedes aegypti*, *Insect Biochem Mol Biol* (2002), 32:1101–1122
- Vaughan, A.M., Aly, A.S. & Kappe, S.H., Malaria parasite pre-erythrocytic stage infection: Gliding and hiding, *Cell Host & Microbe* (2008), 4 (3): 209–18. doi:10.1016/j.chom.2008.08.010
- Vieira, F.G. & 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), 3:476–490
- Vogt, R.G., Prestwich, G.D., Lerner, M.R., Odorant-binding-protein subfamilies associate with distinct classes of olfactory receptor neurons in insects, *J Neurobiol* (1991), 22:74-84
- Warhurst, C.D., The Parasite, *Travelers' Malaria* (2008), pp. 70–1
- Waring, G.L., Morphogenesis of the eggshell in *Drosophila*, *International Reviews of Cytology* (2000), 198, 67–108
- Wilkins, E.E., Howell, P.I. & Benedict, M.Q., IMP PCR primers detect single nucleotide polymorphisms for *Anopheles gambiae* species identification, Mopti and Savanna rDNA types, and resistance to dieldrin in *Anopheles arabiensis*, *Malar. J.* (2006), 5 (1): 125. doi:10.1186/1475-2875-5-125
- Windbichler, N., Menichelli, M., Papathanos, P., A synthetic homing endonuclease-based gene drive system in the human malaria mosquito, *Nature* (2011), 473, 212-215
- Worrall, E., Basu, S., Hanson, K., Is malaria a disease of poverty? A review of the literature, *Tropical Health and Medicine* (2005), (10):1047–59, doi:10.1111/j.1365-3156.2005.01476.x
- WHO, World Health Organisation, Factsheet on the World Malaria Report, *World Malaria Report*, (2013)
- Xu, P.X., Zwiebel, L.J., Smith, D.P., Identification of a distinct family of genes encoding atypical odorant-binding proteins in the malaria vector mosquito, *Anopheles gambiae*, *Insect Mol Biol* (2003), 12:549–560
- Yakob, L., Epidemiological consequences of a newly discovered cryptic subgroup of *Anopheles gambiae*, (2011), *Biol Lett* (2011), 23;7(6):947-9. doi: 10.1098/rsbl.2011.0453
- Zilversmit, M. & Susan, P., *Plasmodium*, Malaria Parasites, Version 05 May 2008, <http://tolweb.org/Plasmodium/68071/2008.05.05> in The Tree of Life Web Project, (2008), <http://tolweb.org/>



