POSITIONAL CLONING AND RNA INTERFERENCE ASSAY DEMONSTRATED

PLASMODIUM-MELANIZING FUNCTION OF A SERINE PROTEASE WITHIN

THE PLASMODIUM ENCAPSULATION 1 LOCUS (Pen1) OF THE MALARIA

MOSQUITO ANOPHELES GAMBIAE

A Dissertation

Submitted to the Graduate School
of the University of Notre Dame
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

by

Lucas Q. Ton

___________________________________
Frank H. Collins, Director

Graduate Program in Biological Sciences
Notre Dame, Indiana
December 2005
Due to its public health impact as the most important vector of malaria in Africa, the *Anopheles gambiae* mosquito and its *Plasmodium* parasites have been the subject of many studies both in the laboratory and out in the fields. For successful transmission of disease, the parasites must undergo a period of development in the mosquito from
gametocytes to sporozoites. During this period, susceptibility or refractoriness is determined by the ability of some mosquitoes to encapsulate ookinetes in a melanin/protein matrix following their penetration of the midgut epithelium. Melanin is deposited onto the surface of the parasites, causing the dead parasites to appear brownish black. Melanin synthesis in insects involves a biochemical pathway named after the first and most important enzyme, phenoloxidase (PO). This enzyme is synthesized in an inactive form and its activity depends on proteolytic activation by a serine protease usually called prophenoloxidase-activating enzyme (PPAE). This serine protease has at least one extra domain called a Clip domain (Jiang and Kanost, 2000). Melanin-associated refractoriness to the primate parasite *P. cynomolgi* B and the rodent parasite *P. berghei* is influenced by at least three Quantitative trait loci (QTLs). A genetic crossing experiment determined that the major QTL lies on chromosome arm 2R between division 8C-8D and this locus was named *Plasmodium* Encapsulation 1 (*Pen1*) to reflect its major role in melanotic encapsulation of parasites (Zheng *et al.*, 1997).

This thesis describes the positional cloning of *Pen1* from the construction of a physical map to the genomic, bioinformatic, molecular and functional analyses supporting the conclusion that *Pen1* encodes a Clip domain serine protease with PPAE function. Specifically, we show that CLIPB17 is found within the *Pen1* region and that it encodes a serine protease with three Clip domains and a serine protease domain with sequence similarity to all known PPAEs. In a selected strain with a fully refractory phenotype (L35 strain), CLIPB17 gene expression is up-regulated at the time of parasite melanization. In the counter-selected strain 4Arr, no such induction is evident. The regulatory region of CLIPB17 contains multiple strain-specific nucleotide
polymorphisms clustered within three putative NF-kB sites that could explain the L35-specific gene induction. We also show that RNAi knockdown abolishes CLIPB17 gene induction in L35 and significantly reduces refractoriness to both *Plasmodium berghei* and *P. cynomolgi*. Melanotic encapsulation represents the fastest and most important endogenous refractory mechanism against pathogens too large or too numerous for other immune mechanisms to cope with (Vey, 1993). Interest in this immune mechanism as a novel malaria control strategy prompted its genetic selection in a laboratory strain. Our determination of *Pen1* molecular identity and function significantly enhances the future prospect to explore this mechanism for controlling one of the most devastating diseases of man.
## CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>TABLES</td>
<td>iv</td>
</tr>
<tr>
<td>FIGURES</td>
<td>v</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>vii</td>
</tr>
<tr>
<td>1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 A brief history of malaria</td>
<td>1</td>
</tr>
<tr>
<td>1.2 The burden of malaria</td>
<td>2</td>
</tr>
<tr>
<td>1.3 Development of <em>Plasmodium</em> in the host and vector</td>
<td>4</td>
</tr>
<tr>
<td>1.4 Vector competence</td>
<td>7</td>
</tr>
<tr>
<td>1.5 A historical perspective on insect immunity</td>
<td>10</td>
</tr>
<tr>
<td>1.6 Recognition of foreignness</td>
<td>12</td>
</tr>
<tr>
<td>1.7 Organs and cells of insect immunity</td>
<td>15</td>
</tr>
<tr>
<td>1.8 Signaling pathways in humoral immunity</td>
<td>16</td>
</tr>
<tr>
<td>1.9 Melanotic encapsulation</td>
<td>20</td>
</tr>
<tr>
<td>1.10 Genetics of mosquito refractoriness to <em>Plasmodium</em></td>
<td>27</td>
</tr>
<tr>
<td>1.11 Aims of thesis</td>
<td>30</td>
</tr>
<tr>
<td>2. MATERIALS AND METHODS</td>
<td>32</td>
</tr>
<tr>
<td>2.1 Mosquito strains and rearing conditions</td>
<td>32</td>
</tr>
<tr>
<td>2.2 Construction of a physical clone map across the <em>Pen1</em> region</td>
<td>32</td>
</tr>
<tr>
<td>2.3 Sequencing BACs in the <em>Pen1</em> region</td>
<td>35</td>
</tr>
</tbody>
</table>
2.4 Bioinformatic analysis of Pen1 sequences……………………………………36
2.5 Semi-quantitative RT-PCR…………………………………………………….37
2.6 PCR cloning of a gene in the polymorphic region…………………………37
2.7 Quantitative real-time RT-PCR analysis of mRNA…………………………38
2.8 Infections with Plasmodium, bacteria, or yeast……………………………40
2.9 PCR Cloning and sequencing of CLIPB17 coding and promoter regions….41
2.10 RNA interference assay of CLIPB17…………………………………………42

3. RESULTS AND DISCUSSION…………………………………………………..43
3.1 Physical mapping and sequencing of the Pen1 region…………………….43
3.2 Sequence analysis of a 528 kb fragment from the Pen1 region………….55
3.3 Assembly of the entire Pen1 region…………………………………………….63
3.4 Analysis of candidate gene from the Pen1 sequence………………………..65
3.5 Bioinformatic analyses of CLIPB17…………………………………………..66
3.6 Expression profile of CLIPB17………………………………………………..71
3.7 Quantitative real-time RT-PCR optimization for CLIPB17……………….77
3.8 qRT-PCR analysis of CLIPB17………………………………………………..80
3.9 Cloning and sequencing of CLIPB17…………………………………………90
3.10 RNAi-mediated effect of CLIPB17 on melanization………………….…95

4. CONCLUSION……………………………………………………………………101

5. REFERENCES……………………………………………………………………104
TABLES

Table 1. List of primers used in experiments..........................................................33
Table 2. List of all genes from 528 kb Pen1 fragment........................................56
Table 3. Summary of BLAST hits from Entrez database queried with CLIPB17.........69
Table 4. Melanized and non-melanized parasites in different genetic backgrounds......99
FIGURES

Figure 1. Complete life cycle of Plasmodium in mosquito and man.......................6
Figure 2. Schematics of the prophenoloxidase cascade........................................22
Figure 3. Domain structure of a typical Clip domain serine protease in insects...........26
Figure 4. Genetic markers defining the *Plasmodium* refractoriness QTLs................29
Figure 5. TaqMan assay principle..............................................................................39
Figure 6. Minimal BAC contig across the *Pen1* region...........................................45
Figure 7. Diagram of the 528 kb *Pen1* sequence...................................................47
Figure 8. Example of intragenic sequence variation................................................49
Figure 9. Percent identity plot between the variant 8N20 and the consensus.................51
Figure 10. Alignment tree showing strain-specific clustering....................................54
Figure 11. RT-PCR showing expression of predicted novel genes..............................58
Figure 12. Micro-synteny analysis of the 528 kb region.............................................60
Figure 13. Annotation of the entire 1.2 Mb *Pen1* region..........................................64
Figure 14. Amino acid sequence and domain structure of CLIPB17...........................67
Figure 15. Multiple sequence alignment of the protease domain of CLIPB17 ..........70
Figure 16. Homology-based structure prediction for the protease domain.................72
Figure 17. RT-PCR assay of developmental expression profile for CLIPB17.................73
Figure 18. RT-PCR assay of infection expression profile for CLIPB17........................76
Figure 19. Amplification and standard curve graphs for S7 and CLIPB17...................79
Figure 20. qRT-PCR assay of non-infected mosquitoes.................................81
Figure 21. qRT-PCR assay of infection expression profile.................................82
Figure 22. Normal and melanized parasites in midguts.................................84
Figure 23. Localization of CLIPB17 in the midgut and abdomen tissues.............85
Figure 24. Temporal expression of CLIPB17 after infection..............................87
Figure 25. CLIPB17 up-regulation is specific for Plasmodium.............................89
Figure 26. Strain-specific mutations in the regulatory region of CLIPB17..........91
Figure 27. Immune-related binding sites within the regulatory region...............92
Figure 28. Effect of double-stranded RNA on expression of CLIPB17.................97
Figure 29. Effect of double-stranded RNA on melanization............................98
ACKNOWLEDGMENTS

First and foremost in the long list of people whom I ought to express my gratitude include the members of my advisory committee, whose patience and encouragement has been beyond human expectations. To Drs. Frank Collins, Joseph O’Tousa, David Severson, and Michael Ferdig: thank you.

My family has also been very supportive, even though they may not understand exactly what I have been doing and why I have been away from them for so long. To them I offer my love, often un-expressed but always un-dying, and my assurance that what I do can help people, if not exactly saving the world.

The work that I present here benefited greatly from our collaboration with Professor Fotis Kafatos and colleagues at the European Molecular Biology Laboratory in Germany. Each of the two trips that I took to the EMBL had been a great experience not only of science but also of life.

People from the lab of Frank and Nora have been supportive friends and colleagues of mine over the years we have been together, especially such old timers like Jim Hogan, Jonathan Kayondo and others too numerous to mention. Katie Cybulski has been really helpful in getting me to meetings and back without much hassle. It has always been an amazement to me every time I realize how she is so equally nice to people who come up just for a chat and to people like me who come up to ask a request that she must drop everything she has been doing to accommodate. Julie and I have spent
a lot of time in our insectary doing the amazing feat of working and talking at the same
time about such healthy subjects as science, politics, family, vacation, and about our
favorite subject of all time: mosquitoes.

Last but not least, I am very grateful for the people from Freimann animal facility
for the time we spent together handling animals from rat to monkey.

I have enjoyed great friendships from a few people met through social dancing.
My dance partners and also partners in crimes include Jenny who is so talented and pretty
that it defies reasons, Greg who hangs out with a lowly graduate student even though he
hails from the lofty stature of a Notre Dame professor, Sarah who graduated and move to
a big city leaving an empty corner in my heart that I am still trying to fill. They and other
social friends have been the gate keepers of my sanity during those times when I was
tempted by the dark side.

But most of all, I dedicate this thesis to my grandmother, who passed away when
I was oversea. I regretted not to be with her during her last moments, but this dedication
will keep the bittersweet memory of her alive for as long as this thesis is appreciated.

With love,

Lucas Q. Ton
1. INTRODUCTION

1.1 A brief history of malaria:

Over a century ago, the world acknowledged Alphonse Laveran and Ronald Ross with two Nobel Prizes in 1907 and 1902 for their fundamental works in dispelling the belief that bad air caused malaria. In 1880, Laveran observed the development of parasites in human blood. A few years later, Ross reported that circular forms of these parasites were found in the gut of mosquitoes four to five days after they fed on infected blood (Ross, 1898). He went on to study in more details the development of avian parasites in mosquitoes, and together with the Italians led by Giovanni Grassi, demonstrated that the protozoan parasites of the genus *Plasmodium* were transmitted by *Anopheles* mosquitoes. This discovery of the mode of transmission of malaria, together with similar reports that mosquitoes also transmitted the filarial worm *Wuchereria bancrofti* (Manson, 1878) as well as the yellow fever and dengue viruses (Lehane, 1991), demonstrated the prominent roles of mosquitoes as vectors of diseases and laid the foundations for medical entomology at the turn of the 20th century. Within a few years, transmission of other important diseases such as African sleeping sickness and Chagas disease by insects were also confirmed scientifically. Another milestone in the history of malaria research was achieved in 1948 when Shortt et al. observed schizonts and merozoites of *P. vivax* in liver tissues, completing the description of all stages in the parasite’s life cycle (Shortt et al., 1949). Other notable events in the turbulent history of this disease include the discovery of the insecticidal properties of Dichloro-diphenyl-
trichloroethane (DDT) in 1939 by the Swiss scientist Paul Muller, the production of the first synthetic anti-malarial drug Chloroquine in the USA during World War II, the success and failure of the global eradication programs of the 1950s and 1960s, the subsequent recognition that eradication was not achievable in many areas of the world and that the goal should instead be malaria control, and the unfortunate re-emergence of malaria during the 1980s and 1990s in areas such as sub-Saharan Africa, where the global eradication program was never started (Nchinda, 1998). Other drugs and insecticides as well as the first synthetic vaccine were developed during this time, but resistance and sustainability issues quickly reduced their effectiveness. In the last decade, recognition of the deteriorating malaria situation has led to renewed attention and the founding of several initiatives such as Roll Back Malaria Partnership (RBM) and Multilateral Initiative on Malaria (MIM) to improve public health system and advance more research toward novel malaria intervention (Amorosa et al., 2005).

1.2 The burden of malaria:

Even today, malaria remains a significant public health burden. Nearly ninety countries containing one third of the world’s population are at risk of the disease (Snow et al., 1999). Most estimates suggest that there are 300 – 500 million clinical cases of malaria annually, causing over 1 million deaths or about 3000 deaths a day (Greenwood et al., 2005). Countries in the sub-Saharan region of Africa have the heaviest malaria burden. The World Health Organization (WHO) reported in 1996 that malaria contributed 2.3% to global diseases but reaching 9.0% of all diseases in Africa, where it ranked third among infectious diseases after pneumococcal acute respiratory infection and tuberculosis (Nchinda, 1998). Most of the deaths occur in African infants under 5
years of age, or are the result of low birth weight children born from mothers who contracted malaria during their pregnancies.

The pathogenesis of malaria includes anaemia, febrile episodes coinciding with erythrocyte destruction, and more severe manifestations such as respiratory distress and cerebral malaria. In a study of clinical cases among African children in areas with high malaria transmission, severe anaemia and cerebral malaria were found most often among complicated manifestations, but respiratory distress was the most dangerous because it was often found in combination with other types of infection (Schellenberg et al., 1999). Besides its direct effect on mortality and morbidity, malaria can have more chronic economic and social costs. Malaria can reduce school attendance or work productivity from the sick and those who must care for them. Evidence suggests that cerebral malaria can result in long term cognitive disabilities and impaired development (Fernando et al., 2003). The economic cost of malaria upon highly endemic countries is also significant, often reducing 2% from yearly economic growth compared to other countries with similar geography but where malaria does not occur (Chima et al., 2003). Furthermore, malaria disproportionately affects poorer individuals, who have a harder time reaching and affording preventive and treatment measures (Barat et al., 2004). It was estimated that malaria alone cost Africa 1.8 billion USD in 1995 (MacLean et al., 1997), reaching 12 billion USD in 2002 (Sachs and Malaney, 2002).

Malaria has been recognized as a re-emerging disease based on statistics collected during the last decade (Nchinda, 1998). A number of factors appeared to be contributing to this resurgence. Social factors included civil disturbances hindering malaria treatment programs and creating susceptible conditions for increased malaria transmission and
epidemics. Increasing travel from non-endemic to endemic areas put more non-immune people at risk. Geological factors such as climate instability from changing rainfall patterns necessitated water development projects such as dams and irrigations which created more breeding sites for mosquitoes. Changes in behavior of mosquitoes and parasites toward insecticides and drugs constituted the most important reasons for the current malaria problem. Over the years of insecticide and drug usage, mosquitoes and parasites have developed resistance to most chemicals. Insecticides such as DDT for residual house spraying was first used in the 1950s and early 1960s to successfully eradicate malaria from Europe and America, but resistance to this chemical has quickly developed, leading to the decision to stop the global eradication program in the late 1960s by the WHO. Although currently banned in the USA and most parts of the world, DDT continues to be used in some parts of Africa with notable success (Mabaso et al., 2004). The use of bednets treated with pyrethroid insecticides has resulted in reduction of malaria mortality and health benefits to children and pregnant women in Africa (Phillips-Howard et al., 2003), but resistance to pyrethroids has also been emerging in this continent (Brooke et al., 2001). Drugs such as Chloroquine and other quinoline compounds once effective as first-line treatment now fail almost everywhere (Wellems and Plowe, 2001). Reduced sensitivity to other drugs such as the antifolates has also been documented (Plowe et al., 2004; Ridley, 2002).

1.3 Development of *Plasmodium* in the host and vector:

Four species of *Plasmodium* account for all infections in man. *P. falciparum*, *P. vivax*, *P. ovale*, and *P. malariae* share a similar life cycle with minor differences (Figure 1). Once injected into the human host by the mosquito bite, all four species penetrate
hepatocytes of the liver. However, *P. falciparum* and *P. malariae* trigger immediate transformation into liver schizonts (schizogony) whereas *P. ovales* and *P. vivax* sporozoites may either trigger immediate schizogony or have a delayed trigger, resulting in resting stages called hypnozoites. Some *P. vivax* strains seem to consist of sporozoites with universally delayed trigger, lasting up to 10 months. Another variance in the life cycle of these human malaria agents stems from the observation that *P. malariae* can also infect higher primates in their natural environments.

When an infected female mosquito bites, she injects a small number of sporozoites (usually <25) from her salivary glands during blood feeding. These travel via the blood stream to the liver where they invade liver cells within an hour. In this environment, the parasites become either hypnozoites or multiply asexually to become many thousands of liver merozoites before the rupture of the infected liver cells, releasing these merozoites into the blood stream where they must rapidly invade erythrocytes (Johnson *et al.*, 1980). Inside the red blood cells (RBC), asexual development continues and the parasites transform into the characteristic ring-stage trophozoites (Francis *et al.*, 1997). The trophozoites further mature into the rounded intra-erythrocytic schizonts, each gives rise to 8-32 merozoites until the infected RBC finally bursts. Disruption of the RBC releases the merozoites, which find their way to another RBC within a few seconds, starting a new cycle of intra-erythrocytic asexual replication. This cycle of asexual reproduction of parasites is synchronized among the infected RBCs, and the synchronized releases of merozoites mark the characteristic cyclic chill and fever episodes classically associated with malaria pathogenesis. While this propagation of asexual schizogony
Figure 1. Complete life cycle of *Plasmodium* in mosquito and man. Oval highlights the time when ookinetes must penetrate the mosquito midgut before transforming into oocysts. Figure is adapted from Vickerman and Cox, 1967.
goes on, some trophozoites develop into sexual forms of the parasites, termed micro and macro gametocytes, which are infective for the female *Anopheles* mosquitoes. Within the blood meal in the mosquito midgut, these gametocytes differentiate into gametes which fuse to form a zygote, restoring the diploidy of the chromosomes for later asexual replication. The zygotes further develop into motile banana-shaped ookinetes that move across between or through midgut epithelium cells to rest beneath the basal lamina (Sinden, 1984). Here they transform into spherical oocysts. Inside the oocysts, threadlike sporozoites are produced through asexual multiplication until the oocysts burst and release thousands of sporozoites into the mosquito body cavity (Pringle, 1965). Some of these sporozoites quickly find their way to the salivary glands where they once again cross through the epithelium. When the mosquito feeds again, a few sporozoites are injected into the human host, continuing the transmission cycle.

1.4 Vector competence:

Transmission of malaria parasites falls under the category of “biological transmission”, which involves an obligatory period of development by the parasites in the mosquito vector. Success of this sporogonic developmental period determines whether an insect is competent. Otherwise, the insect is either insusceptible or refractory to parasitic infection. An insusceptible insect (non-vector) is one in which the parasite cannot successfully infect, while in a refractory insect, its body tissues effectively reduce or abolish an established infection (Sinden *et al.*, 2004). Not all mosquitoes are vectors, and not all vectors can support parasitic development from all *Plasmodium* species. From 422 species of anopheline mosquitoes known to date, more than 70 species can transmit human malaria parasites, but only about 40 species are considered to be
important vectors (Service, 1993). Furthermore, other non-human parasites show
different degrees of compatibility to different mosquitoes. For example, the rodent
parasite *P. berghei* is unable to develop in *Aedes* mosquitoes, while it goes through
sporogony efficiently in *Anopheles*. The avian parasite *P. gallinaceum* is similarly
incompatible to most strains of *Culex* mosquitoes (Sinden *et al.*, 2004). These specific
vector/parasite compatibilities suggest that vector competence is determined by a number
of physiological and genetic factors.

A case in point for a factor affecting competence can be seen in the physiology of
peritrophic membrane (PM) formation. This extracellular membrane starts to form in the
midgut lumen only after the mosquito has taken a blood meal, and surrounds the food
bolus once fully formed, separating it from the midgut epithelium. Although it is not an
absolute barrier preventing parasites from entering the body of susceptible mosquitoes,
the PM can still limit the intensity of infection. In *A. atroparvus*, the PM takes 24 hours
to fully form. When ookinetes of *P. berghei* were fed to this mosquito, no gametocyte
maturation period (about 24 hours) was necessary, and these ookinetes immediately
invaded the midgut epithelium at the time when the PM has not yet fully formed. As the
consequence of side-stepping the PM, more parasites penetrated the midgut (Janse *et al.,
1985). On the contrary, if the PM is artificially thickened, fewer ookinetes successfully
penetrate the gut (Ponnudurai *et al.*, 1988).

The importance of the genetic basis of susceptibility to *Plasmodium* and other
parasites was first recognized by Clay Huff over 60 years ago when he demonstrated that
*Culex pipiens* progenies from a selective mating experiment maintained as a colony for
several generations without a loss in susceptibility to an avian parasite (Huff, 1956).
From the time the first genetic linkage map for *Ae. aegypti* was assembled with 77 morphological and isozyme markers (Munstermann *et al* 1979), a number of investigators have determined the loci affecting susceptibility of different laboratory strains to a number of pathogens, including *Plasmodium* and filarial worm species. Macdonald demonstrated that susceptibility to the human filarial worm *Brugia malayi* was controlled by a single sex-linked locus which he designated $f^m$ (Macdonald and Ramachandran, 1965). However, a different locus on the X linkage group named $f^t$ controlled susceptibility of the same mosquito to a different worm, one that infects dog (McGreevy *et al*., 1974). Kilama and colleagues showed that susceptibility to *P. gallinaceum* was influenced by yet another locus named $pls$, one that was autosomal and dominant (Kilama and Craig, 1969). From these studies, it seemed that susceptibility of *Ae. aegypti* to pathogens was inherited in the Mendelian fashion; However, genetic experiments using a more detailed map based on molecular markers revealed that vector susceptibility is a quantitative trait with multiple interacting loci (QTL), including an additional locus named $fsb2$ for susceptibility to *B. malayi* (Severson *et al*., 1993), and another locus named $pgs3$ to *P. gallinaceum* (Beerntsen *et al*., 1995).

*Anopheles* species are the most important vectors of human malaria. In sub-Saharan Africa, *A. gambiae* is responsible for most transmission of *P. falciparum*, the deadliest among the four human parasites. The importance of this species to public health has prompted tremendous research effort to investigate the genetics and biology of vector competence. The research findings most relevant to this thesis include those on the genetics of mosquito refractoriness to *Plasmodium* and those on the biology of mosquito immune system responsible for that refractoriness. The significance of
mosquito immunity toward refractoriness will be discussed below in the broader context of insect immunity.

1.5 A historical perspective on insect immunity:

Insects form the most successful animal group in terms of species diversity and abundance. To date, about one million species are described, which is more than the combination of all other animal species. They occupy most terrestrial and aerial ecological niches on our planet, and often thrive in microbe-rich environments. Part of their success is likely because they have evolved an efficient immune system to recognize and eliminate pathogenic threats. Even though insects possess only an innate immune system, lacking the specificity from genetic rearrangement seen in adaptive immunity of vertebrates, insect immunity has the advantage of a rapid response to infection, often taking a few hours instead of days for vertebrates to fully respond to their first encounter with pathogens (Gotz and Vey, 1974).

Research on insect immunity began as early as the late nineteenth century, when Louis Pasteur demonstrated that the pebrine disease devastating the French silk industry was caused by a micro-organism infecting the silkworm *Bombyx mori*. He showed that *Nosema bombysis* could pass directly from silkworm to silkworm, and its passage could be prevented with sanitary and uncrowded rearing conditions (reviewed in Brey, 1998). During the first half of the twentieth century, Metchnikoff and others described their observations of insect macrophage-like cells phagocytosing bacteria, or forming multicellular capsules around foreign material too large to engulf, leading to a darkening phenomenon (melanization) at the sites of infection. Based on these studies, many scientists at the time speculated that cellular rather than humoral immunity constituted the
primary mechanism of insect defense. However, the study of humoral immunity took a sharp turn when the first cell-free antibacterial peptides were isolated from the moth *Hyalophora cecropia* (Steiner et al., 1981) and were named Cecropins and Attacins. A host of additional peptides were soon discovered in other insects, confirming that the rapid *de novo* production of short antimicrobial peptides in response to infection is a hallmark of insect humoral immunity (Boman, 1995). When the genes for Attacins were cloned, researchers discovered deca-nucleotide sequences upstream of each coding gene that matched the consensus binding motifs of the vertebrate NF-κB family of transcription factors (Sun et al., 1991). During the following year, a NF-κB factor from *H. cecropia* was isolated, and its binding to Attacin target sequence was subsequently demonstrated (Sun and Faye, 1992). In vertebrates, NF-κB-mediated signaling is responsible for induction of the pro-inflammatory response via expression of antimicrobial molecules, cytokines and other stimulatory molecules that coordinate with the adaptive branch of immunity for efficient response to most pathogens (Pasare and Medzhitov, 2004). The possibility of a functional similarity between vertebrate and insect NF-κB transcription factors marked the beginning of the realization that the basic mechanisms of pathogen recognition and activation of responses might be evolutionary conserved (Hoffmann et al., 1999). Proofs of this conservation have been discovered from genetic and mutagenesis studies of the fruit fly *Drosophila melanogaster* (Silverman and Maniatis, 2001). The recent completion of the genome sequences of the fruit fly and other insects such as the mosquito *A. gambiae* has opened the way to uncover more details of key components and mechanisms of insect immunity in general.
and mosquito immunity in particular (Khush and Lemaitre, 2000). Some of those details will be discussed in the following sections.

1.6 Recognition of foreignness:

Any successful immune response must necessarily begin with the recognition of non-self. For many years, immunologists focused mostly on antigen recognition by the vertebrate adaptive immune system. In 1989, however, Charles Janeway Jr formulated the pattern recognition concept in innate immunity. He hypothesized the existence of pattern recognition receptors (PRRs) with the capacity to detect pathogen-associated molecular patterns (PAMPs) uniquely presented in microbial cells (Janeway, 1989). He speculated that molecules present on a microbe must fulfill some basic criteria to become good PAMPs; namely, it must be widespread among large groups of microbes and it should be fairly evolutionary conserved. This method of recognition enables a few PRRs to potentially recognize a large number of microbes. PAMPs that are known to trigger immune reactions from insects include peptidoglycans (PGNs), lipopolysaccharides (LPS), and β-1,3 glucans. Peptidoglycan is a polymeric molecule present in the cell wall of most bacteria. Gram-positive bacteria have a thick cell wall made up from about 20 to 80 layers of PGN. In contrast, Gram-negative bacteria have only one to three layers, and the PGN is located in the periplasmic space between the plasma membrane and the outer membrane. This outer membrane is a lipid bilayer, of which the outer layer contains a high concentration of LPS, making LPS an essential component of Gram-negative cell membrane. Many fungi contain a cell wall with polymer of β-1,3 glucans not present in any other group of animals, and thus constitute a good fungal target for non-self recognition (Brown and Gordon, 2003).
Peptidoglycan recognition proteins (PGRPs) are the best-known PRRs in insects. It was first purified from *B. mori* as a protein that could bind to PGN and was required for prophenoloxidase cascade activation (Yoshida *et al.*, 1996). A few years later, homologous PGRPs were cloned from a number of insects, including the lepidopteran *Manduca sexta* (Zhu *et al.*, 2003) and the dipteran *D. melanogaster* (Michel *et al.*, 2001). When the genome sequencing of *D. melanogaster* was completed in the year 2000, it was found that this fruit fly had 13 PGRP genes (Khush and Lemaitre, 2000), some of which were up-regulated after bacterial infection (Werner *et al.*, 2000). PGRP-SA was shown to mediate Toll signaling pathway activation in response to infection with Gram-positive bacteria (Michel *et al.*, 2001). PGRP-LC and PGRP-LE were shown to activate another important signaling pathway named IMD in response to Gram-negative infection (Choe *et al.*, 2002; Takehana *et al.*, 2002). It was also shown that PGRP-LC might be important in a phagocytic response against the bacteria *E. coli* (Ramet *et al.*, 2002b) while PGRP-LE played a possible role in triggering melanization (Takehana *et al.*, 2002).

Seven PGRPs were identified in *A. gambiae* genome (Christophides *et al.*, 2002). They were grouped into short (PGRP-S) secreted and long (PGRP-L) membrane-bound forms as seen in *Drosophila*. At least one of them (PGRP-LB) is transcriptionally up-regulated in *Plasmodium*-infected mosquitoes (Dimopoulos *et al.*, 2002). One PGRP was isolated from the haemolymph of the beetle *Holotrichia diomphalia* and was found to bind both PGN and β-1,3 glucan (Lee *et al.*, 2004). Interestingly, it was also shown that binding to β-1,3 glucan activated the prophenoloxidase cascade, but this activation was not triggered with PGN.
Gram-negative binding proteins (GNBPs) are another class of PRR. They are known to bind strongly to the surface of Gram-negative bacteria (Lee et al., 1996), and they are thought to recognize LPS and β-1,3 glucan (Kim et al., 2000), but not PGN and other microbial compounds. However, when the Drosophila GNB1 associates with PGRP-SA, the resulting complex activates Toll pathway upon infection with Gram-positive bacteria (Gobert et al., 2003). Three GNBPs have been identified in Drosophila, while six were detected in Anopheles. Some of them were induced in response to bacterial or Plasmodium infections (Dimopoulos et al., 1998). Another insect PRR that can bind to LPS or β-1,3 glucans is the C-type lectins (CTLs) (Weis et al., 1998). These carbohydrate-binding proteins activate the complement system in vertebrates (Brown and Gordon, 2003). Two CTLs were isolated in the tobacco hornworm M. sexta. They were shown to bind to LPS and initiate the prophenoloxidase cascade (Yu and Kanost, 2004). The Anopheles genome encodes 22 CTLs, two of which have been shown to affect Plasmodium development with RNAi-based knockdown assay. Lack of CTL4 produced melanization-based killing, reducing the number of developing oocysts by 97%. The phenotype of CTLMA2 was similar but not as striking, and when this gene was silenced, the mosquitoes melanized 48% of Plasmodium oocysts (Osta et al., 2004).

Proteins binding to β-1,3 glucans have been identified first in B. mori and then in several other arthropod species (Kanost et al., 2004). These β-1,3 glucan recognition proteins (βGRPs) have been implicated in affecting the prophenoloxidase cascade in several insects (Ma and Kanost, 2000). Recently, one βGRP cloned from the mosquito A. subalbatus showed up-regulation when this mosquito had been infected with bacteria and filarial worms. Silencing of this gene by RNAi strongly inhibited melanization of the
filarial worm *Dirofilaria immitis* (Wang et al., 2005), suggesting that this protein may function as the PRR for melanization of worms in mosquitoes.

1.7 Organs and cells of insect innate immunity:

Once invading organisms have successfully crossed over the chitinous exoskeleton or the epithelial layer of the insect gut, they are then encountered by the humoral and cellular defensive mechanisms of the insect’s blood system in the hemocoelic space. The blood, or haemolymph, consists of a fluid medium in which a variable number of hemocytes are dispersed. Since insects lack a blood vessel system, the haemolymph is separated from insect organs only by their own basal lamina. The circulating hemocytes mediate the cellular responses of the insect immune system. They are classified somewhat differently among the insects, but overall five types of hemocytes have been found in insect blood (Lavine and Strand, 2002). Important cell types include the amoeboid plasmatocytes involving in phagocytosis and encapsulation responses, the granular cells that also participate in phagocytosis and capable of releasing their granules when in contact with a foreign surface, and the oenocytoids believed to involve in melanization. Plasmatocytes are usually the most abundant cells among the hemocytes (Lanot *et al.*, 2001). They mostly resemble mammalian monocyte/macrophage cells, and are capable of receptor-mediated engulfment of invading bacteria as well as the insect’s own apoptotic cells (Lanot *et al.*, 2001; Ramet *et al.*, 2002b). The main characteristic of granulocytes is the presence of numerous membrane-bound vesicles, many of which contain enzymes usually found in lysosomes. The oenocytoids contain activity indicative of phenoloxidase, an enzyme in the melanization pathway (Hillyer *et al.*, 2003).
Besides containing the circulating hemocytes, the haemolymph is also an important medium for humoral immunity by virtue of its direct contact with the collection of cells that made up the fat body. By analogy, insect fat body plays a role equivalent to the mammalian liver, produces and releases humoral factors such as the antimicrobial peptides (AMPs) into the haemolymph for a systemic response to diverse pathogens from bacteria to fungi (Meister et al., 2000).

1.8 Signaling pathways in humoral immunity:

The production of AMPs in *Drosophila* is regulated by two major signaling pathways. The Toll pathway mainly regulates antifungal response and that against Gram-positive bacteria, while Gram-negative bacteria trigger the immune deficiency (IMD) pathway. In the early 1990s, it was shown that the promoter regions of AMPs contained sequence motifs related to mammalian NF-kB response elements found in the promoters of genes regulating inflammatory reactions (Engstrom et al., 1993; Kappler et al., 1993). NF-kB transcription factor is found in the cytoplasm bound to an inhibitory protein, IκB. When signaling cascades induced by an antigen such as LPS or by cytokines such as Interleukin 1 or Tumour Necrotic Factor α are activated, IκB protein becomes phosphorylated and subsequently degraded, freeing NF-kB to translocate into the nucleus where it binds to κB motifs to direct transcription of target genes. It was soon discovered that *Drosophila* has three Rel-homology (homologous to NF-kB) transcription factors and that the events of their activation and translocation are similar to mammalian NF-kB. One of the fly Rel-homology factor, Dorsal, is bound to the IκB-like protein Cactus, and when freed from Cactus, Dorsal translocates to the nucleus to activate genes controlling dorso-ventral polarity during embryogenesis (Ip et al., 1991; Thisse et al., 1991).
Another Rel-homology factor, named Dorsal-related immunity factor (Dif), is utilized the same way but by the immune system to control production of the AMP Drosomycin (Lemaitre et al., 1996). While activity of Dorsal and Dif are controlled by the Toll signaling cascade, the third Rel-homology protein, Relish, is controlled by the IMD pathway (Dushay et al., 1996). Instead of being bound to Cactus, the Relish protein combines the Rel-homology domain with a C-terminal IκB-like inhibitory domain. When IMD pathway is immune activated, the inhibitory domain is proteolytically cleaved and the Rel-homology domain is set free to translocate into the nucleus.

While components downstream of the Toll and Imd receptors are relatively well-characterized, the extracellular recognition and signaling events remain somewhat elusive. For example, the signaling events after PGRP-SA has bound to Gram-positive bacteria remain uncharacterized, although it is likely that serine proteases would be proteolytically activated in sequential fashion, eventually leading to cleavage of the cytokine-like Spatzle to its active form, which can bind to Toll receptor on the membrane of fat body cells. This method of activation is also observed during embryonic patterning when Spatzle is proteolytically cleaved by the sequential actions of three serine proteases named Gastrulation defective, Snake, and Easter (LeMosy et al., 1999). However, in null mutants of these proteins, immune-induced activation of Toll pathway is not affected (Lemaitre et al., 1996), indicating that other serine proteases are involved. At least one of these immune responsive serine proteases must be controlled by Spn43Ac, a serine protease inhibitor of the serpin type. When the Spn43Ac mutants called Necrotic were examined, it was shown that these flies predominantly contained activated Spatzle, and Drosomycin was constitutively expressed (Levashina et al., 1999). When the pathway is
activated by fungi, the initial recognition and signaling events are even less understood. The pattern recognition receptor for fungal antigen is still unknown, but two Gram-negative binding proteins (GNBPs) were implicated from a microarray study of immune-regulated genes (De Gregorio et al., 2002). Genetic screens established that the serine protease Persephone and the serine protease inhibitor Necrotic act upstream of Spatzle in the fungal-mediated signaling events (Ligoxygakis et al., 2002a).

Since its discovery in 1995 from a genetic screen of Drosophila mutants, the IMD pathway has been shown to be important against Gram-negative infection (Lemaitre et al., 1995) but does not affect embryogenesis. This pathway controls the expression of the AMPs Dipterisin, Attacin, Cecropin, Drososin, and Defensin (Engstrom, 1999). The membrane-bound PGRP-LC activates the pathway when it associates with both the bacteria and the Imd protein (Choe et al., 2005). However, PGRP-LC might not be the only receptor because its null mutation phenotype still showed partial effect of IMD-dependant immune activation (Gottar et al., 2002). Support of this observation came with the discovery that PGRP-LE, a haemolymph PGRP, also activates the IMD pathway (Takehana et al., 2002).

The Toll and IMD pathways do not appear to share any intermediate components, and the two pathways control most of the immune gene expression in Drosophila. Flies that are simultaneously mutated in Toll and IMD pathways are totally unresponsive in AMP production. Such flies were subjected to an immune gene oligonucleotide microarray study in which gene expressions of 400 Drosophila immune-regulated genes (DIRGs) were compared under different immune challenges. DIRG expressions from these flies were also compared, under same immune challenges, to that of wild-type flies.
and flies defected in only one of the two pathways (De Gregorio et al., 2002). This study showed that nearly 300 genes were significantly affected (induced or repressed) by a septic injury in the wild-type flies and that about 70% of these were affected in the various mutant flies. Genes controlled by the Toll pathway included several serine proteases and their inhibitors, as well as genes in the Toll pathway itself. Some of these serine proteases have putative functions in melanization by virtue of their sequence similarity to known enzymes of this immune process. The IMD pathway controlled most AMP genes and other enzymes involved in the melanization process, as well as several PGRPs. However, it was also found that some immune-regulated genes were not at all affected in the Toll/IMD double mutants, suggesting that other pathways also regulate the expression of some immune genes.

In addition to Toll and IMD, two other signaling pathways involved in bacterial-induced immunity have been described in Drosophila, namely the Janus kinase (JAK)/signal transducer and activator of transcription (STAT) and the c-Jun N-terminal kinase (JNK) pathways (Boutros et al., 2002). Similar to Toll pathway, JAK/STAT pathway has dual functions in embryonic development and immunity (Zeidler et al., 2000). STAT has been shown to translocate to the nucleus in response to bacterial infection, and it controls the production of secreted opsonins such as Tep1 (Lagueux et al., 2000). Tep1 is also controlled by the Toll pathway, as it is constitutively expressed in Toll gain-of-function mutants. The JNK pathway branches off from the IMD pathway downstream of the Imd protein, and this pathway appears to control cytoskeletal function and wound healing (Ramet et al., 2002a).
Most components of the Toll and IMD pathways have been found in the mosquito \textit{A. gambiae}. There are 10 Toll receptor homologues, six Spatzle-like proteins, and counterparts for all other components of these pathways except for Dif (Christophides \textit{et al.}, 2004). Components of the other immune signaling pathways are not yet well characterized in \textit{Anopheles}. However, it is known that bacterial challenge induces Ag-STAT to translocate into the nucleus and binding to STAT response elements (Barillas-Mury \textit{et al.}, 1999). Although the \textit{Anopheles} TEP1 homolog contains STAT response elements in its promoter region, there is currently no data supporting that it is controlled by the JAK/STAT pathway.

\textit{Anopheles gambiae} TEP1 is a member of a family of Thioester-containing proteins. In vertebrate, this family includes complement factors and the universal protease inhibitor \textalpha{}2-macroglobulin which play central roles in complement activation in immunity. In \textit{Anopheles}, TEP1 has been shown to bind to bacteria to promote phagocytosis (Levashina \textit{et al.}, 2001). More recent data showed that TEP1 could also bind to \textit{Plasmodium} to promote killing of the parasites by melanotic encapsulation of ookinetes (Blandin \textit{et al.}, 2004).

1.9 Melanotic encapsulation:

Melanin synthesis plays important roles in developmental processes such as the tanning of mosquito embryos and physiological processes such as the hardening of the exoskeleton and wound healing (Christensen \textit{et al.}, 2005). More importantly, melanization is an important innate immune response in insects, and once triggered, it results in a rapid response against bacterial and fungal infections, often taking just minutes to a few hours to kill these pathogens (Hillyer \textit{et al.}, 2003). Melanization is also
the most important mechanism in the killing of pathogens too large for phagocytosis, such as parasitoid wasp eggs, Plasmodium parasites, or filarial worms (Vey 1993). In this process, melanin is synthesized from the amino acids tyrosine or phenylalanine through a series of enzymatic and non-enzymatic reactions collectively called the Prophenoloxidase cascade, a description based on the fact that the first and most important enzyme of this cascade is Phenoloxidase (PO). The importance of PO is also reflected from the fact that it is often synthesized as an inactive form named Prophenoloxidase (PPO) in insect hemocytes to prevent adverse effects to these blood cells (Christensen et al., 2005). In insects with many circulating hemocytes in the haemolymph, melanization involves the collaborative accumulation of numerous hemocytes around the pathogen. Hemocytes such as the granular cells and oenocytoids then release components of the PO cascade. Some time later the pathogen becomes surrounded by plasmaticocytes which progressively flatten over the surface, forming an enveloping capsule many cells thick intercalating with newly-formed melanin (Hillyer et al., 2003). In insects with few hemocytes such as the adult mosquitoes, the formation of melanin capsule around pathogens occurs without this multicellular structure, and this process is accordingly termed humoral melanization (Gotz 1986).

Whereas AMP signaling pathways have been primarily dissected with genetic analysis, melanotic encapsulation has been studied with biochemical methods in large insects such as B. mori and M. sexta (Brey, 1998). Melanin synthesis begins with either available tyrosine amino acid, or with an enzymatic conversion of phenylalanine to tyrosine (Figure 2). This amino acid is then hydroxylated with PO to dopa which can be further oxidated to dopaquinone, also by PO. Dopaquinone converts spontaneously to
Figure 2. Schematics of the Prophenoloxidase cascade with its regulation by the serine protease cascade whose last member is usually called Prophenoloxidase-activating enzyme (PPAE).
dopachrome, which is then decarboxylated by Dopachrome conversion enzyme (DCE) to 5,6-dihydroxyindole (DHI). Further oxidation catalyzed by PO converts DHI to indole-5,6-quinone that can polymerize to produce eumelanin. Dopa can alternatively be decarboxylated to form dopamine, which can be further oxidized to produce DHI and melanin (Christensen et al., 2005).

As mentioned above, inactive prophenoloxidase (PPO) undergoes enzymatic proteolysis to become functional as the result of a serine protease cascade that transduces pathogen recognition signal via sequential activation of its protease members (Kanost et al., 2001). The last serine protease of this cascade is usually called prophenoloxidase-activating enzyme (PPAE) in a number of insects. To date, three of such enzymes have been discovered in *M. sexta* (PAP1, PAP2 and PAP3) (Jiang et al., 2003), one in *B. mori* (PPAE) (Satoh et al., 1999), one in the beetle *H. diomphalia* (PPAF1) (Kim et al., 2002), one in the crayfish *P. leniusculus* (ppA) (Wang et al., 2001). No PPAE enzyme has been characterized in *Drosophila*, but at least five putative genes are present in the genome sequence (Adams et al., 2000).

The importance of melanin synthesis and melanization toward mosquito refractoriness against *Plasmodium* and other parasites has prompted several studies to clone and characterize the enzymes of this process. In *A. gambiae*, five serine proteases have been cloned by a conventional Polymerase Chain Reaction (PCR) method using primers designed from conversed regions of several PPAEs from other insects (Gorman et al., 2000). Three of the cloned serine proteases mapped to division 14 on the *A. gambiae* polytene chromosome map and consequently named Sp14A, Sp14D1, and Sp14D2. The other two serine proteases are similarly named Sp18D and Sp22D. From
their domain architecture, it seems that Sp22D can at least be a pattern recognition serine protease because it has two chitin binding domains, one mucin-like region, two low density lipoprotein receptor class A, two scavenger receptor domains in addition to the catalytic protease domain. Some of these domains have been implicated in pathogen recognition leading to PPO activation (Ji et al., 2004). These five serine proteases belong to a large protein family in which the enzyme Trypsin is a member. Trypsin-like serine proteases of the S1 family (Rawlings and Barrett, 1993) play important roles in processes such as digestion, blood coagulation, immunity, and embryonic development (Jiang and Kanost, 2000). A common feature of these serine proteases is that they contain a C-terminal protease domain with a catalytic center composed of histidine, aspartic acid, serine triad. In most S1 family serine proteases, these residues are embedded in the highly conserved sequences of TAAHC, DIAL, and GDSSGDP, respectively (Ross et al., 2003). Within this protease domain, six conserved cysteine residues form three disulphide linkages. Interestingly, all known PPAEs have two extra cysteines, an observation that suggests the short loop formed by these two cysteines could be a signature motif for PPAE. All known PPAEs also have at least one other domain at the N-terminal called a Clip domain (Figure 3). This short domain has six conserved cysteine residues, with the last two residues in tandem, an arrangement similar to that of big Defensin, an antibacterial protein isolated from horseshoe crab (Iwanaga and Kawabata, 1998). The function of Clip domain has not yet been demonstrated, but it seems that this domain plays a role in regulating protein interaction or protein activation. Ample examples can be found in which a protease is inhibited by its own propeptide. The cysteine protease cathepsin L is known to be inhibited by its propeptide, as is also
the case for a serine protease from the subtilisin family (Carmona et al., 1996; Boudreault et al., 1998). Another possible function for Clip domain would be to have antimicrobial properties, inferring from its structure similarity to big Defensin. Indeed, a recombinant Clip domain from *M. sexta* appears to be toxic to the *E. coli* host, as reported in Jiang et al. (2000). In *Drosophila*, the cytokine-like protein Spatzle found in the Toll signaling pathway has a domain structure called the cystine-knot motif consisting of six cysteine residues. However, the topology of this motif is different than that found in the Clip domain. Specifically, no tandem arrangement of the last two residues is found in the cystine-knot motif (McDonald and Hendrickson, 1993). Two more cysteine residues are present in these Clip domain proteases (c-SPs), forming a bridge between the Clip domain and the protease domain. When these c-SPs are activated, the peptide bond between Clip domain and protease domain is proteolytically cleaved, and only the disulphide bridge remains bonded between these domains. Cleaving after positively charged amino acid residues such as arginine and lysine is a characteristic of Trypsin and Trypsin-like serine proteases (Perona and Craik, 1995). All known PPAEs contain either of these two residues at the cleavage site, suggesting that the previous serine protease in the cascade should also be a S1 family member, with or without Clip domain. As of now, no knowledge of the first or the intermediate proteases in the cascade is known. However, a recent study established that a S1 family protease in *M. sexta* functioned as the pattern recognition protease triggering this cascade in response to bacterial challenge (Ji et al., 2004).

In *A. gambiae*, most enzymes in the PPO cascade have been characterized. Multiple PPOs have been studied in hemocyte-like cell lines, and their expression
Figure 3. Domain structure of a typical Clip domain serine protease in insects. Adapted from Jiang and Kanost, 2000.
profiles in mosquito have been accessed. To date, no up-regulation of any PPO gene in response to *Plasmodium* infection has been recorded, although four of the PPO genes showed up-regulation 24 hours after blood feeding (Muller *et al*., 1999). However, evidence of involvement of *Ar. subalbatus* PO I in immunity against filarial worms has been provided (Shiao *et al*., 2001). Despite the number of Clip domain serine proteases cloned in *A. gambiae* (Gorman *et al*., 2000), none has been determined to be the mosquito PPAE responding to *Plasmodium* invasion. Furthermore, no other serine protease upstream from PPAE has yet been cloned by the conventional method described above. Alternatively, a genetic approach has been pursued to delineate the melanization determining factor(s) in this mosquito.

1.10 Genetics of mosquito refractoriness to *Plasmodium*:

Classical genetics of mosquito susceptibility to parasites has been discussed above. During the early years of investigation into the susceptibility of *Ae. aegypti* to various species of *Plasmodium* and filarial worms, it was discovered that susceptibility was not a simple Mendelian trait, but rather polygenic in nature. During the course of research in the last twenty years, it appeared that such quantitative traits also governed mosquito refractoriness to *Plasmodium*. When a standard backcross scheme was performed between a laboratory selected *A. gambiae* strain fully susceptible (4Arr strain) and a strain fully refractory (L35 strain) to most *Plasmodium* species except *P. falciparum* from Africa (Collins *et al*., 1986), the researchers discovered that the killing of *P. cynomolgi* B ookinetes and young oocysts when they reached the basal lamina of the mosquito midgut by a melanotic encapsulation mechanism was in fact a quantitative trait contributed by at least three QTLs (Zheng *et al*., 1997). The authors named these
loci *Plasmodium* Encapsulation 1, 2, and 3 (*Pen1, Pen2, Pen3*) in order of their relative
collection to the trait. Based on microsatellite markers constructed in their previously
published genetic map (Dimopoulos *et al.*, 1996), they determined that the *Pen1* locus
was limited between markers *H290* and *H788*, and was in complete linkage to marker
*H175*. In this experiment, *Pen1* explained ~60% of the trait, while *Pen2* accounted for
~19%. The refractory alleles (from L35 parent) of these three QTLs were largely
inherited in an autosomal dominant fashion with simple additive effect, and all of them
together accounted for ~70% of the variance for melanotic encapsulation of ookinetes
(Figure 4). The association of *Pen1* and refractoriness was further confirmed with similar
genetic experiments in which *Pen1* was also found to be the major locus affecting abiotic
CM-Sephadex beads and *P. berghei* encapsulation (Gorman *et al.*, 1997).

When markers *H290* and *H788* were mapped to the ovarian polytene chromosome
of *A. gambiae* by *in-situ* hybridization (Kumar and Collins, 1994), they were found to
hybridize between division 8C-8D on the right arm of chromosome 2 (Collins *et al*.,
1997). Comparison to known chromosomal locations of the enzymes catalyzing the PPO
cascade and its serine protease signaling cascade revealed no possible co-localization to
*Pen1* or *Pen2* loci, however, the serine proteases at division 14 (Sp14A, Sp14D1,
Sp14D2) that could potentially be PPAEs fell within the *Pen3* locus (Gorman and
Paskewitz, 2001). Therefore, the biological identity for *Pen1* and other loci remains to be
determined.

When a gene is cloned purely from the knowledge of its genetic map position, the
approach is called positional cloning and involves correlating the genetic map with a
physical map constructed by overlapping large-insert clones such as Yeast Artificial
Figure 4. Genetic markers defining the QTLs affecting melanotic encapsulation of *Plasmodium*. Marker ND3B6 defines a smaller *Pen1* locus. More details are discussed in the text.
Chromosome (YAC) or Bacterial Artificial Chromosome (BAC). From this physical map, a transcript map is created to identify protein coding regions. With the recent exponential increase of sequencing power, a rational method to create this transcript map is to sequence the YAC or BAC clones and performing gene prediction. Each predicted gene can then be confirmed by a test of expression, and the likely candidate genes can be further examined functionally. Historically, positional cloning has been used to characterize human diseases and plant phenotypes such as Cystic Fibrosis or a fruit weight phenotype (Riordan et al., 1989; Frary et al., 2000). In these studies, candidate genes have been functionally assayed by transgenic rescue of mutant phenotype, by gene knock-out, or by association of mutant alleles to appropriate phenotype (Korstanje and Paigen, 2002; Salvi and Tuberosa, 2005).

1.11 Aims of thesis:

The region between A. gambiae polytene 8C-8D divisions on chromosome 2R probably represents a large physical region up to 3.0 megabases (Mb) of sequence. Therefore, a genetic strategy was employed to refine the Pen1 locus to a smaller region to minimize the number of candidate genes that must be analyzed later. Toward this end, the 8C-D region of ovarian polytene chromosome was micro-dissected and the resulting fragment was used as a probe to identify BAC clones from a genomic library. Sequences from these clones were examined for the simple nucleotide repeat patterns that could potentially be new microsatellite markers. These new markers were then used to genotype F11 progenies from a single parent mating between the refractory strain L35 and the susceptible strain 4Arr according to the advanced intercross line strategy (Darvasi and Soller, 1995). Analysis from this study revealed that the new microsatellite marker
ND3B6 mapped closer to Pen1 than marker H788 and together with marker H290 represent a smaller Pen1 locus of ~1.2 Mb (Hogan J, unpublished).

This thesis describes the positional cloning of the Pen1 locus as well as subsequent molecular characterization and confirmation of the candidate gene function toward melanotic encapsulation phenotype. The first aim of this work is to identify a set of BAC clones covering the Pen1 genetic interval between the markers H290 and ND3B6. The second aim of this thesis is to obtain nucleotide sequence from the Pen1 region and to analyze the sequence for gene content and other genomic properties. The third aim focuses on accessing candidate gene function toward melanization pathway, while the fourth aim centers on discovering the molecular mutations giving rise to the effect of Pen1 toward Plasmodium refractoriness.
2. MATERIALS AND METHODS

Most materials and methods used in the following experiments are standard and thus will only be described briefly. Relatively novel methods will be described in more details. Oligonucleotide primers used in the experiments are listed in Table 1.

2.1 Mosquito strains and rearing conditions:

The genetically selected *Plasmodium*-refractory strain L35 and the counter-selected *Plasmodium*-susceptible strain 4Arr were described previously (Zheng et al., 1997). Mosquito colonies of each strain were maintained in a separate environmental chamber at 27°C and 80% humidity. Larvae were fed with grounded fish food (Aquaricare) and yeast. Adults were fed 10% high fructose corn syrup.

2.2 Construction of a physical clone map across the *Pen1* region:

Bacterial Artificial Chromosomes (BACs) clones from two genomic libraries of the PEST strain (Mukabayire and Besansky, 1996) that had been end-sequenced (400 – 800 bp of one or both end of the BAC clone were sequenced) were used to construct a clone map across the *Pen1* locus. Each clone from the ND-1 library (Ke and Collins, unpublished) and ND-TAM library (Hong et al., 2003) was present in a pool of 384 clones corresponding to the 384-wells plate used to grow the bacteria. Once a plate pool was determined to contain the target clone, its rows and column pools were screened again to determine the position of the target clone within each plate. Each BAC clone was named according to its position within a plate. For example, BAC clone 25F12 was
<table>
<thead>
<tr>
<th>Role</th>
<th>Primer ID</th>
<th>Sequence (5’-3’)</th>
<th>Direction</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S7.1.F</td>
<td>GGGCAGATCATCATATCTACGT</td>
<td>Forward</td>
<td>S7</td>
</tr>
<tr>
<td></td>
<td>S7.1.R</td>
<td>GCTGTGCAAATTCCTCG</td>
<td>Reverse</td>
<td>S7</td>
</tr>
<tr>
<td></td>
<td>F11.5.F</td>
<td>GCACCAATTCCGAGCAGCTTTTC</td>
<td>Forward</td>
<td>4F11.5</td>
</tr>
<tr>
<td></td>
<td>F11.5.R</td>
<td>ATAGTTGGGTGCAGATGCTGATG</td>
<td>Reverse</td>
<td>4F11.5</td>
</tr>
<tr>
<td></td>
<td>F11.6.F</td>
<td>GACTGCTGTCAGAATGGTTACCA</td>
<td>Forward</td>
<td>4F11.6</td>
</tr>
<tr>
<td></td>
<td>F11.6.R</td>
<td>ACCCTTTTCTTTCCCCCCTATCCAC</td>
<td>Reverse</td>
<td>4F11.6</td>
</tr>
<tr>
<td></td>
<td>F11.7.F</td>
<td>TTCCATCGCTCACACACTCTAAGC</td>
<td>Forward</td>
<td>4F11.7</td>
</tr>
<tr>
<td></td>
<td>F11.7.R</td>
<td>CCATACTGGGCACACAGAGGGAG</td>
<td>Reverse</td>
<td>4F11.7</td>
</tr>
<tr>
<td></td>
<td>F11.8.F</td>
<td>AAGCTTCTCTTGCTAGCTTACTAACAC</td>
<td>Forward</td>
<td>4F11.8</td>
</tr>
<tr>
<td></td>
<td>F11.8.R</td>
<td>CTGGCTTCAATTGCACAGCCTCTT</td>
<td>Reverse</td>
<td>4F11.8</td>
</tr>
<tr>
<td></td>
<td>N17.16.F</td>
<td>CCGCACAACCACAACACACACAT</td>
<td>Forward</td>
<td>11N17.16</td>
</tr>
<tr>
<td></td>
<td>N17.16.R</td>
<td>CCGGCTTGTAATGCTCGGTGGTGATG</td>
<td>Reverse</td>
<td>11N17.16</td>
</tr>
<tr>
<td></td>
<td>J3.1.F</td>
<td>GCTGGAGATGCGCATACATCAA</td>
<td>Forward</td>
<td>22J3.1</td>
</tr>
<tr>
<td></td>
<td>J3.1.R</td>
<td>CAAGCGCAAATCCGATATAACCG</td>
<td>Reverse</td>
<td>22J3.1</td>
</tr>
<tr>
<td></td>
<td>B17.1.F</td>
<td>CAAGCGCAAATCCGATATAACCG</td>
<td>Forward</td>
<td>CLIPB17</td>
</tr>
<tr>
<td></td>
<td>B17.1.R</td>
<td>TATTGTCCGAGATCCAGTCCATAAC</td>
<td>Reverse</td>
<td>CLIPB17</td>
</tr>
<tr>
<td></td>
<td>J3.4.w.F</td>
<td>TTTCCTCCTGTCGCCCTCGAATCACAC</td>
<td>Forward</td>
<td>22J3.4</td>
</tr>
<tr>
<td></td>
<td>J3.4.w.R</td>
<td>GATCTGCTGCGCTCGAAAGCGGC</td>
<td>Reverse</td>
<td>22J3.4</td>
</tr>
<tr>
<td></td>
<td>B17.w.F</td>
<td>ACTCACACTCCGGCACAGAAAGAATC</td>
<td>Forward</td>
<td>CLIPB17</td>
</tr>
<tr>
<td></td>
<td>B17.w.R</td>
<td>AAACCACGTACTCTCTCACTACTC</td>
<td>Reverse</td>
<td>#</td>
</tr>
<tr>
<td></td>
<td>B17.5pr.F</td>
<td>TACTGCGTCTACAAGCCATCTCAC</td>
<td>Forward</td>
<td>#</td>
</tr>
<tr>
<td></td>
<td>B17.5pr.R</td>
<td>CTGACAGGTTTTGCGAGTGTATCT</td>
<td>Reverse</td>
<td>#</td>
</tr>
<tr>
<td></td>
<td>S7.tm.F</td>
<td>AAGCAGAAGAGACCCCGGTTC</td>
<td>Forward</td>
<td>S7</td>
</tr>
<tr>
<td></td>
<td>S7.tm.R</td>
<td>CCGGAATACCAGATCTCTCCA</td>
<td>Reverse</td>
<td>#</td>
</tr>
<tr>
<td></td>
<td>S7.tm.probe</td>
<td>TCACCAGCGTGATCGATGCCC</td>
<td>Reverse</td>
<td>#</td>
</tr>
<tr>
<td></td>
<td>B17.tm.F</td>
<td>CCTCGCTGCTGATGATGATGAT</td>
<td>Forward</td>
<td>CLIPB17</td>
</tr>
<tr>
<td>Sequencing</td>
<td>B17.tm.R</td>
<td>TGATGGCTGTGGCTAGTGCC</td>
<td>Reverse</td>
<td>#</td>
</tr>
<tr>
<td>-------------</td>
<td>------------</td>
<td>----------------------</td>
<td>---------</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>B17.tm.probe</td>
<td>CGGCCTGCAGCCAAGACTATATGA</td>
<td>Reverse</td>
<td>CLIPB17</td>
</tr>
<tr>
<td>J3.4.s1</td>
<td>ACCCTCAAATGCTGTATGTGTGC</td>
<td>Forward</td>
<td>22J3.4</td>
<td></td>
</tr>
<tr>
<td>J3.4.s2</td>
<td>GCCTCGGTTCGTTACACGGGTGAT</td>
<td>Forward</td>
<td>#</td>
<td></td>
</tr>
<tr>
<td>J3.4.s3</td>
<td>AAGCTGTCAGCAGCCGTTTTTCAGTC</td>
<td>Reverse</td>
<td>#</td>
<td></td>
</tr>
<tr>
<td>J3.4.s4</td>
<td>GGTGGGCGATCAGCTACGTGCTAGC</td>
<td>Reverse</td>
<td>#</td>
<td></td>
</tr>
<tr>
<td>B17.s1</td>
<td>CCTGTGCTATGTCAGCAGACC</td>
<td>Forward</td>
<td>CLIPB17</td>
<td></td>
</tr>
<tr>
<td>B17.s2</td>
<td>TGAAGGAGATACATCGCAA</td>
<td>Forward</td>
<td>#</td>
<td></td>
</tr>
<tr>
<td>B17.s3</td>
<td>AGACAGATCGCCGTACCAG</td>
<td>Reverse</td>
<td>#</td>
<td></td>
</tr>
<tr>
<td>B17.s4</td>
<td>GGCGCAATCATATAGTCCT</td>
<td>Reverse</td>
<td>#</td>
<td></td>
</tr>
<tr>
<td>B17.5pr.s1</td>
<td>ATCACTACAACCGTTCCTTCATA</td>
<td>Forward</td>
<td>#</td>
<td></td>
</tr>
<tr>
<td>B17.5pr.s2</td>
<td>TCCAAGATCCAGTTGATGATGC</td>
<td>Forward</td>
<td>#</td>
<td></td>
</tr>
<tr>
<td>B17.5pr.s3</td>
<td>CTCTCTCCAGTTACACCTTTT</td>
<td>Forward</td>
<td>#</td>
<td></td>
</tr>
<tr>
<td>B17.5pr.s4</td>
<td>AGACACGAATGAGTGAGATGTG</td>
<td>Reverse</td>
<td>#</td>
<td></td>
</tr>
<tr>
<td>B17.5pr.s5</td>
<td>ACGGCAGTTTTACGCATACCTTAG</td>
<td>Reverse</td>
<td>#</td>
<td></td>
</tr>
<tr>
<td>B17.5pr.s6</td>
<td>TATGAAGGAGACACCGTTAGTGAT</td>
<td>Reverse</td>
<td>#</td>
<td></td>
</tr>
<tr>
<td>RNAi</td>
<td>T7-GFP.kd.1</td>
<td>TAATACGACACTATAAGGATGCAGATATCCGACCTCCGTGATAG</td>
<td>Forward</td>
<td>GFP</td>
</tr>
<tr>
<td></td>
<td>T7-GFP.kd.2</td>
<td>TAATACGACACTATAAGGATGCAGATATCCGACCTCCGTGATAG</td>
<td>Reverse</td>
<td>GFP</td>
</tr>
<tr>
<td></td>
<td>T7-B17.kd.1</td>
<td>TAATACGACACTATAAGGATGCAGATATCCGACCTCCGTGATAG</td>
<td>Forward</td>
<td>CLIPB17</td>
</tr>
<tr>
<td></td>
<td>T7-B17.kd.2</td>
<td>TAATACGACACTATAAGGATGCAGATATCCGACCTCCGTGATAG</td>
<td>Reverse</td>
<td>CLIPB17</td>
</tr>
</tbody>
</table>
so named because it was in plate 25 at the position where row F intersected with column 12. Screening was performed with Polymerase Chain Reaction (PCR) using a standard condition: (94°C/2 minutes; (94°C/30 seconds, 58°C/30s, 72°C/45s)\textsuperscript{30 cycle}; 72°C/5 minutes), 750 nM of each primer, 2 mM of MgCl\textsubscript{2}, 250 µM each of dNTPs, and 1U of Tag polymerase in reaction volume of 50 µl. Three BACs containing the microsatellite markers \textit{H290}, \textit{H175}, and \textit{ND3B6} were used in the sequence-tagged connector strategy (Rowen \textit{et al.}, 1997) to screen for overlapping clones in both directions. Minimal overlapping was confirmed with Southern hybridization using standard method (Sambrook J., 1989) and the minimal overlapping BAC was \textit{in-situ} hybridized to ovarian polytene chromosome according to established protocol (Kumar and Collins, 1994). A trimming strategy (Hill \textit{et al.}, 2000) was used to remove a 49.2 kb of excess DNA from one end of a BAC to reduce sequence redundancy.

2.3 Sequencing BACs in the \textit{Pen1} region:

The RANDI strategy (Voss \textit{et al.}, 1995) was used to sequence each BAC as described. Briefly, a random (RAN) library with 2 to 5 kb fragments was constructed for each BAC by partial digestion with \textit{Tsp509 I} or \textit{Sau3A} restriction enzymes. A directed (DI) library was also constructed for the same BAC with complete \textit{EcorI} or \textit{BamH1} digestion. All fragments were ligated to pUC18 using T4 Ligase (Invitrogen). Sequencing both ends of all fragments was performed with the AmpliTaqFS core kit (Applied Biosystem) using standard M13 forward and reverse primers labeled with FITC or CY5. All reactions were analyzed on the ARAKIS sequencing system (Wiemann \textit{et al.}, 1995). Raw sequence traces were analyzed with the software package LANE TRACKER (Ansorge \textit{et al.}, 1993). Assembly using the software GENE SKIPPER was
performed after the theoretical 5X coverage of random fragments and end sequences of all directed fragments were obtained. During the assembly process, the contigs assembled from random fragments were used to order all directed fragments along the length of the BAC, and the appropriate directed fragments were used as templates for filling any gap between contigs of random fragments. These gaps were covered by primer walking or transposon-mediated sequencing (Strathmann et al., 1991).

2.4 Bioinformatic analysis of Pen1 sequences:

The assembled sequence from all BACs combined with the sequence obtained from the genome project were annotated with both \textit{ab initio} gene prediction algorithms and alignment algorithms based on sequence similarity. Software that implemented the \textit{ab initio} algorithms included GENSCAN 1.0 (Burge and Karlin, 1997), FGENES 1.0 (unpublished), and GENEID 1.1 (Parra et al., 2000). These were used with default parameters and \textit{D. melanogaster} or human as the organismal option. BLASTX and BLASTP were used to search against the Gene2EST server (Gemund et al., 2001) for available \textit{Anopheles} ESTs. The GENEWISE software that incorporated EST and protein matches in its algorithm was used to predict accurate exon/intron boundaries. Protein domain analysis was performed using InterPro (Apweiler et al., 2000), SMART (Letunic et al., 2004), and Pfam servers. All analyses were viewed with the Artemis tool (Rutherford et al., 2000). A computer program was written to graphically display all genes. To avoid over-prediction (false positive), only genes fulfilling one or more of the following criteria were accepted for further analysis. A gene must be predicted by at least two \textit{ab initio} algorithms, or it must be predicted by one algorithm and is matched to ESTs, cDNAs, and proteins in the public databases, or it must be confirmed with gene
expression assayed by semi-quantitative RT-PCR. tRNA gene were predicted with tRNASCAN-SE v1.21 (Lowe and Eddy, 1997). Cytological locations of putative *Drosophila* orthologues were obtained from FlyBase (www.fruitfly.org).

2.5 Semi-quantitative RT-PCR:

Total RNA was isolated from pools of 8 mosquitoes using the Trizol reagent (Invitrogen) according to manufacturer’s protocol. Isolated RNA was treated with DNA.Free (Ambion) to minimize the chance of contaminating genomic DNA. First strand cDNA synthesis was performed with Superscript III (Invitrogen) and Oligo-dT$_{20}$ primer according to the manufacturer’s recommendation. When possible, gene fragments containing an exon/intron boundary were amplified to distinguish cDNA PCR product from contaminating product from gDNA template. Standard PCR condition was used as described previously. 250 nM of primers amplifying the house-keeping gene ribosomal protein S7 was used as a semi-quantitative control in the same reaction tube with experimental primers. All primers were designed with the Primer3 software (Rozen and Skaletsky, 2000). Forward/Reverse primer sequences are listed in Table 1 for genes that must be expressed to be considered not a false positive from computational predictions. All PCR products were subjected to electrophoresis in 1.5% agarose gel stained with 0.5% Ethidium Bromide and visualized with a CCD camera (Kodak).

2.6 PCR cloning of a gene in the polymorphic region:

Genomic DNA (gDNA) was isolated using the DNeasy kit (Qiagen) with accompanied instructions. Gene 22J3.4 was amplified from gDNA of L35 and 4Arr mosquitoes with a high fidelity polymerase (AmpliTaq Gold High Fidelity, Applied
BioSystem) using primers J3.4.w.F/R. PCR conditions were: 94°C/5m, (94°C/30s, 58°C/30s, 68°C/2m)\(^{35}\) cycles, 68°C/5m. PCR products were electrophoresed in 0.8% agarose gel, purified with QiaQuick columns (Qiagen), and cloned into pCR2.1-TOPO vector with the TA cloning kit (Invitrogen). Conceptually translated amino acid sequences were aligned and bootstrapped trees were constructed with CLUSTALW v1.81 (Thompson \textit{et al.}, 1997) and viewed with TREETOOL v2.0.1 (Chenna \textit{et al.}, 2003).

2.7. Quantitative real-time RT-PCR analysis of mRNA:

Total RNA isolation and first strand cDNA synthesis were performed as described for RT-PCR. The chemistry of the TaqMan sequence detection system on the ABI PRISM 7700 Sequence Detector (Applied BioSystem) was used to quantify cDNA from various experimental conditions. In this chemistry, a polymerase with 5’-3’ nuclease activity (AmpliTag Gold) was used in a PCR assay to degrade the gene-specific TaqMan probe, generating fluorescence detectable by the 7700 Sequence Detector. TaqMan probe for the endogenous control gene S7 (S7.tm.probe) was labeled with the VIC fluorophore on its 5’ end and the experimental CLIPB17 probe (B17.tm.probe) was similarly labeled with FAM. Both probes were labeled with TAMRA on their 3’ ends as the fluorescence quencher dye (Figure 5). Primer and probe sequences were designed with the Primer Express Software v1.5 (Applied BioSystem) (Table 1). PCR reactions were assembled in triplicate in wells of optical 96-well plates (Applied BioSystem) and included primers and probe for either S7 or CLIPB17 genes, 1X TaqMan buffer A, 5.5 mM MgCl\(_2\), 200 µM each of dATP, dCTP, dGTP, 400 µM of dUTP, 0.5U AmpErase UNG, and 2.5U of AmpliTaq Gold DNA polymerase. PCR condition followed 1 cycle at 50°C for 2 minutes (m), 1 cycle at 95°C for 10 minutes, and 50 cycles of (95°C/15s,
Figure 5. Taqman principle. Fluorescence intensity is quantified for each cycle of PCR amplification. (A) After denaturation, primers and a probe anneal to the target. Fluorescence does not occur because of the proximity between the fluorophore and the quencher. (B) During the extension phase, the probe is cleaved by the 5’-3’ nucleosidase activity of a Tag polymerase. Thereby quencher and fluorophore are separated, allowing the fluorescence emission from the reporter dye. FW: forward primer. RV: reverse primer.
Data analysis was assisted with the Sequence Detector program v1.7. Processed data consisted of a list of $C_t$ values from each reaction. The relative standard curve quantification method was employed as described in ABI User Bulletin #2 from Applied BioSystem. A primer optimization step was performed with all possible combination between 50 nM, 300nM, 600 nM, and 900 nM of each primer. Once an optimized primer concentration was determined, a serial dilution of probe concentrations was done to determine the best primers/probe concentration. From these optimizations, a concentration of 100 nM for each S7 primer and 600 nM for each CLIPB17 primer were determined to work best with 100 nM of probe for either gene. Standard curves for both S7 and CLIPB17 were constructed according to instruction using 10-fold dilutions of a 100 ng/µl DNA stock. More details can be searched in User Bulletin #2 and the ABI website.

2.8 Infections with *Plasmodium*, bacteria, or yeast:

Frozen *P. berghei* were thawed at 37°C and injected into a Lobund-Wistar rat weanling intraperitoneally (circa $2 \times 10^7$ infected RBC per inoculum). Parasitemia was checked starting from day four post inoculation and every day after until it reached between 15-20%. When exflagellation was observed in a fresh drop of blood, the rat was anesthetized and placed on mosquito cage for feeding. Mosquitoes were collected at appropriate times for total RNA isolation or for midgut dissection.

Infection with *P. cynomolgi B* was carried out similarly with few exceptions. Parasitemia in a macaque monkey was checked starting from day two post inoculation and every day after until it reached between 0.5-3.0% when the monkey was anesthetized for mosquito feeding. This feeding process was done in the dark to maximize infection
efficiency. Mosquito midguts were dissected 8 to 9 days post infection for oocyst observation.

Septic or yeast infections were carried out by pricking the mosquito thorax with a fine needle (Fine Science) dipped into a suspension of both *Micrococcus luteus* and *Escherichia coli* bacteria cultured overnight, or a suspension of the yeast *Sacharomyces cerevisiae*.

2.9 PCR Cloning and sequencing of CLIPB17 coding and promoter regions:

Cloning of CLIPB17 coding and promoter regions was performed as described previously (Section 2.6) with few differences. Six individual mosquitoes of each strain (L35 or 4Arr) were used for cDNA isolation as described. The coding region of CLIPB17 was PCR amplified from cDNA using primers B17.w.F / B17.w.R (Table 1) and Platinum Taq High Fidelity polymerase (Invitrogen). PCR products were cloned into pGEM-Teasy plasmid vector. The promoter region of CLIPB17 was cloned with primers B17.5pr.F and B17.5pr.R from three individuals of each strain. All sequencing was done on the ABI 3700 DNA Analyzer using Big Dye Terminator v3 chemistry. Sequence data were evaluated and assembled using SEQMAN II software package (Swindell and Plasterer, 1997). Domain characterization for CLIPB17 was performed as described under section 2.4 with some additional analyses. The signal peptide of CLIPB17 was predicted by SignalP (Bendtsen *et al.*, 2004). Basic RNA polymerase II promoter binding site was predicted by Promotor Finder from the BDGP server (Reese, 2001). Other enhancer binding sites were predicted with PROMO (Messeguer *et al.*, 2002) using the TRANSFAC database (May 2005 revision) of all established binding motifs. Sequence alignment was done with CLUSTALW as described in Section 2.6.
2.10 RNA interference assay of CLIPB17:

Fragments of CLIPB17 and GFP were amplified by PCR using appropriate primers with T7 binding site adapted to their 5’ ends (Table 1). PCR products were gel purified by Qiaquick columns (Qiagen) and used to synthesize double stranded RNA (dsRNA) with the Megascript RNAi kit (Ambion). Purified dsRNA were micro-injected into adult mosquitoes essentially as described (Blandin et al., 2002). Mosquitoes were kept in cages with wet tissue papers to keep humidity higher than normal in order to minimize evaporation of mosquito haemolymph through puncture wound.
3. RESULTS AND DISCUSSION

3.1 Physical mapping and sequencing of the *Pen1* region:

A physical map of the *Pen1* region was constructed by screening the ND-1 BAC library (Ke and Collins, unpublished) made from genomic DNA of the PEST strain of *A. gambiae* (Mukabayire and Besansky, 1996). ND-1 library contained more than 12,000 clones of 100 kb in average length, and covered ~5 folds of the mosquito genome. BAC clones 30E5 and 3B6 were identified from PCR screening of successively less complex pools to contain the microsatellite markers *H290* and *ND3B6* respectively (Figure 4). Furthermore, BAC 11N17 containing the marker *H175* was similarly screened. End-sequences of these BACs were used to design new PCR primers to screen overlapping clones. These three BACs were also used to hybridize to all overlapping clones in a Southern hybridization experiment to determine the next minimally overlapping BAC. To avoid jumping to a distance genomic region by mistakenly choosing a chimaeric clone, the minimally overlapping BAC was *in-situ* hybridized to ovarian polytene chromosomes. A trimming strategy was used to remove about 50 kb of DNA from one end of clone 25F12, reducing unnecessary overlap with clone 11N17 (Thomasova *et al.*, 2002). A gap between clones 11B17 and 4L22 was filled by screening the ND-TAM library, a larger BAC library of 12 folds genome coverage (Hong *et al.*, 2003). This work permitted us to develop a contig of 15 clones to represent the *Pen1* region for our subsequent clone-by-clone sequencing method. The *Pen1* region spans six bands from
8C-8D division on the ovarian polytene chromosome and occupies ~0.5% of the mosquito genome (Figure 6).

The proposal and support to obtain the sequence from the Pen1 region has been a part of the coordinated projects directed by the Anopheles genome sequencing consortium (Collins and Hill, 2005). As the consequence of the relatively smaller size of the Pen1 region, we chose to employ a sequencing strategy that, although more costly and slower, enjoyed better accuracy in sequence assembly and was better suited to our resources (see Materials and Methods). We envisioned that we could sequence as many clones from the Pen1 region as we can, stopping when our collaborators from the whole-genome project start to produce traces from their shot-gun sequencing approach. We could then provide both our experience from examining the first large stretch of mosquito sequence and our Pen1 sequence as a dataset for training the assembly pipelines, while they could provide us with assembled sequence if we need to fill any gap in our Pen1 contig.

Five BACs were fully sequenced from the Pen1 physical map and contained 528 kb of un-interrupted sequence (Figure 7). Additional sequences totaling 52 kb were determined as overlaps between these clones and showed an average sequence variation of 0.03%, well within the acceptable range of intra-species polymorphism. In addition, we fully sequenced BAC 8N20, a 129 kb clone overlapping three of the clones from the minimal contig (Figure 7). These three clones showed a high degree of similarity to each other, with their overlaps (21.2 kb in total) differing by only 0.02%. In contrast, the sequence of 8N20 varied extensively from its consensus counterpart by an average of 3.3% in 121.8 kb of aligned sequence. The differences consisted of both nucleotide
Figure 6. Minimal overlapping BACs across the Pen1 region. Clones colored in yellow and green have been sequenced with our RANDI strategy. Sequence corresponding to the clones in pink has been assembled from the trace archive of the whole genome sequencing project. Pen1 region span polytene chromosome division 8C-8D between genetic markers H290 and ND3B6. Marker H175 was genetically mapped closest to Pen1 (Zheng et al., 1997).
changes and short deletions/insertions (indels), and were widely distributed, more so but not exclusively in intergenic regions. For example, the three longest gene-free regions between 11N17.15 and 11N17.16 (8.3 kb in length), 11N17.16 to 22J3.1 (36.6 kb), and 22J3.5 to 22J3.6 (26 kb) showed high overall differences scattered among segments of high sequence conservation. On the other hand, gene-bearing segments tended to be well-conserved (especially 22J3.2 and 22J3.3), except when they were located close to the indels. Figure 8 illustrated the pattern of intragenic sequence variation in parts of the genes 22J3.4/8N20.5 (2.4% variation) and 22J3.5/8N20.6 (4.0% variation). The level of variation in 22J3.4/8N20.5 pair was below the average for 22J3/8N20 divergence, while variation within the other gene pair was higher than the 3.3% average. Considering that 22J3.4 and 22J3.5 were next to each other but 22J3.5 was closer to an indel, we could hypothesize that the above average variation seen between 22J3.5 and its counterpart in BAC 8N20 might be influenced by being closer to a polymorphism hotspot. Indeed, higher variation was again observed for the gene 11N17.16 (data not shown), which was interrupted by a 5.8 kb retrotransposon similar to the BEL element (Accession number 7511879) of *D. melanogaster* (Figure 9).

Two possible explanations for the existence of the variant 8N20 clone can be advanced: that it represents a duplicated chromosomal region picked up by PCR-based screening during the construction of the physical BAC map, or that it represents a high degree of localized polymorphism in *A. gambiae*. *In-situ* hybridization to polytene chromosome showed that 8N20 was located in a single, not very dense polytene band near the telomeric end of subdivision 8C, the same band that clones 11N17, 25F12, and 22J3 mapped to (data not shown). The two flanking BACs, 30E5 and 4F11, hybridized
Figure 7. Diagram of the 528 kb sequence from the Pen1 region, presented in two segments. Predicted genes are in green or red. tRNA genes are shown in black. The sequenced BAC clones are identified as lines on the bottom of each segment, in blue for those containing the consensus sequence and green for the variant 8N20. The sequence axis includes red tickmarks representing 51 potentially new microsatellites. Two arrowheads represent established microsatellites that partially delimit the Pen1 region.
to an adjacent band in 8C. These results strongly argued against the duplication hypothesis.

A haplotype is usually defined as a particular combination of alleles at different loci on the same chromosomal fragment. Accordingly, the variant 8N20 fit the definition of a localized haplotype in *A. gambiae*. We hypothesized that by examining sequences from multiple mosquitoes of different strains, we might detect a phenomenon similar to what was seen between the consensus and the variant 8N20 of the PEST strain. This result would indicate that the variant 8N20 was a real haplotype of the PEST strain rather than a chanced contamination during the BAC library construction process, and that this real haplotype could be traced across multiple mosquito strains. Toward this end, we cloned and sequenced the polymorphic gene 22J3.4/8N20.5, the homolog to *Drosophila* amiloride-sensitive Na$^+$ channel gene (CG4805), from multiple mosquitoes of L35 and 4Arr strains. These strains were selected for their susceptible and refractory melanotic encapsulation phenotype and unrelated to the PEST strain from which the variant 8N20 and the consensus sequence have been derived. Six clones from 4Arr and 11 clones from L35 were recovered after the cloning process and subsequently fully sequenced. When the sequences were aligned, the resulted alignment tree clearly showed that most clones from one strain clustered more closely together than to clones from a different strain (Figure 10). Furthermore, each of the two haplotypes from PEST clustered with 4Arr or L35 clones, respectively, indicating that the consensus haplotype was more likely to be found in L35 and the variant haplotype could be preferably found in 4Arr. Three clones from the L35 strain were observed to cluster in a group situated between the 4Arr and L35 clusters, indicating that a mixed haplotype could be found in some L35 individuals.
Figure 8. Examples of sequence variation with each letter corresponding to three nucleotides (amino acids or Z for introns). The top sequence includes exons 5 and 6 of the polymorphic 8N20.5/22J3.4 interrupted by an intron. The bottom sequence includes the four exons of 8N20.6/22J3.5 interrupted by introns. In each case, the aligned polymorphic sequences are shown by dots if identical. Coding differences are indicated as amino acid changes, whereas silent changes are indicated by tickmarks, + or $, depending on whether they involve one, two, or three DNA changes in the same triplet, respectively.
It is fascinating to speculate on possible genetic explanations for this observation. The presence of an active retrotransposon (data not shown) nearby would suggest that trans-acting mobile elements could be responsible.

While we were sequencing and analyzing the 528 kb sequence, our colleagues from Genoscope were working on a similar project but from the centromeric region of the Pen1 locus (Eiglmeier et al., 2005). This group discovered in a BAC clone the presence of two retrotransposons which could not be found in the 585 kb cognate region from the genome project. One of the retrotransposons was another member of the BEL-like family. The assembled genome sequence revealed that there was one of these complete or partial transposons per 4 mb of sequence. There were also number of insertions found within most complete transposons, indicating that they played an active role in genome shuffling (Eiglmeier et al., 2005). Their potential effect on flanking sequences suggested that their presence and activity could cause a higher frequency of sequence rearrangement in the mosquitoes such as those found in the mixed haplotype group of L35 observed from the alignment (Figure 10).

In summary, observations from analyzing a gene within a segment of polymorphic sequence demonstrated that there were multiple haplotypes in a segment of the Pen1 region found among A. gambiae of different genetic backgrounds. We have speculated that mobile elements might play a role in the dynamics among these haplotypes, but it remains to be determined how long this polymorphic DNA region is, and whether it may have resulted from small (undetectable with current cytogenetic techniques) chromosomal rearrangements or from retroelement actions (Charlesworth et al., 1994). An intriguing possibility arising from the analysis of Figure 10 was that this
Figure 9. Percent identity plot (PIP) between BAC 8N20 (jagged black line) versus the consensus sequence. Percent sequence identity is shown by colored blocks in pink (>98%), yellow (95-98%), green (90-95%), blue (80-90%), and purple (10-80%). White blocks between 1.6-1.7 kb, 88.9-90.1 kb, and 93.8-94.6 kb represent sequences present only in 8N20. The white block between 20.1 and 25.8 kb contains a retrotransposon similar to *D. melanogaster* BEL (Accession number 7511879). Exons of the indicated genes are shown in black numbered boxes.
extensive polymorphism may be correlated with mosquito refractoriness to *Plasmodium*. It should be noted that PEST mosquitoes exhibit both melanization susceptible and refractory phenotypes, and this phenotypic heterozygocity may be influenced by the presence or absence of a refractory-determining consensus haplotype.

Whether this extensive local polymorphism is unique to the *Pen1* region is another interesting question that has implications toward the melanization phenotype. If this type of haplotypic variation is exclusively found within the *Pen1* region, it would argue in favor of them having a role influencing refractoriness. One way to answer this question was to look for the presence of similar polymorphic regions throughout the mosquito genome. Our communications with the whole-genome sequencing team revealed that they have encountered regions displaying characteristics of being highly polymorphic among the PEST individuals used to construct the sequencing libraries. They have catalogued 726 regions of the genome (average length 28 kb) where mate pairs (the two end-sequences of a clone) with distance violation or orientation violation were found (Collins, 2005). These violations reflected the possibility that diverged haplotypes had been incorporated into adjacent locations during the assembly process. Furthermore, there were also a number of small scaffolds (144 scaffolds) between 20-300 kb in length that while highly similar to larger scaffolds, also exhibited variation of at least a few percents, suggesting that these small scaffolds belong to an alternate haplotype. From these observations, we concluded that the local polymorphism seen in the *Pen1* region was reflected throughout the genome of *A. gambiae*. 
Figure 10. Phylogenetic tree of the complete gene 22J3.4 from 4Arr (red), L35 (blue), and PEST laboratory strains. Six clones from strain 4Arr and 11 clones from strain L35 were aligned with CLUSTALW. Three clones from the L35 strain cluster in a group showing a haplotype more similar to 4Arr clones. The PEST sequences (P22J3 and P8N20) are from BAC 22J3 (consensus sequence, purple) and BAC 8N20 (variant, green) respectively. Bootstrap values are associated with the tree nodes.
3.2 Sequence analysis of a 528 kb fragment from the Pen1 region:

We have annotated the 528 kb consensus sequence using a gene prediction pipeline that included both *ab initio* and similarity-based computational gene prediction methods with manual curation of all generated data. The analysis resulted in the prediction of 48 genes, of which 46 were putative protein-coding and two were possibly tRNA-coding. We have provisionally named the protein-coding genes sequentially, in a telomere-to-centromere order and according to the BAC in which they were fully contained (Figure 7). Thirty nine of these genes (85%) were predicted with both *ab initio* algorithms and algorithms that also incorporated similarity to proteins or coding sequences from other organisms. Seven putative genes were predicted only on the basis of the *ab initio* approach, but four of them have been confirmed by reverse-transcription test of expression (Table 2 and Figure 11). Expression of the other three putative novel genes was not detected in adult mosquitoes. Whether these predicted genes represent false positives, or whether they are expressed only in developmental stages other than adult, remains to be tested.

The average length of the genes in this sequence (including introns but not 5’ and 3’ untranslated regions) was 1.98 kb, and their average density was one gene per 11.5 kb. Gene distribution along this sequence was variable, with gene-dense segments punctuated by gene-free stretches. For instance, the 36 kb region between the gene 11N17.16 and 22J3.1 was predicted to be non-coding, whereas a nearby 19 kb region contained five genes (11N17.11 to 11N17.15). The average gene density in this sequence, combined with the reported *A. gambiae* haploid DNA content of 260 Mb based on C₀t reassociation kinetics (Besansky and Powell, 1992), allowed us to predict a total of approximately
### TABLE 2

**LIST OF ALL GENES FROM 528 kb *Pen1* FRAGMENT**

<table>
<thead>
<tr>
<th>Anopheles gambiae 8C (2R)</th>
<th>Drosophila melanogaster</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene</td>
<td>Encodes putative</td>
</tr>
<tr>
<td>30E5.1</td>
<td>Ion transporter</td>
</tr>
<tr>
<td>30E5.2</td>
<td>GPCR</td>
</tr>
<tr>
<td>30E5.3</td>
<td>TPR phosphoprotein</td>
</tr>
<tr>
<td>30E5.4</td>
<td>Transcription factor</td>
</tr>
<tr>
<td>30E5.5</td>
<td>GPCR</td>
</tr>
<tr>
<td>30E5.6</td>
<td>GPCR</td>
</tr>
<tr>
<td>30E5.7</td>
<td>Glycerol kinase</td>
</tr>
<tr>
<td>30E5.8</td>
<td>Topoisomerase</td>
</tr>
<tr>
<td>30E5.9</td>
<td>Transcription factor</td>
</tr>
<tr>
<td>30E5.10</td>
<td>-</td>
</tr>
<tr>
<td>30E5.11</td>
<td>-</td>
</tr>
<tr>
<td>30E5.12</td>
<td>Muscle motor</td>
</tr>
<tr>
<td>4F11.1</td>
<td>Muscle binding protein</td>
</tr>
<tr>
<td>4F11.2</td>
<td>-</td>
</tr>
<tr>
<td>4F11.3</td>
<td>RHO small GTPase</td>
</tr>
<tr>
<td>4F11.4</td>
<td>Transcription factor</td>
</tr>
<tr>
<td>4F11.5</td>
<td>Expressed</td>
</tr>
<tr>
<td>4F11.6</td>
<td>Expressed</td>
</tr>
<tr>
<td>4F11.7</td>
<td>Not detected in adult</td>
</tr>
<tr>
<td>4F11.8</td>
<td>Not detected in adult</td>
</tr>
<tr>
<td>4F11.9</td>
<td>-</td>
</tr>
<tr>
<td>4F11.10</td>
<td>Cytoskeletal regulator</td>
</tr>
<tr>
<td>11N17.1</td>
<td>-</td>
</tr>
<tr>
<td>11N17.2</td>
<td>sRNP</td>
</tr>
<tr>
<td>11N17.3</td>
<td>Deoxynucleoside kinase</td>
</tr>
<tr>
<td>11N17.4</td>
<td>-</td>
</tr>
<tr>
<td>11N17.5</td>
<td>V-ATPase</td>
</tr>
<tr>
<td>11N17.6</td>
<td>V-ATPase</td>
</tr>
<tr>
<td>11N17.7</td>
<td>Translation Init. Factor</td>
</tr>
<tr>
<td>11N17.8</td>
<td>FGF signaling promoter</td>
</tr>
<tr>
<td>11N17.9</td>
<td>Cell-adhesion protein</td>
</tr>
<tr>
<td>11N17.10</td>
<td>Tachykinin receptor</td>
</tr>
<tr>
<td>11N17.11</td>
<td>Transcription factor</td>
</tr>
<tr>
<td>11N17.12</td>
<td>Acetyltransferase</td>
</tr>
<tr>
<td>11N17.13</td>
<td>Ribosomal protein</td>
</tr>
<tr>
<td>11N17.14</td>
<td>Not detected in adult</td>
</tr>
<tr>
<td>11N17.15</td>
<td>Chitin binding protein</td>
</tr>
<tr>
<td>11N17.16</td>
<td>Expressed</td>
</tr>
<tr>
<td>22J3.1</td>
<td>Expressed</td>
</tr>
<tr>
<td>22J3.2</td>
<td>5´nucleotidase/apyrase</td>
</tr>
<tr>
<td>22J3.3</td>
<td>5´nucleotidase/apyrase</td>
</tr>
<tr>
<td>22J3.4</td>
<td>Sodium channel</td>
</tr>
<tr>
<td>22J3.5</td>
<td>-</td>
</tr>
<tr>
<td>22J3.6</td>
<td>Tyr/Ser/Thr phosphatase</td>
</tr>
<tr>
<td>22J3.7</td>
<td>5-oxoprolinase</td>
</tr>
<tr>
<td>22J3.8</td>
<td>Differentiation regulator</td>
</tr>
</tbody>
</table>
Figure 11. RT-PCR assay showing expression of genes in the Pen1 region predicted without similarity to any known protein. Upper band from each panel is PCR product of S7 control. Lower band is experimental gene. Gene expression from female (F), bloodfed female (FB), and infected female (FI) from each strain (L35 or 4Arr) were assayed.
14,800 genes in the A. gambiae genome, a number very similar to the 13,601 genes reported for D. melanogaster (Adams et al., 2000).

Annotation of the Adh region from D. melanogaster revealed a large number of genes (8%) nested within the introns of other genes (Ashburner et al., 1999). Because gene finding computer programs do not efficiently predict this class of genes, a prediction must usually rely on supporting evidence obtained from cDNA and EST homologies or expression profiling. We predicted one nested gene (4F11.5) within a 28 kb intron of the gene 4F11.4 (Figure 12). The 4F11.4 gene product resembled a transcription factor from the fruit fly (AAL28078.1) and was transcribed from the opposite strand of 4F11.5. Even though the putative 4F11.5 protein showed no protein similarity to any database entry, its prediction was supported by RT-PCR expression (Figure 11). Similarly, a tRNA-Thr gene was found within an intron of gene 30E5.10, but no molecular evidence has yet supported this prediction. We also observed two pairs of adjacent genes that seemed to have duplicated after the divergence of the Drosophila and Anopheles lineages. One pair of mosquito genes (11N17.5 and 11N17.6) shared 66% amino acid identity and encoded proteins highly similar to the V-ATPase subunit α from a number of organisms. The other recently duplicated genes in our sequence were transcribed from opposite strands and encoded proteins that were 59% identical. They belonged to the Apyrase/5’-nucleotidase family of proteins that facilitates hematophagy by inhibiting aggregation of the host platelets when arthropods feed on blood (Mans et al., 1998).

The genome of the fruit fly D. melanogaster had been sequenced and annotated at the time we finished sequencing our 528 kb fragment (Adams et al., 2000), allowing us to
Figure 12. Micro-synteny analysis of 528 kb sequence from the Pen1 region of A. gambiae. Picture displays all protein-coding genes in two blocks, 0-280 kb and 280-528 kb from telomere (T) to centromere (C) orientation. Genes are number as described in the text and colored green for orienting forward and red for reverse direction. The polytene chromosomal location of this sequence (Ag 2R:8C) is seen at the lower left corner, while microsyntenic fruit fly clusters are shown at the lower right. Putative orthologues in both species are connected and shown in bold font. Drosophila genes are prefixed with CG. Of the six syntenic cluster, one each maps to 3L and 2R, and four to 3R in Drosophila. Individual cytogenetic location in the fruit fly is listed to the right of each box for genes not in a cluster. The x represents the absence of a fly homologue. Two pairs of duplicated mosquito gene are shown in brackets.
perform a one-way comparative genomics simulation in which we traced our *A. gambiae* genes to their best matches in the fruit fly genome to detect local clustering and/or micro-synteny (gene order conservation) between the sequences. It should be noted that we cannot label the fruit fly best matches to our *A. gambiae* genes as true orthologues (related by a speciation event) in the absence of a complete mosquito genome. Thus, our one-way search turned up what we can generally call homologues (sharing significant level of amino acid similarity). However, when a mosquito gene has the highest similarity to a single fruit fly gene followed by mediocre or no match to others (including other members of a gene family), then the probability is high that they are true orthologues. We have identified 38 mosquito genes showing strong matches to a single fruit fly gene (Table 2). Out of these matches, 20 genes had putative orthologues in the 3R chromosomal arm of the fruit fly. From the remainder, seven genes mapped to the fruit fly 3L, five to the X chromosome, and two each to 2L and 2R. This comparison suggested that mosquito arm 2R and the fruit fly 3R retained sufficient similarity of gene content after 250 million years of evolution. Indeed, by comparing in-situ localization at random positions of putative *A. gambiae* and *D. melanogaster* orthologues on polytene chromosomes, another group of scientists reached a similar conclusion that the chromosomal arms of these two dipterans possessed between 41% and 73% of homologous gene content (Bolshakov *et al.*, 2002). This study also suggested the existence of limited micro-synteny (local gene order conservation) between the two species. By comparing a small cluster of genes in *A. gambiae* chromosome 2R with the fruit fly genome, our study was better suited to examine the extent of local synteny. From 20 mosquito genes (clustered in a 265 kb tract) with putative orthologues in the
fruit fly 3R, four clusters on the fruit fly 3R could be traced to. Five mosquito genes (occupying about 50 kb of sequence) matched with a 50 kb cluster at the fruit fly 98A polytene chromosome division (Figure 12). Gene order from these matches was conserved, but extra genes could be seen hyphenated within both fruit fly and mosquito clusters. Four neighboring mosquito genes (in 20 kb of sequence) could be traced to a cluster of 60 kb at division 91E with one transposition in gene order. Hyphenation was again seen in two other small clusters at the fruit fly 85E and 88C divisions, with three genes inserted at 85E and four genes inserted at 88C. Two other clusters were found outside of fruit fly 3R, located at division 69A-B on chromosome arm 3L and 58E on arm 2R, respectively. In summary, our observations suggested that although some local synteny was seen between the two species, their micro-syntenic relationship might not extend beyond circa 50 kb.

When the first draft of *A. gambiae* genome was assembled in 2002, it was discovered to span a circa 278 Mb of DNA, somewhat larger than suggested earlier (Holt *et al.*, 2002). In agreement with our result, comparison of this genome with 179 Mb of the fruit fly genome also led to the conclusion that homology could be seen at whole chromosomal arm or near whole arm level, but the level of micro-synteny was not significant enough to determine the position of orthologues of one’s distant neighbors in the other organism (Zdobnov *et al.*, 2002). This has both disappointing and exciting implications to future studies of insect genomes. The disappointment is that one cannot generally use synteny knowledge from a well-studied organism such as *D. melanogaster* to investigate a novel genome of recent divergence, but this inability also highlights the need to sequence and analyze more genomes from organisms that are important to our
welfare, as is being done for another mosquito of public health significance, Aedes aegypti (Severson et al., 2004).

Our 528 kb sequence did not contain any candidate that could be one of the immune-related genes discussed in the Introduction. Thus, we needed to look beyond our sequence to the rest of the Pen1 region for better candidates. However, our analysis of the 528 kb sequence from the Pen1 region of A. gambiae allowed us a keyhole look at the genome of this medically important species. Multiple haplotypes seen in this region were also observed throughout the genome, and might be caused by the high activity of mobile genetic elements, or might be the result of rapid rate and number of chromosomal rearrangements as seen by cytogenetics and other molecular methods (Sharakhov et al., 2002; Collins, 2005).

3.3 Assembly of the entire Pen1 sequence:

We stopped sequencing more BACs from the Pen1 contig when the availability of assembled raw sequence from the whole-genome project became imminent. Instead we concentrated our resources onto analyzing the available sequence to provide a detailed picture of both the sequence and the gene map to help train the whole-genome assembly and analysis pipeline (Holt R, personal communication). When the un-annotated whole-genome sequence became available, we made use of that resource to assemble the rest of the Pen1 region to reach marker ND3B6. Our entire Pen1 region spanned 1.2 mb of sequence and is represented by 14 BAC clones (Figure 13). Since our Pen1 sequence was published (Thomasova et al., 2002), our colleagues from Genoscope published their own analysis of a 448 kb contig plus a single BAC (137 kb) located in a region centromeric of Pen1, between markers ND3B6 and H788 (Eiglmeier et al., 2005). As a
Figure 13. Diagram of the 1.2 Mb Penl region, presented in two segments. Predicted genes are in red (orienting from telomeric to centromeric) and blue (centromeric to telomeric). Genes are provisionally named after their best prediction program. CLIPB17 is highlighted in red, and is the best candidate for Penl function. A GC% graph is drawn across the Penl regions, highlighting unusual places where GC content drops below the cut-off value. The Penl region is represented by 14 BAC clones, shown as lines.
result, the combination of works on the *Pen1* region and its vicinity established this region as the best studied genomic segment in the *A. gambiae* genome comparable to the *Adh* region in the fruit fly genome (Ashburner et al., 1999).

3.4 Analysis of candidate genes from the *Pen1* sequence:

Gene prediction for the un-annotated sequence obtained from the genome project followed the method that had been used on the 528 kb fragment (see Materials and Methods). Forty six more genes were predicted purely from computational algorithms without the molecular support from a test of gene expression (Figure 13). Among the genes were four probable candidates for *Pen1* function predicted from their domain structure. They included Fgen72 with a serine protease and triosephosphate isomerase domains, Fgen77 with an immunoglobulin and a fibronectin domain, Fgen90 with three Clip domains and a trypsin-like serine protease domain, and GeneID90 with a serine protease domain. All candidates with a serine protease domain could potentially function as enzymes in the serine protease cascade that leads to proteolytic activation of prophenoloxidase (PO), the main enzyme in melanin synthesis and melanization of *Plasmodium*. However, the domain structure of the gene Fgen90 strongly suggested that it was a Clip domain serine protease with prophenoloxidase activating function. When the genome sequence of *A. gambiae* was later annotated, Fgen90 was renamed CLIPB17 to reflect that it was a member of the Clip domain serine protease family. *A. gambiae* genome encoded 41 Clip domain serine proteases divided into four subfamily provisionally named A, B, C, and D. However, CLIPB17 clustered into a hybrid group that also contains members of subfamily A and C (Christophides et al., 2002). To date, only two of these CLIPs from subfamily B (CLIPB14, CLIPB15) have been proven to
affect mosquito immunity against pathogens (Volz et al., 2005) but it seems likely that other members of this subfamily also have roles in insect immunity. Our analysis of CLIPB17 suggested that it was likely a prophenoloxidase activating enzyme (PPAE), as will become more apparent following further bioinformatics and molecular analyses described below.

3.5 Bioinformatic analyses of CLIPB17:

CLIPB17 protein is encoded by three exons separated by two small introns. The first exon encoded the signal peptide as predicted by the program SignalP. The second exon encoded two of the three Clip domains, and the last exon contained the third Clip domain and the protease domain. Promoter Finder predicted a basic RNA Polymerase II promoter binding site 927 bp upstream from the first amino acid. A polyA signal was found 253 bp downstream from the coding region (Figure 14).

CLIPB17 protein had all the features of a typical PPAE, including a signal peptide, three of the Clip domain, one extra cysteine located between the Clip domains and the protease domain, a proteolytic cleavage site after the amino acid lysine, and a protease domain with all the characteristic conserved amino acids including two extra cysteines distinguishing the known PPAEs from CLIPs with different functions. CLIPB17 protein was somewhat larger than the average PPAE, probably due to the extra Clip domains. The definite function for Clip domain remains to be elucidated, but available information suggested that this domain was important for protein regulation (discussed in the Introduction). During the process of melanin synthesis, toxic intermediate quinoid compounds are generated, and these must be tightly controlled to avoid a systemic toxic shock (Christensen et al., 2005). It is conceivable that Clip
Figure 14. Amino acid sequence and domain structure of A. gambiae CLIPB17 serine protease. CLIPB17 has 467 amino acid subdivided into three basic domains: a signal peptide, three Clip domains, and a Trypsin-like S1 family protease domain. Italic amino acids represent the signal peptide. Each of the three Clip domains is enclosed in a box (colored pink). Amino acids of the protease domain are shown in brown. One cysteine (in green) from the peptide remains bonded with the other cysteine from the protease domain (connected with the green arrow) when CLIPB17 is activated by proteolysis after Lys231 (red arrow block). Three disulphide linkages are formed within the protease domain (connected by red arrows). The catalytic triad of the protease domain is shown in blue. Two other cysteines (shown in yellow) are found in all the known insect prophenoloxidase-activating enzymes. RNA Pol II binding site was predicted by the program Promotor Finder. Signal peptide was predicted by SignalP.
domain plays a role toward regulating the protease activity of CLIPB17 to control downstream biochemical reactions in the PPO cascade. It is interesting to note that CLIPB17 has the highest number of the Clip domain among all members of this family in \textit{A. gambiae}.

The orthologue for CLIPB17 in \textit{Drosophila} (CG9733) was located at polytene chromosome division 99E5, very close to the established syntenic block at 98A to which five of the genes from our 528 kb \textit{PenI} fragment could be traced to (see section 3.2). This observation suggested that orthology for CLIPB17 could potentially be inferred both on the ground of their protein sequence similarity and their presence nearby or within the syntenic block between mosquito 8C-8D and its counterpart within the fruit fly 98A-99E chromosomal divisions, despite the significantly larger distance between 98A and 99E.

CLIPB17 also had significant sequence similarity to PPAEs from a number of organisms such as \textit{B. mori}, \textit{M. sexta}, and the beetle \textit{H. diomphalia} (Table 3). Multiple sequence alignment with other Clip domain serine proteases revealed that although the overall sequence similarity among the family members are low (25-35\%), alignment around conserved amino acid remained highly identical (Figure 15). These islands of highly converged sequence were presumably important for structural stability and functional activity of these proteins.

Besides revealing information on overall sequence similarity, multiple sequence alignment was also the fastest method to distinguish between serine protease and serine protease homolog (SPH), which has been defined as a serine protease with a mutated catalytic triad, usually with just one amino acid mutation. These homologues have been discovered by sequence alignment in a number of insects, including \textit{M. sexta}.
# TABLE 3

**SUMMARY OF BLAST HITS FROM ENTREZ DATABASE QUERIED WITH CLIPB17**

<table>
<thead>
<tr>
<th>Organism</th>
<th>Gene</th>
<th>GO Terms</th>
<th>% ID</th>
<th>E-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>D. melanogaster</em></td>
<td>CG9733</td>
<td>GO:0008439</td>
<td>27%</td>
<td>8e-28</td>
</tr>
<tr>
<td><em>D. melanogaster</em></td>
<td>CG4920 (easter)</td>
<td>GO:0004252</td>
<td>28%</td>
<td>3e-26</td>
</tr>
<tr>
<td><em>B. mori</em></td>
<td>PPAE-3</td>
<td></td>
<td>32%</td>
<td>6e-25</td>
</tr>
<tr>
<td><em>M. sexta</em></td>
<td>PAP-3</td>
<td></td>
<td>28%</td>
<td>6e-23</td>
</tr>
<tr>
<td><em>D. melanogaster</em></td>
<td>CG1102</td>
<td>GO:0008439</td>
<td>30%</td>
<td>1e-22</td>
</tr>
<tr>
<td><em>H. diomphalia</em></td>
<td>PPAF-1</td>
<td></td>
<td>28%</td>
<td>2e-21</td>
</tr>
<tr>
<td><em>A. gambiae</em></td>
<td>Sp14D2</td>
<td></td>
<td>30%</td>
<td>6e-20</td>
</tr>
<tr>
<td><em>M. musculus</em></td>
<td>Kallikrein B</td>
<td>GO:0007596</td>
<td>31%</td>
<td>5e-19</td>
</tr>
<tr>
<td><em>R. norvegicus</em></td>
<td>Elastase 2</td>
<td>GO:0004252</td>
<td>28%</td>
<td>4e-19</td>
</tr>
<tr>
<td><em>A. gambiae</em></td>
<td>Sp14D1</td>
<td></td>
<td>29%</td>
<td>3e-18</td>
</tr>
</tbody>
</table>

Note: BLAST hits from Entrez database queried with the amino acid sequence of *A. gambiae* CLIPB17. Homologues include the fruit fly CG9733 at 99E5 chromosomal division, CG4920 at 88E12 division, the silkworm prophenoloxidase-activating enzyme 3 (PPAE-3), the tobacco hornworm prophenoloxidase-activating protein 3 (PAP-3), the fly CG1102 serine protease, the beetle prophenoloxidase-activating factor 1 (PPAF-1), the mosquito Clip-domain serine proteases from division 14 (Sp14D1, Sp14D2), and the mammalian serine proteases. GO:0008439 (prophenoloxidase activator), GO:0004252 (serine protease activity), GO:0007596 (blood clotting).
Figure 15. Multiple sequence alignment of the protease domain of CLIPB17 (fgen90) performed by ClustalW version 1.81. This alignment includes known Clip domain serine proteases from a number of insects, some of which are known PPAEs. Several islands of high sequence conservation are seen scattered throughout the alignment, and are presumably important for the structure and function of these proteins.
(Yu et al., 2003), *D. melanogaster* (Ross et al., 2003), and *A. gambiae* (Dimopoulos et al., 1997). In *M. sexta*, they played a role as cofactors to enhance prophenoloxidase activity of PPAEs.

Protein structure of a Clip domain serine protease in the horseshoe crab has been examined with X-ray crystallography (Muta et al., 1993), enabling homology-based structure prediction of CLIPB17 (Figure 16). The protein model for CLIPB17 illustrated specific folding of this polypeptide to enable the proximity of the catalytic triad within the substrate binding clef. Presumably, this folding pattern was guided by conserved amino acids and was stabilized by the highly conserved disulphide linkages.

3.6 Expression profile of CLIPB17:

The conclusion from bioinformatic analyses described above suggested that CLIPB17 has a high probability to encode for Pen1 function. To further demonstrate this probability, we undertook to profile mRNA expression of this gene through out the major developmental stages from both *Plasmodium*-refractory (L35) and susceptible (4Arr) mosquitoes. Total RNA was collected from developmental stages from embryo to adult and subjected to RT-PCR experiment using the expression level of ribosomal protein S7 as a semi-quantitative control. Several observations could be drawn from the result shown in Figure 17. Higher level of expression could be seen in L35 mosquitoes at most developmental stages than in 4Arr. Secondly, expression was up-regulated in L35 pupae but not 4Arr. S7 control expression remained at similar level throughout the developmental stages. Detection of gene expression for CLIPB17 provided the first molecular evidence to confirm our previous bioinformatic gene prediction and warranted further experiments with this gene. The relatively higher expression in L35 was
Figure 16. Homology-based structure prediction for the protease domain of CLIPB17 using the SwissProt database. To simplify the model, only peptide backbone for all amino acid is shown, except for the side-chain groups of the catalytic triad (HIS47, ASP94, SER182) showned here to illustrate their proximity within a substrate binding clef.
Figure 17. Developmental expression profile of CLIPB17 assayed by semi-quantitative RT-PCR using expression level of the housekeeping gene S7 as control. Expression of CLIPB17 can be visually compared between the *Plasmodium*-refractory strain L35 and the susceptible strain 4Arr through out developmental stages from embryo to adult. At most stages, expression level in L35 is higher than in 4Arr. Interestingly, expression level of L35 mosquito at the pupae stage is up-regulated (further discussion in the text).
intriguing to us and suggested that L35 mosquitoes might be in a constant higher state of readiness to response to future immune challenges. Two observations from the literature supported this hypothesis. From morphological and expression profiling analysis, it was suggested that the refractory and susceptible strains have broad expressional and physiological differences, such as the differences related to the production and detoxification of reactive oxygen species (ROS) (Kumar et al., 2003). The concentration of ROS in refractory mosquitoes was consistently higher than that detected in the susceptible strain. It was further concluded that these refractory mosquitoes (L35) were in a constant state of oxidative stress, and more importantly, inhibition of ROS reduced the level of melanization of implanted Sephadex beads. From these observations, it could be envisioned that the constant oxidative stress in the refractory mosquitoes had an indirect effect to melanization via causing a higher expression of immune-related genes to respond more quickly to challenges. At this time, expressional up-regulation found exclusively in pupae of the refractory strain remains unexplained. We hypothesize that further analysis of the regulatory regions of CLIPB17 between refractory and susceptible mosquitoes will shed light upon this observation (see later data).

To examine the role of CLIPB17 toward Plasmodium infection, RT-PCR experiment was performed to measure gene expression level at 24 hours after the mosquitoes were exposed to P. berghei-infected blood. The result strongly indicated that this gene was up-regulated upon Plasmodium infection in refractory mosquitoes at the time when the ookinetes actively penetrated the midgut epithelium on their way to the basal lamina (Figure 18). Taken together with putative function suggested by sequence
homology, the demonstration of expressional up-regulation upon immune challenge established CLIPB17 to be a strong *Pen1* candidate for further analysis.
Figure 18. RT-PCR expression of L35 and 4Arr mosquitoes infected with *P. berghei* measured at 24 hpi (hours post infection). Sugar-fed mosquitoes (naive) were used as control. CLIPB17 expresses at a higher level in L35 than in 4Arr, and is up-regulated in L35 in response to *P. berghei* infection.
3.7 Quantitative real-time RT-PCR optimization for CLIPB17:

To analyze gene expression of CLIPB17 more critically, we employed the relatively novel technique of quantitative real-time RT-PCR (qRT-PCR) that had been proven to be a sensitive and reproducible assay. In this assay, proprietary PCR chemistry (Taqman chemistry) and primer design software (Primer Express Software v 1.5) from Applied Biosystem were used to achieve sensitive and accurate results following gene-specific optimization steps. Sensitivity was optimized in preliminary experiments that included the optimization of primer and probe concentrations to produce an earliest amplification time and highest maximum amplified product given constant amount of RNA template. Better accuracy stems from the fact that the signal to be measured is only available when two primers and one fluorescent probe simultaneously anneal to the right template. Three simultaneous hybridization events across a small sequence (100-200 bp) reduce the chance of non-specific annealing giving rise to fall positive result. More detail of this assay was described in the Materials and Methods, and Figure 5 illustrated the chemical events happening during the PCR reaction. Our preliminary experiments clearly showed that for CLIPB17, optimal primer concentration is 600 nM, while that for the probe is 100 nM. Similarly for S7 transcript, 100 nM each for both primer and probe produced the best condition (data not shown). Further preliminary experiments were also performed to construct a standard curve plotting changes in C$_t$ value as the function of ten-fold dilutions of templates (Figure 19). These plots could be used to derive the template concentration from C$_t$ data obtained from each experiment.
Figure 19. Amplification plots and standard curve graphs of S7 and CLIPB17. (A) Amplification plots for ten-fold dilutions from $10^5$ to $10^2$ pg of S7 RNA template. (C) represents the standard curve plot of the change in C$_t$ value according to the dilutions seen in (A). (B) plots the same ten-fold dilutions but for the gene CLIPB17, and (D) shows the standard curve for this gene. Each ten-fold dilution of the template should ideally correspond to 3.3 C$_t$ value change, and this is reflected by the slope of the standard curve equation. For CLIPB17, this slope is -3.37, very close to the ideal value. For S7, the slope is -3.0 but it is still within acceptable range, indicated by a high R$^2$ value of >0.99.
3.8 qRT-PCR analysis of CLIPB17:

Using qRT-PCR, we re-analyzed the difference in CLIPB17 expression level between L35 and 4Arr that had been observed earlier. Our result from this experiment confirmed that transcript level of this gene was significantly higher in L35 than in 4Arr mosquitoes without an immune challenge (Figure 20). For each experiment, a pool of eight mosquitoes was used for total RNA isolation to minimize gene expression variation among individual mosquito of the same genetic background. In other word, when we compared two expression levels, we were comparing the mean expression level of 8 mosquitoes in one group against the mean expression level of 8 different mosquitoes. Consequently, only one or two experimental replications were required to generate statistically significant data. Most of our subsequent experiments have been performed more than once unless indicated otherwise. Even with the experiments that could only be done once to date, we were confident that the result would agree well with future replications.

We also used a GFP-expressing *P. berghei* parasite (Franke-Fayard *et al.*, 2004) to infect both strains to determine whether expression in L35 was up-regulated, as seen previously. The result from Figure 21 strongly indicated that CLIPB17 was up-regulated in the refractory mosquitoes specifically after recognition of *Plasmodium* infection. It was not up-regulated following an un-infected blood meal, indicating that pattern receptor-mediated recognition of *Plasmodium* might trigger both the proteolytic activation of the signaling cascade involving serine proteases, as previously known for other insects, and the induction of gene expression from these proteases, as shown here with the putative PPAE from the *Penl* region. Our result constituted a significant finding
Figure 20. Quantitative real-time RT-PCR of CLIPB17 from L35 and 4Arr mosquitoes not exposed to malaria infection. Expression level in L35 is consistently higher than in 4Arr from data collected in two independent experiments (Exp1 and 2).
Figure 21. qRT-PCR expression profile of L35 and 4Arr mosquitoes that have been fed with either sugar (female), non-infected blood (BF24h), or *P. berghei* infected blood (Pb24hpi). Three independent experiments are shown. Up-regulation of CLIPB17 is only seen in infected L35 mosquitoes.
because it again implicated the role of CLIPB17 toward *Plasmodium* encapsulation seen in L35 mosquitoes (Figure 22).

For a gene with possible *Pen1* function, CLIPB17 should localize to the organ used as the entry point by *Plasmodium* to gain access to mosquito tissue, namely the midgut, or its localization should indicate a function toward immunity. The fat body probably the most immune-active organ in the mosquito. It is known that most antimicrobial peptides that must be able to gain access to the hemolymph quickly for a systemic response are produced in the fat body cells (Dimopoulos, 2003). Hemocytes in mosquito species such as *Armigeres subalbatus* are also important for the immunity of that mosquito against diverge pathogens from bacteria to parasites (Hillyer *et al.*, 2003). As the consequence of having comparatively few hemocytes (Paskewitz *et al.*, 1989), melanotic encapsulation response seen in *A. gambiae* strain L35 may be entirely a humoral immune reaction. Cells of the midgut and/or the fat body are good candidates as the source of gene expression up-regulation seen with CLIPB17. To assay where expression of CLIPB17 was localized within the mosquito, midguts from L35 individuals were separated from the abdomen section (including most fat body tissue) and the rest of the carcasses (head and thorax portion) before these tissues were subjected to RNA isolation. Gene expression was detected in both the midgut and the abdomen that included fat body tissue (Figure 23). Furthermore, transcription was up-regulated after these L35 mosquitoes were infected with *P. berghei*, confirming that these tissues are immune active and responsive to immune signaling when challenged with *Plasmodium*. Two other members of Clip domain serine protease from subfamily B (CLIPB14, CLIPB15) have been recently characterized (Volz *et al.*, 2005). They were shown to be
Figure 22. Midguts from L35 (A) and 4Arr (B) mosquitoes were dissected 12 days after they were infected with GFP-expressing *P. berghei*. Pictures were taken at 200X magnifications. Parasites in L35 are encapsulated in a melanin matrix. Surviving parasites in 4Arr midgut express GFP strongly at all developmental stages except gametocytes.
Figure 23. Localization of CLIPB17 at the midgut and fat body tissues of L35 mosquitoes. qRT-PCR also detected up-regulation at these tissues when the mosquito is infected with *Plasmodium*. 
induced by Gram-negative and *Plasmodium* challenges, but inductions were detected equally in both a melanotic susceptible and the refractory strain L35, in contrast to what is seen with CLIPB17. Furthermore, they were shown to be expressed exclusively in hemocytes and their gene products secreted into the haemolymph. As suggested by the contrast to CLIPB17 in expression profile and localization, they were found not to affect ookinete melanization in the L35 strain but rather functioned in limiting the number of parasites successfully transformed into oocysts. When they were knock-downed by RNA interference, increased number of living oocysts was seen in midguts of a susceptible strain, while increased number of melanized oocysts was seen in the refractory strain L35. It is interesting to observe that members of the same subfamily would have seemingly different functions yet still have effects on refractoriness against *Plasmodium*. It demonstrates that the interaction between the mosquito and *Plasmodium* may necessarily be much more complex and interactive than previously thought, and in response to this complexity, the immune system may employ more than one way to limit parasite survival.

In addition to our interest in the localization of CLIPB17 within the mosquito body, we were also interested in the temporal expression pattern of CLIPB17 in response to *Plasmodium* infection. The timing of melanization of parasites is peaked around 24 hours post infected blood meal. Thus, it was reasonable to speculate that expression of CLIPB17 would also reach maximum around this time frame. We observed that CLIPB17 gene was up-regulated as early as 12 hours post blood meal, and expression gradually returned to basal level during the next three days (Figure 24). The early expressional up-regulation around 12 hours post infection (hpi) was somewhat
Figure 24. Temporal expression of CLIPB17 measured in L35 mosquitoes from the time of infection with *P. berghei* (Pb) to 3 days later. Up-regulation is detected as early as 12 hours post blood meal (12hpi), and expression gradually returns to basal level during the next three days.
unexpected. It might be the result of non-synchronized infection of the midgut, in which some parasites started to invade by 12 hpi, or it might just reflect the fact that what we measured were mRNA transcript and not the actual CLIPB17 protein, whose expression might be preceded by transcriptional up-regulation, but which would become more concentrated in the next 12 hours in time for the encapsulation response. We also could not exclude the possibility that the stress of blood feeding induces a quick expression of CLIPB17 detectable at 12 hpi but that expression level would return to basal by 24 hpi if no recognition of *Plasmodium* had taken place. In any case, more replications of this experiment would have to be done to extract a more accurate temporal expression pattern of this gene.

In light of our increasing realization that the immune system plays a much more complex and interactive role with encountered pathogens, we were interested to profile CLIPB17 expression in response to different type of pathogens such as the Gram-negative *Escherichia coli* and Gram-positive *Micrococcus luteus* bacteria, and also the yeast *Saccharomyces cerevisiae*. When measured at 12 hours, 24 hours and 48 hours post infection, no significant up-regulation of CLIPB17 was observed in response to any of these elicitors (Figure 25). Although more replications and infection methods should be attempted, the available result from this experiment intriguingly pointed to a specific response of CLIPB17 limited to *Plasmodium* infection. If it proves to be the case, we may then regard CLIPB17 as an example of a species-specific and functionally adapted protein unique to mosquito immunity against the malaria parasite.
CLIPB17 expression is not regulated by bacteria or yeast infection

Figure 25. CLIPB17 expression profile from L35 mosquitoes was measured at 12 hours, 24 hours, and 48 hours after these mosquitoes were punctured in the thorax with either a sterile needle, a needle dipped in a suspension of *E. coli* and *M. luteus*, or a needle dipped in a yeast suspension. No regulation was seen under these conditions.
3.9 Cloning and sequencing of CLIPB17:

Our result from qRT-PCR analysis clearly pointed to the conclusion that at least one of the mechanisms regulating CLIPB17 protein activity must involve transcriptional regulation. To further explore this possibility, we have cloned and sequenced both the coding region and the 5’ regulatory region of this gene from both L35 and 4Arr strains by a PCR-based approach. A 1.6 kb fragment was amplified from reverse-transcribed cDNA from multiple individuals of each strain and cloned into a plasmid vector. Sequence analysis of these clones revealed no difference between the two strains in the translated amino acids, and confirmed previous computational predictions of exon/intron boundaries for this gene (data not shown). This high level of sequence conservation was in contrast to the highly variable genes seen in other parts of the Pen1 region such as the gene 22J3.4 examined in a previous section of this thesis. We then turned our attention to obtaining sequence data for the 5’ regulatory region of CLIPB17. We cloned this gene from multiple individuals of each strain and sequenced multiple clones from each individual. Alignment of the sequences revealed all strain-specific mutations as shown in Figure 26. Most importantly, there were three putative Dorsal/NF-kB binding sites predicted by the TRANSFAC database (Meseguer et al., 2002) at places with multiple nucleotide changes between the refractory and susceptible strains. These putative response elements were located at -191, -258, and -701 from the transcriptional start site within a RNA Polymerase II promoter binding site predicted by the program Promoter Finder (Figure 27). There were also five E74A binding sites, with three of them located within a 66 bp sequence deleted in the 4Arr consensus. Two other E74A sites were oriented within a palindromic sequence. Other nucleotide changes were present but their
Figure 26. Sequence alignment of the 5’ regulatory region of CLIPB17. Strain specific differences are observed between the refractory strain L35 and the susceptible strain 4Arr. Three putative NF-kB binding sites are co-localized with many mutations. Five E74A binding sites are found within this sequence, and three of them are located within a 66 bp deletion in the 4Arr sequence. Two E74A sites are located within a palindromic sequence, indicated by arrows. NF-kB site is a response element for Rel family transcription factors of the *Drosophila* Toll and IMD immune signaling cascades. E74A are binding sites for E74 transcription factors of the Ets superfamily required for pupal development (Urness and Thummel, 1995). Dotted lines represent segments with no sequence difference. Red color bases represent mutation from L35 sequence. Site prediction was performed using the program PROMO (Messeguer et al., 2002) that queries the TRANSFAC database of existing binding site matrix.
Figure 27. Immune-related transcription factor binding sites within a 1.6 kb upstream from the coding region of gene CLIPB17. This sequence shows strain-specific mutations between L35 and 4Arr such as a 66 bp deletion in 4Arr. NF-kBs are involved in Toll and IMD immune signaling cascade (Silverman and Maniatis, 2001). STATs are transcription factors from the JAK/STAT signaling pathway implicated in immune responses (Barillas-Mury et al., 1999). IRF are interferon regulatory factors often involving in enhancing the activity of other factors such as STAT or NF-kB (Smith et al., 2005). E74A is an ecdysone-responsive transcription factor involving in pupal development and metamorphosis (Fletcher and Thummel, 1995). This diagram was created with VectorNTi suite version 8.0.
functional implication remained obscure. Half of the nucleotide changes were clustered into the three NF-kB sites, while the rest were more randomly scattered. Functions of the corresponding transcription factors for these binding sites were described in the figure legends (Figure 27). The mosquito genome encodes two Rel homology NF-kB-like factors named Rel1 and Rel2. Rel1 is homologous to the *Drosophila* Dorsal/Dif transcription factor, while Rel2 is more similar to the factor Relish in the fruit fly (Christophides et al., 2004). As described in the Introduction, both of these factors were involved in immune signaling cascades affecting most aspects of insect immunity (De Gregorio et al., 2002). A striking example of this is the growing realization that the Toll pathway controls melanization in the fly haemolymph (Naitza and Ligoxygakis, 2004). This conclusion came about after a study of a serine protease inhibitor (serpin Spn27A) indicated that its depletion from the haemolymph during the melanization process is dependant on a functional Toll pathway (Ligoxygakis et al., 2002b). In Toll pathway mutants, Spn27A depletion and PO activation are blocked. In the mosquito *A. gambiae*, RNAi knockdown of Cactus (the Rel1-associated inhibitory protein in the cytoplasm) in a susceptible mosquito completely reversed its melanization phenotype to refractory, indicating for the first time that the mosquito Toll pathway played a role in encapsulation of *Plasmodium* via Rel1-mediated expressional up-regulation (Frolet et al.). The authors then speculated that any melanization pathway gene whose expressional up-regulation could be measured in a Cactus-deficiency background must be controlled by the mosquito Toll pathway. The opsonin TEP1 was discovered to be one such gene. It was induced after injection of dsCactus and its regulatory sequence contained at least one NF-kB site (the allele of TEP1 found in refractory mosquitoes has two of these sites).
sequence analysis revealed that there were three NF-kB sites upstream of CLIPB17, all of them showed multiple nucleotide changes between L35 and 4Arr mosquitoes. When this observation is combined with the up-regulation of L35-specific CLIPB17 upon Plasmodium challenge observed with qRT-PCR, we can confidently propose that CLIPB17 may be another of the genes (besides TEP1) whose up-regulation is mediated by mosquito Toll/Rel1 signaling. Specifically, we propose a model in which parasite recognition by pattern receptors, including such protein as TEP1, will lead to proteolytic activation of at least two cascades, one that involves the extracellular protease cascade transducing the activation signal to PPAE(s) in the haemolymph, and one that activates the intracellular Toll pathway to induce gene expression of such proteases, including the putative PPAE CLIPB17 in the refractory mosquitoes. At least one of the genes involved in these signaling cascades must fail to function or to be up-regulated in the 4Arr background, and that failure may be because of insufficient binding caused by mutations in the NF-kB sites upstream of the CLIPB17 allele in these mosquitoes.

In our RT-PCR experiment when we followed the expression profile of CLIPB17 throughout the major developmental stages, we observed that this gene was significantly up-regulated during the pupae stage. We could now explain that the higher expression level differentially seen in pupae of L35 mosquitoes but not those of 4Arr could be the result of E74A transcription factor regulation. In Drosophila, this Ecdysone-sensitive transcription factor is known to be highly activated in pupae to regulate its development and metamorphosis (Fletcher and Thummel, 1995). Thus the extra E74A sites found specifically in the L35 sequence (Figure 26) might explain how the L35 allele of CLIPB17 could direct higher gene expression at this developmental stage.
3.10 RNAi-mediated effect of CLIPB17 on melanization:

One of the most remarkable stories in biology recently has been the discovery that an unusually double-stranded form of RNA mediates silencing of eukaryotic genes via degradation of their cognate mRNAs. First discovered in the nematode *Caenorhabditis elegans* (Fire *et al*., 1998), RNAi is now known to work efficiently in insects (Hammond *et al*., 2001) including the mosquito *A. gambiae* (Blandin *et al*., 2002). When we considered the most effective functional assay to prove whether CLIPB17 has *Pen1* function, we had limited options from either the transgenic approach or the genotype/phenotype association approach. We had previously shown the association between specific mutations in the regulatory region of CLIPB17 to the susceptible strain. To complement that data, we opted for a transgenic approach to confirm gene function. We could employ either stable germ-line transformation with a transposable element or a transient knockdown with RNAi in adult mosquitoes. To date, transformation in *A. gambiae* remained troublesome, and we thus decided toward using RNAi. We synthesized and micro-injected a double-stranded RNA (dsRNA) fragment of CLIPB17 into female L35 mosquitoes at two to three days old. After an incubation period of three to four days, these mosquitoes were fed on rats infected with *P. berghei*, or on a macaque monkey infected with *P. cynomolgi* B. The mosquitoes were then allowed an appropriate period of time for the parasites to develop into oocysts before midguts were dissected and examined for both live and melanized parasites. To assay the molecular effect of double-stranded CLIPB17 (dsB17), we compared gene expression by qRT-PCR in L35/dsGFP versus L35/dsB17 at 24 hours post infection. At this time, transcript level in L35/dsB17 dropped back to the basal level seen in naive L35 used as control, while L35/dsGFP
expression level remained elevated, indicating that dsB17 indeed exhibited a specific RNAi effect to CLIPB17 in this mosquito (Figure 28). We also observed the expected effect of dsB17 upon refractoriness by melanotic encapsulation (Figure 29). Specifically, non-melanized *P. berghei* oocysts were seen in the midgut of L35/dsB17 background. More data were collected from this experiment and a quantitative summary is presented in Table 4. T-test statistics were performed on the mosquito samples from each background (L35/dsGFP to serve as non-specific dsRNA control versus the experimental L35/dsB17), and statistically significant reduction of *P. berghei* refractoriness from ~100% to 50% was reported as the result of dsB17 knockdown effect. This effect was also seen in the case of *P. cynomolgi* B infection, but at a lower magnitude. Specifically, dsB17 knockdown caused 23% reduction in the melanization ability of L35 mosquitoes against this parasite.

Our data from this experiment complemented earlier expression and sequence data to support the conclusion that serine protease CLIPB17 encodes the function described genetically as the *Pen1* locus in the mosquito strain refractory to *Plasmodium* infection. Molecularly, this serine protease is one of the prophenoloxidase-activating enzymes (PPAEs) regulating melanin synthesis. The genome of insects such as *A. gambiae* and *D. melanogaster* encode multiple genes that are good candidates for a PPAE because of their structural similarity to that protein from other insects. We determined for the first time that one of those candidate genes has PPAE function by additionally showing that it is mapped physically to the genetic locus known to affect melanization, and that the ability of some mosquitoes to melanize *Plasmodium* correlates with its higher gene expression at the time of midgut invasion and transformation of ookinetes.
Effect of double-stranded RNA injection on gene expression of CLIPB17 in L35 mosquitoes

Figure 28. Effect of dsB17 injection on transcript level of CLIPB17 in L35 mosquitoes. At 24 hours post infection (hpi), *P. berghei*-infected mosquitoes of L35/dsB17 background exhibit a reduction in transcript level back to the basal level seen in naive L35. Two independent experiments (Exp) are reported.
Figure 29. Bright field and FITC views of the same field in a midgut of L35/dsCLIPB17 background. Both melanized (top) and live oocysts (bottom) are visible in these pictures. The presence of live oocysts in the refractory midgut in which induction of CLIPB17 is prevented demonstrate visibly the direct effect of CLIPB17 gene expression level upon the refractory phenotype of this mosquito.
TABLE 4

NUMBER OF MELANIZED AND NON-MELANIZED PARASITES IN DIFFERENT GENETIC BACKGROUND

**Plasmodium berghei** data

<table>
<thead>
<tr>
<th>Strain/injection (background)</th>
<th>Midguts #</th>
<th>Mel Par</th>
<th>Oocyst</th>
<th>Total Par</th>
<th>Par/midgut [mean(range)]</th>
<th>% Mel</th>
<th>P-value (t-test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L35/dsGFP</td>
<td>39</td>
<td>242</td>
<td>1</td>
<td>243</td>
<td>6.2 (1-20)</td>
<td>99.5</td>
<td></td>
</tr>
<tr>
<td>L35/dsB17</td>
<td>39</td>
<td>121</td>
<td>127</td>
<td>248</td>
<td>6.5 (1-37)</td>
<td>48.8</td>
<td>0.000252</td>
</tr>
<tr>
<td>4Arr</td>
<td>40</td>
<td>0</td>
<td>350</td>
<td>350</td>
<td>8.7 (1-36)</td>
<td>0.0</td>
<td></td>
</tr>
</tbody>
</table>

**Plasmodium cynomolgi** B data

<table>
<thead>
<tr>
<th>Strain/injection (background)</th>
<th>Midguts #</th>
<th>Mel Par</th>
<th>Oocyst</th>
<th>Total Par</th>
<th>Par/midgut [mean(range)]</th>
<th>% Mel</th>
<th>P-value (t-test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L35/dsGFP</td>
<td>23</td>
<td>76</td>
<td>0</td>
<td>76</td>
<td>3.5 (1-10)</td>
<td>100.0</td>
<td></td>
</tr>
<tr>
<td>L35/dsB17</td>
<td>23</td>
<td>51</td>
<td>15</td>
<td>66</td>
<td>2.9 (1-7)</td>
<td>77.2</td>
<td>0.001079</td>
</tr>
</tbody>
</table>

Note: Numbers of melanized and normal oocysts in L35 mosquitoes injected with either a control dsGFP fragment (L35/dsGFP genetic background) or with the experimental dsB17 (double-stranded RNA fragment from the gene CLIPB17). The ratio between melanized parasites (Mel Par) and total number of parasite (Total Par) defines percent melanization (% Mel) for each genetic background listed under Strain/injection. T-test was performed on normal oocyst number between samples in each genetic background.
into oocysts, and most importantly, by showing via transcript inhibition its direct effect on melanization.
4. CONCLUSION

More than a hundred years after Ronald Ross reported his observation of the black pigments found in the guts of some mosquitoes in the field (Ross, 1898), we described in this thesis a gene affecting the synthesis of the melanin pigments causing these “Ross spores” in a laboratory selected strain of A. gambiae (L35 strain). These mosquitoes were selected for the ability to encapsulate Plasmodium parasites with melanin polymers, resulting in the pigmented appearance of dead parasites seen by Ross. When this ability was investigated genetically, it was discovered to be affected quantitatively by a number of loci, including a major locus named Pen1 located in a genomic region corresponding to division 8C-8D on polytene chromosome arm 2R. To clone this gene from its genetic position, we constructed a physical clone map with 14 BACs across the Pen1 locus to serve as templates for sequencing across this region. We have sequenced six BACs and assembled five of them into a 528 kb fragment representing about half of the Pen1 region. The sequence from one entire BAC clone was found to be part of an alternate haplotype with as much as 3.3% sequence variation from the 528 kb consensus. This type of haplotype variation was later found throughout the genome of A. gambiae, suggesting that this mosquito possesses a genome with an unusually high degree of sequence variation in addition to the frequent occurrences of chromosomal inversion polymorphism characterized by cytogenetics. The combination of these chromosomal inversions and local sequence polymorphisms is probably the driving force behind the speciation and adaptation events giving rise to geographical and
seasonal variation of malaria transmission seen among members of the *A. gambiae* species complex (Michel et al., 2005; Sharakhov et al., 2002). Although 528 kb represented less than half a percent of the genome, it was the longest sequence fragment at the time, prompting us to perform sequence comparison to the genome of the fruit fly *D. melanogaster*. A small number of genes from our *Pen1* region could be traced to their orthologues on division 98A in the fruit fly genome, but at the chromosomal arm level, more than half of the *Pen1* genes had orthologues on arm 3R of the fruit fly. This conclusion from our analysis enabled us to predict that there could be a small degree of micro-synteny between the genomes of the mosquito and the fruit fly. Besides permitting a glimpse of the genome structure, the assembled sequence of the *Pen1* region allowed us to search for candidate genes by computational methods. Gene prediction and database searches revealed a candidate gene with sequence similarity to prophenoloxidase-activating enzymes from a number of insects. These enzymes are known to regulate the biochemical production of the melanin that mosquitoes use to encapsulate a variety of pathogens, including *Plasmodium* species. We analyzed further to show that this serine protease (CLIPB17) was expressed in the midgut and abdomen at the time *Plasmodium* ookinetes were moving across the midgut to rest at the basal lamina. More interestingly, we showed that CLIPB17 transcript level was up-regulated at this time when parasites were exposed to the immune system in the haemolymph, and that transcript induction was specific for the refractory L35 mosquitoes infected with *P. berghei* and *P. cynomolgi*. Further studies showed that CLIPB17 transcript up-regulation was also a specific response against *Plasmodium* infection, but not against infections by bacteria or fungi. Sequence analysis of the regulatory region of this gene indicated that expression up-
regulation might be induced by the mosquito’s Toll pathway equivalence. Toll pathway affects many processes in *Drosophila*, including the regulation of antimicrobial peptide synthesis and melanization (Naitza and Ligoxygakis, 2004). In the mosquito, knocking down gene expression of a Toll pathway inhibitor caused complete reversal from a susceptible background to become refractory (Frolet *et al.*), demonstrating its role toward regulation of genes involving in the melanin synthesis process. We showed that there were multiple mutations located in the promoter region of CLIPB17 allele from the susceptible mosquitoes. Moreover, these mutations clustered into three sites corresponding to putative NF-kB response elements and thus could affect binding to mosquito Rel homology transcription factors (Rel1 and Rel2) similar to those used in the Toll and IMD pathways. We performed RNAi knockdown experiment to directly test for the role of CLIPB17 in affecting the ability of L35 mosquitoes to melanize parasites. Our result clearly showed that without transcript up-regulation, a number of parasites escaped from being melanized in L35. All together, these data definitely linked *Pen1* function to the CLIPB17 allele from a laboratory selected strain of *A. gambiae*. The experiments and interpretations presented here also demonstrate for the first time in this medically important mosquito a positional cloning study linking a gene only described genetically to its molecular identity.
5. REFERENCES


Dimopoulos, G., D. Seeley, A. Wolf, and F. C. Kafatos. "Malaria Infection of the Mosquito Anopheles Gambiae Activates Immune-Responsive Genes During


