Chapter 1
Review of literature
1.1 Brief history of Mosquito Vectors

Mosquitoes are insect vectors responsible for the transmission of parasitic and viral infections to millions of people worldwide, with substantial morbidity and mortality. Infections transmitted by mosquitoes include malaria, yellow fever, chikungunya and other arboviruses [1]. An understanding of mosquito evolution, distinguishing features, and the insect life cycle is important for disease investigation as well as for designing and implementing effective methods for disease control.

Mosquitoes belong to the class Insecta, order Diptera and family Culicidae. The two subfamilies are Anophelinae (which includes the genus *Anopheles*, the mosquito vector for malaria) and Culicinae (which includes the genera *Aedes* [Stegomyia], *Culex*, *Mansonia* and *Haemagogus*, the mosquito vectors for arboviruses and some filariases) [1]. Each subfamily has hundreds of species within it, although only a few dozen bite humans and therefore are capable of serving as disease vectors. Planning disease control measures requires identification of the mosquitos in a particular geographic region.

The mosquito life cycle consists of four stages: egg, larva, pupa, and adult. The full life cycle usually takes about 14 days, but the duration varies with temperature and species.

1.2 Anopheline mosquitos

Malaria is transmitted among humans by female mosquitoes of the genus *Anopheles* (Figure 1.1). Female mosquitoes take blood meals to carry out egg production, and this is the link between the human and the mosquito hosts in the parasite life cycle. The successful development of the malaria parasite in the mosquito depends on several factors. The most important is ambient temperature and humidity. Higher temperatures accelerate the parasite growth in the mosquito [2]. Secondly, the *Anopheles* should survive long enough to allow the parasite to complete its cycle in the mosquito host [3]. There are approximately 3,500 species of mosquitoes grouped into 41 genera. Human malaria is transmitted only by females of the genus *Anopheles*. Among approximately 430 *Anopheles* species, only 30-40 transmit malaria (i.e., are "vectors") in nature. Anophelines are distributed worldwide except Antarctica. Malaria is transmitted by different *Anopheles* species, depending on the region and the environment.
1.2.1 Anopheles stephensi

*A. stephensi* is the major vector of malaria caused in humans from Middle East and South Asia region, and belongs to the same subgenus as *A. gambiae*, the main malaria vector from Africa [4]. *A. stephensi* is widely used in laboratories in Europe and the United States to propagate malaria agent, *Plasmodium berghei*, a species that only infects rodents. Three biological forms are found in India, Iran and surrounding regions which are distinguished on the basis of differences of eggs. The type form, *A. stephensi stephensi*, was reported to inhabit urban areas and *A. stephensi mysorensis* and *A. stephensi intermediate* are common in rural areas and generally both have been described to be zoophilic, preferring cattle in rural areas and humans in urban areas [5]. This mosquito usually rests in different buildings, cattle sheds or underground shelters such as wells. Favourable breeding sites of *A. stephensi* are wells, cisterns, empty containers, roofs gutters, hoof prints of animals, rice fields, fountains and many more [6, 7]. Larval breeding is significantly higher during the rainy season [7].

Adult females and males become sexually mature by the 2nd night of life. Although *A. stephensi* could live up to 26 days in laboratory conditions, all adults are dead in about 11 days in natural conditions. Females lay about 100 eggs during one oviposition and maximum nine oviposition can occur [7]. *A. stephensi* is a disease vector of rodent malaria (*P. berghei*, *P. yoelli*), simian malaria (*P. cynomolgi*) and human malaria (*P. falciparum*) [8].

1.2.2 Anopheles gambiae

*A. gambiae* is the major vector of *P. falciparum* in Africa and is one of the most efficient malaria vectors in the world. Every year, more than 500 million people become severely ill with malaria, with most cases and deaths occurring in sub-Saharan Africa. Asia, Latin America, the Middle East and parts of Europe are also affected. *A. gambiae* is strongly anthropophilic species, which means it, prefers to feed on humans and so it is more likely to transmit the *Plasmodium* parasites from one person to another. The adult females can live up to a month (or more in captivity) but most probably do not live more than 1-2 weeks in nature. Males live for about a week, feeding on nectar and other sources of sugar. The *A. gambiae* female, mates only once and stores sperm for subsequent egg production. Adult
females lay 50-200 eggs per oviposition and it takes 2-3 days for eggs to hatch. Larvae develop in wide range of temporary bodies of water and pass three moult before becoming pupae. These mosquitoes can develop from egg to adult in as little as 5 days but usually take 10-14 days in tropical conditions. There is a complex of around six morphologically indistinguishable species very close to A. gambiae and they are called A. gambiae complex. [http://www.allmosquitos.com/mosquito-species/anopheles-gambiae.html].
1.3 *Plasmodium* life cycle

The malaria parasite is protists with a complex life cycle requiring an insect vector and a human host. There are approximately 156 named species of *Plasmodium* which infect various species of vertebrates. Four species are considered true parasites of humans, as they utilize humans almost exclusively as a natural intermediate host: *P. falciparum*, *P. vivax*, *P. ovale* and *P. malariae*. However, there are periodic reports of simian malaria parasites being found in humans, most reports implicating *P. knowlesi*.

![Complete life cycle of *Plasmodium*. Image courtesy CDC](http://www.cdc.gov/malaria/about/biology/)

**Figure 1.1:** Complete life cycle of *Plasmodium*. Image courtesy CDC

http://www.cdc.gov/malaria/about/biology/
There are three phases in its life cycle: the pre-erythrocytic cycle, the erythrocytic cycle, and the sporogonic cycle. The malaria parasite life cycle involves two hosts. During a blood meal, a malaria-infected female *Anopheles* mosquito inoculates sporozoites into the human host. Sporozoites infect liver cells and mature into schizonts, which rupture and release merozoites. (Of note, in *P. vivax* and *P. ovale* a dormant stage [hypnozoites] can persist in the liver and cause relapses by invading the bloodstream weeks, or even years later). After this initial replication in the liver (exo-erythrocytic schizogony), the parasites undergo asexual multiplication in the erythrocytes (erythrocytic schizogony). Merozoites infect red blood cells. The ring stage trophozoites mature into schizonts, which rupture releasing merozoites. Some parasites differentiate into sexual erythrocytic stages (gametocytes). Blood stage parasites are responsible for the clinical manifestations of the disease.

The gametocytes, male (microgametocytes) and female (macrogametocytes), are ingested by an *Anopheles* mosquito during a blood meal. The parasites’ multiplication in the mosquito is known as the sporogonic cycle. While in the mosquito's stomach, the microgametes penetrate the macrogametes generating zygotes. The zygotes in turn become motile and elongated (ookinetes) which invade the midgut wall of the mosquito where they develop into oocysts. The oocysts grow, rupture, and release sporozoites, which make their way to the mosquito's salivary glands. Inoculation of the sporozoites into a new human host perpetuates the malaria life cycle. (Courtesy: CDC website)
1.4 Factors Involved in Malaria Transmission and Malaria Control

Understanding the biology and behaviour of Anopheles mosquitoes can help understand how malaria is transmitted and can aid in designing appropriate control strategies. Factors that affect a mosquito's ability to transmit malaria include its innate susceptibility to Plasmodium, its host choice, and its longevity. Factors that should be taken into consideration when designing a control program include the susceptibility of malaria vectors to insecticides and the preferred feeding and resting location of adult mosquitoes.

1.4.1 Preferred Sources for Blood Meals

One important behavioural factor is the degree to which an Anopheles species prefers to feed on humans (anthropophily) or animals such as cattle (zoophily). Anthropophilic Anopheles are more likely to transmit the malaria parasites from one person to another. Most Anopheles mosquitoes are not exclusively anthropophilic or zoophilic. However, the primary malaria vectors in Africa, A. gambiae and A. funestus, are strongly anthropophilic and, consequently, are two of the most efficient malaria vectors in the world.

1.4.2 Life Span

Once ingested by a mosquito, malaria parasites must undergo development within the mosquito before they are infectious to humans. The time required for development in the mosquito (the extrinsic incubation period) ranges from 10 to 21 days, depending on the parasite species and the temperature. If a mosquito does not survive longer than the extrinsic incubation period, then she will not be able to transmit any malaria parasites.

It is not possible to measure directly the life span of mosquitoes in nature. But indirect estimates of daily survivorship have been made for several Anopheles species. Estimates of daily survivorship of A. gambiae in Tanzania ranged from 0.77 to 0.84 meaning that at the end of one day between 77% and 84% will have survived [9]. Assuming this is constant through the adult life of a mosquito, less than 10% of female A. gambiae would survive longer than a 14-day extrinsic incubation period.
Control measures that rely on insecticides (e.g., indoor residual spraying) may actually impact malaria transmission more through their effect on adult longevity than through their effect on the population of adult mosquitoes.

1.4.3 Patterns of Feeding and Resting

Most *Anopheles* mosquitoes are active at dusk or dawn or nocturnal (active at night). Some *Anopheles* mosquitoes feed indoors (endophagic) while others feed outdoors (exophagic). After blood feeding, some *Anopheles* mosquitoes prefer to rest indoors (endophilic) while others prefer to rest outdoors (exophilic). Biting by nocturnal, endophagic *Anopheles* mosquitoes can be markedly reduced through the use of insecticide-treated bed nets (ITNs) or through improved housing construction to prevent mosquito entry (e.g., window screens). Endophilic mosquitoes are readily controlled by indoor spraying of residual insecticides. In contrast, exophagic/exophilic vectors are best controlled through source reduction (destruction of the breeding sites).

1.4.4 Insecticide Resistance

Insecticide-based control measures (e.g., indoor spraying with insecticides, ITNs) are the major way to kill mosquitoes that bite indoors. However, after prolonged exposure to an insecticide over several generations, mosquitoes, like other insects, may develop resistance, a capacity to survive even after contact with an insecticide [3, 10]. Since mosquitoes can have many generations per year, high levels of resistance can arise very quickly [10]. Resistance of mosquitoes to some insecticides has been documented just within a few years after the insecticides were introduced [11]. There are over 125 mosquito species with documented resistance to one or more insecticides [12]. The development of resistance to insecticides used for indoor residual spraying was a major impediment during the Global Malaria Eradication Campaign. Judicious use of insecticides for mosquito control can limit the development and spread of resistance. However, use of insecticides in agriculture has often been implicated as contributing to resistance in mosquito populations [3]. It is possible to detect developing resistance in mosquitoes and control programs are well advised to conduct surveillance for this potential problem [13].
1.5 Control measures (Genetic manipulation)

Absence of an efficient vaccine [14], evolution of drug resistance in the parasites [15-17], and insecticide resistance in the mosquitoes [3, 12] accentuate the need of an effective malaria control strategy. Human immunization against parasite proteins through transmission blocking vaccines (TBVs) is one such strategy. Bacterial and fungal-based mosquito control methods are other alternatives but these suffer from major difficulties in practical application [18, 19]. Transgenic mosquitoes could provide another control method [20], but successful application in field will require designing of appropriate vector-parasite study model [21]. Ito et al.[20], showed transgenic expression of an antiparasitic peptide SM1 in mosquitoes leading to impairment in \textit{P. berghei} development. However, the peptide failed to show such activity against \textit{P. falciparum} [21], the human malaria parasite. Study of naturally occurring \textit{P. falciparum}-resistant \textit{A. gambiae} mosquitoes revealed a \textit{Plasmodium}-responsive gene, \textit{Anopheles Plasmodium}-responsive leucine-rich repeat 1 (APL1) [22], which could form a potent target for the transgenic approach against \textit{P. falciparum}. Many other antiparasitic and/or immunologically active genes like SRPN6 [23] from \textit{A. gambiae} and \textit{A. stephensi}, TEP1 [24] and leucine-rich repeat protein (LRIM1) [25] from \textit{A. gambiae} have also been identified recently. Moreover, availability of \textit{A. gambiae} genome sequence [26] has improved the chances of discovery of more such potential genes in this insect.
1.6 Global statistics

In 2010, there were 106 malaria-endemic countries and approximately 3.3 billion people at risk for infection, worldwide (Figure 1.2). The same year, there were an estimated 216 million episodes of malaria and some 655,000 malaria deaths – of which 91 percent were in the African region. Approximately 86 percent of malaria deaths globally were of children under 5 years of age.

Figure 1.2: This map shows an approximation of the parts of the world where malaria transmission occurs. Image courtesy http://www.cdc.gov/malaria/about/distribution.html
Figure 1.3: Shows global distribution of malaria vectors. Image courtesy http://www.cdc.gov/
1.6.1 Scenario in Indian subcontinent

*Anopheles stephensi* is a major malaria vector in India along with other anopheline mosquitoes like *A. culicifacies, A. fluviatilis A. minimus, A. dirus* and *A. annularis*. Factors that may cause outbreaks include an increase in vector breeding sites, migration of infected people into a vector-rich area populated with susceptible individuals, arrival of new efficient vectors, breakdown of vector control measures, resistance of the parasites to treatment and resistance of the vectors to insecticides. World malaria report 2012 published by World Health Organization shows cases of malaria in India (Figure 1.4) (Table1).

![Figure 1.4: Shows distribution of confirmed malaria cases per 1000 population in India alone. Image courtesy: WORLD MALERIA REPORT-2012](image-url)
Table 1.1: Shows the epidemiological profile of malaria cases in India with High and low transmission

<table>
<thead>
<tr>
<th>Epidemiological profile*</th>
<th>2010</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>High transmission (≥ 1 case per 1000 population)</td>
<td>273 000 000</td>
<td>22</td>
</tr>
<tr>
<td>Low transmission (0-1 case per 1000 population)</td>
<td>832 000 000</td>
<td>67</td>
</tr>
<tr>
<td>Malaria-free (0 cases)</td>
<td>137 000 000</td>
<td>11</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>1 242 000 000</td>
<td></td>
</tr>
</tbody>
</table>

**Parasites and vectors**

- Major plasmodium species: *P. falciparum* (51%) *P. vivax* (49%)
- Major anophiles species: *A. stephensi, cilicifacies, fluvatilis, minimus, dirus, annularis*

*Data was taken from WORLD MALERIA REPORT-2012

1.7 Expressed Sequence tags

Expressed sequence tags (ESTs) are fragments of mRNA sequences derived through single sequencing reactions performed on randomly selected clones from cDNA libraries [27]. ESTs provide direct evidence for all the sampled transcripts from a tissue or cells [28]. They are one of the most important resources for exploring the whole transcriptome of cell or tissue. They are usually 400-800 bp long and can be sequenced from 5’ as well as 3’ end of the gene. ESTs are primarily used for gene discovery and therefore there has been an exponential growth in EST data in NCBI database.

1.7.1 Generation of EST sequences

Messenger RNA (mRNA) sequences from a cell or tissue represent the total number of expressed genes. As mRNAs cannot be cloned directly they are converted into a double stranded cDNA sequences using reverse transcriptase enzyme. These cDNAs are cloned to make libraries which represent the subset of expressed genes in a cell or tissue. Subsequently these clones are then sequenced randomly to obtain 5’ or 3’ ESTs. The resultant ESTs can be full length or partial depending upon the length of the sequenced cDNA. EST sequences obtained are redundant as many mRNAs can be produced for a single gene. Barbazuk *et al.* [29] has provide a detailed instruction on cDNA library construction and normalization.
1.7.2 Errors associated with EST generation

EST sequence is a very short copy of mRNA and contains errors especially at the end regions (Figure 1.5)

![Figure 1.5: A characteristic of EST sequence. Theoretically, it contains vector sequence at both the ends and possibly may contain repeat or low complexity regions like polyA. Image courtesy Nagaraj et al., 2007 [30].]

As ESTs are sequenced only once they are prone to sequencing errors. Generally the quality of the sequence is bad towards both the ends. Usually towards the start 50-100 bp have poor quality then it improves gradually and again goes down at the end (Figure 1.6) [30].

![Figure 1.6: Phred quality score are plotted as a function of sequence length for a hypothetical EST sequence. Image courtesy Nagaraj et al., 2007 [30].]
1.7.3 EST pre-processing

EST pre-processing improves the quality of data and increases the efficacy of further analyses. Often a part of the vector is also sequenced along with EST (Figure 1.5). These vector sequences must be trimmed off before EST clustering. Repetitive elements, such as LINEs (Long interspersed elements), SINEs (Short interspersed elements), LTRs (Long terminal repeat) and SSRs (Short simple repeats), can lead to erroneous assembly of sequences. Therefore, they should be removed before assembling EST sequences. Many EST sequences may also contain polyA tails which are not part of the genomic sequence. These sequences should also be trimmed of before clustering and assembly process [30].
**Figure 1.7:** Generic steps involved in EST analysis. 1. Raw EST sequences are checked for vector contamination, low complexity and repeat regions, which are excised or masked. Low quality, singleton and very short sequences are also removed. 2. ESTs are then clustered and assembled to generate consensus sequences (‘putative transcripts’). 3. DNA database similarity searches are carried out to assign, identify homologues and sign possible function. 4. Putative peptides are obtained by conceptual translation of consensus sequences. 5. Protein database similarity searches are performed to assign putative function(s). The analysis is extended to functional annotation followed by visualization and interpretation of results. The steps enclosed by the grey box alone are implemented in the currently available pipelines. Image courtesy [30].
1.7.4 EST clustering and assembly

EST clustering and assembly is done to reduce the overall redundancy present in the data. ESTs belonging to single gene or transcript are assembled into a unique cluster. An EST cluster contains ESTs of different length which are then can be assembled to get a contiguous sequence (contig). A simple way of EST clustering is by measuring pair-wise sequence similarity between them. These similarity values are then used as seed to decide which EST to cluster together [31]. Phrap [32] and CAP3 [33] are the most extensively used programs for sequence assembly and clustering.

1.7.5 Functional annotation of ESTs

Once the high quality EST sequence is obtained, it can be translated to get a putative polypeptide by using tools like BLASTX [34] and ESTscan [35]. The functional annotation of these putative or hypothetical polypeptides can be done using BLAST2GO [36] or InterProScan [37]

1.7.6 Application of ESTs

From last several years ESTs are used successfully for the gene discovery process. ESTs were first used to construct maps of human genome [38]. 3’ Untranslated regions of ESTs were used as STS (sequence tagged sites) for rapid chromosome assignment. The gene index of the human genome was also assessed using EST data. EST databases are also used for gene structure prediction [39, 40], to predict alternative splicing [41-44], to characterize SNPs[45-47] and also to find out tissue or disease specific expression [48]. ESTs have also been used to explore the salivary gland transcriptome of Anopheles stephensi with and without Plasmodium infection [49].