2.1. History of Malaria

Malaria is a potentially life threatening parasitic disease caused by parasites of the genus *Plasmodium* transmitted by the female *Anopheles* mosquitoes and is a major public health problem to the humanity. Malaria is an ancient disease and symptomatic description of the disease was evident in a Chinese document named *Nei Ching* (The Canon of Medicine) from China in 2700 BC. Documentation on malaria was also evident from the clay tablets from Mesopotamia in 2000 BC and Egyptian papyri in 1570 BC (Cox, 2010). By the 4th century BC Romans and Greeks recognised malaria with the symptoms of intermittent fevers, with high incidence during the rainy seasons. This disease was also evidenced by the Hindu texts as far back as the 6th century BC. In the *Susruta Samhita*, a Sanskrit medical treatise, the symptoms of malarial fever were described and attributed to the bites of certain insects.

The early Greeks, including Homer in about 850 BC, Empedocles of Agrigentum in about 550 BC and Hippocrates in about 400 BC, were well aware of the characteristic poor health, malarial fevers and enlarged spleens seen in people living in marshy places. For over 2500 years the idea that malaria fevers were caused by miasmas rising from swamps persisted and it is widely held that the word malaria comes from the Italian *mal'aria* meaning bad air. (Cox, 2010)

2.2. Malaria: Global Scenario

According to the World Malaria Report, 2010 (WHO, 2010a), the disease is present in tropical countries with more than 90% occurrence in Sub-Saharan Africans (Fig. 2.1). During the past 100 years, the area of malaria risk has reduced from around 50% to 25% of the earth’s land surface, (Hay et al., 2004). In 2010, number of countries endemic to malaria risk fell from 140 to 106, 7 countries were in the state of prevention of reintroduction phase and 2 countries (Morocco and Turkmenistan) certified as free of malaria. However, because of demographic changes, the number of people exposed to malaria has increased substantially over the same time and currently 3.3 billion people (half of the world population) are at risk of malaria globally. In 2009, an estimated 225 million cases of malaria were reported globally with 7,81,000 deaths, down from the estimated 244 million cases and 9,27,000 malaria related deaths in 2005 (WHO, 2010a). Among all malaria endemic regions, the African region harbour the highest percentage for both malaria cases (78%) and deaths (91%). More than 85% of the
mortality and morbidity caused by malaria is attributed to the children under five years of age. (WHO, 2010a). The global burden of disease analysis uses a summary measure of disease burden, disability adjusted life year (DALY), which combines information on mortality and non-fatal health outcomes into a single time based matrix. Malaria was ranked as the fourth leading cause of burden of disease in low-income countries in 2004 estimates, (WHO, 2008a) but this estimate varies widely from country to country. (Table 2.1)

Table 2.1: Leading cause of burden of disease (DALYs) in low income countries, 2004

<table>
<thead>
<tr>
<th>Rank</th>
<th>Disease or injury</th>
<th>DALYs (Millions)</th>
<th>% of total DALYs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Lower respiratory infections</td>
<td>76.9</td>
<td>9.3</td>
</tr>
<tr>
<td>2</td>
<td>Diarrhoeal diseases</td>
<td>59.2</td>
<td>7.2</td>
</tr>
<tr>
<td>3</td>
<td>HIV/AIDS</td>
<td>42.9</td>
<td>5.2</td>
</tr>
<tr>
<td>4</td>
<td>Malaria</td>
<td><strong>32.8</strong></td>
<td><strong>4</strong></td>
</tr>
<tr>
<td>5</td>
<td>Prematurity and low birth weight</td>
<td>32.1</td>
<td>3.9</td>
</tr>
<tr>
<td>6</td>
<td>Neonatal infections and other</td>
<td>31.4</td>
<td>3.8</td>
</tr>
<tr>
<td>7</td>
<td>Birth asphyxia and birth trauma</td>
<td>29.8</td>
<td>3.6</td>
</tr>
<tr>
<td>8</td>
<td>Unipolar depressive disorders</td>
<td>26.5</td>
<td>3.2</td>
</tr>
<tr>
<td>9</td>
<td>Ischaemic heart disease</td>
<td>26</td>
<td>3.1</td>
</tr>
<tr>
<td>10</td>
<td>Tuberculosis</td>
<td>22.4</td>
<td>2.7</td>
</tr>
</tbody>
</table>

2.3. Malaria burden in South-East Asia Region

WHO South-East Asia region (SEARO) includes 11 countries including India, of which 10 countries (except Maldives) are malaria endemic. Approximately 77% population in the region is under constant threat of malaria, with 20% at high risk. With an estimated 2.4 million confirmed malaria cases and 3320 deaths in 2009, this regions is regarded as the second highest malaria affected region by WHO. A whopping 97% of the reported cases in the region are contributed by 3 countries viz., India (65%), Myanmar (20%) and Indonesia (12%), most of which are due to *P. falciparum*. Five countries of this region viz., Bhutan, the Democratic People's Republic of Korea, Nepal, Sri Lanka and Thailand witnessed almost 50% reduction of the reported malaria cases during 2000-2009 of which DPR Korea and Sri Lanka are, in the pre-elimination stage. (WHO, 2010a)
Figure 2.1: World-wide distribution of malaria (Source: www.cdc.gov)
2.4. Malaria burden in India

Although the South-East Asian region is contributing only 15% and 6% of global malaria cases and deaths respectively, India is playing a major role in South-East Asia by contributing 65% of malaria cases to the region making it a major public health problem in the country. About 95% Indian population resides in malaria endemic areas and 80% of malaria reported in the country is confined to areas consisting 20% of population residing in tribal, hilly, difficult and inaccessible areas. (NVBDCP, 2010). In 2010, an estimated 1.49 million malaria cases and 767 malaria related deaths were recorded in India (NVBDCP, 2010). Up to late nineties the average annual morbidity due to malaria was around 2 million which, however, declined gradually from 2002 onward along with reported death cases and currently is maintained at around 1.5 million cases (Table 2.2).

Due to the diverse topographical, ecological and climatic conditions of India, there is a great heterogeneity in the transmission of malaria and hence the cases. The climate varies from tropical monsoon in the South and North-east of the country to temperate in the North. According to Bansal et al., (1988), the climates of India can be differentiated into 6 major climatic groups, viz., hot and dry, warm and humid, moderate, cold and cloudy, cold and sunny, and composite. According to a recent code of Bureau of Indian Standards, the country may be divided into five major climatic zones, hot and dry, warm and humid, moderate, cold and composite (BIS, 2005). All these climatic zones have different vegetation type, temperature and rainfall patterns over the country. Such climatic diversity influences the distribution of vectors and species of malaria parasite; as a result, malaria in India different paradigms, including forest/tribal malaria, urban/slum malaria, industrial malaria, and plains malaria. (Das et al., 2012).

Historically, the highest incidence of malaria in India was reported in the 1950s, with an estimated 75 million cases and 0.8 million deaths per year (Dash et al., 2008). The launch of the National Malaria Control Program (NMCP) in 1953 resulted in a significant decline in the number of reported cases to <50,000 and no reported mortality, by 1961. Despite its near elimination in the mid-1960s, malaria resurged to approximately 6.45 million cases in 1976 (Sharma, 1996). Since then, confirmed cases have gradually decreased to 1.6 million cases and ~ 1100 deaths in 2009. Recently, it
has been suggested that the malaria incidence in India is between 9 and 50 times greater than reported (reviewed in Hay et al., 2010b), with a 13-fold under-estimation of malaria-related mortality (Dhingra et al., 2010).

Table 2.2: Malaria Epidemiological Situation in India (1995 – 2010)
(Source: www.nvbdcp.gov.in)

<table>
<thead>
<tr>
<th>Year</th>
<th>Population (in '000)</th>
<th>Total Malaria Cases (million)</th>
<th>P. falciparum cases (million)</th>
<th>Pf %</th>
<th>API</th>
<th>Deaths due to malaria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995</td>
<td>888143</td>
<td>2.93</td>
<td>1.14</td>
<td>38.84</td>
<td>3.29</td>
<td>1151</td>
</tr>
<tr>
<td>1996</td>
<td>872906</td>
<td>3.04</td>
<td>1.18</td>
<td>38.86</td>
<td>3.48</td>
<td>1010</td>
</tr>
<tr>
<td>1997</td>
<td>884719</td>
<td>2.66</td>
<td>1.01</td>
<td>37.87</td>
<td>3.01</td>
<td>879</td>
</tr>
<tr>
<td>1998</td>
<td>910884</td>
<td>2.22</td>
<td>1.03</td>
<td>46.35</td>
<td>2.44</td>
<td>664</td>
</tr>
<tr>
<td>1999</td>
<td>948656</td>
<td>2.28</td>
<td>1.14</td>
<td>49.96</td>
<td>2.41</td>
<td>1048</td>
</tr>
<tr>
<td>2000</td>
<td>970275</td>
<td>2.03</td>
<td>1.05</td>
<td>51.54</td>
<td>2.09</td>
<td>932</td>
</tr>
<tr>
<td>2001</td>
<td>984579</td>
<td>2.09</td>
<td>1.01</td>
<td>48.20</td>
<td>2.12</td>
<td>1005</td>
</tr>
<tr>
<td>2002</td>
<td>1013942</td>
<td>1.84</td>
<td>0.90</td>
<td>48.74</td>
<td>1.82</td>
<td>973</td>
</tr>
<tr>
<td>2003</td>
<td>1027157</td>
<td>1.87</td>
<td>0.86</td>
<td>45.85</td>
<td>1.82</td>
<td>1006</td>
</tr>
<tr>
<td>2004</td>
<td>1040939</td>
<td>1.92</td>
<td>0.89</td>
<td>46.47</td>
<td>1.84</td>
<td>949</td>
</tr>
<tr>
<td>2005</td>
<td>1082882</td>
<td>1.82</td>
<td>0.81</td>
<td>44.32</td>
<td>1.68</td>
<td>963</td>
</tr>
<tr>
<td>2006</td>
<td>1072713</td>
<td>1.79</td>
<td>0.84</td>
<td>47.08</td>
<td>1.66</td>
<td>1707</td>
</tr>
<tr>
<td>2007</td>
<td>1087582</td>
<td>1.51</td>
<td>0.74</td>
<td>49.11</td>
<td>1.39</td>
<td>1311</td>
</tr>
<tr>
<td>2008</td>
<td>1119624</td>
<td>1.53</td>
<td>0.77</td>
<td>50.81</td>
<td>1.36</td>
<td>1055</td>
</tr>
<tr>
<td>2009</td>
<td>1150113</td>
<td>1.56</td>
<td>0.84</td>
<td>53.72</td>
<td>1.36</td>
<td>1144</td>
</tr>
</tbody>
</table>
| 2010 | 1151788              | 1.49                          | 0.77                         | 52.12| 1.3 | 767                  

2.5. Malaria situation in North-east India

The North-eastern region of India (NE India), comprising of 8 states viz. Assam, Arunachal Pradesh, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim and Tripura, is situated between N 22°04′ to N 29°31′ and E 89°48′ to E 97°25′ covering an area of 2,62,179 KM² and a population of 45.6 million (2011 census report). The region is sharing international boundaries with China and Bhutan (in the North), Myanmar (in the East) and with Bangladesh in the south. The NE India region accounts for 7.8% of Indian land mass and 3.8% of the Indian population.

In NE India, 25.2% of the population belongs to tribal representing many ethnic and cultural groups with different belief, culture, tradition, social need and life styles which has a direct role in malaria transmission over the region (Dutta et al., 1999; Singh, 1994).
Topographically this region can be broadly divided into 3 geographical entities, viz. the Himalayas, the Indo-Brahmaputra basin along with the Barak, Manipur and the Tripura valleys and the hill ranges bordering Myanmar and Bangladesh (North-eastern Plateau). The northern Himalayan ranges are low and usually do not exceed 600 m in height. The areas above 1500 m bordering China are free from malaria but the foothills and adjacent hills are highly malarious. The Indo-Brahmaputra basin (alluvial plains) is sandwiched between the Himalayas in the North and the hill ranges extending from Myanmar in the east to the Garo hills in the South. This region, including the Barak, Manipur and the Tripura valleys, has variable endemicity of malaria. The hill ranges along the Indo-Myanmar border, lying in the north south direction, extends up to Indo-Brahmaputra basin are hyper endemic for malaria. The hill ranges bordering Bangladesh in the South (lies east-west) which include the Naga, Mizo, North Cachar, Cachar, Karbi, Jaintia, Khasi and Garo hills have large areas with hyper- to holo- endemic malaria (Mohapatra et al., 1998). About 45% land mass of NE India is covered by tropical evergreen forests which, along with the forest fringe areas are highly endemic to malaria (Mohapatra et al., 2001).

The climate of NE India is conducive for malaria transmission. The South west monsoon brings heavy rainfall during June to September, however, the rain occurs more or less throughout the year with average 200 – 250 rainy days. Except for the north-eastern plateau, the NE region experiences more than 2000mm rainfall per year with a maximum of 11,500 mm in Cherrapunji - Mowsynram region (Meghalaya) to a minimum of 1000 mm in the Manipur valley (Mohapatra et al., 1998).

The Region can be divided into three isothermic zones - in the greater and lesser Himalayas, the temperature ranges between -20°C in winter to 15°C to 20°C in summer; in the Northern part, except for greater and lesser Himalayas, along the sub Himalayan areas temperature lies within 4°C in winter to 21°C in summer; while the Indo-Brahmaputra basin experiences 18-20°C in winter and 30-35°C in summer. The relative humidity of the Region is more than 60% throughout the year.

Due to the diverse topographical, climatic and socio-economic status, the Region has varied malaria endemicity. The NE region falls within the 20 % “High risk population” group of India for malaria accounting for 80% cases in India. High malaria burden in the NE India can be guessed by the fact that nearly 4 per cent of country's
population living in this part contributes 10 to 12% of the malaria cases every year in national scenario (Patra and Dev, 2004). The malaria situation in NE India in last 5 decades is shown in Fig. 2.2. Of the eight north-eastern states of India, Assam, Arunachal Pradesh and Meghalaya contributes to the maximum cases of malaria, whereas Arunachal Pradesh records majority of malaria deaths.

According to NVBDCP, the total malaria cases reported from NE India were 1,73,467 with 290 confirmed deaths due to malaria in the year 2011 (NVBDCP, 2011).

2.6. Human Malaria parasites

The causative organism of malaria is a protozoan parasite of genus *Plasmodium* belonging to the phylum *Apicomplexa* (Greenwood et al., 2005). There are more than 200 identified species of the genus *Plasmodium* that are parasitic to reptiles, birds and mammals (Rich and Ayala, 2006). Globally, four *Plasmodium* species have been well known to cause human malaria, namely, *P. falciparum*, *P. vivax*, *P. ovale*, and *P. malariae*. A fifth species, *P. knowlesi*, has recently been documented recently to cause human infections in many countries of Southeast Asia (Dhanescar et al., 2009; Singh et al., 2004). Of the five *Plasmodium* species known to infect humans, *P. falciparum* causes by far the greatest morbidity and mortality (Snow et al., 2005). All malaria parasites infecting humans, except *P. knowlesi*, were assumed to arose from the great apes (Chimpanzees, western Gorillas and bonobos) (Duval et al., 2010; Escalante and Ayala, 1994; Escalante et al., 1995; Jeffares et al., 2007; Krif et al., 2010; Liu et al., 2010; Prugnolle et al., 2010; Rich et al., 1998). In case of *P. knowlesi*, it was found that the species was derived from an ancestral parasite population that existed prior to human settlement in South-east Asia, and underwent significant population expansion approximately 30,000–40,000 years ago associating with wild monkeys as a reservoir (Lee et al., 2011).

In North-east India, though *P. falciparum* and *P. vivax* are the two common species of malaria parasite with the predominance (60 - 80%) of the former (Mohapatra et al., 1998), presence of all four malaria parasites has been recorded. One case of *P. ovale* was reported in 2003 from Titabor, Jorhat, Assam (Prakash et al, 2003) and a focus of *P. malariae* from Arunachal Pradesh was reported by Mohapatra et al., (2008).
Figure 2.2: Malaria epidemiological data of North-east India from 1961-2010
(Source http://nvbdcp.gov.in/)
2.7. Vectors of Human malaria

More than a Century ago, in 1897, Nobel laureate Sir Ronald Ross made the path breaking discovery that the malaria (avian malaria) parasites are transmitted by mosquitoes (Mitchell, 1907; Ross, 1897). Before this, in 1879, Patrick Manson showed that mosquitoes (later identified as belonging to genus Culex) transmitted human filarial worms (Capanna, 2006). Since then mosquitoes have been enjoying the unquestionable honour of having been the first insects to be associated with the transmission of a disease, thus regarded as a medically important taxa. In 1898, Grassi and his colleagues reported that the mosquito, Anopheles claviger, was the carrier of human malaria (Bastianelli, 1898) and ever since the epidemiology of malaria was known as a complex interplay among man, Plasmodia and Anopheline mosquitoes.

The genus Anopheles includes a total of 465 formally recognised species and more than 50 unnamed members of species complexes are recognised as distinct morphologically or biological species of the genus (Harbach, 2012). Of these, approximately 70 Anopheles species have been considered to be competent human malaria vectors (Service and Townson, 2002) of which 41 recognised Anopheles species including 15 Anopheles species complexes are now-a-days regarded as dominant malaria vector world-wide (Fig. 2.3) (Hay et al., 2010a). In South-east Asia, 28 Anopheles species comprising 10 Anopheles species complexes are considered as malaria vector (Fig. 2.4) (Subbarao, 1998).

In India, a total of 58 species of Anopheles are found (Rao, 1984; Subbarao, 1998) of which 9 species have been incriminated as malaria vectors (Lal et al., 2000; Sharma, 1999). Six primary vectors include An. culicifacies s.l. (rural vector), An. fluviatilis s.l. (vector in the foothills), An. stephensi (urban vector), An. minimus s.l. (vector in the foothills and forest fringes), An. dirus s.l. (vector in the forests) and An. sundaicus s.l. (coastal vector in the Andaman and Nicobar islands) while An. annularis, An. phillipinensis/nivipes and An. subpictus are recognised as secondary vectors. Of these 9 vectors of malaria in India, except An. stephensi, all are species complexes.

The NE India, being a part of the "Indo-Burma hot-spot" (Myers et al., 2000), is represented by its rich fauna and flora (Chatterjee et al., 2006). Malhotra and Mahanta (1994) reported a total of 130 species of mosquitoes, including 37 Anophelines and 93 Culicines, in 12 genera from NE India. Another survey enumerated 191 mosquito
Figure 2.3: World-wide distribution of dominant malaria vectors (Source: www.cdc.gov)

Figure 2.4: Distribution of Malaria vectors in SE Asia
(Source: http://www.searo.who.int/en/Section10/Section21/Section340_4267.htm)
species (45 Anophelines, 146 culicines) in 14 genera from this region (Dutta et al., 2003) of which *Anopheles dirus, An. fluviatilis* and *An. minimus* complex mosquitoes are regarded as main malaria vectors in NE India (Prakash et al., 2006). Recently, Bhattacharyya et al., (2010) incriminated *An. nivipes* as a malaria vector in Nagaland.

### 2.7.1. *Anopheles minimus* complex

The Minimus complex includes three formally named species. *An. minimus* (earlier species A), *An. harrisoni* (earlier species C) and *An. yaeyamaensis* (earlier species E). Present day taxonomy of *An. minimus* complex has been dealt with and resolved by painstaking work of Chen et al., (2002), Harbach et al., (2006, 2007) and Somboon et al., (2010). *An. minimus* was considered to be the most important malaria vector in the entire sub-Himalayan belt. Due to the large scale application of DDT under NMEP (National Malaria Eradication Program), this species was thought to have been disappeared from this Region. However, it again re-emerged in the eighties and incriminated as a vector of malaria in Assam (Kareem et al., 1985 ; Prakash et al., 1996), Arunachal Pradesh (Dutta and Barua, 1987), Nagaland (Bhatnagar et al., 1982) and Mizoram (Das and Baruah, 1985). Presence of only 1 species of Minimus complex, i.e., *An. minimus* s.s. (species A) has been reported from NE India (Prakash et al., 2006). *An. minimus* is found throughout the year with higher densities between March-August. It is endophagic, endophillic and highly anthropophillic in nature. It bites on humans throughout the night with a peak occuring between 01:00 to 04:00AM. It breeds in slow-flowing streams with grassy banks with perceptible flow of water (Dev, 1996). It is considered susceptible to DDT (Prakash et al., 1998).

### 2.7.2. *Anopheles fluviatilis* complex

It is a species of hills and foothills but also found in the irrigated plains. Cytotaxonomic studies of fixed inversions in polytene chromosomes have identified three chromosomal forms within the Fluviatilis Complex, *An. fluviatilis* S, T, and U, informally recognized as sibling species (Subbarao et al., 1994). However, the taxonomic status of the Fluviatilis complex is still not clear and complicated by the observation of the presence of molecular variants within the complex, including species X in Orissa, India (Manonmani et al., 2003; Naddaf et al., 2002) and form V recorded only from Iran (Chen et al., 2006; Manguin et al., 2008). However, recently the occurrence of *An. fluviatilis* form V in Iran was doubted by Naddaf et al., (2010). The Minimus and Fluviatilis complexes are phylogenetically (Garros et al., 2005) as well as morphologically closely related and often misidentified (Singh et al., 2010). Based on morphological and molecular data of *An. fluviatilis* S and *An. harrisoni* (minimus C),
these two species were deemed to be conspecific (Chen et al., 2003, 2006; Garros et al., 2005). However, this conclusion was refuted by Singh et al., (2006) who found pairwise distances of 3.6% and 0.7% for the ITS2 and 28S-D2/D3 loci, respectively, between the species. In a thorough review, Chen et al., (2006), concluded that this complex consists of two species, species T (with intraspecific variations, including the putative species Y) and species U, and two forms, X (different from species S) and V. Singh et al., (2006) removed *An. harrisoni* (as *An. minimus* C) from synonymy with *An. fluviatilis* S, and reported that *An. fluviatilis* X is synonymous with the latter species. Therefore, according to Singh et al., (2006) the Fluviatilis complex includes species S, T, U, and form V. As currently interpreted, *An. fluviatilis* S is distinct from *An. harrisoni*, which does not occur in India.

Among the 3 species' of this complex *An. fluviatilis* S is more anthropophilic, endophagic and known to be a highly competent malaria vector (Nanda et al., 2005) than T and U (Subbarao et al., 1994). *Anopheles fluviatilis* complex is a clear water breeder found in small streams, irrigation channels with grassy margins. It is basically a post monsoon species prevalent between November and July transmitting malaria in winter season and susceptible to the commonly used insecticides, e.g., DDT (Kareem et al., 1985).

### 2.7.3. Anopheles dirus complex

*A. dirus* complex mosquitoes, encompassing at least 7 species, are regarded as very efficient and important malaria vectors associated with the sylvatic environment of the South-east Asian region including NE India (Obsomer et al., 2007; Oo et al., 2004). In India, the distribution of *An. dirus* complex mosquitoes encompasses tropical rain forest areas of NE India, Andaman-Nicobar islands and Western Ghats in south-western peninsular India (Bhat, 1998). The taxonomy of this mosquito complex was resolved recently and all the species now have morphological descriptions and formal Latin names (Sallum et al., 2005a), and their distributions in South-east Asia have been mapped (Baimai, 1987; Obsomer et al., 2007).

#### 2.7.3.1. Taxonomic status of Anopheles dirus complex

The Dirus complex mosquitoes belongs to the Leucosphyrus subgroup under Leucosphyrus Group of Neomyzomyia Series of subgenus Cellia of *Anopheles* genus (Diptera: Culicidae) (Harbach, 2004) and includes at least 7 isomorphic species (Sallum
et al., 2005a). This Leucosphyrus group as such is regarded medically important since six species belonging to this group, namely, viz., \textit{An. balabacensis} Baisas, \textit{An. latens} Sallum & Peyton, \textit{An. leucosphyrus} Döntz, \textit{An. baimaii} (species D of Dirus complex), \textit{An. dirus} (dirus A of Dirus complex) Peyton & Harrison and \textit{An. sulawesi} Koesoemawinanggoen are considered very efficient and highly competent vectors of human malaria in the South-east Asia. Recently, two members of the Leucosphyrus group, viz, \textit{An. latens} (= \textit{leucosphyrus}) and \textit{An. cracens} (species B of dirus complex) have been incriminated as vectors for transmitting \textit{P. knowlesi} in Malaysia (Vythilingam et al., 2006; 2008). Several species of the Leucosphyrus group have also been reported vectoring simian malaria (Eyles et al., 1963; Fooden, 1994; Tsukamoto et al., 1978; Warren and Wharton 1963; Wharton and Eyles, 1961; Wharton et al., 1962).

The Leucosphyrus group includes 3 subgroups, namely, Leucosphyrus, Hackeri and Riparis subgroup (Sallam et al., 2005a). Besides Dirus complex mosquitoes, the Leucosphyrus subgroup includes the Leucosphyrus complex, \textit{An. baisasi} Colless, and the geographical form named Con Son. The Dirus Complex comprises seven species (Fig. 2.5). Peyton & Harrison (1979) relied heavily on morphological characters of the adult, pupal, and larval stages to describe \textit{An. dirus}. Their elevation of \textit{An. takasagoensis} to species status (Peyton & Harrison, 1980) and separation of this taxon from both \textit{An. dirus} and \textit{An. balabacensis} relied heavily on cross-mating and cytogenetic, as well as morphological evidence. Thereafter, the discovery of most members of the Dirus Complex was achieved primarily with crossing studies, cytogenetic studies, polytene chromosome banding patterns (Baimai et al., 1987; Baimai et al., 1988a, 1988b, 1988c; Poopittayasataporn & Baimai, 1995; Sawadipanich et al., 1990), and allozyme data (Green et al., 1992). The species status of species A, B, C and D was reinforced by population genetic evidence using allozymes (Green et al., 1992). Cytogenetic, allozyme and crossing studies indicate that \textit{An. dirus} A and C are very closely related (Baimai et al., 1987; Baimai et al., 1988a; Green et al., 1992). More recently, polymerase chain reaction (PCR) based methods were developed in Thailand using species specific primers to differentiate \textit{An. dirus}, \textit{An. cracens} Sallum and Peyton, \textit{An. scanloni} Sallum and Peyton, and \textit{An. baimaii} Sallum and Peyton, of the Dirus Complex (Huong et al., 2001; Walton et al., 1999). Similarly, a species diagnostic polymerase chain reaction assay based on differences in the ITS2 sequences of the
rDNA was developed to separate populations of *An. dirus* A and *An. dirus* D from China (Xu et al., 1998). Walton et al., (1999) showed that ITS2 sequence of Chinese ‘species D’ of Xu & Qu (1997) is distinct from Thailand specimens and thus it may represent an unrecognized species of the complex.

The Leucosphyrus complex includes *An. leucosphyrus* (Dn. le), *An. latens* (Sallum & Peyton), *An. introlatus* Colless, and *An. balabacensis* (Baisas). The Riparis subgroup consists of *An. riparis* King & Baisas, *An. cristatus* King & Baisas, *An. macarthuri* Colless, and the Negros Form. Sallum et al., (2005b) transferred *An. elegans* to the Dirus complex and renamed the earlier Elegans subgroup as the Hackeri subgroup to reflect the change. Currently, the Hackeri subgroup includes *An. hackeri* Edwards, *An. pujutensis* Colless, *An. mirans*, *An. sulawesi*, and *An. recens* Sallum & Peyton. The species previously called *An. elegans* from Sri Lanka (Mendis et al., 1983) and Southern India (Tewari et al., 1987) has now been renamed *Anopheles mirans* and is not a part of the Dirus complex (Sallum et al., 2005b). Therefore, according to the latest classification and formal naming by Sallum et al., (2005a), the Dirus complex includes the following 7 species, *An. dirus* Peyton & Harrison, *An. cracens* Sallum & Peyton, *An. scanloni* Sallum & Peyton, *An. baimaii* Sallum & Peyton, *An. elegans* (James), *An. takasagoensis* Morishita, *An. nemophilous* Peyton & Ramalingam (Fig. 2.5, Table 2.3).

<table>
<thead>
<tr>
<th>Kingdom</th>
<th>Animalia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phylum</td>
<td>Arthropoda</td>
</tr>
<tr>
<td>Class</td>
<td>Insecta</td>
</tr>
<tr>
<td>Order</td>
<td>Diptera</td>
</tr>
<tr>
<td>Family</td>
<td>Culicidae</td>
</tr>
<tr>
<td>Sub-Family</td>
<td>Anophelinae</td>
</tr>
<tr>
<td>Genus</td>
<td>Anopheles</td>
</tr>
<tr>
<td>Sub-Genus</td>
<td>Cellia</td>
</tr>
<tr>
<td>Species complex</td>
<td>Anopheles dirus s.l.</td>
</tr>
<tr>
<td>Species</td>
<td>Anopheles dirus (Species A)</td>
</tr>
<tr>
<td></td>
<td>Anopheles cracens (Species B)</td>
</tr>
<tr>
<td></td>
<td>Anopheles scanloni (Species C)</td>
</tr>
<tr>
<td></td>
<td>Anopheles baimaii (Species D)</td>
</tr>
<tr>
<td></td>
<td>Anopheles elegans (Species E)</td>
</tr>
<tr>
<td></td>
<td>Anopheles nemophillous (Species F)</td>
</tr>
<tr>
<td></td>
<td>Anopheles takasagoensis</td>
</tr>
</tbody>
</table>
Figure 2.5: Species in the *Anopheles dirus* complex.
(Adapted from Obsomer et al., 2007)
2.7.3.2. Geographical distribution of *Anopheles dirus* complex

The Dirus complex includes seven species that vary from highly competent malaria vectors to non-vectors of human malaria in tropical evergreen rainforests, cultivated forests, and forest fringes throughout Southeast Asia (Oo et al., 2004; Sallum et al., 2005a). *Anopheles dirus* (=*An. dirus* species A) has a wide distribution in eastern Asia—it is known to occur in Myanmar, Thailand, Cambodia, Laos, Vietnam, and Hainan Island (China) (Fig. 2.6). *Anopheles cracens* (=*An. dirus* B) is known from southern peninsular Thailand, peninsular Malaysia, and Sumatra (Indonesia) (Baimaii et al., 1988d). *Anopheles scanloni* (=*An. dirus* C) occurs in a relatively narrow area along the borders of southern Myanmar and western and southern Thailand, and appears to be intimately linked to limestone environment. In Thailand, *An. scanloni* populations are restricted to “islands” of limestone karst habitat that support high levels of population structure (O’Loughlin et al., 2007). *Anopheles baimaii* (= *An. dirus* D) occurs in areas from south western China (Yunnan Province), western Thailand, Myanmar, borders of Nepal (Darsie et al., 1990) and Bangladesh (Baimaii et al., 1988e) to NE India and the Andaman Islands of India (Sallum et al., 2005a). *Anopheles elegans* (=*An. dirus* E) is restricted to hilly forests of Western Ghats in South-western India (Sawadipanich et al., 1990; Tewari et al., 1987). *Anopheles nemophilous* (=*An. dirus* F) has a patchy distribution along the Thai–Malay peninsula and Thai border areas with Myanmar and Cambodia. Finally, *An. takasagoensis* is restricted to Taiwan (Peyton and Harrison, 1980). There seems to be no major geographic or topographic reasons that justify the current species distribution pattern of this mosquito complex within its distribution range. Therefore, speciation probably does not result from adaptation to a specific environment but the species might have been isolated for some time. Their distribution most probably reflects geo-morphological changes which have occurred in the past and are today not evident (Baimaii et al., 1988e).

2.7.3.3. Biology and behaviour of *Anopheles dirus* complex mosquitoes

Members of *An. dirus* complex contains both vector and non-vector taxa with differential biology and behaviour. *An. dirus* s.s. and *An. baimaii*, regarded as primary vectors of human malaria, are highly anthropophilic, exophagic as well as endophagic and exophilic (Dutta et al., 1996; Nandi et al., 2000; Prakash et al., 2005; Trung et al., 2005; Vy wholeg et al., 2003). *An. dirus* s.l. mostly prefers to bite humans even in the
Figure 2.6: Geographical distribution of *Anopheles dirus* complex mosquitoes
(Adapted from Manguin et al., 2008)
presence of numerous cattles (Dutta et al., 1996), less anthropophilicity however was also recorded sporadically (Toma et al., 2002; Tun-Lin et al., 1995). Although monkeys are considered as the natural hosts for An. dirus s.l. (Rosenberg and Maheshwary, 1982; Yang, 1983), more specimens were collected on human bait on ground as compared to the collections on monkey bait on canopy (Eyles et al., 1964). Studies have shown that biting rhythm of members of Dirus complex is species-specific, for example, An. dirus has a tendency to bite between 20:00 and 23:00 h and An. baimaii from 22:00 h to 02:00 h in Thailand (Baimaii et al., 1988d; Sallum et al., 2005a). However, in India biting of An. baimaii starts as early as 18:00 hrs and ~13% biting occurs during I quarter of night (Dutta et al., 1996; Prakash et al., 2005). An. scanloni is also anthropophilic and plays a focal role in malaria transmission of both P. falciparum and P. vivax in Thailand (Manguin et al., 2008; Peyton, 1989; Sallum et al., 2005a). This is an early evening biter with peak activity occuring between 18:00-19:00 h (Baimaii et al., 1988d). Evidence is lacking on involvement of An. cracens (restricted to the Thai-Malaysian peninsular) and An. elegans (only present in hill forests of south western India) in transmitting human malaria (Baimaii, 1989; Sallum et al., 2005a). However, recently, An. cracens has been found to transmit knowlesi malaria in Malaysia (Vythilingam et al., 2006; 2008).

An. dirus complex mosquitoes are primarily exophillic and exophagic mosquitoes (Oo et al., 2003; Prakash et al., 1997a; Rosenberg and Maheshwary, 1982) but enter open shelters to feed (Scanlon et al., 1967). Open houses in the close vicinity of jungle or inside jungle showed no significant biting differences between indoors and outdoors (Dutta et al., 1996; Gould et al., 1966; Tun-Lin et al., 1995; Vythilingam et al., 2003). However in some instances, even higher rate of indoor biting was also recorded (Dutta et al., 1992; Willkinson et al., 1978). Resting places are generally difficult to find and occurs mainly in holes, wells (Oo et al., 2003; Paing et al., 1989) and vegetation such as bushes, tree holes (Kyí, 1970), thick grass, under the surfaces of leaves (Paing et al., 1979), branches near the ground and under the foots of trees (Yang, 1983). Attempts to locate the daytime resting places of An. dirus have met with little success with the exception of a study in a broken evergreen forest in Dibrugarh district, Assam, India (Prakash et al., 1997b), where 45 man-hours of searching led to the detection of 20 An. dirus s.l. resting in moist dark corners of large tree trunks up to the height of 1.5 meters;
in 7 hours searching 2 individuals were found resting in creepers at a height of 1.8 meters. The closest resting individual was ~150 meters from the village. The day-resting habit of this mosquito complex can be separated from night-resting habits (Elias and Dewan, 1982). Females rest early in the evening in outdoor vegetation, fences and wood stacked around houses (Dutta et al., 1996) but after feeding, they fly almost immediately back to the forest (Gould et al., 1966) with none found even at dawn around dwellings (Willkinson et al., 1978). Eyles et al. (1964) caught more An. dirus s.l. in the canopy than on the ground and when Willkinson et al. (1978) released hundreds of females in predawn darkness, they flew immediately upwards into trees suggesting that they might rest in the canopy of trees.

Larvae of the Dirus complex inhabit both primary and secondary larval habitats with mostly small, shallow, usually temporary, mostly shaded bodies of fresh, stagnant (or very slowly flowing) water, such as pools, puddles, small pits (e.g. gem pits), animal footprints (e.g. elephant footprints), wheel ruts, hollow logs, streams and even wells located in primary, secondary evergreen or deciduous forests, bamboo forests and fruit or rubber plantations (Baimaii 1988d; Meek, 1995; Oo et al., 2002; Prakash et al., 1997b, 2002) with clear or turbid water (Dutta et al., 2010). Larval habitats with nitrogenous wastes, due to elephant and buffalo excreta or rotten leaves, appeared to be more productive for An. baimaii (Prakash et al., 2002). These species are most abundant during the rainy (monsoon) season due to the larval requirement and oviposition preference for small temporary pools (Baimaii 1988d; Dutta et al., 2010; Vythilingam et al., 2003). Larvae of An. dirus s.s. and An. baimaii were also found to show adaptations to temporary habitats in Bangladesh (Rosenberg, 1982) and in Thailand (Willkinson et al., 1978). Larvae of An. baimaii was also reported from wells close to houses in Mudon area of Myanmar throughout the year (Htay-Aung et al., 1999; Oo et al., 2002; Paing et al., 1989; Tun-Lin et al., 1987).

Different biotic and abiotic factors such as temperature, rainfall, season, topography, soil type, relative humidity, pH and salinity of breeding water, shade and land cover play an important role in the occurrence, distribution, seasonality, behaviour and vectorial status of An. dirus complex mosquitoes. These factors play with components of the mosquito life cycle on one hand and interact with different parasite cycle on the other (Obsomer et al., 2007). Favourable environmental conditions of
dense vegetation, humid soil, high relative humidity and shade, coupled with the
presence of permanent suitable larval habitats or primary sites and natural hosts such as
monkeys, appear to persist in forest or forest fringes and forested foothills and deep
inside the forests during the dry season makes An. dirus complex mosquitoes as a
predominant Anopheles species in the forest settings of South East Asia along with
north east India transmitting malaria over all this region (Manguin et al., 2008; Obsomer
et al., 2007; Sinka et al., 2011).

2.7.3.4. Status of Anopheles dirus s.l. as malaria vector in SE Asia including NE
India

The statement of Rosenberg et al., (1990) “The danger from An. dirus s.l. is not
only that it is very resistant to control within its habitat but that it is an extraordinarily
efficient vector, so long-lived and anthropophilic that only a small population is
necessary to maintain high malaria endemicity” gives a fairly good idea about the
vectorial status of members belonging to Dirus complex. Only during last 50 years, the
role of An. dirus complex in transmission of malaria has been assessed (Obsomer et al.,
2007). Due to its biological attributes and its close association with the forest eco-
systems, the An. dirus s.l. mosquitoes transmit malaria in the whole SE Asia region
along with NE India and thus, is regarded as a primary vector over the area.

Both An. dirus s.s. and An. baimaii are regarded as malaria vectors in Myanmar,
Thailand, China, and Indochina (Peyton and Harrison, 1979), including NE India (Dutta
et al., 1989a; Prakash et al., 2006). These species were incriminated as a vector in
Myanmar (Htay-Aung et al., 1999), Thailand (Rattanarithinkul et al., 1996; Rosenberg
et al., 1990), Lao PDR (Sidavong et al., 2004), Vietnam (Erhart et al., 2005; Sang et al.,
2008), Cambodia (Durnez et al., 2011), Nepal (Darsi et al., 1990), Bhutan (Yangzom et
al., 2012) and in NE India (Das and Baruah, 1985; Dutta et al., 1989a; 1989b; Prakash
et al., 2001).

Sporozoite rates of An. dirus s.l. were recorded to vary with season and location,
the highest rates recorded in October (7.8%) in Assam, India (Prakash et al., 2001) and
rates up to 14% in forested sites Myanmar (Kyi, 1970). Rosenberg et al., (1990) found
high variation between villages 800 m away from each other with a sporozoite rate three
to four times greater at the site of lower abundance. Sporozoites of P. vivax and P.
falciparum have been commonly detected in An. baimaii and An. dirus s.s. Baimai et
al., (1984) reported sporozoites in *An. scanloni* and *An. crascens* with slight differences between species in relation to the parasite. *An. dirus s.s.* developed *P. vivax* and *P. falciparum* oocysts more readily than *An. crascens* and *An. scanloni*. *An. elegans*, *An. nemophilous*, and *An. takasagoensis* probably only transmit simian malaria (Baimai et al., 1984). *An. crascens* has been implicated in the transmission of *knowlesi* malaria in Malaysia (Vythilingam et al., 2006).

Activities such as colonization of new land for agriculture, logging, mining, resettlement of populations or military and forest surveillance inside the forest expose peoples to the high transmission risks worsening the malaria epidemiology situation worsen in the whole SE Asian region. Rosenberg, 1982 noted that the invasion of the jungle by human settlers in Bangladesh probably increased the density of *An. dirus s.l.* mosquitoes by providing hosts and the small transitory pools that are preferred for oviposition. In north east India, the Indian military and para-military forces are being deployed along the strategically important International borders of Bangladesh, Myanmar, China and Bhutan. Most of these border areas are forested and pockets of malaria, where the movement in to the infested area is a regular phenomenon which makes them vulnerable to malaria infection (Dhiman et al., 2010; Patra and Dev, 2004).

Literature search revealed that chloroquine resistant *P. falciparum* have been commonly associated with *An. dirus s.l.* (Das and Baruah, 1985; Dutta et al., 1989b; Gould et al., 1966; Yang, 1983).

*An. dirus s.l.* was reported to be susceptible to DDT (Mittal et al., 2004; Paing et al., 1989; Prakash et al., 1998; Van Bortal et al., 2008; Wilkinson et al., 1978) although possible resistance to DDT and type II pyrethroids alpha-cypermethrin and lambda-cyhalothrin was also found from Northern Thailand and Central Vietnam (Van Bortal et al., 2008). However, due to exophilic behaviour and excito-repellency activity of the insecticide, females of *An. dirus s.l.* avoid insecticide treated walls (Ismail et al., 1975; Wilkinson et al., 1978) or even avoid the sprayed huts by biting more outdoors after residual spraying (Rosenberg and Maheswary, 1982; Suwonkerd et al., 1990) and high sporozoite rates may persist after application of DDT (Yang, 1983).

### 2.8. Malaria control strategies

Malaria is too complex a disease to be addressed in a single approach and any attempt to do so is fraught with danger. The World Health Organisation (WHO)
identified malaria as a priority health issue in the world and attempted to address the deteriorating malaria situation by holding a ministerial level meeting in 1992 in Amsterdam to develop strategies (Global Malaria Control Strategy) to control malaria (WHO, 1993) which was further extended to Roll Back Malaria (RBM) initiative launched by Director General of WHO in 1998 with a goal of halving malaria deaths by 2010 and halving again by 2015 (Nabarro and Tayler, 1998; Narsimhan and Attaran, 2003). In this new initiative the WHO, United Nations Children’s Fund (UNICEF), United Nations Development Programme (UNDP), World Bank (WB) and various collaborators join forces to fight malaria. The initiative identified four basic technical elements to achieve the objective: (1) disease management through early diagnosis and prompt treatment, (2) disease prevention through selective and sustainable preventive measures, (3) early detection and containment of malaria epidemics, (4) strengthening of local capacities in basic and applied research to permit and promote the regular assessment of malaria situation (WHO, 2004a). However, it is important to customize these strategies to the prevailing ecological and epidemiological viz, hypoendemic, mesoendemic, hyperendemic and holoendemic malaria conditions of an area considering the immune status of the population inhabited (Shiff, 2002).

**2.9. Vector control tools**

Vector control has been historically successful in eradication of malaria in various parts of the world (Harrison, 1978). Therefore, the Global Malaria Control Strategy prioritized the need of selective and sustainable preventive measures for reducing malaria transmission (WHO, 1993). Rozendaal, (1997) described various methods on the use of insecticides, insecticide-treated materials, biological control agents, insect growth regulators (IGRs), environmental management, and personal protection methods used in mosquito vector control.

Adult mosquito control is the most important part of any vector borne disease control program which is mainly accomplished by application of chemical insecticides such as Dichlorodiphenyltrichloroethane (DDT) against adult stages throughout the world after the discovery of its insecticidal potential in twentieth century (Breman, 2001). Various issues related with environment and human health associated with application of DDT, opened up a debate between use or ban DDT in pest control (Dash et al., 2007). Based on the increasing scientific evidences, WHO in September 2006
gave a clean chit to use of DDT to combat malaria in Africa and other areas where the vectors are still susceptible to DDT (WHO, 2006a). However, the debate on the use of DDT is still continuing until a more effective, affordable, and safe alternative tool is made available.

Stable formulations of insecticides such as DDT, malathion, deltamethrin, propoxure, fenitrothion, lambda-cyhalothrin etc are applied to the interior sprayable surfaces of houses, commonly known as Indoor Residual Spray (IRS), to kill adult mosquitoes (WHO, 2006b). Apart from IRS, WHO also recommended insecticide products (mainly pyrethroids) for the treatment of mosquito nets typically known as Insecticide Treated Net (ITN) for malaria vector control which can markedly reduce the endophagic nocturnal biting Anopheles by its deterrence, excito-repellence activity or even kill mosquito with its residual insecticidal activity. Likewise the LLINs (long Lasting Insecticidal Nets) such as Olyset® (full recommendation by WHO), Perma-Net® (interim recommendation by WHO) were also used as vector control methods due to its long lasting nature (3–4 years for polyester nets and 4–5 years for polyethylene nets).

Use of chemical larvicides to eliminate or reduce the vector populations by killing the larvae is another vector control tool. Chemicals like petroleum oils (Gratz and Pal, 1988), Paris Green (Copper Acetoarsenite) (Rozendaal, 1997), several organophosphate larvicidal formulations viz, temephos and fenthion (Sharma et al., 1996) are used as chemical larvicides in malaria vector control programs. However, due to the non-target toxicity of Paris Green and fenthion as well as synthetic pyrethroids, these are not used now-a-days in larval control (WHO, 2006c). Two bacterial species, *Bacillus thuriengiensis israelensis* (*Bti*) and *Bacillus sphaericus* (*Bs*) have been widely demonstrated to be effective larvicides against mosquitoes as the protoxins in parasporal crystals, and the spore coat of these bacteria, function as stomach poisons in the mosquito larval midgut. Due to the effectiveness at relatively low doses, safety to humans and non-target organisms, low cost of production and lower risk of resistance development in mosquitoes (Yap, 1985), several formulations of *Bti* are used against *An. gambiae, An. albimanus, An. sinensis, An. stephensi, An. maculatus, An. sundaicus* and *An. maculipennis* (Becker and Margalit 1993; Das and Amalraj, 1997; Mittal, 2003) since the discovery of the mosquito larvicidal activity of *Bti* (serotype H-14) in 1977.
Apart from the bacterial biocontrol agents, predators, particularly fishes like *Gambusia*, *Tilapia*, *Poeciliia*, *Aphanius dispar dispar*, fungal pathogens like *Metarhizium anisopliae*, *Beauveria tenella*, *Lagenidium giganteum*, and *Chrysosporium lobatum* have shown promise as larvicides (Scholte et al., 2003a), although commercial formulations are not presently available for mosquito control (Scholte et al., 2004). Scholte et al., (2003b) reported that adult *An. gambiae* is susceptible to *Beauveria bassiana* and *M. anisopliae*, while Mohanty et al., (2008) reported that *M. anisopliae* 892 was effective against larvae of *An. stephensi*. Other biological control agents include the nematode *Romanomermis culcivorax* and the *Azolla* plant (Lacey and Lacey 1990). However, large scale application of these pathogens have not found a place in any national vector borne disease control programme.

Phytochemicals derived from natural plant products such as crude aqueous extract of fruit pods of *Swartzia madagascariensis* against *An. gambiae* (Minjas and Sarda, 1986), petroleum ether extract of six plants *Acorus calamus*, *Ageratum conyzoides*, *Annona squamosa*, *Bambusa arundanasia*, *Madhuca longifolia*, and *Citrus media* against *An. gambiae* (Sujatha et al., 1988), crude extract of *Yucca aloifolia* against *An. stephensi* (Thomas et al., 1994), *Solanum nigrum* against *An. culicifacies* species A, C and *An. stephensi* (Raghavendra et al., 2009a), emulsified neem oil formulation against late instars of *An. stephensi* (Dua et al., 2009) are among the promising larvicidal agents that can be used for larval control (Raghavendra et al., 2011). However, very few plant products have shown promise for use in large-scale vector control in field, till date.

Although larval control of malaria vector mosquitoes is a proven and effective preventive method (Walker and Lynch, 2007), yet high coverage of breeding sites is required to achieve significant success, which is a major operational and logistical challenge in many ecological settings (Pampana, 1969) and especially difficult for mosquitoes like *An. dirus s.l.* breeding deep inside forests.

### 2.10. Newer strategies for vector control

Due to the severe demerits like health hazards, damage to the environment and local ecology associated with the application of chemical insecticides like DDT, potential risk of developing insecticide resistance in vector populations, operational and logistic problems in implementation of larval control strategies and the complex biology
and behaviour of the malaria vector mosquitoes, newer approaches like integrated vector management (IVM) and genetic control methods are under way.

IVM provides a great opportunity by integrating several tools from ecological methods to achieve a stronger impact on vector control where insecticides are used as a final option to bring malaria transmission below the threshold level (WHO, 2008b). IVM emphasises on evidence based combinations of vector control methods along with effective coordination of the control activities of all sectors that have an impact on vector borne diseases, including health, water, solid waste, sewage disposal, housing and agriculture (WHO, 2004b).

Owing to certain successes in insect pest management e.g., eradication of screw worm flies and Mediterranean fruit flies from Latin America by using sterile insect technologies (Jayaraman, 1997), genetic manipulation of the disease vector species is considered as a promisable vector control strategy which basically includes genetic engineering and sterile insect techniques, the two most potential and tested genetic modification techniques. However, the molecular technological advances started only after the complete genome sequences became available (Alphey et al., 2002). Genetically engineered mosquitoes include transgenic techniques which allows producing the genetically modified vectors that are either resistant or refractory to infection with \textit{Plasmodium} parasite (Christophides, 2005; Riehle et al., 2003). The methodologies for producing transgenic mosquito can be either biological or physical (Atkinson et al., 2001). Biological methods include use of viruses like Densovirus (Carlson et al., 2006), symbiotic microorganisms like \textit{Wolbachia} (Hoffman et al., 2011), transposable elements (TE) such as \textit{piggyBac, minos, mariner or mos1, herves and hermes} (Allen et al., 2001; Aresnburger et al., 2005; Catteruccia et al., 2000; Coates et al., 1998; Grossman et al., 2001) for mosquito transformation. Physical methods includes microinjection, biolistics, (Huynh and Zieler, 1999; Mialhe and Miller 1994), electroporation (Leopold et al., 1996), etc., to deliver the foreign macromolecules into the target tissue.

Technologies like DNA recombinant technology and RNAi in combination with genetic transformation (Shin et al., 2003) may allow the identification, isolation, and characterization of new genes and genetic elements that inhibit \textit{Plasmodium} invasion in mosquito or reduce life span of malaria vector. Developing transgenic variants of all the
important vector species would be extremely challenging, and limitation will be the means to spread these transgenes through wild populations. Success in the development of genetically modified anopheline mosquitoes has been reported in recent years (Catteruccia et al., 2000; Grossman et al., 2001; Ito et al., 2002; Jacobs-Lorena 2003; Riehle et al., 2003).

Sterile insect technique (SIT) provides another ecologically safe program (Benedict and Robinson, 2003), in which rearing, sterilization, and release of genetically modified male mosquitoes result in reduction in female population (Catteruccia et al., 2005). This technique is suggested in relatively isolated area and with a single vector species. However, the major issues related to SIT are (1) establishment of sophisticated production units, (2) production of large-scale sterile insects, (3) requirement of highly trained human resource, (4) sexing of mosquitoes among others. Despite these limitations, it is the only program that has solely achieved the eradication of some parasitic diseases of veterinary importance, and equally, not yet effective for parasites of public health importance (Alphey et al., 2010; Bowman, 2006). In 1970s, SIT has been tested against the *An. stephensi*, in India and *An. albimanus* in El Salvador but unfortunately both programmes were interrupted due to some nontechnical reasons in India and eruption of civil war in El Salvador, although El Salvador successfully controlled malaria in some parts of the country (Klassen 2009; Ragheb, 2007). Pilot studies with mosquitoes are currently being undertaken in Sudan, with assistance from the International Atomic Energy Agency (Knols et al., 2007).

Other genetic control techniques like cytoplasmic incompatibility, incompatibility due to chromosomal factors, chromosome translocations, conditional lethal, meiotic drive, compound chromosome, etc., are proposed for alternative vector control strategies but were not extensively tested (Pal and LaChance, 1974).

### 2.11. Issues in malaria vector control

Malaria control with the existing arsenal of anti-malarials is problematic due to the emergence of multi-drug resistance in *Plasmodium falciparum* (WHO, 2010b). Vector control is, therefore, a cost-effective and practical approach to reduce the burden of malaria (Hemingway and Craig, 2004). Conventional vector control measures such as indoor residual spraying and insecticide treated nets are effective but operationally difficult, logistically demanding and relatively less effective against exophilic and
exophagic vectors such as An. baimaii (Obsomer et al., 2007). Due to such problems, novel vector control strategies such as genetically modified mosquito and sterile insect techniques are being developed. Whether using conventional methods or genetic-based methods of vector control, comprehensive knowledge and understanding of the biology, proper identification and distribution, genetic population structure and gene flow regime of the vector species is essential for effective vector control. Proper identification of vectors is highly important as misidentification of vectors will lead to failure of vector control as was reported in Vietnam, misidentification of An. minimus as An. varuna led to the suggestion of wrong vector control methods (Van Bortel et al., 2001). Information on population structure and gene flow provides insights into the potential spread of insecticide resistance in a vector population (Van Bortel et al., 2003). Another factor in understanding population structure is genetic drift. Small population size increases the probability that gene frequencies will shift by random change, that is, random genetic drift. Genetic drift affects the entire genome, including those traits that are relevant for vectorial capacity and control (Tabachnick and Black, 1996).

2.12. Identification of Anopheles dirus species complex

As mentioned earlier, targeted vector control is dependent on the understanding of the biology and behaviour of the vector species along with proper identification and distribution of the species. Among Diptera in many genera, like Drosophila, Simulium, Anopheles, Aedes, Sciara and Chironomus, sympatric populations within a species do not interbreed, suggesting the existence of biological species. These morphologically similar but genetically different, reproductively isolated sympatric populations within a taxon are known as cryptic species or sibling species or isomorphic species and the taxon as a "species complex" (Subbarao, 1998). The term sibling species, which are found in other animal groups also, was introduced by Mayr in 1942 (Mayr, 1970). The existence of cryptic species of Anopheles was first realized in the 1930s from the work of Grassi, Roubaud and others (Hackett, 1937). Before that time, it had been a mystery that although malaria was more prevalent in some parts of Europe than others, the apparent vector, Anopheles maculipennis, occurred throughout Europe. This phenomenon of ‘Anophelism without malaria’ was explained with the recognition that An. maculipennis exists as a complex of cryptic species that differ in their capacity to transmit malaria. The discovery of species complex adds a new dimension to vector
control. Members of the complexes are genetically isolated by either pre-mating or post-mating barriers, hence the genetic structure of each species differs from the other and, thus, has to be taken into account for any types of control strategies. Failure to recognize sibling species of anopheline taxa may result in failure to distinguish between a vector and a non-vector, hence the assessment of the impact of control measures may be seriously misleading (Coluzzi, 1988).

A number of methods have been developed to distinguish between the species of *An. dirus* complex. These methods, from basic morphology to cross-mating experiments and the more recent molecular methods such as PCR, are described below.

### 2.12.1 Morphological Identification

After the first morphological description given by Peyton and Ramalingam, (1988), morphological identification keys for *An. dirus s.l.* mosquitoes were developed by Rattanarithikul & Panthusiri (1994). The most significant adult character is:

"There is no accessory sector pale (ASP) spot on the costa and usually none on the subcosta (an occasional male in some species will exhibit an ASP spot on the costa, but always in less than 6% of specimens)"

Additional adult characters are

"Presector dark (PSD) spot on wing vein R with one or more pale spots on at least one wing (except occasional specimens of *nemophilous*), hind tarsomere 4 with a distinct basal pale band or dorsal patch"

Recently, the morphological characters along with the diagnostic characters and identification keys for female and fourth-instar larvae were described in a comprehensive review (Sallum et al., 2005a).

Although the morphological identification keys are available, it is not possible at present to distinguish between all members of the *An. dirus* complex morphologically, hence a number of other methods have been employed to differentiate them (Plate - A).

### 2.12.2. Cytogenetic and Cross-mating Experiments

The recognition of all the sibling species of the *An. dirus* complex are based on results from crossing experiments. The unidirectional F1 hybrid male sterility observed between the Bangkok strain (identified as *An. dirus s.s.* by Peyton and Harrison, 1979) and Perlis form strain was the first evidence that *An. dirus* is a complex of two sibling species (Hii, 1985). The Bangkok strain was designated as species A and the latter as
species B (Subbarao, 1998). Mitotic karyotype of these two species were described by Baimai et al., 1981. Likewise, species C, D, E, F and An. takasagoensis of An. dirus complex were also identified (Baimai et al., 1981, 1987, 1988b; Sawadipanich et al., 1990; Wibowo et al., 1984). The mitotic karyotypes of all members of An. dirus complex and their distribution in SE Asia were illustrated by Baimai, (1989). However, this method has several practical disadvantages such as extensive intraspecific variation in the amount and distribution of heterochromatin on the sex chromosomes (Baimai and Traipakvasin, 1987), need to have good chromosomal preparation and experience etc that preclude their large-scale application.

2.12.3. Isoenzyme or Allozyme analysis

Isoenzyme analysis is a biochemical method where the protein molecules in a gel matrix are exposed to an electrical field (Brewer, 1970). The method was described by Mahon et al., (1976) to distinguish between four members of the An. gambiae complex and in 1978, Miles published a Biochemical Key to identify six members of the complex. The advantage of the isoenzyme electrophoresis technique over cytogenetic analysis is that it does not require a specific sex or larval stage (Miles, 1978), however, the specimens for electrophoresis need to be fresh or stored in liquid Nitrogen. Although isozymes have been used to differentiate four members (A,B,C,D) of An. dirus complex to various degrees with some success (Green et al., 1992), use of allozymes has now largely been superseded by DNA-based methods of identification.

2.12.4 PCR based identification of Anopheles dirus complex

Since DNA based techniques are more specific, accurate and generally not limited to specific developmental stages or to a specific sex and have less stringent requirements for preservation of materials, typically by desiccation or in ethanol, they have replaced other sibling species identification methods.

One DNA-based strategy is to use DNA repeat sequences, which vary either in copy number or sequence between species, as probes in a hybridization assay, commonly known as DNA hybridization. It is a straight forward approach and has been applied to several Anopheles species complexes such as An. gambiae, An. dirus, An. quadrimaculatus, An. punctulatus and An. farauti complexes (Audtho et al., 1995; Cooper et al., 1991; Gale & Crampton, 1987; Hill et al., 1991; Hill and Crampton, 1994; Panyim et al., 1988). Problems with the use of this methodology include DNA cross-
hybridization, variation in copy number of repeated sequences across the geographic range of a species, sensitivity and the need to calibrate the amount of DNA loaded on a squash or dot-blot, and the difficulty of finding species-specific sequences that discriminate among all members of a complex. More recently, due to its reliability, reproducibility and ease of use, there has been a move towards methods that utilize the polymerase chain reaction (PCR).

The PCR (Mullis et al., 1986) is an in-vitro method developed for the enzymatic synthesis of specific DNA sequences, using two oligonucleotide primers that hybridize to opposite strands and flank the region of interest in the target DNA (Saiki et al., 1985). For PCR based species identification a range of diagnostic markers, viz, microsatellites, RAPD, PCR-RFLP and r-DNA ITS2 are used. Due to the high mutation rate, microsatellites appears to be a good marker for distinguishing very closely related species. However, the alleles are not exclusive to a species and even the use of several loci might not produce a conclusive result. Unlike microsatellites, randomly amplified polymorphic DNAs (RAPDs) have the great advantage that they can be applied to organisms where there is no prior molecular information about the genome (Williams et al., 1990). For example, they have been used successfully to identify species in the *Anopheles albitarsis* complex (Wilkerson et al., 1995), to distinguish *An. gambiae* from *An. arabiensis* (Wilkerson et al., 1993). This technique was further modified to develop species specific sequence characterized amplified region (SCAR) markers to identify four species (A, B, C and D) of *An. dirus* complex (Manguin et al., 2002). However, low reproducibility of the technique, dominant nature of RAPD alleles and the difficulty in assigning homology to amplified fragments complicate the interpretation of RAPD results (Walton et al., 1999b).

The species-specific nucleotide differences are exploited successfully in the development of diagnostic polymerase chain reaction (PCR) assays. These assays have the advantage of being able to distinguish species using a specific DNA sequence where the two species differ by as little as only one nucleotide pair. A number of different repeat DNA families are potentially good targets for such work, including the extra chromosomal mitochondrial, chloroplast and kinetoplast genomes, and the tandemly repeated nuclear gene families like the histones, the 5s RNA genes and the ribosomal RNA (rRNA) genes. Nuclear sequences like the rDNA are preferred to organelle
genomes like mtDNA because the latter have the propensity to be transferred across species boundaries by rare interspecies hybridization events or at stable interspecies hybridization zones. Collins & Paskewitz, (1996) have advocated the use of ribosomal DNA (rDNA) to differentiate cryptic species of Anopheles. The rDNA is a multicopy gene family that exists as one or more tandem arrays of many transcriptional units per cell (Gerbi, 1985), where concerted evolution rapidly spreads mutations to all members of the gene family, even if arrays are located on different chromosomes (Dover, 1982; Gerbi, 1985). In most organisms it is relatively easy to use primers designed to anneal to conserved parts of the ribosomal RNA genes to amplify an intervening variable region (Beebe and Soul, 1995). Paskewitz and Collins, (1990) applied this PCR method to distinguish members of the An. gambiae complex based on species-specific nucleotide sequences in the rDNA intergenic spacer regions. This method is useful in identifying species within the complex regardless of life stage and sex using either extracted DNA or fragments of a specimen. This method uses a universal forward primer that anneals to the DNA of all species and a series of species-specific primers that can act as a reverse primer and the method as a whole is known as allele specific PCR (ASPCR) (Scott et al., 1993). subsequently this method was used to identify sibling species of all mosquito complexes and groups of mosquitoes such as An. claviger (Kampen et al., 2003 ), An. pulcherrimus (Djadid et al., 2003), An. quadrimaculatus (Levine et al., 2004), An. maculipennis (Proft et al., 1999; Simsek et al., 2011), An. crucians (Wilkerson et al., 2004), An. punctulatus (Henry-Halldin et al., 2011), An. fluviatilis (Manonmani et al., 2001), An. nuneztovari (Fritz et al., 1994), An. minimus (Phuc et al., 2003), An. superpictus (Oshaghi et al., 2008), An. hyrcanus (Li et al., 2005), An. annularis (Walton et al., 2007a), An. maculatus (Ma et al., 2006; Walton et al., 2007b), An. sundaicus (Alam et al., 2006). The ASPCR based on the fixed nucleotide differences of the rDNA ITS2 region among five members of An. dirus complex was first explored successfully by Walton et al, (1999a) on the mosquitoes collected from Thailand. The applicability of the method on the identification of An. dirus complex mosquito in NE India was further validated successfully by Prakash et al., (2006).

Apart from ASPCR based on rDNA-ITS2, other PCR based methods such as PCR-RFLP of ITS2 region in An. culicifacies s.l. (Manonmani et al., 2007; Van Bortel et al., 2002), An. minimus group (Van Bortel et al., 2000), An. hyrcanus group (Gao et
al., 2004), PCR-RFLP of D3 region in *An. culicifacies* s.l. (Goswami et al., 2006), *An. annularis* group (Alam et al., 2007); multiplex PCR based on D2 region in *An. culicifacies* s.l. (Raghavendra et al., 2009b), multiplex PCR based on D3 region in *An. culicifacies* s.l. (Singh et al., 2004a), *An. fluviatilis* s.l. (Singh et al., 2004b) were also used for sibling species identification. Recently, Real Time PCR (PT-PCR) based assays were also developed to identify members of *An. gambiae* s.l. (Bass et al., 2007) and *An. funestus* species group (Vezenegho et al., 2009).

2.13. Molecular population genetics, genetic diversity and Phylogeography in vector control

The study of population genetics, genetic diversity and phylogeography of mosquito species is of great importance for malaria vectors control. Population genetics can reveal the historical processes that have influenced the current distribution of vector populations. Information about the level of gene exchange within and between populations can be obtained from population genetic studies. The primary focus of such studies is to identify population demes, lineages or molecular forms that may exhibit differential involvement in malaria transmission (Cohuet et al., 2010; Donnelly et al., 2002). Estimates of intraspecific genetic diversity and effective population size ($N_e$) can provide insight into the relative importance of drift or selection in driving important phenotypes such as insecticide resistance and vector competence. Comparing estimates of $N_e$ in *Anopheles* species before and after insecticide application can be a useful genetic monitoring tool (Czeher et al., 2010; Wondji et al., 2005). Characterization of barriers to gene flow can help predict the spread of genes involved in parasite refractoriness and/or insecticide resistance and in planning the release of genetically modified or sterile vectors where the scale of success will depend on the extent to which the released vectors mate with wild counterparts (Catteruccia, 2007; Marshall et al., 2008; Pinto et al., 2003). An understanding of the demographic history of *Anopheles* mosquitoes at deeper time scales can provide insights into responses to past geological and climatic changes and colonization events that led to current geographic distributions. At more recent time scales, population genetics can delineate isolated vector populations that may be appropriate target for local eradication (Chen et al., 2011; Dusfour et al., 2007; Foley and Torres, 2006; Garros et al., 2005; Gutiérrez et al.,
2009; Loaiza et al., 2010; Morgan et al., 2009, 2010; Oshaghi et al., 2006; O’Loughlin et al., 2008; Pedro et al., 2010; Zarowiecki et al., 2011a). Removing the effects of natural selection and population history may help identify genes involved in mosquito–pathogen co-evolution, which are potential targets for malaria control interventions (Lehmann et al., 2009). Molecular population genetics studies may, by advancing our knowledge of Anopheles speciation, lead to a better understanding of vector evolutionary history and malaria-transmission dynamics (Lehmann et al., 2009; Zarowiecki et al., 2011b). Additionally, if transgenic strategies are to be used to control Plasmodium transmission it is necessary to have an understanding of Anopheles genetic population structures which can provide practical information about the number of release points for transgenic mosquitoes (Loaiza et al., 2012; Tabachnick and Black IV, 1995).

2.14. Population genetics and phylogeography

Population genetics is the study of genetic variation by describing the changes in allele frequency for a particular trait over time within and between populations and the analysis of the forces responsible for the variation. These forces are mainly mutation, migration, recombination, selection and genetic drift. Therefore, population genetics is closely related to evolutionary genetics because evolution depends heavily on changes in gene frequencies. With the advent of molecular biology tools and access to DNA sequencing data, the classical population genetics, which was prospective in nature and focused on the evolution forward in time of a population under the Mendelian hereditary mechanism from various genetic parameters, has altered the focus of population genetics towards the retrospective theory to answer the question of how through evolution the population arrived at its presently observed state using the coalescent theory (Ewens, 2010; Kingman, 1982, 2000).

Phylogeography is the study concerned with the principles and processes governing the geographical distribution of genealogical lineages or gene trees especially at the intraspecific level and focuses on relationships among populations within species or among closely related species. 'Phylogeography', the term coined by Avise et al., (1987), deals with estimating and quantifying the spatial and temporal components of intraspecific population structure and interpreting the evolutionary and
ecological processes responsible for it. As a sub discipline of biogeography, phylogeography emphasizes historical aspects of contemporary spatial distributions of gene lineages (Avise, 1996). Because it focuses on the population–species interface and demographic/evolutionary history, phylogeography has provided valuable contributions to studies of geographic variation, speciation, historical biogeography, conservation biology, biodiversity research/taxonomy, palaeoecology, palaeoclimatology and volcanology (Beheregaray, 2008).

2.14.1. Evolutionary theories

When studying population genetics, evolutionary processes must be taken into account to explain the current patterns of DNA polymorphisms. The choice of model of evolution will depend on the type of data collected. For DNA sequence data an infinite-sites model is used, which assumes that most of the time new mutations will arise at sites that were previously monomorphic (Watterson, 1975).

The neutral theory of molecular evolution suggests that most polymorphisms observed at a molecular level are selectively neutral, so that their frequency dynamics in a population are determined by random genetic drift (Kimura, 1968). Examining departures from neutrality has proved to be a valuable tool in the analysis of molecular data. Departures from neutrality can be caused by selection, hitchhiking, recombination and rates of nucleotide substitution, as well as population expansion or substructure (Hartl and Clark, 2006). The effect of selection and recombination can be minimised by choosing marker from a neutral loci such as mitochondrial DNA (mtDNA), although evidences are therethat selection does take place at some mitochondrial loci, and that genetic hitchhiking, where mtDNA will ‘hitchhike’ along with a cytoplasmic factor, also occurs (Hey, 1997; Turelli et al., 1992).

2.14.2. Hardy-Weinberg principle

Population genetics is based around the Hardy-Weinberg principle (Hartl and Clark, 2006). It states that the alleles frequencies in an ‘ideal’ diploid population will be in a state of equilibrium, according to the equation \( p^2 + 2pq + q^2 = 1 \), where \( p \) = frequency of A allele and \( q \) = frequency of a allele. Various assumptions governing the Hardy-Weinberg principle are:

(i) The organism is diploid.
(ii) Reproduction is sexual.

(iii) Generations are non-overlapping.

(iv) The population is large enough to be unaffected by random gene changes (i.e., genetic drift).

(v) There is no gene flow (immigration or emigration).

(vi) No mutations occur or there is mutational equilibrium.

(vii) Reproduction is random (independent of genotype).

(viii) Natural selection is not acting on a particular phenotype.

This means that if a population is found to be departing from Hardy-Weinberg equilibrium it must be violating one or more of these assumptions. The study of population genetics and phylogeography attempts to use departures from Hardy-Weinberg equilibrium to examine genetic drift, gene flow, mutation and natural selection.

2.14.3. Forces affecting genetic variation

Evolution is frequently considered as a two-step process (Mayr, 1988):

1. Genetic variation is randomly created by new mutations, also accompanied by random recombination
2. Natural selection and/or genetic drift act upon the existing variants

However, the truth of the first point has been questioned by Bradshaw, (1991). When inferences are based on DNA sequence data, it is necessary to understand the evolutionary forces shaping the diversity of the studied population/s that are selection, mutation, migration (gene flow) and genetic drift. Natural selection, genetic drift, and mutations tend to increase genetic differentiation among populations, while migration tend to reduce it.

Natural selection is defined as any consistent difference in fitness among phenotypically different classes of biological entities (Futuyma, 2009). Genetic variation promotes natural selection and genetic inheritance transmits adaptive traits from one generation to the next. One way natural selection alters allele frequencies within populations is by reducing the frequency of deleterious alleles within the population over time. Individuals with deleterious phenotypes do not survive and reproduce as well as those who lack the deleterious allele. As a result, the frequency of
the deleterious allele is reduced over time. Just as selection serves to reduce the frequency of one allele within a population, it can also serve to increase frequencies of others.

There are three modes of natural selection—directional, stabilizing, and disruptive (Futuyma, 2009). Directional selection results in a shift in the frequency of one or more traits in a particular direction. For example, many insect species have gradually increased the amount of pigment found within their cells in order to blend in with the environment and escape predation. Change associated with directional selection is relatively common in changing environments. Disruptive selection serves to increase the frequency of extreme phenotypes and decrease the frequency of average phenotypes within populations. Stabilizing selection acts against extreme phenotypes and selects for the average phenotype in the population. Stabilizing selection is characteristic of a stable, unchanging environment.

**Genetic drift** is defined as the change in allele frequencies over time due to chance and chance alone (Avise, 2004; Futuyma, 2009). Drift is more when the population size is small. The smaller the population, the greater the chance of having random deviations from the average. Such random events may lead to a change in allele frequencies within the population. Drift decreases diversity within a population because it tends to cause the loss of rare alleles, reducing the overall number of alleles. The founder effect and a population bottleneck are two examples of random drift that can have profound effects in small populations (Futuyma, 2009; Hartl and Clark, 2006). The founder effect occurs when an entire population descends from a small number of individuals. Such a situation produces a population whose gene pool is a tiny sample of the original and the frequency of any allele within this population is likely to be different from that of the original population. In the case of a population bottleneck, all organisms within the population have descended from a few individuals.

**Mutation** is defined as any heritable change in the genetic material (Hartl and Clark, 2006). Therefore, mutation includes a change in the nucleotide sequence as well as the formation of a chromosome rearrangement, such as inversion and translocation. Mutations that do not influence the ultimate reproductive fitness of an organism are called neutral mutations and give rise to what are called neutral genes and neutral alleles. Although neutral alleles may not be important for the evolutionary change, they
are of extreme importance to geneticists tracing the evolution of populations and species.

*Migration* (gene flow) implies not only the movement of individuals into new populations but also the introduction of new alleles into the population due to it (Futuyma, 2009; Hartl and Clark, 2006). *Gene flow* is the passage and establishment of genes typical of one population in the gene pool of another by natural or artificial hybridization and backcrossing resulting in changes of allele frequencies within populations over time. In general, gene flow tends to increase variation within local populations, but reduces variation between adjacent populations.

*Non-random mating* occurs when individuals those are more closely (inbreeding) or less closely related mate more often than would be expected by chance for the population. It increases the homozygosity of a population and its effect is generalized for all alleles. Inbreeding per se does not change the allelic frequencies but, over time, it leads to homozygosity by slowly increasing the two homozygous classes.

Of all the different forces, genetic drift and migration are most important for population genetics and phylogeographical inferences, as most of the molecular markers used for such analysis avoid other forces such as recombination and selection. Drift has a stronger effect in populations with small effective population \((N_e)\) sizes due to the increased probability of any particular allele going to the next generation (Nei et al., 1975). In subdivided populations, the random nature of drift will tend to cause fixation of different alleles thereby reducing diversity. The degree of differentiation created by drift is counteracted by the amount of migration or gene flow. If there is more than one migrant per generation \((N_m>1)\), this will be sufficient to counteract drift and prevent population differentiation (Slatkin, 1987).

Wright (1931) and Fisher (1930) independently described mathematically the simple population model which became standard for most circumstances. The main feature of the Wright–Fisher model is the persistence of a single population of constant size, with random mating among individuals (panmixia).

### 2.14.4. Models of population structure

Real populations are different from Wright–Fisher populations, as they often have complex geographies and consist of many populations that are connected by gene
flow. Individuals in most real populations are more likely to reproduce with nearby individuals than with distant individuals. To consider population structure four models have been developed from which migration can be modelled in different ways appropriate to the system being studied. (Reviewed in Hey and Machado, 2003).

The earliest model of population substructure was Wright’s *Island model* (Wright, 1931) which considered a population as an infinite number of 'demes' or 'islands' between which varying amounts of gene flow may occur. Wright proposed a series of statistics to describe the genetic structure of populations:

To measure inbreeding within and between subpopulations three measurements are needed:

- $H_I$ = the heterozygosity of an individual in a subpopulation
- $H_S$ = the expected heterozygosity of an individual in an equivalent random mating subpopulation
- $H_T$ = the expected heterozygosity of an individual in an equivalent random mating total population

These measurements are used to derive three coefficients:

1). The inbreeding coefficient. The reduction of heterozygosity of an individual due to non-random mating within a subpopulation:

$$F_{IS} = \frac{(H_S - H_I)}{H_T}$$

2). The fixation index. The reduction of heterozygosity of a subpopulation:

$$F_{ST} = \frac{(H_T - H_S)}{H_T}$$

3). The overall inbreeding coefficient. The reduction in heterozygosity of an individual relative to the total population (due to non-random mating within subpopulations):

$$F_{IT} = \frac{(H_T - H_I)}{H_T}$$

$F_{ST}$ is an important measure of population structure and is widely used in diversity studies. It gives an overall level of genetic divergence among subpopulations. The value of $F_{ST}$ can in theory be from 0 to 1. (Wright, 1978) suggested the following quantitative guidelines to interpreting $F_{ST}$:

- 0 to 0.05 – little genetic differentiation
0.05 to 0.15 – moderate genetic variation
0.15 to 0.25 – great genetic variation
above 0.25 – very great genetic differentiation

\(F_{ST}\) is also related to the level of migration between subpopulations by the equation

\[F_{ST} = \frac{1}{4Nm + 1}\]

where \(N\) is the deme size and \(m\) is the number of migrants per generation. Island models are good for understanding the effects of small population size and limited gene flow on rates of genetic drift and levels of divergence between island populations (Wright, 1940).

Unlike island model, Stepping-stone model specifically includes a spatial element, with individual populations only able to exchange genes with adjacent populations. Stepping-stone model can be one dimensional (with populations in a line), two- or three dimensional (Kimura and Weiss, 1964). If a stepping-stone model is taken to the extreme, then every individual is restricted in its local movement where gene flow is more likely between nearest neighbours in a two dimensional landscape. This model is known as Isolation by distance model (Wright, 1943). In nature, not only do individuals move between populations, but also individual populations come and go over time (\(t\)) with the colonization and extinction of entire populations which is considered in the Metapopulation model of population structure (Slatkin, 1977; Wade and McCauley, 1988). Figure 2.7 depicts different models of population structure schematically.

### 2.14.5. Summary statistics

The genetic information stored in the sequence data can be generated by simple calculations based on the nucleotide polymorphisms, known as summary statistics. A commonly used measure of diversity is Nei’s gene diversity statistics, also known as haplotype diversity. This statistics measures the probability that two alleles or haplotypes drawn at random from the population will be different from each other. It is equivalent to the expected heterozygosity for diploid data, however this diversity measure does not account for molecular distances between alleles or haplotypes. Gene diversity (Nei, 1987, p. 180) is estimated as
Figure 2.7: Models of population structure (a. Island model of migration, b. Stepping-stone model, c. Isolation by distance model, and d. metapopulation model) (Adapted from Hey and Machado, 2003).
where \( n \) is the number of gene copies in the sample, \( k \) is the number of alleles or haplotypes and \( p_i \) is the sample frequency of the \( i\)-th haplotype. Another measure of diversity is the **nucleotide diversity**, \( \pi \), which describes the probability that two copies of the same nucleotide drawn at random from a set of sequences will be different from one another and is analogous to Nei's gene diversity. The commonly used estimator for \( \pi \) is shown below

\[
\pi = \frac{n}{n-1} \sum \pi_{ij}
\]

(Nei, 1987 (p. 257); Nei and Miller, 1990)

where \( n \) is the number of sequences, \( x_i \) and \( x_j \) the frequencies of the \( i\)-th and \( j\)-th sequences respectively and \( \pi_{ij} \) the proportion of different nucleotides between them.

The standard measure for examining genetic diversity is the population parameter theta (\( \theta \)), which is represented as

\[
\theta = 4N_e \mu
\]

(Nei, 1987 (p. 257); Tajima, 1993)

where \( N_e \) is the effective population size and \( \mu \) is the mutation rate per nucleotide site per generation. The above equation is specific for diploid loci; however, a more general form of the equation that considers loci with their inheritance pattern can be represented as follows

\[
\theta = 2nN_e \mu
\]

where \( n \) represents the number of heritable copies of the locus per individual, which is 2 for diploid loci, 0.5 for Y chromosome and mtDNA and 1.5 for the X chromosome.

There are several methods available for calculating \( \theta \) estimates. \( \theta_s \) is estimated from the infinite-site equilibrium relationship between the number of segregating sites (S) (Watterson, 1975):

\[
\theta = \frac{S}{a_1}, \text{ where}
\]

\[
a_1 = \sum_{i=1}^{n-1} \frac{1}{i}
\]
On the other hand, $\theta\pi$ is based on the infinite-site equilibrium relationship between the mean number of nucleotide differences in all pairwise comparisons (Tajima, 1983):

$$\theta = E(\pi)$$

Here, as with all summary statistics, a neutral model of evolution is assumed and under a neutral model of evolution $\theta$ is equal to the nucleotide diversity ($\pi$) which means that allele frequencies are assumed to be at equilibrium and there are no changes in population size or gene flow is constant. In real populations this is difficult to find. Different estimates of $\theta$ can give an indication of how a sample departs from the neutral model. For example, the ratio between $\theta\pi$ and $\theta_s$ (Tajima's D) will change after a change in population size, as $\theta_s$ would quickly respond to the change whereas $\theta\pi$ would reflect the past population size for a longer time (Tajima, 1989).

Summary statistics also include measures of population differentiation such as $F_{ST}$. $F_{ST}$ is a measure of the proportion of variance between populations compared with the amount of variance in the total sample. Differentiation between populations is related to gene flow so $F_{ST}$ is also an estimate of migration through the equation

$$F_{ST} = \frac{1}{1 + 4Nm}$$

where $Nm$ is the number of migrants. The derivation of this relationship assumes migration-drift equilibrium in a population so it cannot be used reliably to infer gene flow in a population that is under expansion or faced bottleneck or has had changing gene flow (Whitlock and McCauley, 1999).

Another commonly used way of representing diversity for various types of molecular data is the mismatch distribution which is the frequency distributions of the numbers of segregating sites in all possible pairwise comparisons. The mismatch distribution is a good example of a descriptive summary statistics as the mean of the distribution provides a simple description of the overall diversity, the shape of the distribution infers about the population history. A smooth and unimodal mismatch distribution indicates a period of rapid population growth or recent population expansion due to the concentration of coalescent events at the base of the underlying star-like genealogy, whereas a ragged, multi-modal distribution indicates a long term stable population that has undergone stochastic lineage loss (Rogers and Harpending, 1992; Slatkin and Hudson, 1991). To distinguish between these two types of
distribution a **raggedness statistics** \( r \) is used that is simply the sum of the squared difference between neighbouring peaks (Harpending, 1994) and is estimated by the equation

\[
    r = \sum_{i=1}^{d+1} (x_i - x_{i-1})^2
\]

where, \( d \) is the maximum number of differences between alleles and \( x_i \) is the relative frequency of \( i \) pairwise differences. From mismatch distribution time of population expansion \( t \) can also be estimated using

\[
    t = \frac{\tau}{2u},
\]

where \( u = \mu \times \text{number of base pairs sequenced} \times \text{generation time} \) and \( \mu \) is the mutation rate (Harpending et al., 1993).

### 2.14.6. The neutral theory of molecular evolution

This theory, proposed by Kimura, (1968) and King and Jukes, (1969) and later developed in detail by Kimura, (1983), states that standing molecular variation in the population is largely selectively neutral, while most new mutations are highly deleterious. Most evolution would proceed by random fixation or extinction of neutral variants due to the action of drift with advantageous mutations vanishing rarely. At equilibrium, there is a balance between mutation and random genetic drift and population size has no effect on it. If evolution occurs by stochastic sampling processes, one can predict rates of change based simply on the rate of input of neutral mutations (\( \mu \)), and sequence evolution should be clocklike. The neutral theory explains much molecular polymorphism data very well; however, there are many examples of natural selection at the molecular level (Kreitman and Akashi, 1995).

The 'neutral theory' largely has been superseded by the 'nearly neutral theory' (Ohta, 1992), which predicts that much of the variation we see is due to weakly selected mutations. The strength of selection is a function of population size; a mutation will be effectively neutral if its selective disadvantage is less than \( 1/2 \Ne \) (Ohta and Gillespie, 1996).

### 2.14.7. Tests of neutrality

In a standard neutral model genetic variation is assumed to be selectively neutral, populations are assumed to be panmictic (randomly mating), constant in size and in mutation-drift equilibrium (Wayne and Simonsen, 1998). Point mutations in
DNA sequences are typically modelled using the infinite-site model (Kimura, 1969) in which each new mutation occurs at a site that has not previously mutated. Deviations from this assumptions may lead to rejection of the null hypothesis.

Various tests have been derived to test departures from the neutral evolutionary model. For nucleotide data, one of the most popular tests is Tajima's D-test (Tajima, 1989). Tajima's D is the scaled difference in the estimate of $\theta = 4N_e\mu$ ($N_e$ = effective population size, $\mu$ = mutation rate per generation) based on the number of pairwise differences ($\theta\pi$) and the number of segregating sites ($\theta_s$) in a sample of nucleotide sequences. It is defined as

$$D = \frac{\theta\pi - \theta_s}{\sqrt{Var(\theta\pi - \theta_s)}}$$

Under neutrality, Tajima's D is expected to be zero. Significantly positive values of this statistics indicates population subdivision or balancing selection, whereas, negative values indicates positive selection or population growth or expansion. This test is based on the infinite-site model without recombination, appropriate for short DNA sequences or RFLP haplotypes. Likewise Tajima's D, Fu's Fs test of selective neutrality is based on infinite-site model without recombination and estimates the probability of observing a random sample with a number of alleles equal to or smaller than the observed value under given the observed level of diversity and the assumption that all the alleles are selectively neutral (Fu, 1997). If we call this probability $\hat{S}$, then

$$F_s = \ln\left(\frac{\hat{S}}{1 - \hat{S}}\right)$$

A negative value of Fs is the evidence for an excess number of alleles, as would be expected from a recent population expansion or from genetic hitchhiking. A positive value of Fs is the evidence for an deficiency of alleles, as would be expected from a recent population bottleneck or from over dominant selection. Fu, 1997 noticed that Fs statistics was more sensitive to population expansion and genetic hitchhiking, which generally lead to large negative Fs values, than Tajima's D. Fu and Li's ($D^*, F^*, D, F$) tests are also based on comparing different estimates of the parameter $\theta = 4N_e\mu$ based on mutations on internal and external branches of the geneology (Fu and Li, 1993). Assuming that selection will purge deleterious mutations, those mutations present are
likely to have arisen recently and are found close to the tips of the genealogy. Similarly mutations in the internal branches are likely to be older and selectively neutral, although a recent mutation conferring a selective advantage could increase to a high frequency and therefore, appears internal. If purifying selection is acting on a locus, or if there has been a recent population expansion, an excess of external mutations would exist. There are several similar tests based on slightly different test statistics such as Fay & Wu' H (Fay and Wu, 2000) and E (Zeng et al., 2006). A likelihood ratio test of a similar problem was described in Galtier et al., (2000). All the above tests are based on the distributions of allele frequency and applicable to single locus.

Another class of neutrality test is based on the fixation rates and pattern of fixation of mutations in coding and non-coding regions of a DNA sequence, known as $D_n/D_s$ or $K_a/K_s$ test. Hill & Hastie, (1987) and Hughes & Nei, (1988) were the first to use the ratio of non-synonymous differences ($D_n$) to synonymous differences ($D_s$) among DNA sequences as a test for positive selection. The ratio of $D_n/D_s$ is often known as $\omega$ which if significantly greater than 1 indicates diversifying selection and if significantly less than 1, indicates purifying selection.

Another class of neutrality test utilizes data from intraspecific polymorphism and interspecific divergence. The McDonald - Kreitman (MK) test (McDonald and Kreitman, 1991) examines the neutral prediction that, in the absence of selection, the ratio of non-synonymous to synonymous substitutions will be the same within and between species. This test uses a two by two contingency table, which is usually evaluated by a $G$ test, although Fisher’s exact test may be used if the sample size is very small. The Hudson - Kreitman - Aguade (HKA) test (Hudson et al., 1987) is a goodness-of-fit test which evaluates the neutral hypothesis that levels of variation within species (polymorphism) and variation between species (divergence) at two (or more) loci will be correlated. The HKA test statistic $X^2$ follows a $\chi^2$ distribution under the null hypothesis. The test is usually applied to silent variation, that is, synonymous sites. The HKA test can be applied to both DNA sequence and RFLP data, and has been used commonly to test for departures from neutral expectations where multiple loci are involved.
2.14.8. Coalescent Theory

The development of phylogeography has led to the development of a body of analytical and statistical methods of ‘coalescent theory’ that explores connections between population demography and gene genealogies within and among closely related species (Hudson, 1991; Kingman, 1982; Tajima, 1983). Unlike traditional ‘forward looking’ population genetic theory designed to predict changes in allele frequencies over time (Ewens, 1972), coalescent theory projects gene genealogies backwards in time until the most recent common ancestor (MRCA) of the sample is reached and estimates parameters such as historical population sizes, divergence times and migration rates, while accounting for the stochastic processes of lineage sorting within and among gene trees (Fig. 2.8) (Wakeley, 2008). This relationship provides a framework for testing different models of population history (reviewed in Emerson et al., 2001).

Coalescent theory has been considered as a basic approach in statistical phylogeography which is based on methods that make both explicit statistical links between process, prediction and test and also incorporates a diverse array of processes and histories (Knowles, 2009; Knowles and Maddison, 2002). These include methods that calculate the probability of the data using likelihood-based or Bayesian approaches. One of the most widespread uses of the coalescent is as a simulation tool. Using the coalescent it is possible to simulate samples from a wide variety of models and use these simulations as a basis for statistical testing of the models (reviewed in Rosenberg and Nordborg, 2002). In recent years numerous computer programmes have been developed (reviewed in Excoffier and Heckel, 2006; Kuhner, 2008) which use a maximum likelihood or bayesian approach to determine the best value for a particular demographic parameter such as θ, migration (m), growth rate (g) and divergence time (t) from coalescent probabilities and to disentangle complex evolutionary histories, e.g., distinguishing divergence with gene flow from retention of ancestral polymorphism (Carstens & Knowles, 2007). Usually a Markov Chain Monte Carlo (MCMC) sampling method is used to reduce the number of genealogies analysed. However, potential drawbacks of the statistical phylogeography are that an inappropriate choice of evolutionary model may lead to the erroneous interpretation of the results and generally simple evolutionary and demographic models may fail to elucidate the complexity of evolution in natural population (Neilsen and Beaumont, 2009).
Figure 2.8: The basic principle behind the coalescent.

a. The complete genealogy for a population of ten haploid individuals is shown (diploid populations of $N$ individuals are typically studied using a haploid model with $2N$ individuals). The black lines trace the ancestries of three sampled lineages back to a single common ancestor. b. The sub genealogy for the three sampled lineages. In the basic version of the coalescent, it is only necessary to keep track of the times between coalescence events ($T(3)$ and $T(2)$) and the topology — that is, which lineages coalesce with which. $N$, number of allelic copies in the population; $n$, sample size. (Adapted from Rosenberg and Nordborg, 2002).
2.14.9. Maximum likelihood and Bayesian methods

Maximum likelihood and Bayesian methods are statistical analytical approaches that allow the incorporation of the evolutionary model into estimates of phylogenetic relationships and various demographic parameters, and, therefore, provide more accurate and robust estimates (Holder and Lewis, 2003). Methods have been developed to determine the model of evolution that best fits the data (Posada and Crandall, 2001). Maximum likelihood method assigns a likelihood score to a particular hypothesis (e.g., the topology of a phylogenetic tree) based on how well the specified evolutionary model best fits the data and the hypothesis with the highest likelihood score is identified. Since only one hypothesis is identified as the most likely and there is no indication of the degree of error associated with the hypothesis, further bootstrapping steps are required to determine degree of error which is time consuming and computationally intensive (Holder and Lewis, 2003). On the other hand, Bayesian analysis not only incorporate evolutionary model parameters, but also allows the incorporation of prior information regarding the assumed distribution of various parameters based on background knowledge (Beaumont and Rannala, 2004; Holder and Lewis, 2003). The product of the probability distribution of the prior (the assumed distribution of parameter values prior to analysis) and the likelihood (a conditional distribution based on a statistical model that measures the probability of the various parameter values of the given data) creates a joint distribution. By manipulating this joint distribution, the probability of various parameters given the data may be investigated and a posterior distribution of the parameter values inferred (Beaumont and Rannala, 2004).

2.15. Molecular population genetics and phylogeography studies on *Anopheles dirus* complex mosquitoes

The first study of *An. dirus* population structure was carried out using enzyme electromorphs (Green et al., 1992). This study looking at *An. dirus* species A, B, C, and D in Thailand suggested, using F-statistics, complete assortative mating and therefore supported the integrity of each species. A subsequent study on *An. dirus* A, C and D was carried out using the mitochondrial cytochrome oxidase I (COI) (Walton et al., 2000). Mosquitoes were collected from 14 sites in Thailand, Myanmar and Bangladesh. Within species FST values found that the two populations of *An. dirus* C were
genetically quite distinct whereas no population structure was detected among populations of *An. dirus* A or D. Using a hierarchical AMOVA pairwise comparisons *An. dirus* C was found to be distinct from both *An. dirus* A and *An. dirus* D. Mantel test results suggested that there was no isolation by distance among all A populations, all D populations or all A and D populations combined. A starburst shaped minimum spanning haplotype network, smooth unimodal mismatch distribution curve and negative values of neutrality tests for *An. dirus* A and D suggested a recent population expansion, either demographic or by a selective sweep. A greater genetic diversity was observed for *An. dirus* D then species A suggesting that population expansion occurred first in species D and subsequently in species A.

Overall, these results were surprising as it was thought previously that *An. dirus* A and C were very closely related (see section ‘Taxonomic status of *Anopheles dirus* complex’). The genetic similarity between *An. dirus* A and D was not expected. The most likely explanations were that historical introgression occurred before population expansion when both populations had limited distribution, or that a selective sweep of mtDNA originating from *An. dirus* D passed into *An. dirus* A. Ongoing gene flow between the two species is unlikely due to limited range overlap and unfit hybrids. However, ongoing geneflow could not be ruled out given the data available. In this study homoplasy caused by hypervariable sites caused some problems in the data analysis.

A further study was carried out by Walton et al., (2001) using microsatellites to examine the populations previously studied using mitochondrial locus. Pairwise comparisons found a high level of differentiation between each species. Evidence of population structure was found within species D. The two populations of species C showed high differentiation, close to the between species level suggesting the possibility of two species existing. The differentiation between species A and D supported the theory of historical introgression of mtDNA between the two species.

Later, O’loughlin et al., (2008) carried out an in depth population genetics and phylogeographical study on *An. dirus* A and *An. baimaii* (species D) based on mtDNA COI and COII sequences from 21 different populations from mainland SE Asia and reported a more complex population history of the two species. This study included one
population of *An. baimaii* from Assam and found that this population bears the highest genetic diversity and expanded ~ 0.3 MYBP (million years before present) with a stable and oldest population history indicating that this region may be a possible Pleistocene forest refugium. This study also reported presence of genetic structure in *An. baimaii* and in one population of *An. dirus*. A significant Mantel test in *An. baimaii* populations indicated an isolation by distance model of population structure suggesting restricted gene flow, however estimates of gene flow was not possible due to the signal of expansion. It was also concluded that the population expansion in the two species was not due to human expansion, but rather fits to the Pleistocene climate change.

In contrast to *An. dirus* and *An. baimaii*, *An. scanloni* (*An. dirus C*), which is restricted to limestone curst habitat in Thailand was found to be highly diverged using mtDNA COI and COII sequences (O’loughlin et al., 2007). One fixed polymorphism (A →G at position 337) at rDNA locus was also observed between northern and western populations, however, no post mating barrier was found from cross-mating studies indicating speciation in species C of *An. dirus* complex mosquito.

Recently, an in depth phylogeographical study including *An. dirus*, *An. baimaii* and *An. scanloni* from the mainland SE Asia using mitochondrial, microsatellite and nuclear sequence markers revealed that *An. baimaii* had a more confined westerly distribution until it spreads eastwards making secondary contact ~62 kyr with a closely related species, *An. dirus*, on the Thai-Myanmar border (Morgan et al., 2010). This resulted in introgression of mtDNA from *An. baimaii* into *An. dirus* accompanied by a selective sweep of mtDNA supporting the views of Walton et al., 2000 and O’loughlin et al., (2008). Using highly analytical approach like coalescent based IMa analysis (Hey and Nielsen, 2007), this study revealed a recent divergence (within last 1.5 MY) for all the three species e.g., 192-877 kya (Kilo year) for the divergence of *An. scanlini* from *An. baimaii*, 209 - 932 kya for *An. scanlini* from *An. dirus* and 163 kya - 1.53 MY for *An. baimaii* from *An. dirus* based on nuclear sequence data. This study also included one population of *An. baimaii* from Assam and found that this population has substantially higher genetic diversity and significantly differentiated from other *An. baimaii* populations in the study and divergence from northern Thailand populations was estimated 117-535 kya in the late Pleistocene. These findings further supported the observation of O’loughlin et al., (2008) as the long term environmental stability of
the NE India as a potential Pleistocene refugial region and strongly pointed out that *An. baimaii* in this region is the origin of all extant populations. Although bidirectional gene flow was found to occur between *An. baimaii* and *An. dirus* due to secondary contact, speciation between these two species was attributed to prolonged allopatric isolation in separate forest refugium without secondary contact. The recognition of NE India along with the Indo-Myanmar biodiversity hot spot area as a putative forest refugial region was also postulated based on a phylogenetic study of the Neocellia Series of *Anopheles* mosquitoes (Morgan et al., 2009) and a comparative phylogeography study including 9 species of Anopheles mosquitoes (Morgan et al., 2011).

2.16. Choice of molecular markers

The use of PCR has made it easier to access the genome and investigate the molecular genetics of any species, individual or population size samples by using different molecular markers. These markers range from the traditional tools like cytogenetics and isozymes to the ‘classical genetic markers’ (mitochondrial DNA and cDNAs) to methods used to detect and identify single nucleotide polymorphisms (SNPs) and finally to highly polymorphic markers (RAPDs, microsatellite DNAs and AFLPs) (Norris, 2002). One of the greatest advantages of this wide variety of genetic markers is that researchers may choose to utilise any combination of markers or techniques to address different questions regarding anopheles biology, malaria vector control and malaria transmission by anopheline mosquitoes (Taylor et al., 2001). These molecular markers have proven useful in a wide variety of applications including molecular taxonomy, evolutionary systematics, population genetics, genetic mapping, and a variety of molecular diagnostics.

2.16.1 Markers for *Anopheles* species Identification

Different markers used for identification of cryptic species were mentioned earlier (refer to section 2.12.4 PCR based identification of *Anopheles dirus* complex). Of these the allele specific PCR (ASPCR) method, based on the length differences and fixed substitutions among ribosomal DNA (rDNA) ITS2 region, has been used widely for most of the *Anopheles* species identification (Collins and Paskewitz, 1996; Loaiza et al., 2012; Walton et al., 1999a). rDNA has long been considered a useful marker for comparative evolutionary and phylogenetic studies (Mindell and Honeycutt, 1990; Schlotterer et al., 1994). The basic organization of rDNA is conserved throughout
eukaryotes. Eukaryotic ribosomal RNA genes are arranged in tandemly repeated clusters, with each cluster containing the genes for 18S-, 5.8S-, and 28S-like ribosomal RNAs. The genes are separated by several spacers, namely, the NTS (nontranscribed spacer), and ITS1 and ITS2 (internal transcribed spacers). The NTS separates neighboring repeat units; ITS1 is located between the 18S- and 5.8S-like coding regions; ITS2 lies between the 5.8S- and 28S-like genes (Fig. 2.9) (Gerbi, 1985). Ribosomal repeats are usually localized to one or a few chromosomes and form part of the nucleolar organizers (NO). The nucleotide sequence of the spacer regions are often much more polymorphic between species than within species. This makes this region of the genome useful for delineating molecular differences between cryptic species of Anopheles by length or sequence polymorphisms (Reviewed in Norris, 2001).

The population structure of malaria vectors has also been investigated with the ITS2 marker, but to a much lesser extent than mitochondrial markers. Overall, patterns of genetic divergence by the ITS2 seem to be consistent with the geographical origin of populations, and this may indicate that concerted evolution is acting intraspecifically regardless of significant intraindividual variation (Reviewed in Loaiza et al., 2012). Also, even though natural selection and chromosomal recombination may affect the accuracy of this spacer to identify genetic discontinuities (i.e., species and population boundaries), a recent study by Alquezar et al., (2010) strongly supported its evolutionary cohesion across interbreeding populations, thus, reinforcing its value for taxonomic studies in malaria vectors. The rDNA-ITS2 is now-a-days regarded as a molecular key for the identification of eukaryote biological species (Coleman, 2003) and the phylogenetic property of the secondary structure of ITS2 region was recently elucidated to resolve taxonomic problems (Coleman, 2003; Keller et al., 2010).

2.16.2 Markers for population genetics and phylogeography

In population genetics and phylogeography, the connection between demographic features (such as migration, population size, natural selection, historical events etc.) and the distribution of molecular genetic variants of a species or group of species are investigated and from which inferences about the biology of the organism can be made (Burk et al., 1998). Thus, by examining genetic markers with appropriate rates of change, and, therefore, suitable signals, information can be obtained about
Figure 2.9: Structure of a r-DNA cistron showing the ITS regions.
almost any population and evolutionary process through the hierarchy of life. Based on the levels of molecular change that would provide information at different levels of population biology, the genetic markers used for population genetics and phylogeography study can be broadly divided into **genotypic**, most commonly used in the form of multiple microsatellite loci scored in samples of individuals. In sexual species, these arrays are reshuffled at each generation, and, therefore, are useful for the shortest- and finest-scale population processes, such as individual identification and tracking, parentage and relatedness of interacting individuals; **genic**, such as mitochondrial DNA (mtDNA) and other single-locus markers that can be analysed as individual genes with frequencies and geographic distributions which is under the effect of drift, selection gene flow and founder effect. These properties change on larger spatial and temporal scales than genotypic arrays, and are effective markers of gene flow and population history, even in species with limited genetic variation; and, **gene genealogies**, in which analysis of mutations in sequence of mtDNA or nuclear markers (single copy nuclear markers) and their evolutionary relationships determined by the mutation rate, selection and population parameters, such as $N_e$ and changes in $N_e$ are highly informative about the longer-term processes of phylogeography, speciation and deeper taxonomic phylogenetic reconstruction within and among species (Reviewed in Sunnucks, 2000).

The choice of a right marker to address a specific question of evolutionary biology is always important and critical since different markers may give different molecular signals owing to its different mutation rate, recombination rate, mode of inheritance and genome location, which all must be considered for adequate resolution of different evolutionary processes (Rosenberg and Nordborg, 2000; Wandeler et al., 2007). Choice of a proper marker is also central to reduce errors and ambiguities in estimates of population structure and gene flow (Bossart and Prowell, 1998). Some of the frequently used molecular markers in the field of population genetics and phylogeography studies of *Anopheles* are reviewed below.

### 2.16.2.1 Mitochondrial DNA

The first studies considering mtDNA variation employed restriction fragment length polymorphisms (RFLP) (Avise et al., 1979; Brown & Wright, 1979). Since then mtDNA has been used extensively as a tool for inferring the evolutionary and
demographic past of both populations and species due to the uniparental (maternal) mode of inheritance, non-recombining nature, high rate of mutation, high copy numbers, nearly neutral fashion of evolution (as being involved in basic metabolic functions), clock-like evolution and the availability of conserved primers and PCR protocols (Brown et al., 1979; Kocher et al., 1989). After publishing of mitochondrial genome of *An. gambiae* in 1993 (Beard et al., 1993), both coding (i.e. NADH dehydrogenase subunit 5 (ND5) and cytochrome oxidase subunits I and II (COI and COII)) and non-coding (16S and 12S RNAs) regions have been frequently targeted for *Anopheles* phylogenetics, population genetics and phylogeography studies (Reviewed in Loaiza et al., 2012; Norris, 2002). However, the utility of mtDNA as a marker in population genetics and phylogeographic studies have been debated due to the inherent genetic and biological properties *viz*, clonality (maternal inheritance), neutrality, clock-like evolution, introgression and selective sweep of mitochondrial genome (Ballard and Kreitman, 1995; Ballard & Whitlock 2004; Galtier et al., 2009; Zink and Barrowclough, 2008). Apart from such biological limitations, technical limitations such as illegitimate amplification of *Numts* (nuclear sequence of mitochondrial origin) or mitochondrial pseudogenes in some species can seriously confound population genetics and phylogeographic studies (Bensasson et al., 2001; Zhang and Hewitt, 1996). Despite such undoubted problems, mtDNA remains one of the most powerful and reliable tools for detecting population structure and inferring population history due in no small part to its high mutation rate and smaller effective population size than nuclear DNA that leads to more rapid lineage sorting and divergence between populations following their isolation (Zink and Barrowclough, 2008). Clearly, mtDNA is the most convenient and cheapest solution when a new species has to be genetically explored in the wild (Galtier et al., 2009). In *An. dirus* s.l. both COI and COII gene sequences of mtDNA were used to estimate population genetics structure and demographic history (Morgan et al., 2010; O’loughlin et al., 2008; Walton et al., 2000).

2.16.2.2 Microsatellite loci

Nuclear simple tandem repetitive DNA, more commonly known as microsatellite DNA (msats), are present widely throughout eukaryote genomes, have high mutation rates, obeys Mendelian inheritance and follows apparently simple mode of evolution (Zhang and Hewitt, 2003), have been developed and used primarily to
estimate population structure and gene flow among malaria vectors. Contemporary gene flow among Anopheles populations are generally assessed using microsatellites because, due to their high mutation rate, they can detect differentiation even in weakly structured species (Reviewed in Loaiza et al., 2012). However, large-scale geographical differences are better detected using mtDNA polymorphisms because, due to maternal inheritance, $N_e$ for mtDNA is a quarter of that of nuclear markers and fix new alleles faster than nuclear markers, and genetic drift may produce a strong signal of spatial population processes (Ballard & Whitlock, 2004). Several studies indicate, though, that microsatellite loci, having a larger $N_e$ and longer coalescent time than mtDNA, also depict signals of demographic history rather than current genetic exchange between Anopheles populations (Lehmann et al., 1997; Morgan et al., 2010; Ndo et al., 2010). In An. dirus s.l. 11 microsatellite markers were isolated and characterized by Walton et al., (2000b) and further used to study population structure and demographic history among three species (sp A, sp C and sp D) of the complex (Morgan et al., 2010; Walton et al., 2001). However, concerns have been raised to the potentiality of mutations in the flanking regions of microsatellite loci causing inaccuracies in determination of the fragment sizes (Walton et al., 1998). Zink and Barrowclough, (2008) have pointed out that microsatellites are challenging to study due to their complex mutation process; microsatellite mutation rates may vary across loci and across alleles within the same locus, and comparative analyses across taxa are often made impractical due to mutation bias and different rates of evolution among lineages (Rubinsztein et al., 1999). Extremely high evolutionary rates, size homoplasy and other genotyping artefacts (e.g. null alleles and allele dropouts) also pose severe limitations to the analysis of microsatellites (Rubinsztein et al., 1999).

2.16.2.3 Nuclear sequence based markers

Since phylogeographic studies enquires about the estimates of pattern and process of genetic variation, for many valid reasons, mtDNA is likely the marker of choice to estimate the pattern of population history due to its rapid coalescence time. In contrast, for robust estimates of demographic variables, multiple independent loci are needed to reduce the process error associated with coalescence (Zink and Barrowclough, 2008). In fact, mtDNA estimates of population size ($N_e$), gene flow ($mN_e$), population growth (e.g. mismatch distributions), and divergence times are
associated with large confidence intervals (Edwards & Beerli, 2000), and the solution to that problem is to analyse multiple nuclear genes to reduce the coalescent process error.

Various classes of nuclear sequence based markers are available and being developed for both intraspecific and interspecific phylogeography study (Brito and Edwards, 2009; Zhang and Hewitt, 2003). The development of conserved exon - primed - intron - crossing (EPIC) PCR primers to facilitate amplification of low- or single-copy nuclear loci has promoted population genetics studies in past decades (Sunnucks, 2000; Zhang and Hewitt, 2003). Although it is easier to develop conserved PCR primers for mtDNA, the case is very complicated for nuDNA except for a limited success achieved with primers amplifying ribosomal ITS region (Hills and Dixon, 1991). However some nuclear markers based on coding regions, introns and SNP were successfully applied to detect population genetic and phylogeographic structure in a number of Anopheles species (Dixit et al., 2010; Mirabello and Conn, 2008; Morlais et al., 2004; Reidenbach et al., 2009; Rona et al., 2010). In An. dirus complex, Morgan et al., (2010) developed 3 nuclear coding markers viz, est6, ninaE and hsp82 to assess the gene flow pattern among three species of the complex.

The primers to amplify nuclear sequence based markers for intraspecific studies need to satisfy the following requirements:

(i) they should be evolutionarily well conserved across different taxonomic groups, (ii) their amplicons should be of reasonable size, (iii) target sequences should be highly variable at the intraspecific level, and (iv) target sequences should be single-copy or low-copy in the nuclear genome. However, several processes of molecular evolution of nuclear DNA counteract the search for such nuclear primers. These include gene duplication or amplification, production of pseudogenes, intron losts, sliding or size change, etc. (Zhang and Hewitt, 2003). Nuclear DNA has a lower mutation rate than mtDNA, which minimizes the back and parallel mutations that can reduce phylogenetic resolution in mtDNA data. Hence, in data with low homoplasmy like nuDNA, even a single fixed difference can provide a statistically strong result at the intraspecific level, regardless of bootstrap support (Harris and Hey, 1999). Although nuDNA sequence based markers have advantages in estimating population demographic processes, issues like recombination, selection (non-neutrality), heterozygosity, insertion/deletion polymorphism, low divergence and polytomy, gene-specific variation in rate and
history, PCR and sequencing difficulty are always associated with it and need to be critically cared to get reliable estimates on phylogeography of a species (Zhang and Hewitt, 2003). However, there are a number of statistical methods (such as four-gamate test) to estimate the minimum number of recombination events for a given nuclear DNA sequence data (Hudson and Kaplan, 1985). Another technical challenge in determination of proper allelic phase can also be handled by both probabilistic or computational methods (such as using PHASE program) (Stephens et al., 2001) and empirical methods (cloning).

2.17. Computational analysis in molecular genetics

As reviewed earlier, although there is a need to use varied and high resolution molecular markers to assess the pattern and process of genetic variations of a species over its geographical range of distribution, proper analytical methods are also important to analyse and document these genetic data. These analytical methods are implemented in various software packages and computer programs, which sometimes rely on massive computations. There are lots of such programs used for data analysis in population genetics and phylogeography studies, whose principles, statistics they compute, their assumptions and their limitations are reviewed in Excoffier and Heckel (2006) and Kuhner (2009). Based on the task and analyses performed, different softwares and computational programs can be broadly divided into the following three categories (Excoffier and Heckel, 2006).

Multi-purpose programs, which compute descriptive statistics (such as degree of heterozygosity, number of polymorphic loci, theta estimates etc.), population comparison and genetic structures and demographic inference and neutrality tests within and between populations. Software packages and computer programs such as Arlequin, DnaSP, FSTAT, GDA, Genepop, MEGA, MSA can be grouped into this category. The assumptions made by the multipurpose packages are difficult to summarize, because they implement various computations that are based on different methodologies. The computations of most summary statistics or tests of selective neutrality that are computed from diploid samples of DNA sequences assume a known gametic phase.

Individual-centred programs represent a recent development of population genetics, where the main focus of the analysis is on individuals and the very recent
history of a population. BAPS, GeneClass, Geneland, Structure are some of the programs that can be categorized under individual-centred program that are used to detect recent immigrants among samples analyzed at various multi-allelic markers or allocate individuals to predefined populations (as in GeneClass) or to 'virtual' populations (as in BAPS, Geneland and Structure), for which allele frequencies are also iteratively estimated. Most individual-centred programs assume Hardy-Weinberg equilibrium (HWE) within ‘populations’ and assume that loci are unlinked (except for Structure program).

Specialized programs, generally intend to infer some demographic parameters such as effective population size ($N_e$) or gene flow ($Nm$) rates from genetic data under a specific evolutionary scenario. Computer programs like BATWING, IM, IMa, IMa2, BEAST, LAMARC and Migrate-n are mostly used in phylogeographical studies to estimate parameters for divergence times, $N_e$ and gene flow among populations. Most of these programs use a Bayesian framework for parameter inference which facilitates the incorporation of some prior knowledge on the parameters of interest, often leading to improved estimations. In most of these specialized programs a strict evolutionary model is assumed and the genetic data used are assumed to be free from recombination. Constant population size and constant migration rate among the studied populations are also another assumption of these types of programs. Some of the software packages and computer programs commonly used in analysing genetic data for population genetics and phylogeographical inferences are listed in Table 2.3.
### Table 2.4: List of computer programs used for population genetics and phylogeography study
(Adapted and modified from Excoffier and Heckel, 2006)

<table>
<thead>
<tr>
<th>Name</th>
<th>Version</th>
<th>Platform</th>
<th>Data handled</th>
<th>Data format</th>
<th>Analytical applications</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arlequin</td>
<td>3.5</td>
<td>Win/Linux/Mac</td>
<td>DNA, SNP, STR, MULT, FREQ</td>
<td>Specific, GENEPOP</td>
<td>1, 2, 3, 4, 5, 7, 8</td>
<td>Excoffier and Lischer, 2010</td>
</tr>
<tr>
<td>DnaSP</td>
<td>5.0</td>
<td>WIN</td>
<td>DNA, SNP</td>
<td>MEGA, NEXUS, NBRF/PIR, FASTA, PHYLIP,</td>
<td>1, 3, 4, 5, 7, 8</td>
<td>Librado and Rozas, 2009</td>
</tr>
<tr>
<td>FSTAT</td>
<td>2.9.3.2</td>
<td>DOS/Win</td>
<td>STR, MULT</td>
<td>Specific, GENPOP</td>
<td>1, 2, 3, 7, 8</td>
<td>Goudet, 1995</td>
</tr>
<tr>
<td>GDA</td>
<td>1.1</td>
<td>Win/Mac</td>
<td>AFLP, MULT</td>
<td>NEXUS, BIOSYS</td>
<td>1, 2, 3, 7</td>
<td>Lewis and Zaykin, 2001</td>
</tr>
<tr>
<td>Genepop</td>
<td>4.1</td>
<td>Win/Mac/Linux</td>
<td>STR, MULT</td>
<td>Specific</td>
<td>1, 2, 3, 7</td>
<td>Rousset, 2008</td>
</tr>
<tr>
<td>MEGA</td>
<td>5.0</td>
<td>Win/Mac/Linux</td>
<td>DNA, DIST</td>
<td>specific, CLUSTAL, NEXUS, PHYLIP, GCG, FASTA, NBRF/PIR</td>
<td>1, 3, 4</td>
<td>Tamura et al., 2011</td>
</tr>
<tr>
<td>MSA</td>
<td>4.0</td>
<td>DOS/Mac/Linux</td>
<td>STR, MULT</td>
<td>EXCEL</td>
<td>1, 3</td>
<td>Dieringer and Schlötterer, 2003</td>
</tr>
<tr>
<td>BAPS</td>
<td>5.3</td>
<td>Win/Mac/Linux</td>
<td>MULT</td>
<td>Specific, GENPOP</td>
<td>3, 9</td>
<td>Corander et al., 2008</td>
</tr>
<tr>
<td>GeneClass</td>
<td>2.0</td>
<td>Win</td>
<td>MULT</td>
<td>GenePop, FSTAT, GENETIX</td>
<td>9</td>
<td>Piry et al., 2004</td>
</tr>
<tr>
<td>Geneland</td>
<td>4.0.0</td>
<td>Win/Mac/Linux</td>
<td>SNP, AFLP, STR, MULT</td>
<td>Specific</td>
<td>3, 9, 10</td>
<td>Guillot et al., 2005</td>
</tr>
<tr>
<td>Structure</td>
<td>2.2</td>
<td>Win/Mac/Linux</td>
<td>MULT</td>
<td>Specific</td>
<td>2, 9</td>
<td>Pritchard et al., 2000</td>
</tr>
<tr>
<td>BATWING</td>
<td>DOS/Mac/Linux</td>
<td>SNP, STR</td>
<td>Specific</td>
<td></td>
<td>5, 6, 7, 8, 9</td>
<td>Wilson et al., 2003</td>
</tr>
<tr>
<td>IM, IMa, IMa2</td>
<td>DOS/Mac/Linux</td>
<td>DNA, STR, hapSTR</td>
<td>Specific</td>
<td></td>
<td>5, 6, 7, 8, 9</td>
<td>Hey, 2010; Hey and Nielsen, 2004, 2007</td>
</tr>
<tr>
<td>BEAST</td>
<td>1.7.1</td>
<td>Win/Mac/Linux</td>
<td>DNA, MULT</td>
<td>NEXUS</td>
<td>5, 6, 7, 8</td>
<td>Drummond et al., 2012</td>
</tr>
<tr>
<td>LAMARC</td>
<td>2.1.6</td>
<td>DOS/Mac/Linux</td>
<td>DNA, SNP, STR</td>
<td>Specific, PHYLIP, Migrate</td>
<td>5, 6, 7, 8, 9</td>
<td>Kuhner, 2006</td>
</tr>
<tr>
<td>Migrate-n</td>
<td>3.2.17</td>
<td>DOS/Mac/Linux</td>
<td>DNA, SNP, STR, MULT</td>
<td>Specific, PHYLIP</td>
<td>5, 6, 7, 8</td>
<td>Beerli and Felsenstein, 2001</td>
</tr>
</tbody>
</table>


AFLP: amplified fragment length polymorphism; DNA: DNA-sequence data; DIST: distance matrix; FREQ: frequency data (e.g., allele frequency); hapSTR: linked SNP and STR (short tandem repeat) markers; MULT: multi-allelic markers, for which no particular mutation model is assumed; RAPD: random amplified polymorphic DNA; SNP: single nucleotide polymorphism; STR: also called microsatellites; Win: Windows.