Vector Evolutionary Genomics

Fine-scale Evolutionary Genetic Insights into *Anopheles gambiae* X-chromosome

Understanding the genetic architecture of individual taxa of medical importance is the first step for designing disease preventive strategies. To understand the genetic details and evolutionary perspective of the model malaria vector, *Anopheles gambiae* and to use the information in other species of local importance, we scanned the published X-chromosome sequence for detailed characterization and obtain evolutionary status of different genes. The telocentric X-chromosome contains 106 genes of known functions and 982 novel genes. Majorities of both the known and novel genes are with introns. The known genes are strictly biased towards less number of introns; about half of the total known genes have only one or two introns (Fig. 51). The extreme sized (either long or short) genes were found to be most prevalent (58% short and 23% large). Statistically significant positive correlations between gene length and intron length as well as with intron number and intron length were obtained signifying the role of introns in contributing to the overall size of the known genes of X-chromosome in *An. gambiae*. We compared each individual gene of *An. gambiae* with 33 other taxa having whole genome sequence information. In general, the mosquito *Aedes aegypti* was found to be genetically closest and the yeast *Saccharomyces cerevisiae* as most distant taxa to *An. gambiae* (Fig. 52). Further, only about a quarter of the known genes of X-chromosome were unique to *An. gambiae* and majorities have orthologs in different taxa (Fig. 53). A phylogenetic tree was constructed based on a single gene found to be highly orthologous across all the 34

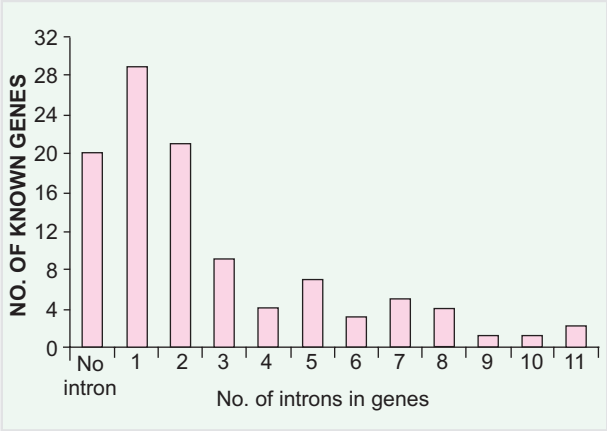

Fig. 51: Distribution of *An. gambiae* X-chromosome known genes according to the number of introns

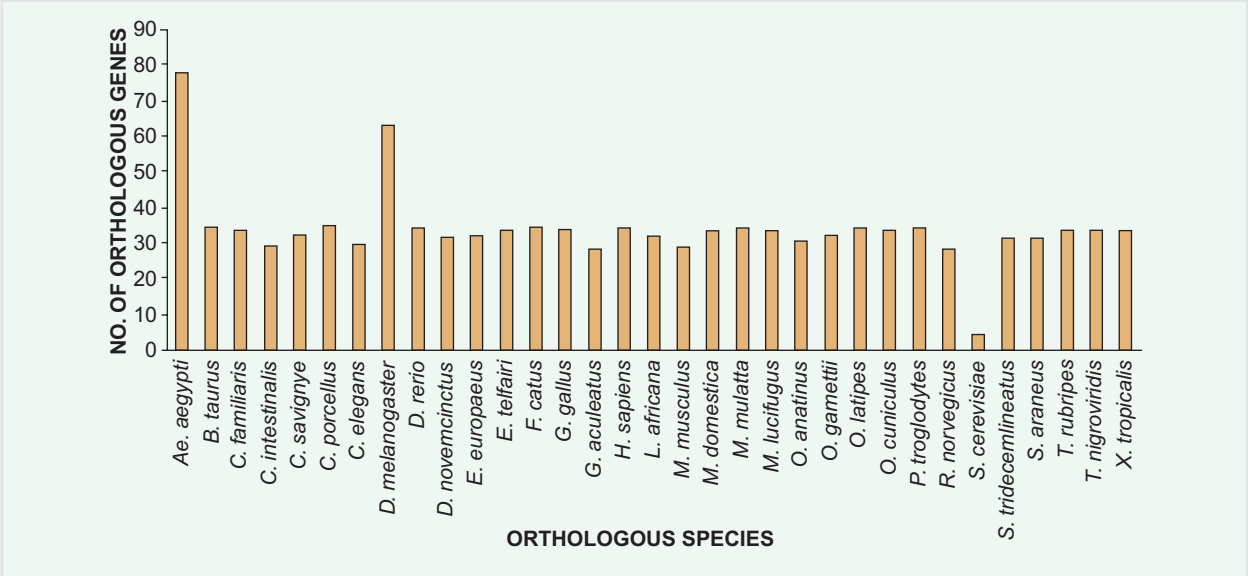

Fig. 52: Distribution of different taxa showing number of shared genes with *An. gambiae*

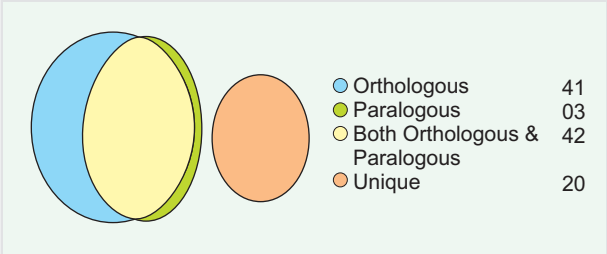

Fig. 53: Distribution of different gene types (based on homology prediction) in X-chromosome of *An. gambiae*

taxa (Fig. 54). Evolutionary relationships among 13 different taxa were inferred which corroborate the previous and present findings on genetic relationships across various taxa.

An Evolutionary Genetic Insight of Insecticide Resistance Gene Families in *Anopheles gambiae*

Insecticide resistance mechanism developed by malaria vector species is one of the major obstacles in vector control strategies and disease control and is known to be genetically controlled. Three major gene families (*Cyp*, *Gst* and *Coe*) are determined for the insecticide resistance mechanisms that encode various proteins to metabolize endogenous as well as exogenous compounds in insects. Since, insecticides are in excessive use (and misuse) in the

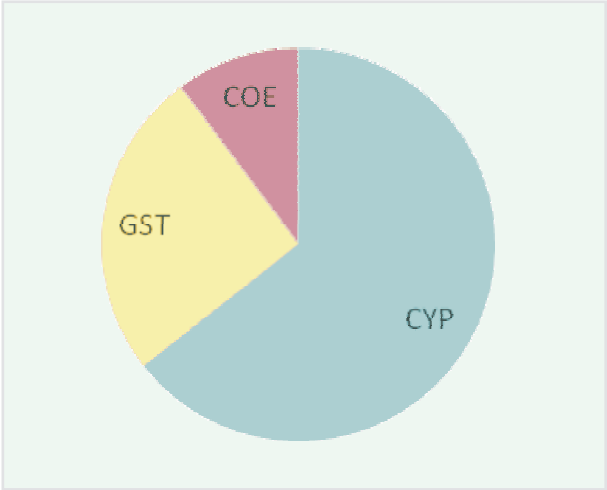

Fig. 55: Contribution of each insecticide resistance gene family in genome of *An. gambiae*

field putting enormous pressure for the evolution of more suitable and efficient insecticide resistance mechanisms in insects, it is important to have fair knowledge on how genetic basis of insecticide resistance genes evolve in these three different gene families. This is enormous importance to malaria research, as vector control and thus to control malaria has been grossly hampered by emergence and evolution of insecticide resistance in malaria vectors. We herewith studied the contribution of all three insecticide resistance gene families by utilizing

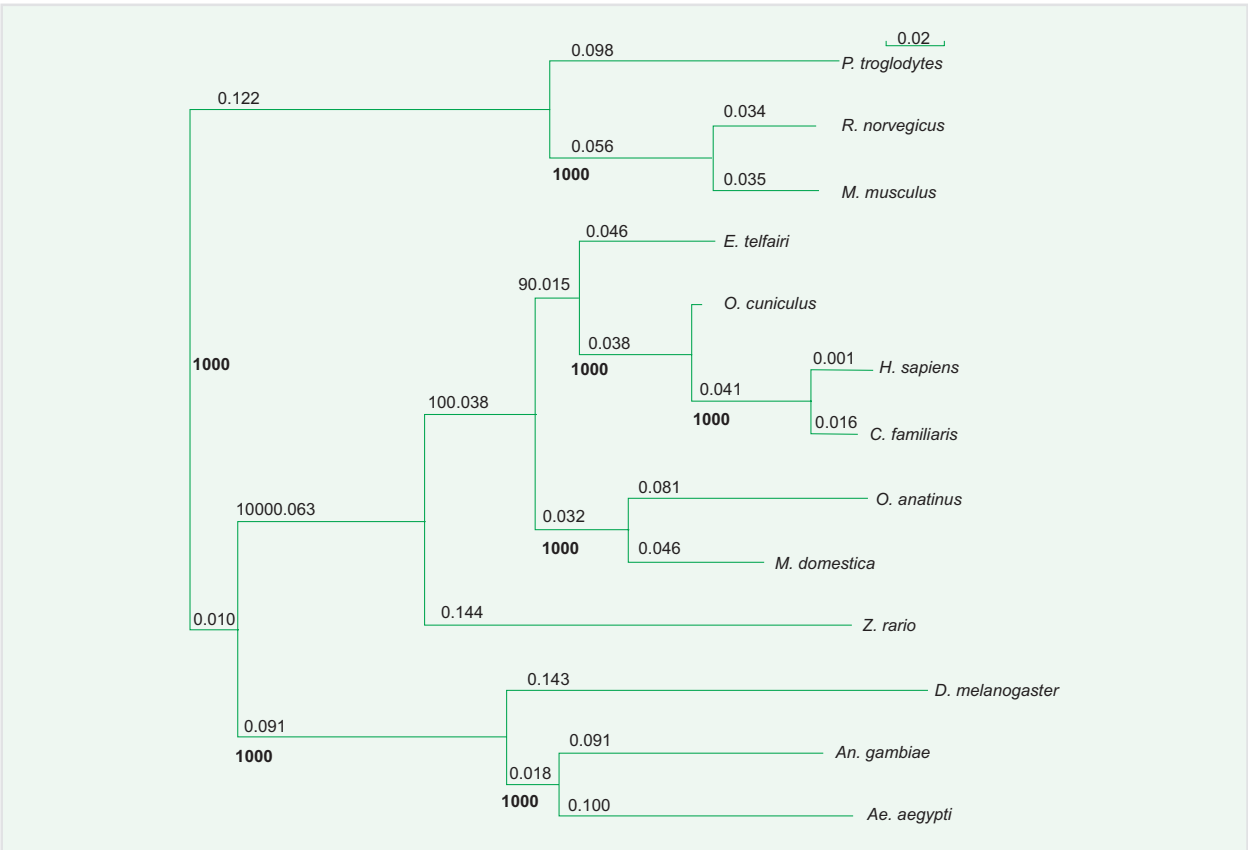

Fig. 54: Phylogenetic tree with bootstrap values (in bold font) and branch length (in normal font) in 13 different taxa

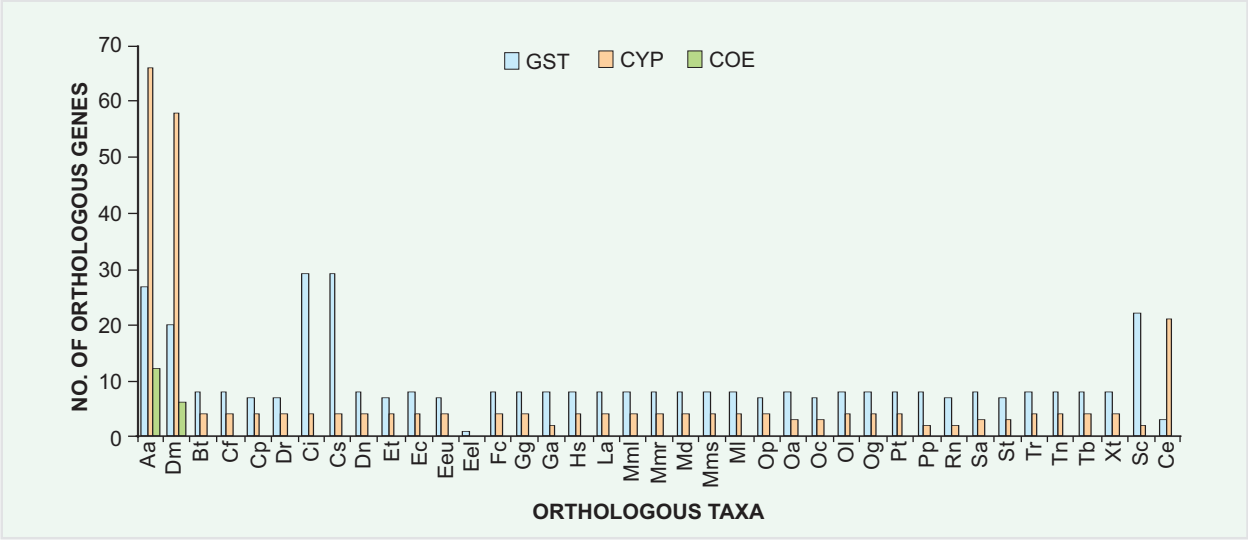

Fig. 56. Orthology of genes of insecticide resistance gene family *Cyp, Gst* and *Coe* to 39 different taxa

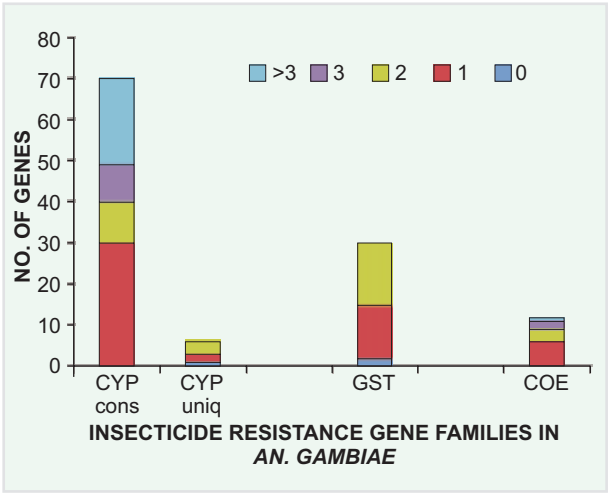

Fig. 57. Number of genes of each insecticide resistance gene family of *Anopheles gambiae* classified on the basis of intron number

genome sequence information of African malaria vector *An. gambiae* (Fig. 55). The pattern of conservation of insecticide resistance genes across various taxa (Fig. 56) and organization of introns in the genes (Fig. 57) have been determined to infer the present evolutionary status of the three gene families. We mapped each individual gene of all three insecticide resistance gene families in all chromosomes (Fig. 58) and measured distribution of genes across chromosomes (Fig. 59). Further, phylogenetic relationships were reconstructed within each gene families (Figs 60–62) and correlated the location of the genes with their position on chromosomes that provide the evidence for mode of expansion of gene families in genome. The results, as a whole in different gene families provide clues to evolutionary mechanisms evolve differently in each gene families of vectors. The knowledge on the

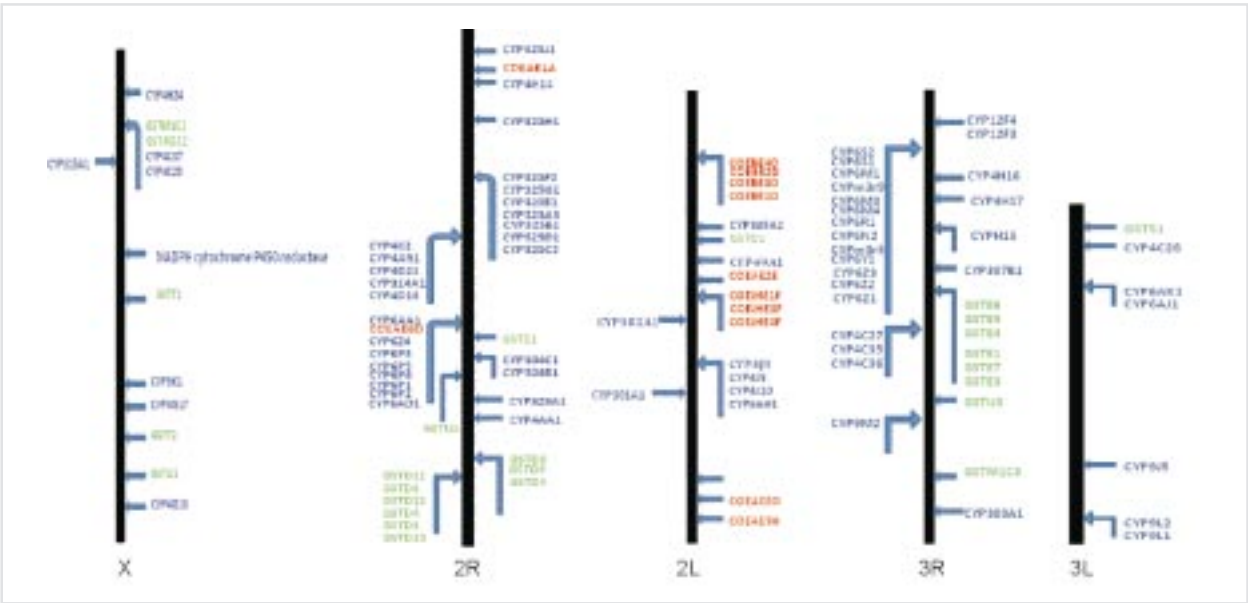

Fig 58: Location of different genes of insecticide resistance gene families *Cyp, Gst* and *Coe* on the chromosomes of *An. gambiae*

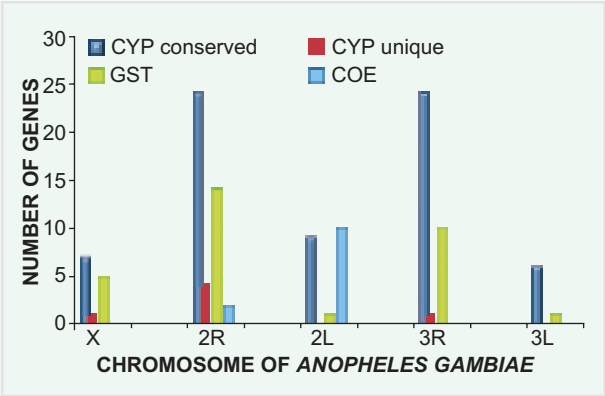

Fig. 59: Chromosomal distribution of genes of three gene families *Cyp*, *Gst* and *Coe* in *An. gambiae*

evolutionary architecture of these three insecticide resistance gene families in *An. gambiae* will be helpful to understand the genetic basis of resistance in malaria vectors of local and focal importance.

Phylogenetic Reconstruction of Indian Malaria Vectors using Multilocus DNA Sequences

Understanding the evolutionary status of closely related species of health importance is the first step in disease management. Malaria is one of the deadliest vector borne diseases, and the taxonomic status of several malaria vectors of *Anopheles* genus and Celia subgenus has been well-documented based on morphological features and gene sequences of ribosomal and mitochondrial regions. However, phylogenetic studies based on multilocus nuclear DNA sequences in *Anopheles* are still in dearth. Moreover, no concrete information on molecular phylogenetic status of major Indian malaria vector species is available till date; though India majorly contributes to the global malaria vector species diversity. We screened the available whole genome sequence information of *An. gambiae* to find six orthlogous nuclear genetic regions and sequenced these regions in six species of Indian malaria vectors. The sequence information of seven species of *Anopheles* (six Indian, and *An. gambiae*) was utilized to reconstruct phylogenetic trees for each individual genetic regions and the time of divergence among these species was calculated based on COII gene sequences. Although tree topologies with COII, ITS2 and one of the nuclear gene responsible for Carboxyl Esterase (*Coe*) genes mirror-imaged each other, for no other genetic region, similar tree topologies in all the seven species were observed (Fig. 63 a–h). Although Indian malaria vectors show gene-specific tree topologies which might be due to differential function-dependant evolutionary cons-traints, in principle, the reconstructed phylogenetic status follow the pattern based on morphology and that of the COII and ITS2 genetic regions. These results on one hand provide evidence for robustness of COII and ITS2 for phylogenetic inference in closely related species, on

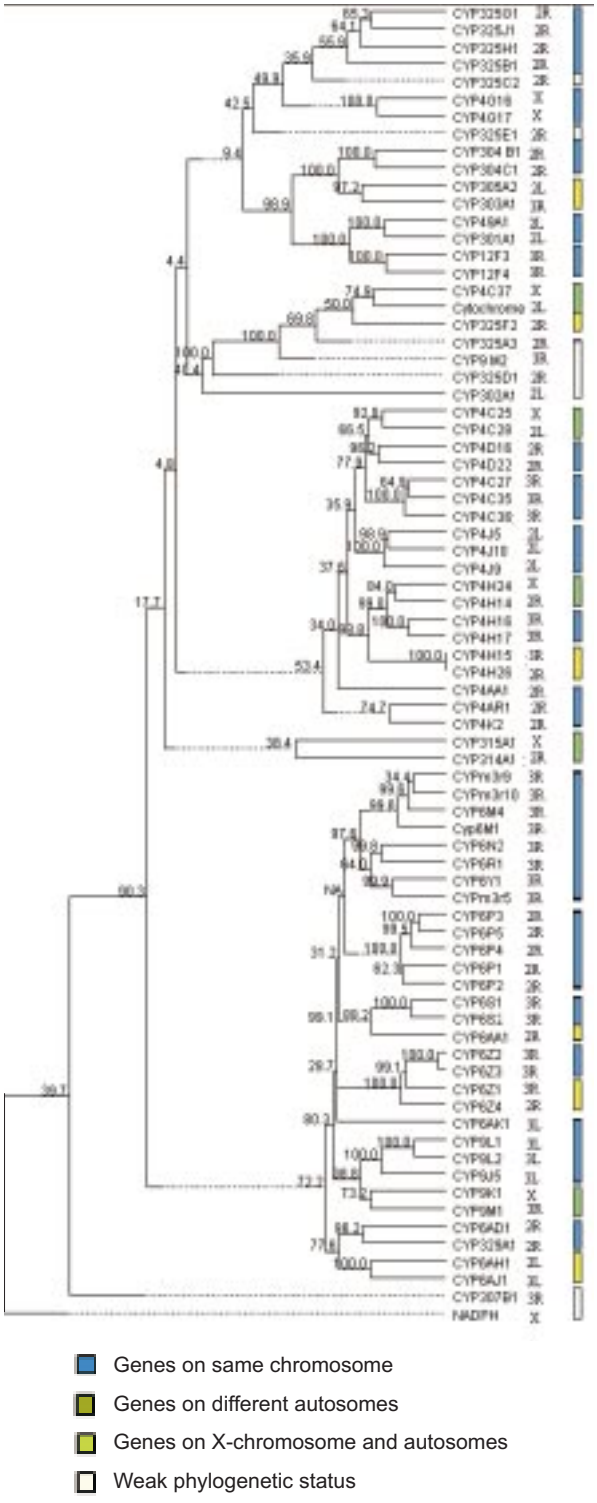

Fig. 60: Neighbor-Joining phylogenetic tree among the genes of *Cyp* gene family

the other hand signifies the utility of multilocus DNA sequence information in dissemi-nating gene-specific from global tree topologies. The divergence times calculated between species further corroborates the earlier theories about the major radiation of species belonging to *Cellia* subgenus in cretaceous period (Fig. 64). The information could be utilized in malaria management not only in India but also in places where some or all of these species are endemic.

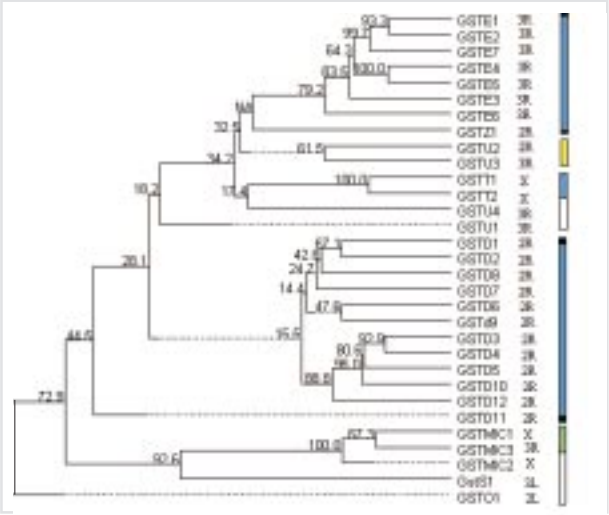

Fig. 61: Phylogenetic relationship based on neighbor-Joining method between genes of *Gst* gene family

Population Genomics of Indian *Anopheles minimus*

Malaria spreads through mosquitoes, belonging to genus *Anopheles* and ability to transmit malaria is uniquely present within this genus, with only ~30 out of 500 species, are major vectors. In addition to this,

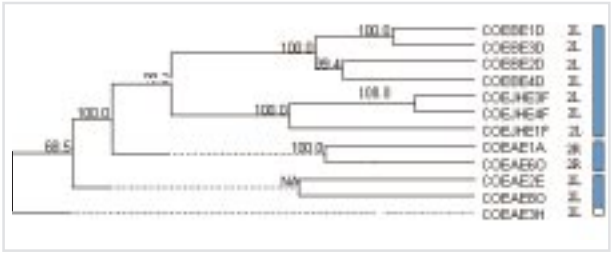

Fig. 62: Neighbor-Joining phylogenetic tree showing phylogenetic relationship among all genes of *Coe* gene family

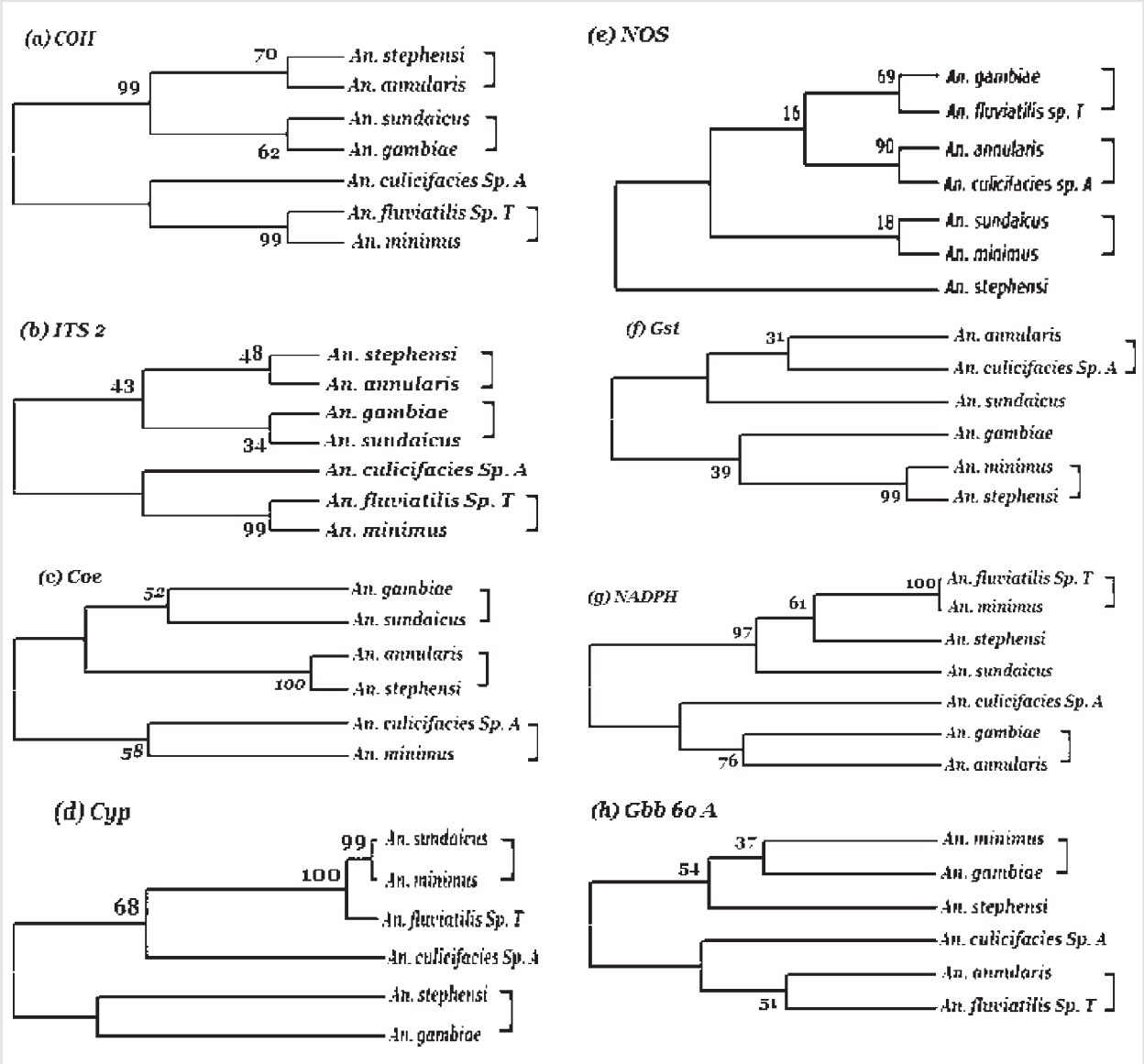

Fig. 63: The consensus tree resulting from maximum parsimony analysis based on the (a) mitochondrial Cytochrome oxidase II (COII) gene sequences (b) nuclear non-functional *ITS2* sequences (c) *Coe* gene sequences (d) *Cyp 450 4 G 16* gene sequence (e) *NOS* gene sequences (f) *Gst* gene sequence data (g) NADPH gene sequence data (h) *Gbb 60 A* gene sequences data. Tree was obtained by Max-mini branch and bound search option of MEGA 4.1.

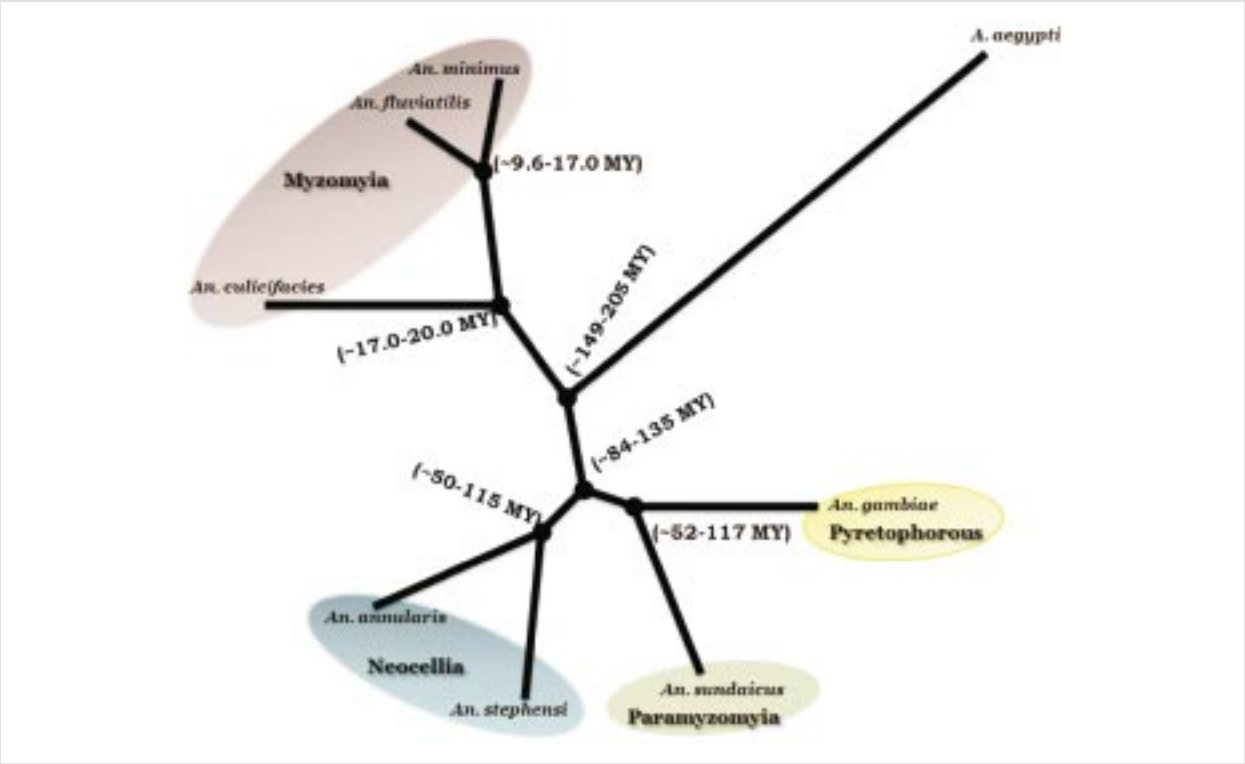

Fig. 64: Phylogenetic status and divergence time estimations of six Indian malaria vectors. Values in parentheses indicate the approximate divergence time between two vector species. Divergence time between *Anopheles* and *Ae. aegypti* (~145–200 MY) has been taken as the calibration point for estimation of divergence times

differential vectorial capacity within members of the same *Anopheline* species complex also exits. The possible reason behind this anomaly is the difference in the genetic makeup and population history of different members of a species complex. Without the knowledge of these aspects of a vector population, one cannot devise new measures to control malaria spread. However, till now very few studies have been done to explore the genetic structure and population history of vector species populations. To this respect we very recently have initiated the population genetic structure and demographic history of one of the main malaria vectors of our country, i.e. *An. minimus* species A. We have applied a comparative genomic approach in designing the putatively neutral nuclear DNA markers in the genome of this Indian vector species by taking the available genome sequence of *An.*

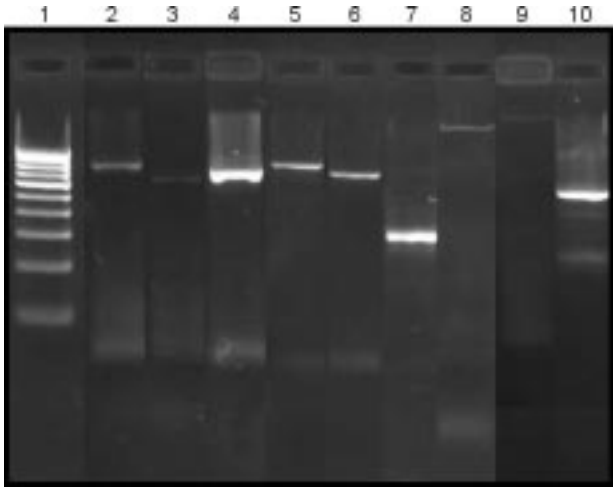

Fig 65: PCR amplified DNA fragments in nuclear genome of *An. minimus*

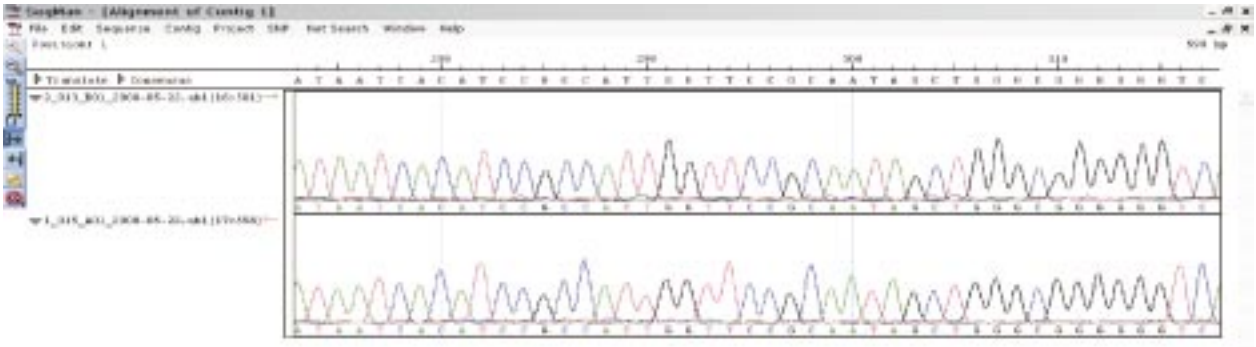

Fig. 66: Sequence chromatogram of a DNA fragment in Indian *An. minimus*

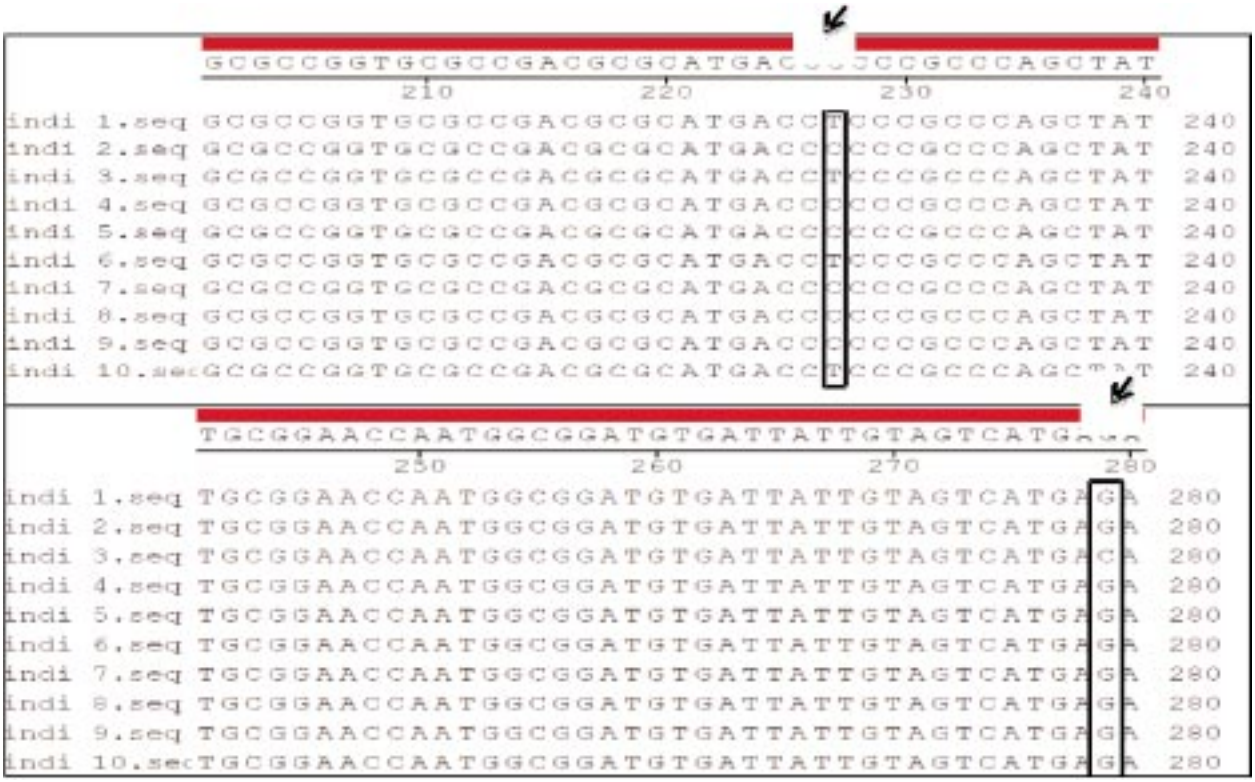

Fig. 67: DNA sequence alignment and SNP detection in Indian *An. minimus*

gambiae as reference (Fig. 65). The sequences for these markers have been generated in the lab (Fig. 66). Till now we have developed and studied three fragments as putatively neutral fragments in the genome of *An. minimus* taken from five different population samples of this species from the northeastern states of India. With the help of computer programmes we have detected 23 single nucleotide polymorphisms in the genome of *An. minimus* (Fig. 67). Neucleotide diversity, Tajima's D and Fu and Li's D were calculated for all the fragments. Except P22 fragment (for Moregaon population), statistical neutrality was observed in majority of the cases implying that all the three fragments could be considered as putatively neutral markers for population genetic studies in *An. minimus*.

❑