

Fine-scale evolutionary genetic insights into *Anopheles gambiae* X-chromosome

Hemlata Srivastava¹, Jyotsana Dixit¹, Aditya P. Dash², Aparup Das

Evolutionary Genomics and Bioinformatics Laboratory, National Institute of Malaria Research, Sector 8, Dwarka, New Delhi-110 077, India. ²Present address: World Health Organization, Southeast Asian Regional Office, New Delhi, India. ¹Equal contributions. Email: aparup@mrcindia.org

Received 28 April 2009; revised 10 May 2009; accepted 15 May 2009.

ABSTRACT

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 detail 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. 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*. Further, only about a quarter of the known genes of X-chromosome were unique to *An. gambiae* and majorities have orthologs in different taxa. A phylogenetic tree was constructed based on a single gene found to be highly orthologous across all the 34 taxa. Evolutionary relationships among 13 different taxa were inferred which corroborate the previous and present findings on genetic relationships across various taxa.

Keywords: *Anopheles gambiae*; Comparative Genomics; Evolution; Malaria; Orthologous Genes; X-chromosome

1. INTRODUCTION

Determination of genetic architecture of different taxa in a vector borne disease model, helps not only in understanding the genetic pattern of host-parasite-vector interaction but also adds in devising methods for control measures. Further, characterization of different genes in the entire chromosome leads to identification of novel genes of essential functions and evolutionary process that governs these genes in populations. These determinations of genetic architecture should start with deep understanding and evolutionary inference of each individual gene of known function at the chromosomal level. The detail knowledge on the relative size of the genes [1], differential compositions of coding and non-coding elements in each gene [2] and contribution of non-coding DNA to the average length of the gene [3] could easily be evaluated with such kind of studies. This is further important when scanning is performed on chromosome-to-chromosome basis, so that differential genetic composition in each chromosome of a species can be compared [4]. Further, comparing genes among different taxa exploits both similarities and differences of different organisms to infer how Darwinian natural selection might have acted upon on these elements in the course of evolution. Considering the genomes as “bags of genes” and measuring the fraction of orthologs shared between genomes could provide vital information on the evolutionary history of the genes [5]. Also, if any particular gene is found to be conserved in many organisms, reconstruction of phylogeny among these organisms is possible. However, such kinds of studies are possible only when adequate genome information is at hand. Fortunately, many organisms have been fully sequenced in recent years providing opportunities to fine-scale understanding of the genetic architecture of species of medical and agricultural importance and comparison across species and taxa [6].

To this respect, malaria is a devastating disease with global cases of about 300 to 500 million infections per year and deaths of about one and half millions [7]. Global efforts to eradicate malaria failed immaturely and scientist and policy makers are now focusing on the control of this disease. However, emergence and spread of drug-resistant parasites and insecticide-resistant vectors have seriously hampered the efforts and put new challenges to tackle with the situation [8]. This situation invites a close and deep genetic understanding of both parasites and vectors and devise new methods for malaria control.

The whole genome sequence information of the mosquito *Anopheles gambiae* [9], the principal vector of malaria in Africa is available in the public domain. However, the vector species are not all the same across different malaria endemic zones in the globe as different other species of the genus *Anopheles* are of local importance. Since controlling the malaria vector is one of the finest strategies to control malaria, understanding the genetic composition of different endemic *Anopheles* species in different localities is the need of the hour for development of effective malaria control strategies. Keeping in view that the whole genome sequence information is only available for one of the malaria vector *An. gambiae*, utilization of such information might lead to understanding the genetics of vector potentiality, insecticide resistance, etc. and extend the information to other species of local and focal importance. In addition, due to lack of genome information in other *Anopheles* species, the information from *An. gambiae* could be utilized in designing genetic markers for evolutionary studies in genes and populations of vectors in the malaria endemic zones of the globe.

We herewith utilize the whole genome sequence information of *An. gambiae* to characterize the whole X-chromosome for different gene compositions and fine-scale study of each individual genes of known function. We performed homology searches of each of the known genes of *An. gambiae* X-chromosome in 33 taxa with published whole genome sequences and also constructed phylogenetic tree. The results not only provide detail understanding on differential compositions of genetic elements in *An. gambiae* X-chromosome, but also would help in developing genetic markers to study genetic diversity and population histories of other species of *Anopheles* of local importance.

2. MATERIAL AND METHODS

The *An. gambiae* genome comprises of 3 pairs of chromosomes, 2 pairs of autosomes and a pair of sex chromosome. Whole genome sequence information is available at the public domain for the pest strain of *An. gambiae* [9]. We used the Ensemble web database (www.ensembl.org) from release 45-June 2007, to re-

trieve genetic information on the telocentric X- chromosome. We started our scanning for different genes from one end of the X-chromosome and proceeded till we reached the end. We looked for genes that have known functions (known genes) and also genes that are completely new (novel genes) following the classifications provided at the Ensemble database. Due to functional importance, we deeply characterized only the known genes leaving apart the novel genes. For the convenience of further analysis, we classified the known genes based on length of nucleotide bases as, Class 1 (0 to 1 kb); Class 2 (1-2 kb); Class 3 (2-3 kb); Class 4 (3-4 kb); Class 5 (4-5 kb) and Class 6 (above 5 kb). The composition of different genes in the *An. gambiae* X- chromosome was determined as per information in the Ensemble web database. Further, information on 33 other taxa with whole genome sequence information was also available at the Ensemble web database. We utilized this information to infer X-chromosome genes of *An. gambiae* having orthologs (orthologs are genes derived from single ancestral gene in last common ancestor of compared species) and paralogs (paralogous genes develop by gene duplication in the similar lineage) across 33 different taxa. The genes of *An. gambiae* with no ortholog or paralog were considered to be unique genes to this species. Three criteria were followed to define the candidate orthologous genes suggested by [8] in the Ensemble web database. First, the sequences of genes should have highest level of pair-wise identity when compared with genes in the other genome. Second, pair-wise identity should be significant (E, the expected fraction of false positives should be smaller than 0.01) and third, the similarity extends to at least 60% of one of the gene. We followed similar procedures to classify the orthologous genes in *An. gambiae* X-chromosome. Out of the many orthologs found, one particular gene was found to be present in all the 34 different taxa (included *An. gambiae*) presently studied. We constructed an un-rooted neighbor-joining (NJ) tree to infer the evolutionary status of different taxa at this conserved gene (AGAP001043). However, due to high sequence dissimilarity, only 13 taxa could be utilized for a meaningful phylogenetic tree construction. Length of each branch and bootstrapped values for each internal node were also estimated using VEGA ZZ software downloaded from internet (<http://www.ddl.unimi.it/vega/index.htm>). For all statistical analyses, the free version of 'analyze-it' a Microsoft Excel add-in was used.

3. RESULTS

Scanning of the whole X chromosome of *An. gambiae* revealed the presence of 1088 genes, out of which 982 were novel and 106 were genes of known functions. Due to functional relevance, the known genes were further analyzed. These genes were classified based on size (see materials and methods). Out of the 106 known genes,

most (62 genes, 58%) were small and thus fall under Classes 1 and 2. Thus, majority of the known genes in *An. gambiae* X-chromosome are small in size (Figure 1). Only 18% of the total known genes falls under Classes 3, 4, and 5, whereas 23% comes under Class 6 (more than 5 kb). Thus, the distribution of known genes in *An. gambiae* seems to be quite uneven, as small and large sized genes constitute 81% of the total known genes in the *An. gambiae* X-chromosome (Figure 1).

Since genes in the eukaryote genome are often found to bear introns (non-coding part of the gene, flanked on each side by the coding parts) and considering *An. gambiae* as a higher eukaryote, we determined the distribution of exons (coding part of the gene) and introns of each known gene (Figure 2). The distribution of genes with different number of introns is shown in Figure 2. It is interesting to note that almost two-third of the known genes (79 genes, 74.52%) have either no or very less number (maximum of three) of introns. Genes having more than three introns contribute to only 17% of the total known genes of X-chromosome of *An. gambiae*. Further, we looked for size of each intron and exon in each gene and calculated the average intron and exon length and their ratio (Figure 3). The average ratio of exon to intron was higher in genes with less number of introns (Figure 3), as compared to genes with more number of introns. Thus, it seems that the average length of introns in a gene increases with the increase in number of introns. In order to test this hypothesis, we calcu-

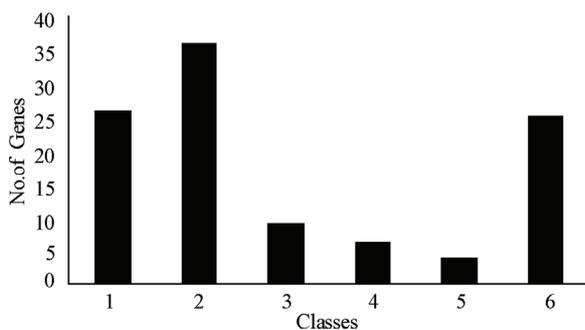


Figure 1. Classification of known genes of *An. gambiae* X-chromosome based on size (nucleotide base pair).

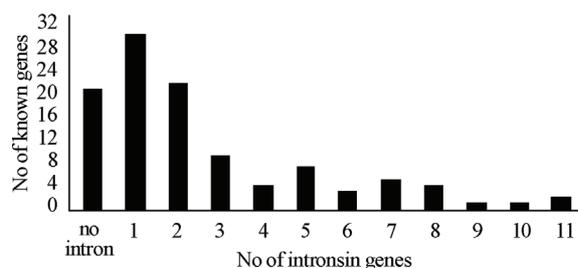


Figure 2. Distribution of *An. gambiae* X-chromosome known genes according to the number of introns.

lated Pearson's correlation coefficient (r) which was found to be positive and highly statistically significant ($r=0.99$, $P<0.0001$). Moreover, in order to test the hypothesis if the accumulation of introns has considerably contributed in increasing the length of the gene in general, we calculated r value between intron length and gene length which was found to be positive and highly statistically significant ($r=0.49$, $P<0.0001$) as well. Thus, it is clear that introns play a major role in the overall length of genes in the X-chromosome of *An. gambiae*.

As many as 86 known genes out of 106 total known genes of *An. gambiae* X-chromosome were found to have homologs (both orthologs and paralogs) across 33 different taxa. Out of these 86 genes, 41 have only orthologs, 3 have only paralogs and 42 have both orthologs and paralogs. No homologs could be detected in the rest 20 genes, thus are considered unique to *An. gambiae* (Figure 4). Thus, in total, 83 (41+42) have orthologs and 45 (3+42) have paralogs in the X-chromosome known genes of *An. gambiae*. However, the distribution of orthologs varies across 33 different taxa; *Aedes aegypti* seems to bear most of the *An. gambiae* homologs (78 out of 83 orthologs) and the yeast *S. cerevisiae* bears the least (6 out of 83 orthologs) (Figure 5). These are at the highest and lowest ends of the homology prediction of 83 X-chromosome genes of *An. gambiae*.

Since majority of the known genes of X-chromosome of *An. gambiae* are either short or large in size (Figure 1), we were interested to know the distribution pattern of different types of genes based on homology predictions (orthologous, paralogous and unique) in different classes based on size (Figure 1). The details of such distribution are shown in Figure 6, which seems to be random. Whereas the orthologous genes are slightly abundant in Classes 2 and 6, the paralogous genes show a clear pattern of decreasing abundance from Class 1 to Class 6. In contrast unique genes are found to be much prevalent in Class 2 and least in Class 6 (Figure 6). We further looked at the distribution of these three types of genes (orthologous, paralogous and unique) based on the number of introns they possess (Figure 7) and found that

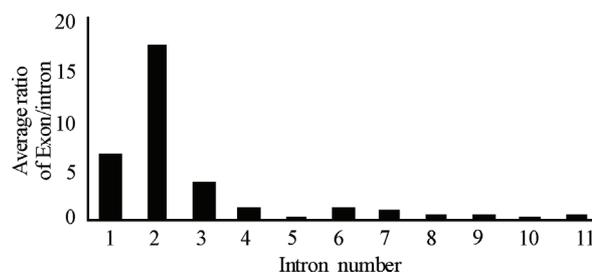


Figure 3. Average exon to intron ratio of *An. gambiae* X-chromosome known genes.

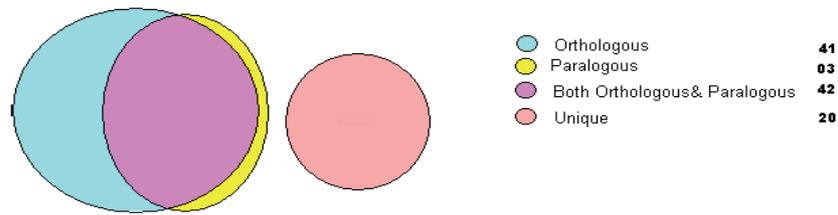


Figure 4. Distribution of different gene types (based on homology prediction) in X-chromosome of *An. gambiae*.

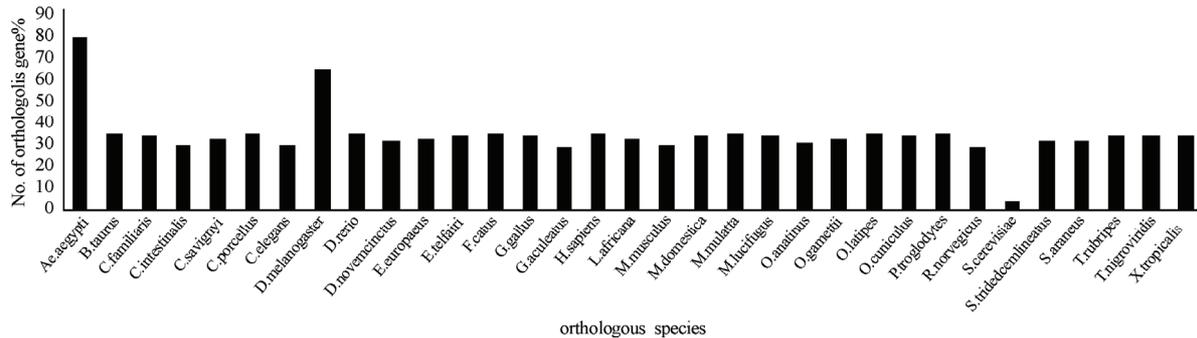


Figure 5. Distribution of different taxa showing number of shared genes with *An. gambiae*.

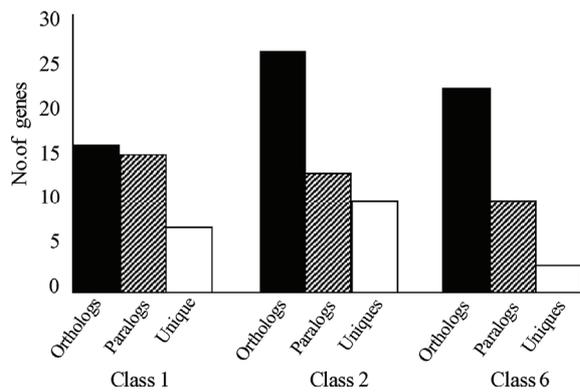


Figure 6. Distribution of orthologous, paralogous and unique genes of *An. gambiae* X-chromosome across different classes as in **Figure 1**.

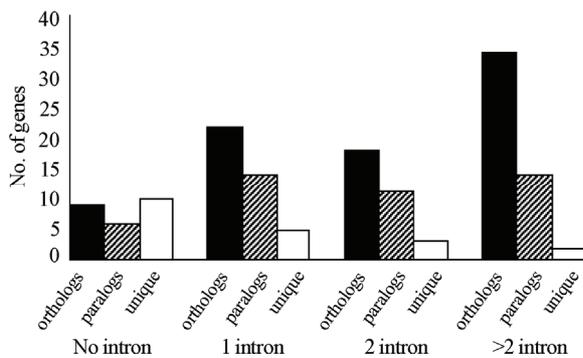


Figure 7. Distribution of orthologous, paralogous and unique genes of *An. gambiae* X-chromosome based on intron number.

most of the orthologous genes are biased towards more number of introns (>2) and majority of unique genes are without introns. There seem to be no biasness on the distribution of paralogous genes based on the number of introns, although a comparatively less number of paralogous genes were found to be without introns. In contrast, most of the unique genes are found to be without intron and the number decreases with increase in the number of introns. However, no significant correlation was found to exist between these two variables ($r=-0.04$, $P=0.88$).

In the process of homology prediction, we found a single gene (AGAP001043) of *An. gambiae* having orthologs in all other 33 taxa. This situation provided us an opportunity to look for evolutionary history, based on this conserved gene across all the 34 taxa. For this we constructed un-rooted neighbor joining (NJ) tree with only the coding sequences (exons) considering all 34 taxa, but due to lots of non-homologous sequences, no proper alignment among sequences could be obtained. Thus we went on deleting taxa with more number of non-homologous stretches and ended-up with only 13 taxa where meaningful alignment and construction of a phylogenetic tree was possible (**Figure 8**). Length of each branch leading to taxa was calculated as also the strength of each internal node through estimation of bootstrapped trees (**Figure 8**). Although in most cases, known closely related species came together in a single clade, (e.g. *Aedes aegypti*, *An. gambiae* and *Drosophila melanogaster*), the separation of human and chimpanzee branch is somehow an interesting observation (**Figure 8**). The strength of

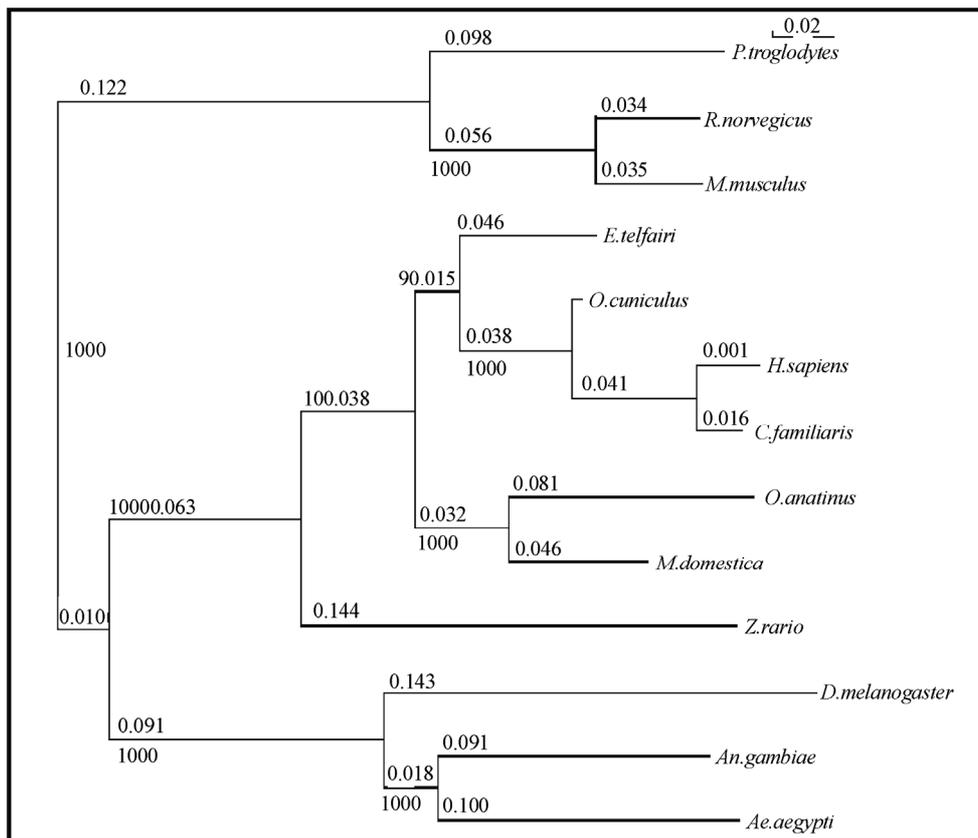


Figure 8. Phylogenetic tree with bootstrap values (in bold font) and branch length (in normal font) in 13 different taxa.

each internal node was found to be absolute in all the cases, signifying the phylogenetic tree is robust enough to be considered significant.

4. DISCUSSION

Although about two third part of the world is colonized by different species of mosquitoes and various species of the genus *Anopheles* is of great abundance, whole genome sequence information from only a single species, *An. gambiae* is so far available at the public domain. This situation possesses great constraint on researchers working on species of non-African importance. The preliminary objective of this study is to understand genomic architecture of *An. gambiae* through fine-scale dissection of the X-chromosome and utilize the information in developing nuclear DNA markers for estimating genetic diversity to infer both population demography and natural selection in species of importance in non-African malaria vectors. To start with, we have studied the X-chromosome, since it has several advantages over the autosomes. For example, X-chromosome is known to bear a rich resource of easily accessible genetic data, and provides a unique tool for population genetic studies [10]. Further, majority of population genomic studies

have been undertaken on the model organism, *Drosophila* utilizing X-linked genes [11,13].

We have considered genes of known function for detail characterization and analyses as they are the store house of important identifiable characters. More than a half of the known genes in *An. gambiae* X-chromosome are of short length and about 40% of the total known genes are engaged in housekeeping functions. Housekeeping genes are known to be often compact in size, which is attributed to selection for economy in transcription and translation [14,15]. Although a direct correlation between the function and size of the genes was not found (data not shown), abundance of shorter genes in *An. gambiae* might be due to the fact that most of the genes are engaged in housekeeping functions. However, long-sized genes (of more than 5kb length) were also abundant to certain extent (20%) in this chromosome. In contrast, genes of intermediate length (2kb to 5kb) were very less in occurrence. As far our knowledge goes, no study has reported this trend in the distribution of variable gene length across a single chromosome in any organism.

Majority of genes in *An. gambiae* X-chromosome shows a clear tendency of having either no or a very few (about three) introns (**Figure 2**). It has been suggested

that intronless gene families can evolve rapidly either by gene duplication or by reverse transcription/integration [16]. In *Oryza* and *Arabidopsis*, intronless genes have no homology and perform species-specific functions that are unique in respective species [17]. Thus, the findings of biasness towards either no or a few introns in majority of genes of known functions point towards the role of natural selection in reducing the size or presence of introns, thus reducing the overall transcriptional cost for effective expression. Findings of genes with very few numbers of introns in eukaryotic genome in general [18], further corroborate our contention.

The present study revealed the presence of only four genes with an appreciably high number of introns (9-11). Three of these four genes belong to nicotinic acetylcholine receptor gene families [19] which are the known targets of insecticides, neonicotinoids [20], as well as naturally derived spinisyns [21]. The presence of such high number of introns in these genes shows that they require more variability in their protein products and more number of introns make these genes more versatile in their expression by increasing the chances of alternative splicing and less 5' splice site biasness [22]. These genes are of much importance as far as growing insecticide resistance in almost all species of malaria vectors is concerned. Understanding of molecular mechanisms of resistance and molecular evolutionary studies should be targeted towards these genes for better understanding on evolution of insecticide resistance in malaria vectors.

We found strong positive correlations between gene length and intron length and intron length and intron number, which clearly suggests that size of the X-chromosome known genes in *An. gambiae*, is somehow dependent on accumulation of introns. These findings are in complete agreement with the general pattern observed in eukaryotic genome, where enlargement of genome size and decrease in genome compactness with increase in number and size of introns was supposed to be a general pattern during evolution of eukaryotes [5].

Based on evolutionary status, known genes in *An. gambiae* fall in all three categories (orthologous, paralogous and unique). Further, there is a clear distinction on the size, intron number and non-coding DNA content in different gene categories. Orthologous genes are comparatively larger in length and also contain more number of introns. These observations corroborate the fact that conserved or orthologous genes show more number of introns [23], possibly due to their slow evolution rate. Further, these types of genes tend to gain more and lose fewer introns [24]. In contrast to the characters of orthologous genes, no clear pattern was observed for the paralogous genes. This observation fits to the findings in the malaria parasite, *Plasmodium falciparum* and *P. yoelli yoelii*, where duplicate genes or paralogous genes in both the species exhibit a dramatic acceleration

of intron gain/loss and protein evolution in comparison with orthologous genes, suggesting increased directional and/or relaxed selection in duplicate genes [25]. In contrast, the unique genes are mostly intron-less and might be evolving at a faster rate losing the non-coding portions; hence homologs of these genes are no more recognizable [26].

We detected a close genetic affinity at the X-chromosome known genes among three taxa of insects; *An. gambiae*, *Ae. aegypti* and *D. melanogaster*. This is not surprising, as three of them belong to the similar order Diptera. Similar observations have been made in earlier studies [27] as also among *Drosophila* species with comparative genomic approaches [6]. Although *S. cerevisiae* was found to be farthest in homology, six genes of *S. cerevisiae* show homology to *An. gambiae* genes. The function of these genes was found to be tyrosyl-t RNA synthetase, peroxiredoxin dependent peroxidase and glutathione reductase, one was putative membrane bound O-acyltransferase. However, these genes possess very few introns. Although it is apparent that gene size has increased through accumulation of introns in higher taxa, it is surprising that these six genes remained conserved in the long evolutionary process from yeast to insects. It is probable that, these genes might be under some functional constraint and natural selection saved their compositions across different lineage.

We considered a gene present in all the 34 taxa to infer evolutionary history through phylogenetic tree construction. The pattern in the tree was mostly found to comply with general patterns of phylogenetic relationships among organisms. The placement of *An. gambiae*, *D. melanogaster* and *Ae. aegypti* in a single clade further corroborates our observation of close genetic relatedness among these three taxa (see above). However, it was surprising to note that human and chimpanzee are placed in two separate clades. Although, human and chimpanzee are genetically close to each other [28], gross differences at several genes have also been observed [29]. Included in these genes is a gene responsible for malaria susceptibility in humans and chimpanzee [30]. These two taxa differ in their susceptibility to malaria due to human specific loss of N-glycolylneuraminic acid, which is present in primates. It might then be true that genomic difference between two species only can be traceable at selected genes and the presently studied gene is one of them. This gene codes for a protein which is involved in signal transduction process and participates in transmission of developmental information by associating with alpha or beta catenin. Whereas basic function of this gene is conserved, little functional differences exist across almost all taxa. Further in depth study revealed that however, the gene function is conserved in taxa falling in a particular clade found here, whereas few differences were observed across clades in the phylogenetic tree (Figure 8). Thus, functional con-

ervation of this gene is exactly overlaps with the position of taxa in the phylogenetic tree.

In conclusion, the present study not only provides fine-scale views to the genetic architecture of the X-chromosome of the model malaria vector of African importance, but also reveals several interesting features on evolutionary insights into genes and taxa of different taxonomic status. The information is of great importance, especially to the population geneticists, to understand genetic diversity and infer the respective roles of demography and natural selection in evolution of genes in different *Anopheles* species populations of local importance.

5. ACKNOWLEDGEMENTS

Extramural funding from the Indian Council of Medical Research (ICMR), New Delhi in the form of an Ad-hoc research grant to AD is thankfully acknowledged. We thank Dr. Neena Valecha for her kind support to HS during the initial phase of the study. Lily Basu, Deepshikha Lal, Garima Goyal and Suchita Singh helped in organizing the initial work elements.

REFERENCES

- [1] Hahn, M. W., Han, M. V., and Han, S. G., (2007) Gene family evolution across 12 *Drosophila* genome, Public Library of Science Genetics, **3**, 1–12.
- [2] Matthee, C. A., Eick, G., Willows, M. S., Montgelard, C., Pardini, A. T., and Robinson, T. J., (2007) Indel evolution of mammalian introns and the utility of non coding nuclear markers in eutherian phylogenetics, *Molecular Phylogenetics and Evolution*, **42**, 827–837.
- [3] Cardazzo, B., Bargelloni, L., Toffolatti, L., and Patarrello, T., (2003) Intervening sequences in paralogous genes: A comparative genomic approach to study the evolution of X chromosome introns, *Molecular Biology and Evolution*, **20**, 2034–2041.
- [4] Gazave, E., Bonet, T. M., Fernando, O., Charlesworth, B., and Navarro, A., (2007) Patterns and rates of intron divergence between humans and chimpanzees, *Genome Biology*, **8**, 1–13.
- [5] Huynen, M. A. and Bork, P., (1998) Measuring genome evolution, *Proceedings of National Academy of Sciences USA*, **95**, 5849–5856.
- [6] Clark, A. G., Eisen, B. M., Smith, D. R., Bergman, C. M., Oliver, B., Markow, T. A., *et al.*, (2007) Evolution of genes and genomes on the *Drosophila* phylogeny, *Nature*, **450**, 203–218.
- [7] WHO. (2005) World malaria report. <http://www.rbm.who.int/wmr>.
- [8] Zakeri, S., Afsharipad, M., Raeisi, A., and Djadid, N. D., (2007) Prevalence of mutations associated with antimalarial drugs in *Plasmodium falciparum* isolates prior to the introduction of sulphadoxine-pyrimethamine as first-line treatment in Iran, *Malaria Journal*, **6**, 1–2.
- [9] Holt, R. A., Subramanian, G. M., Helpert, A., Sutton, G. G., Charlab, R., Nusskern, D. R., *et al.*, (2002) The genome sequence of the malaria mosquito *Anopheles gambiae*, *Science*, **298**, 129–149.
- [10] Stephen, S. F., (2004) The X chromosome in population genetics, *Nature Reviews Genetics*, **5**, 43–51.
- [11] Vogl, C., Das, A., Beaumont, M., Mohanty, S., and Stephan, W., (2003) Population subdivision and molecular sequence variation: Theory and analysis of *Drosophila ananassae* data, *Genetics*, **165**, 1385–1395.
- [12] Bains, J. F., Das, A., Mousset, S., and Stephan, W., (2004) The role of natural selection in genetic differentiation of worldwide populations of *Drosophila ananassae*, *Genetics*, **168**, 1987–1998.
- [13] Das, A., Mohanty, S., and Stephan, W., (2004) Inferring the population structure and demography of *Drosophila ananassae* from multilocus data, *Genetics*, **168**, 1975–1985.
- [14] Castillo-Davis, C. I., Mekhedov, S. L., Hartl, D. L., Koonin, E. V., and Kondrashov, F. A., (2002) Selection for short introns in highly expressed genes, *Nature Genetics*, **31**, 414–418.
- [15] Vinograd, A. E. (2004) Compactness of human housekeeping genes: Selection for economy or genomic design, *Trends in Genetics*, **20**, 248–253.
- [16] Jain, M., Tyagi, A. K., and Khurana, J. P., (2006) Genome wide analysis, evolutionary expansion, and expression of early auxin-responsive SAUR gene family in rice (*Oryza sativa*), *Genomics*, **88**, 360–371.
- [17] Jain, M., Khurana, P., Tyagi, A. K., and Khurana, J., (2007) Genome-wide analysis of intronless genes in rice and Arabidopsis, *Functional & Integrative Genomics*, In Press.
- [18] Simpson, A. G. B., Macquarrie, E. K., Roger, A. J., (2002) Eukaryotic evolution: Early origin of canonical introns, *Nature*, **419**, 270.
- [19] Jones, A. K., Grauso, M., and Sattelle, B. D., (2004) The nicotinic acetylcholine receptor gene family of the malaria mosquito, *Anopheles gambiae*, *Genomics*, **85**, 176–187.
- [20] Matsuda, K., Buchingham, S. D., Kleier, D., Rauh, J. J., Grauso, M., and Sattelle, D. B., (2001) Neonicotinoids: Insecticides acting on insect nicotinic acetylcholine receptors, *Trends in Pharmacological Sciences*, **22**, 573–580.
- [21] Bond, G. J., Marina, C. F., and Williams, T., (2004) The naturally derived insecticide spinosad is highly toxic to *Aedes* and *Anopheles* mosquito larvae, *Medical and Veterinary Entomology*, **18**, 50–56.
- [22] Manuel, I., David, P., and Scott, W. R., (2007) Coevolution of genomic intron number and splice sites, *Trends in Genetics*, **23**, 321–325.
- [23] Jordan, I. K., Marino-Ramirez, L., Wolf, Y. I., and Koonin, E. V., (2004) Conservation and co-evolution in the scale-free human gene co-expression network, *Molecular Biology and Evolution*, **21**, 2058–2070.
- [24] Carmel, L., Rogozin, I. B., Wolf, Y. I., and Koonin, E. V., (2007) Evolutionarily conserved genes preferentially accumulate introns, *Genome Research*, **17**, 1045–1050.
- [25] Castillo-Davis, C. I., Bedford, T. B. C., and Hartl, D. L., (2004) Accelerated rates of intron gain/loss and protein evolution in duplicate genes in human and mouse malaria parasite, *Molecular Biology and Evolution*, **21**, 1422–1427.
- [26] Stirling, B., Yang, Z. K., Gunter, L. E., Tuskan, G. A., and Bradshaw, H. D., (2003) Comparative sequence

analysis between orthologous regions of the Arabidopsis and Populus genomes reveals substantial synteny and microcollinearity, Canadian Journal of Forest Research, **33**, 2245–2255.

- [27] Bohbot, J., Pitts, R. J., Kwon, H. W., Rutzler, M., Robertson, H. M., and Zwiebel, L. J. (2007) Molecular characterization of *Aedes aegypti* odorant receptor gene family, *Insect Molecular Biology*, **16**, 525–537.
- [28] Uddin, M., Wildman, D. E., Liu, G., Xu, W., Johnson, R. M., Hof, P. R., *et al.*, (2004) Sister grouping of chimpanzees and humans as revealed by genome-wide phylogenetic analysis of brain gene expression analysis, *Proceedings of National Academy of Sciences*, **101**, 2957–2962.
- [29] Gilad, Y., Man, O., and Glusman, G., (2005) A comparison of the human and chimpanzee olfactory receptor gene repertoires, *Genome Research*, **15**, 224–230.
- [30] Martin, M. J., Rayner, J. C., Gagneux, P, Barnwell, J. W., and Varki, A., (2005) Evolution of human-chimpanzee differences in malaria susceptibility: Relationship to human genetic loss of N-glycolylneuraminic acid, *Proceedings of National Academy of Sciences*, **102**, 12819–12824.