Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 3, Issue, 11, pp.066-073, October, 2011

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

ISSN: 0975-833X

RESEARCH ARTICLE

COMPARATIVE EVOLUTIONARY ANALYSIS OF INSECTS USING MITOCHONDRIAL GENES

***Habeeb, S. K. M., Sirish Kumar, Sanjayan, K. P., Anuradha, V and Praveena, A.**

Department of Zoology & Department of Bioinformatics, Guru Nanak College, Chennai, India

ARTICLE INFO

Article History:
Received 7th July, 2011
Received in revised form
19th August, 2011
Accepted 24th September, 2011
Published online 15th October, 2011

Key words:
mtDNA,
16srRNA,
12srRNA,
Phylogeny,
COI & Bioinformatics.

ABSTRACT

In this study, a comparative analysis of select insect mitochondrial DNA (mtDNA) representing five insect orders (Diptera, Lepidoptera, Hemiptera, Ephemeroptera and Coleoptera) consisting of 24 different species in an effort to study a common set of genes and to understand the evolution of mitochondrial genome with respect to Cytochrome Oxidase subunit 1 (CO1), 16srRNA and 12srRNA gene sequences was done. To compare the similarity between closely related insect mitochondrial genome sequences, pairwise distance matrix was constructed using the MEGA tool. CO1, 16srRNA and 12srRNA sequences were used to construct a phylogenetic tree to determine the relationship among five insect orders. All the three sequences yielded a tree with branching patterns reflecting the expected pattern as insect species belonging to different orders were put into separate clades with an exception *S. graminum* & *M. destructor* being branched together in to a separate group. Based on the sequence similarity, insect species belonging to five different orders in general appear to be closely related. The tRNA species were identical in most species with an exception that two of the twenty four species had an additional tRNA gene. Based on this study we conclude that, although the gene types are very similar across these insect orders, significant differences in GC content perhaps suggest multiple mitochondrial ancestors.

Copy Right, IJCR, 2011, Academic Journals. All rights reserved

INTRODUCTION

Mitochondrial DNA (mtDNA) has been one of the most widely used molecular markers for phylogenetic studies in animals, because of its simple genomic structure (Avise, 2004). Animal mitochondrial genomes (mtgenomes) are small, circular DNA with length ranging from 14,000 bp to 17,000 bp (Boore 1999; Cameron, 2007). They usually encode 37 genes (13 protein coding, 22 transfer RNA, and 2 ribosomal RNA genes). The number of complete mtgenomes has steadily been on the rise with the technical feasibility of sequencing their entirety (Hwang, 2001; Yamauchi, 2004). This increasing availability of mtgenome data invites comparative study. Among insects, the maximum number of mitochondrial genomes has been characterized in the order Diptera (Cameron, 2007; Shao, 2007). The mitochondrial genome (mt-genome) is one of the largest sets of homologous genes which can be compared across animal taxa and has become an effective data source for resolving deep-level phylogenetic problems (Cameron, 2006; Dellaporta, 2006). Within Insecta, more than two hundred mtgenomes are available now in databases and mtgenomes have been shown to resolve intra ordinal relationships, such as in Diptera (Cameron, 2007), Hymenoptera (Cameron, 2008), and Orthoptera (Fenn, 2008).

Corresponding author: habeeb_skm@yahoo.co.in

There are many possible ways of using mt-genomes in phylogenetic analyses, for example by using different genes, amino acid sequences or nucleotide sequences. Using the nucleotide sequences of all available genes clearly has been shown to be the best way to extract a phylogenetic signal from Mt-genomes (Cameron, 2007; Fenn, 2008). Though mtDNA sequence data have proved valuable in determining phylogenetic relationships, the choice of gene is also of great significance (Simon, 1994; Lunt, 1996). The size and structure of cytochrome oxidase subunit 1 (COI) gene has been well conserved in the animal groups analyzed so far, a feature which makes it especially suitable for evolutionary studies (Lunt, 1996). In recent years, comparison of mitochondrial DNA sequences has been used for population genetics and phylogenetic studies in several dipterans of medical, veterinary and economic importance (Hall, 2001; Zehner, 2004; Cummings, 2005; Segura, 2006; Angella, 2007). It has become apparent that other mitochondrial gene regions can also provide insights concerning deeper divergences, as shown by a study that employed 16S rDNA sequences to examine the affinities of major gastropod lineages (Thollesson, 1999). Therefore, the nucleotide sequence of the small-subunit (12s) and large subunit (16s) ribosomal RNA (srRNA) are proving extremely valuable for probing phylogenetic relationships among distantly related taxa because many regions remain conserved or semiconserved over large

periods of time. Woese (1987) has suggested that the ribosomal RNAs may be the ultimate molecular clocks. There is a large body of evidence that suggests that the rate of molecular evolution in general is relatively constant with time (Wilson, 1977; Ayala, 1986). Because the srRNA molecule is common to virtually all life forms, it becomes possible to test the clock for a single molecule across many phyla. The 16srRNA gene, which encodes the mitochondrial large ribosomal subunit (mt LSU) in animals, has been employed extensively to explore phylogenetic relationships in arthropods at most phylogenetic levels (Flook, 1995), familial level (Black, 1994) and the genus level and below (DeSalle, 1992; Bond, 2001). The wide range in utility of 16s at various taxonomic levels suggests that the differential rates of molecular evolution within 16s, due to varying functional constraints, greatly affect its phylogenetic utility.

Hence based on these factors we have decided to compare the mitochondrial genomes of different insect orders among 24 species by studying their gene content, base composition, and the utility of the mitochondrial gene products Cytochrome Oxidase subunit 1, 16srRNA & 12srRNA gene sequences for assessing the relationship between the different orders by carrying out a evolutionary analysis using the bioinformatics tools.

MATERIALS AND METHODS

National Center for Biotechnology Information (NCBI)

The data for the study was obtained from the NCBI database (Geer, 2010). A search with the keyword "mitochondrion insect" in the NCBI platform resulted in 243 entries spread across various Orders. Out of the several insect orders we selected twenty four species from five orders namely Diptera, Coleoptera, Hemiptera, Lepidoptera and Ephemeroptera based on the condition one species per family and five species per insect order. There were only four entries for Ephemeroptera and hence we selected all the four under this category. Five species each were selected from the other four orders. The mtDNA was downloaded and their proposition was determined. The list of species selected for analysis and their information is provided in the table. 1. From this mtDNA we selected the nucleotide sequences of the protein coding gene Cytochrome Oxidase subunit 1, and the ribosomal RNAs such as the 16srRNA and the 12srRNA otherwise called as large subunit and small subunit respectively. This after the literature study revealing that these selected biological entities have an important role in the species identification. The protein table option of NCBI helped us to narrow down on the CO1, 16srRNA and 12srRNA and tRNA sequences.

Molecular Evolutionary Genetic Analysis (MEGA)

The nucleotide sequences of the CO1, 16srRNA and 12srRNA were aligned using the ClustalW option present in the MEGA4 software (Tamura, 2007). Once the alignment is done, nucleotide composition, analysis of various sites such as conserved sites, parsimonious informative sites, variable sites etc., are performed using the highlight section of the sequence data explorer window of the MEGA tool. The results were imported in to Microsoft Excel 2007 and statistical calculations pertaining to the nucleotide bases were performed for all the gene types. Further, the aligned sequences were used to construct Phylogenetic trees individually for all the three gene types using the neighbor-joining algorithm (Saitou, 1987) involving 500 replicates of bootstrap analysis (Felsenstein, 1985). The evolutionary distances were computed using the p-distance method and are in the units of the number of base differences per site. The analysis involved 24 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 1519 positions in the final dataset of the CO1, 992 positions for 16srRNA and 554 positions for the 12srRNA sequences.

RESULTS

Gene Content

An analysis of the content of mtDNA for all the species was done and variation among the orders was determined. The species selected for the study and their genome composition is shown in table 1. It was observed that among the selected data for analysis R.*magnicornis* had highest genome size and a bigger by almost 1300 bp than an average genome size among the Coleoptera but lesser coding percent of 63. The same can be said about *N. viridula* of Hemiptera with a coding percent of 65 and higher in genome size than the average genome by approximately 1200 bp (table.1). Thus, it reflects the fact that the complexity of the organism does not correlate with the genome size of an organism. There were a total of 13 protein coding genes and two ribosomal RNA genes in all the insect species and all the organisms had 22 tRNA genes in common with an additional tRNA gene coding for isoleucine and methionine in A.*issoria* and P.*youi* respectively.

The mean and standard deviation was calculated among the insect orders and it was found that the average mtDNA size of Coleopteran species to be 16204.4bp the largest and the Lepidopteran species had the lowest average mtDNA size at 15387.2 bp. However, best sequence conservation was found to be among the Lepidopteran species as they had a smallest standard deviation of 113.37, the largest was calculated to be 938.86 for the Hemiteran species. The GC content was highest in Ephemeroptera and lowest for Lepidoptera at 29.5 ± 2.51 and 19 ± 0.70 respectively as shown in Table 2.

Base Statistics of COI, 16srRNA & 12srRNA

The mean number of thymines (T) was 36.91 in COI genes, which was higher in 16srRNA & 12srRNA at 42.3 & 40.75 respectively. Similar was the case when the analysis is extended to adenine content. However, Cytosines were found to be more frequented in COI genes (17.10), two times higher in number than the 16srRNA & 12srRNA at 6.9 & 7.67 respectively (table 3). The combined GC content predicted to be higher in COI than the others, which was expected from the protein coding gene. The AT content was almost found to be uniform among the data. 12srRNA had the very less sequence conservation percent in comparison to the other gene sequences closely followed by 16srRNA. It is worth a fact to see that the sequence conservation of the CO1 gene sequences was relatively higher than the other two RNA coding genes. On the other hand the percent variation site and the parsimoniously informative sites were lower in the CO1 sequences than the other two.

Table 1. Data taken for the study

Species	Family	Accession No	Size (bp)	GC	AT	Coding %	tRNA Genes
Hemiptera							
Schizaphis graminum	Aphididae	NC_006158	15721	16	84	69	22
Nezara viridula	Pentatomidae	NC_011755	16889	23	77	65	22
Dysdercus cingulatus	Pyrrhocoridae	NC_012421	16249	22	78	67	22
Geocoris pallidipennis	Lygaeidae	NC_012124	14592	24	76	75	22
Physopelta gutta	Largidae	NC_012432	14935	25	75	73	22
Diptera							
Drosophila simulans	Drosophilidae	NC_005781	14972	22	78	74	22
Aedes albopictus	Culicidae	NC_006817	16665	20	80	67	22
Aedes aegypti	Culicidae	NC_010241	16655	21	79	67	22
Mayetiola destructor	Cecidomyiidae	NC_013066	14759	15	85	73	22
Bactrocera tryoni	Tephritidae	NC_014611	15925	27	73	70	22
Lepidoptera							
Manduca sexta	Sphingidae	NC_010266	15516	18	82	72	22
Antheraea yamamai	Saturniidae	NC_012739	15338	19	81	73	22
Diatraea saccharalis	Crambidae	NC_013274	15490	19	81	72	22
Acraea issoria	Nymphalidae	NC_013604	15245	20	80	73	23**
Helicoverpa armigera	Noctuidae	NC_014668	15347	19	81	72	22
Coleoptera							
Tribolium castaneum	Tenebrionidae	NC_003081	15881	28	72	69	22
Chrysochroa fulgidissima	Buprestidae	NC_012765	15592	30	70	71	22
Psacothea hilaris	Cerambycidae	NC_013070	15856	23	77	70	22
Rhopaea magnicornis	Scarabaeidae	NC_013252	17522	24	76	63	22
Apatides fortis	Bostrichidae	NC_013582	16171	32	68	68	22
Ephemeroptera							
Parafronurus youi	Heptageniidae	NC_011359	15481	33	67	72	23*
Davidius lunatus	Gomphidae	NC_012644	15913	29	71	70	22
Ephemera orientalis	Ephemeridae	NC_012645	16463	27	73	67	22
Siphlonurus immanis	Siphlonuridae	NC_013822	15529	29	71	72	22

*Additional Met tRNA gene **Additional Ile tRNA gene

Tabel 2. Size & GC Content

Order	Size	GC %
Hemiptera	15677.2 ± 938.86	22 ± 3.53
Diptera	15795.2 ± 903.33	21 ± 4.30
Lepidoptera	15387.2 ± 113.37	19 ± 0.70
Coleoptera	16204.4 ± 764.55	27.4 ± 3.84
Ephemeroptera	15846.5 ± 454.19	29.5 ± 2.51

Table 3. Gene Wise Base Statistics

Parameter	CO1	16sr RNA	12sr RNA
T	36.91 ± 0.72	42.3 ± 0.68	40.75 ± 0.65
C	17.10 ± 0.62	6.9 ± 0.38	7.67 ± 0.50
A	31.17 ± 0.50	37.45 ± 0.91	37.28 ± 1.07
G	14.83 ± 0.33	13.35 ± 0.65	14.28 ± 0.85
Size Including gaps	1578	1502	913
Conserved site	689	352	168
Variable site	872	1104	702
Parsimony informative site	749	932	583
Overall distance	0.752	0.443	0.664

Table 4. Maximum Composite Likelihood Estimate of the Pattern of Nucleotide Substitution

	CO1				16sr RNA				12sr RNA			
	A	T	C	G	A	T	C	G	A	T	C	G
A	-	9.62	4.45	5.04	-	12.24	2.52	6.1	-	9.73	2.41	7.19
T	8.09	-	10.2	3.86	10.84	-	3.25	4.5	8.78	-	5.34	3.86
C	8.09	22.14	-	3.86	10.84	15.77	-	4.5	8.78	21.5	-	3.86
G	10.55	9.62	4.45	-	14.69	12.24	2.52	-	16.4	9.73	2.41	-
R = si/sv	0.828				0.467				0.75			

Nucleotide Substitution Pattern

Each entry in table 4 shows the probability of substitution (r) from one base (row) to another base (column). Rates of different transitional substitutions are shown in bold and those of transversionsal substitutions are shown in *italics*. The transition/transversion rate ratios are k_1 = 1.304 (purines) and k_2 = 2.301 (pyrimidines), k_1 = 1.355 (purines) and k_2 = 1.288 (pyrimidines) and k_1 = 1.864 (purines) and k_2 = 2.212 (pyrimidines) and the overall transition/transversion bias R = 0.828, 0.467 & 0.750 for CO1, 16srRNA and 12srRNA sequences respectively.

Distance Matrix

COI

Analysis of the distance matrix table of the COI sequences revealed a maximum distance value of 1.222 between the species *A.fortis & R.magnicornis* followed by values 1.171, 1.162, and 1.156 between S.*immanis & E.orientalis, A.fortis & P.hilaris and D.saccharalis & D.simulans* respectively. The distance matrix is shown in Fig 1. While the minimum distance was recorded between *H.armigera & D.simulans* with a distance value of 0.404 followed closely by *M.sexta & D.simulans, H.armigera & M.destructor and M.destructor & S.graminum* with values of 0.431, 0.435 and 0.440 respectively. An interesting aspect was found with *A.aegypti & A.albopictus* species belonging to same family of Culicidae, showing a distance of 1.103 a relatively higher than the distance between species of different orders. The average distance between the species was found to be 0.752.

S.immamis & E.orientalis with a value of 1.071 followed by A.fortis & T.castaneum, S.immanis & P.youi, C.fulgidissima & D.simulans, P.youi & D.saccharalis and P.youi & A.issoria with a value of 0.727, 0.688, 0.675, 0.674 and 0.674 respectively. Whereas the minimum distance between the species was found to be 0.221, 0.222, 0.244, 0.249 and 0.260 for the species, *D.saccharallis & A.yamamai, A.issoria & P.hilaris, D.saccharalis & A.aegypti and D.saccharalis & D.simulans* respectively. A distance of 0.429 was recorded between the species *A.aegypti & A.albopictus* (Fig 2). The average distance separating the species from each other recorded was 0.443.

12srRNA

The maximum distance of 1.750 was recorded between the species *A.aegypti & A.albopictus,* both species belonging to the same family of Culcidae in the Order Diptera and interesting prediction, followed by *B.tryoni & D.simulans,*

Species	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
S.graminum																							
N.viridula	0.633																						
D.cingulatus	0.539	0.732																					
G.pallidipennis	0.517	0.747	0.742																				
P.gutta	0.681	0.978	0.930	0.940																			
D.simulans	0.565	0.525	0.498	0.578	0.754																		
A.albopictus	0.655	0.534	0.603	0.587	0.735	0.557																	
A.aegypti	0.689	0.550	0.631	0.631	0.693	0.619	1.103																
M.destructor	0.440	0.477	0.545	0.607	0.645	0.458	0.526	0.516															
B.tryoni	0.734	0.642	0.708	0.746	0.833	1.156	0.848	0.787	0.737														
M.sexta	0.456	0.530	0.495	0.538	0.597	0.431	0.604	0.597	0.491	0.775													
A.yamamai	0.490	0.517	0.455	0.475	0.589	0.440	0.630	0.650	0.445	0.801	0.598												
D.saccharalis	0.557	0.608	0.552	0.658	0.671	0.605	0.746	0.711	0.541	0.832	0.748	0.805											
A.issoria	0.527	0.624	0.550	0.638	0.679	0.500	0.634	0.712	0.462	0.759	0.680	0.811	0.854										
H.armigera	0.484	0.579	0.482	0.534	0.648	0.404	0.575	0.589	0.435	0.755	0.568	0.504	0.748	0.556									
T.castaneum	0.857	0.824	0.782	0.867	0.717	0.896	0.851	0.899	0.981	0.818	1.021	1.016	0.979	0.989	0.968								
C.fulgidissima	0.741	0.816	0.715	0.830	0.907	0.830	0.774	0.910	0.753	0.799	0.880	0.718	0.848	0.837	0.777	1.132							
P.hilaris	0.618	0.629	0.555	0.700	0.681	0.607	0.688	0.721	0.598	0.741	0.602	0.601	0.755	0.835	0.619	1.048	0.922						
R.magnicornis	0.633	0.590	0.588	0.549	0.755	0.613	0.619	0.634	0.508	0.817	0.642	0.541	0.711	0.668	0.604	0.897	0.948	0.755					
A.fortis	0.903	0.787	0.850	0.966	0.861	1.095	0.902	0.990	1.067	1.027	0.985	0.939	0.947	0.970	0.838	0.909	0.925	1.162	1.222				
P.youi	0.861	0.753	0.715	0.760	0.763	0.788	0.932	0.929	0.748	0.924	0.856	0.843	1.056	0.894	0.879	1.038	0.856	0.886	0.826	0.990			
D.lunatus	0.871	0.898	0.780	0.880	0.801	0.878	0.969	0.906	0.845	0.883	0.922	0.774	1.021	0.945	0.845	0.869	0.969	0.938	0.835	0.948	0.920		
E.orientalis	0.749	0.556	0.665	0.665	0.682	0.771	0.760	0.706	0.689	0.802	0.704	0.730	0.942	0.755	0.782	0.935	0.799	0.771	0.735	0.828	0.897	0.882	
S.immanis	0.697	0.631	0.695	0.671	0.708	0.726	0.872	0.878	0.718	0.825	0.830	0.849	0.851	0.834	0.727	1.046	0.925	0.754	0.735	1.099	1.038	1.011	1.171

Fig 1: Distance Matrix of COI

	Species	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	S.graminum																							
2	N.viridula	0.358																						
3	D.cingulatus	0.336	0.361																					
4	G.pallidipennis	0.365	0.490	0.445																				
5	P.gutta	0.404	0.554	0.567	0.575																			
6	D.simulans	0.271	0.354	0.321	0.381	0.396																		
7	A.albopictus	0.274	0.325	0.328	0.373	0.433	0.327																	
8	A.aegypti	0.275	0.288	0.308	0.348	0.390	0.280	0.429																
9	M.destructor	0.340	0.336	0.361	0.368	0.439	0.330	0.327	0.320															
10	B.tryoni	0.349	0.370	0.335	0.354	0.400	0.520	0.447	0.402	0.397														
11	M.sexta	0.270	0.366	0.276	0.406	0.391	0.273	0.271	0.256	0.330	0.392													
12	A.yamamai	0.297	0.428	0.336	0.379	0.411	0.345	0.325	0.302	0.368	0.475	0.378												
13	D.saccharalis	0.279	0.338	0.287	0.369	0.413	0.260	0.278	0.249	0.293	0.392	0.221	0.222											
14	A.issoria	0.287	0.384	0.298	0.355	0.403	0.274	0.284	0.244	0.318	0.390	0.346	0.344	0.282										
15	H.armigera	0.263	0.378	0.283	0.335	0.372	0.297	0.290	0.267	0.283	0.412	0.343	0.320	0.278	0.297									
16	T.castaneum	0.455	0.513	0.418	0.425	0.531	0.509	0.362	0.332	0.415	0.512	0.443	0.479	0.429	0.414	0.476								
17	C.fulgidissima	0.588	0.513	0.449	0.329	0.505	0.675	0.543	0.506	0.576	0.543	0.563	0.648	0.577	0.573	0.588	0.658							
18	P.hilaris	0.354	0.402	0.336	0.293	0.386	0.298	0.290	0.273	0.349	0.370	0.369	0.419	0.317	0.350	0.371	0.562	0.615						
19	R.magnicornis	0.440	0.416	0.304	0.341	0.406	0.337	0.320	0.322	0.371	0.351	0.387	0.486	0.383	0.373	0.360	0.554	0.631	0.360					
20	A.fortis	0.616	0.565	0.522	0.416	0.488	0.584	0.500	0.492	0.619	0.487	0.567	0.630	0.539	0.619	0.545	0.727	0.613	0.506	0.601				
21	P.youi	0.506	0.513	0.435	0.440	0.456	0.644	0.538	0.542	0.594	0.508	0.603	0.641	0.674	0.674	0.628	0.522	0.522	0.531	0.538	0.523			
22	D.lunatus	0.574	0.493	0.431	0.490	0.510	0.532	0.486	0.500	0.545	0.511	0.505	0.495	0.523	0.525	0.545	0.515	0.629	0.427	0.497	0.592	0.537		
23	E.orientalis	0.472	0.491	0.441	0.431	0.507	0.506	0.483	0.485	0.576	0.520	0.553	0.617	0.599	0.545	0.539	0.578	0.642	0.516	0.506	0.532	0.613	0.618	
24	S.immanis	0.426	0.524	0.419	0.399	0.493	0.500	0.462	0.485	0.507	0.497	0.528	0.581	0.578	0.497	0.517	0.568	0.634	0.534	0.463	0.568	0.688	0.555	1.071

Fig. 2. Distance Matrix 16srRNA

	Species	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	S.graminum																							
2	N.viridula	0.548																						
3	D.cingulatus	0.456	0.614																					
4	G.pallidipennis	0.578	0.658	0.758																				
5	P.gutta	0.576	0.609	0.740	0.700																			
6	D.simulans	0.495	0.592	0.476	0.674	0.578																		
7	A.albopictus	0.515	0.509	0.440	0.563	0.523	0.913																	
8	A.aegypti	0.546	0.600	0.476	0.612	0.541	0.891	1.750																
9	M.destructor	0.323	0.476	0.327	0.447	0.429	0.530	0.423	0.400															
10	B.tryoni	0.700	0.674	0.702	0.728	0.596	1.444	0.887	0.824	0.645														
11	Manduca_sexta	0.440	0.470	0.402	0.528	0.516	0.615	0.530	0.506	0.306	0.838													
12	A.yamamai	0.500	0.487	0.426	0.518	0.461	0.573	0.563	0.565	0.330	0.763	0.591												
13	D.saccharalis	0.466	0.462	0.492	0.534	0.583	0.584	0.593	0.598	0.305	0.903	0.529	0.633											
14	A.issoria	0.479	0.445	0.386	0.483	0.423	0.556	0.527	0.506	0.402	0.714	0.474	0.563	0.548										
15	H.armigera	0.506	0.448	0.423	0.514	0.521	0.605	0.547	0.523	0.327	0.772	0.522	0.609	0.576	0.390									
16	T.castaneum	0.606	0.693	0.717	0.720	0.740	0.785	1.000	1.013	0.478	0.885	0.587	0.490	0.570	0.592	0.608								
17	C.fulgidissima	0.911	0.769	0.729	0.750	0.610	0.776	0.732	0.653	0.597	0.688	0.682	0.573	0.643	0.630	0.564	0.863							
18	P.hilaris	0.630	0.477	0.500	0.552	0.610	0.570	0.687	0.704	0.436	0.744	0.457	0.385	0.490	0.490	0.474	0.610	0.588						
19	R.magnicornis	0.702	0.569	0.535	0.615	0.429	0.806	0.930	0.826	0.500	0.700	0.628	0.538	0.663	0.568	0.635	0.667	0.615	0.500					
20	A.fortis	0.725	0.557	0.602	0.657	0.627	0.735	0.787	0.782	0.600	0.675	0.656	0.577	0.704	0.587	0.701	0.821	0.814	0.589	0.750				
21	P.youi	1.065	0.650	0.645	0.692	0.632	0.974	1.096	1.127	0.772	0.829	0.988	1.064	1.000	0.859	0.944	1.023	0.826	0.740	0.925	0.540			
22	D.lunatus	0.938	0.679	0.590	0.595	0.557	0.902	0.791	0.888	0.590	0.848	0.747	0.758	0.768	0.758	0.745	0.835	0.736	0.717	0.835	0.731	0.803		
23	E.orientalis	0.954	0.657	0.550	0.673	0.613	0.803	0.600	0.628	0.533	0.688	0.770	0.753	0.865	0.663	0.789	0.988	0.828	0.604	0.942	0.642	1.208	0.705	
24	S.immanis	0.955	0.755	0.587	0.748	0.628	0.816	0.671	0.684	0.547	0.707	0.824	0.816	0.860	0.680	0.872	0.977	0.897	0.670	1.111	0.692	1.275	0.819	0.722

Fig. 3. Distance Matrix 12srRNA

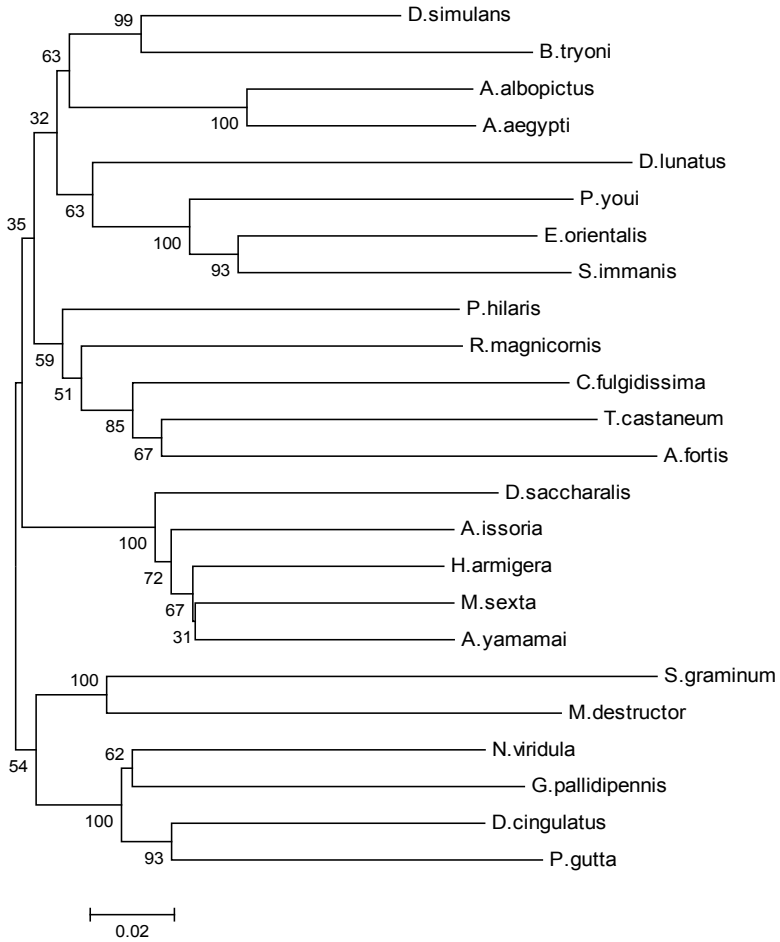

Fig 4. Phylogenetic Tree COI

S.immanis & P.youi, E.orientalis & P.youi and P.youi & A.aegypti with distance values of 1.444, 1.275, 1.208 and 1.127 respectively. The species *D.saccharalis & M.destructor* had a distance of 0.305 between them followed by M.*sexta* & *M.destructor, M.destructor & S.graminum, H.armigera & M.destructor* and *M.destructor & D.cingulatus* with a distance values of 0.306, 0.323, 0.327 and 0.327 respectively (Fig 3). The average distance separating the species was 0.664.

Evolutionary tree involving COI Seqeunces

The tree derived from the CO1 sequences (Figure. 4) placed their host insect species in to taxonomic groups similar to that

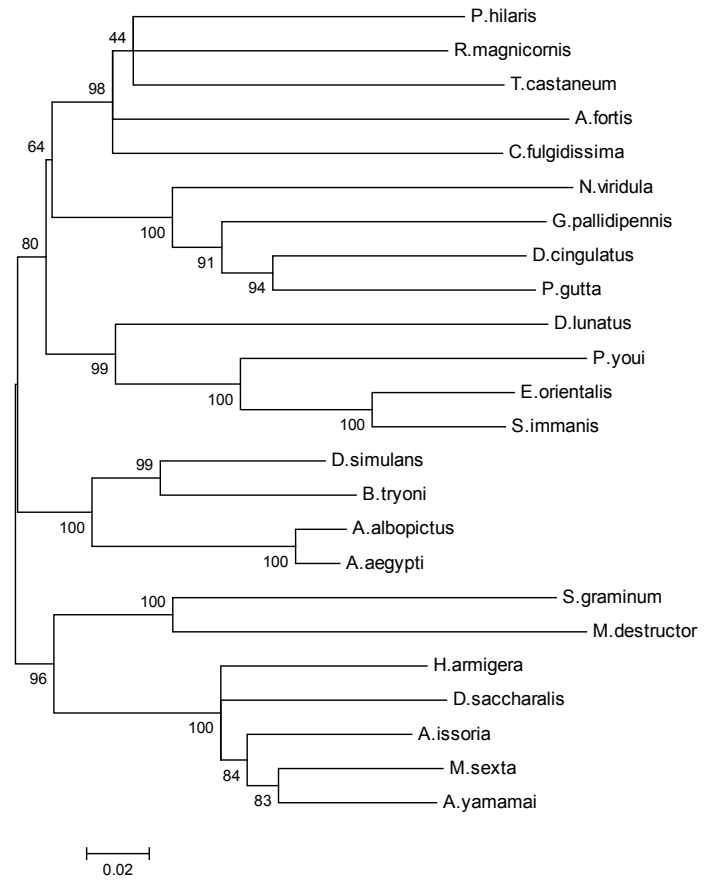

Fig. 5. Phylogenetic Tree 16srRNA

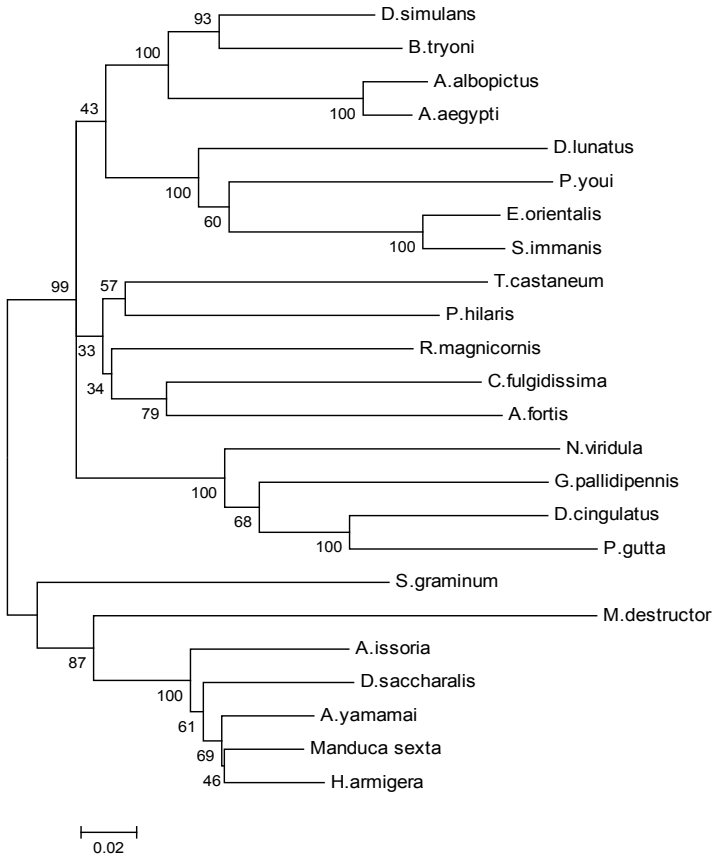

Fig. 6. Phylogenetic Tree 12srRNA

of morphologically derived characteristics, with an exception of *M.destructor* (Diptera) which was branched together with the Hemiptera species *S.graminum* (Hemiptera) with a 100% bootstrap value suggesting a close relationship between these two and a good indication about the evolution. All the other species were placed in to their respctive taxonomic groups. As expected the *A.aegypti* and *A.albopictus* were placed under a single roof with a bootstrap value 100% and D*.simulans* and *B.tyroni* with a bootstrap value of 99%, very high intra Order bootstrap values than the others. The two species with lowest bootsrap value of 32% were the lepidopteran M.*sexta* and *A*.*yamamai*. Further analysis of the CO1 tree gives us a glimpse that, the Diptera and Coleoptera shared best relationship with a bootstrap value of 32% than the other Orders.

Evolutionary tree involving 16srRNA Seqeunces

The phylogenetic tree obtained based on the large subunit sequences (16srRNA) is shown in figure 5. Similar results were obtained with the 16srRNA sequences as that of the CO1 sequences were a 100 % bootstrap relationship between the A.*aegypti* & A.*albopictus* (Diptera) and M.*destuctor* (Diptera) & S.*graminum* (Hemiptera). Further, the Ephemeroptera species' E.*orientals* & S.*immanis* also shared 100% value, hence placed under a branch. The analysis with these sequences was in tune with that of the CO1 sequences analysis. Strengthening the fact that the mitochondrial 16sRNA and CO1 are the molecular markers for species identification.

Evolutionary tree involving 12srRNA Seqeunces:

Simialr analysis was extended to mitochondrial small subunit sequnces (12srRNA) and thetree is shown in figure 6. The species belonging to the genus Aedes again were found have 100% bootstrap value alongwith the *Hemipteran P.gutta & D.cingulaus, Ephemeropteran E.orientalis & S.immanis.* This time the interspecies bootstrap was 87% between *M.destructor & S.graminum.* All the trees constructed using differenct sequence data resulted in almost identical results.

DISCUSSION

The insect genomes are closed circular and have the bacterial universal genetic code. The genome size varies from 17522 bp (Rhopaea magnicornis) to 14592 bp (Geocoris pallidipennis) (Table.1). The mtDNA from Hemiptera, Coleoptera and Diptera have minor differences in size within each order. For example, Chrysochroa fulgidissima is 1,930 bp shorter than Rhopaea magnicornis. Variation in mitochondrial size is due in some cases to variation in the repeat length of noncoding regions. However, despite the size differences, the gene order is not altered. The mtDNA size in the orders Diptera, Hemiptera, Coleoptera & Ephemeroptera are similar, but Lepidoptera. The mean size of mtDNA was higher for Coleoptera at 16204.4 bp and the lowest was Lepidoptera at 15387.2 bp. The lowest mean GC content was found to be among the Lepidopteran species with a mean value as low as 19% where as the highest was recorded among the Ephemeropteran species and the value stood at 29.5%. As observed in Cochliomyia hominivorax (Diptera), mtDNA has a bias in nucleotide composition, which perhaps has led to an AT rich genome (Lessinger, 2000). This bias is higher at the third position in the codons of protein coding genes due to lower selection and mutation (Jermiin, 1994). For all other orders, the A+T composition is very close to the mean observed in insect mtDNA, at 77% (Stewart, 2005). From our comparative studies, differences in the genes for tRNA have been observed. The typical gene content for the insect mtDNA was observed for all 24 insect species. Almost all species of insect mtDNA encode large and small rRNA, 22 tRNAs, and 13 to 14 polypeptides. As many as one additional tRNA coding gene was found to be present in *Parafronurus youi* & *Acraea issoria*, coding for methionine & isoleucine respectively.

GC content was higher in CO1 than the others, which was expected from the protein coding gene. Within a long region of genomic sequence, genes are often characterized by having a higher GC-content in contrast to the background GC-content for the entire genome. Evidence of GC ratio with that of length of the coding region of a gene has shown that the length of the coding sequence is directly proportional to higher G+C content (Pozzoli, 2008). The AT content was almost found to be uniform among the data. This has been pointed to the fact that the stop codon has a bias towards A and T nucleotides, and, thus, the shorter the sequence the higher the AT bias (Wuitschick, 1999). The evolutionary relationship between the species was done by constructing phylogenetic trees using the MEGA tool based on the neighbor-joining method. Separate trees were constructed with three different datasets. It was inferred from the trees that all the sequence datasets yielded almost similar trees with the members of the order forming a clade a monophyly. An interesting aspect noted from these trees was the hemipteran species *S.graminum* and the Dipteran *M. sexta* favoring to form a seperate clade as they were branched together seperately with a heavy bootstrap value close to 100 and as expected the Dipteran species A.*aegypti* and *A.albopictus* were placed under a single roof with a bootstrap value 100% and *D.simulans* and *B.tyroni* with a bootstrap value of 99%, very high intra Order bootstrap values than the others. Further research involving comparative genome analysis of a large number of insect orders with substantial number of insect species will bring light on the possibility of multiple mitochondrial ancestors.

REFERENCES

Angilla Aline F, Luis HS Gil, Luis HP da Silva, Paulo EM Ribolla, (2007). Population structure of the malaria vector Anopheles darlingi in Rendonia, Brazilian Amazon, based on mitochondrial DNA. Mem. Inst. Oswaldo Cruz 102, 953-958.

Avise, 2004. Molecular Markers, natural history and evolution. 2nd edition. Sunder land, Massachusetts: sinauer Associates; 2004

Ayala, F. J. 1986. On the virtues and pitfalls of the molecular evolutionary clock. *J. Hered.,* 77: 226-235.

Black, W.C. and J. Piesman, 1994. Phylogeny of hard- and soft-tick taxa (Acari: Ixodida) based on mitochondrial 16S rRNA sequences. Proceedings of the National Academy of Sciences, USA 91:10034–10038.

Bond, J.E., M.C. Hedin, M.A. Ramirez and B.D. Opell. 2001. Deep molecular divergence in the absence of morphological and ecological change in the Californian

coastal dune endemic trapdoor spider *Aptostichus simus*. *Molecular Ecology,* 10: 899–910.

Boore, 1999. Animal Mitochondrial genomes Nucleic Acids. *Res*., 27: 1767-1780

Cameron S. L., Lambkin C. L., Barker S. C. and Whiting M. F. 2007. A mitochondrial genome phylogeny of Diptera: whole genome sequence data accurately resolve relationships over broad timescales with high precision. *Syst. Entomol.,* 32, 40–59

Cameron SL, Beckenbach AT, Dowton M and Whiting MF, 2006. Evidence from mitochondrial genomics on interordinal relationships in insects. *Arthropod Systematics & Phylogeny,* 2006, 64:27-34.

Cameron SL, Dowton M, Castro LR, Ruberu K, Whiting MF, Austin AD, Diement K and Stevens J, 2008. Mitochondrial genome organization and phylogeny of two vespid wasps. *Genome*, 2008, 51:800-808.

Cummings and Krafsur 2005. Spatial diversity in Mitochondrial Cytochrome c oxidase in houseflies. *Med.Vet. Entomol.,* 19, 53-59.

Dellaporta SL, Xu A, Sagasser S, Jakob W, Moreno MA, Buss LW, Schierwater B, 2006. Mitochondrial genome of Trichoplax adhaerens supports placozoa as the basal lower metazoan phylum. *Proc. Natl. Acad. Sci.,* USA 2006, 103:8751-8756.

DeSalle, R., J. Gatesy, W. Wheeler and D. Grimaldi. 1992. DNA Sequences from a fossil termite in Oligo-Miocene amber and their phylogenetic implications. *Science*, 257:1933–1936.

Felsenstein J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*, 39:783-791.

Fenn JD, Song H, Cameron SL and Whiting MF, 2008. A preliminary mitochondrial genome phylogeny of Orthoptera (Insecta) and approaches to maximizing phylogenetic signal found within mitochondrial genome data. *Mol Phylogenet Evol.,* 2008, 49:59-68

Flook, P.K. and C.H.F. Rowell, 1995. The effectiveness of mitochondrial rRNA gene sequences for the reconstruction of the phylogeny of an insect order (Orthoptera). *Molecular Phylogenetics and Evolution,* 8:177–192

Geer LY, Marchler-Bauer A, Geer RC, Han L, He J, He S, Liu C, Shi W and Bryant SH. 2010. The NCBI BioSystems database. Nucleic Acids Res. 2010 Jan; 38 (Database issue): D492-6.

Hall MJ, Edge W, Testa JM, Adams ZJ and Ready PD, 2001. Old world screwworm fly, Chrysomya bezziana, occurs as two geographical races. *Med. Vet. Enomol.,* 15, 393-402.

Hwang UW, Park CJ, Yong TS and Kim W. 2001. One-step PCR amplification of complete arthropod mitochondrial genomes. *Mol Phylogenet Evol.,* 19:345–352.

Jermiin LS, Graur D, Lowe RM and Crozier RH, 1994. Analysis of directional mutation pressure and nucleotide content in mitochondrial cytochrome b genes. *J. Mol. Evol.,* 39: 160-73.

Lessinger AC, Martins JAC, Lemos TA, Kemper EL, da Silva FR, Vettore AL, Arruda P and Azeredo-Espin AM 2000. The mitochondrial genome of the primary screwworm fly Cochliomyia hominivorax (Diptera: Calliphoridae). *Insect Mol Biol.,* 9: 521-9.

Lunt DH, Zhang DX, Szymura JM and Hewitt GM. 1996. The insect cytochrome oxidase I gene: evolutionary patterns and conserved primers for phylogenetic studies. *Insect Mol. Biol.,* 5, 153-165.

Pozzoli U, Menozzi G and Fumagalli M, 2008. Both selective and neutral processes drive GC content evolution in the human genome. *BMC Evol. Biol.,* 8: 99.

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, volume 4, issue 4, pp. 406-425, July 1987.

Segura MD, Callejas C, Fernández MP and Ochando MD 2006. New contributions towards the understanding of Phylogenetic relationships among economically important fruit flies (Diptera: Tephritidae). *Bull. Entomol. Res.,* 96: 279-288.

Shao and Barker, 2007. Mitochondrial genomes of parasitic arthopods: implications for studies of population genetics and evolution. *Parasitology,* 134: 153-167

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H. and Flook, P 1994. Evolution, Weighting and phylogenetic utility of nmitochondrial gene sequence and a compilation of conserved polymerase chain reaction primers. *Ann. Entomol. Soc. Am.,* 87: 651-701

Stewart JB and Beckenbach AT 2005. Insect mitochondrial genomics: the complete mitochondrial genome sequence of the meadow spittlebug *Philaenus spumarius* (Hemiptera: Auchenorrhyncha: Cercopoidae). *Genome*, 48: 46-54.

Tamura K, Dudley J, Nei M and Kumar S 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution,* 24: 1596-1599.

Thollesson, M., 1999. Phylogenetic analysis of Euthyneura (Gastropoda) by means of the 16S rRNA gene: use of a fast gene for ┐ higher-level phylogenies. Proc. R. Soc. London B 266, 75–83.

Wilson, A.C., Carlson, S.S. and White, T. J. 1977. Biochemical evolution. *Annu. Rev. Biochem.,* 46:473-639.

Woese, 1987. Bacterial evolution. *Microbial Rev.,* SI: 221-271.

Wuitschick JD, Karrer KM, 1999. Analysis of genomic G + C content, codon usage, initiator codon context and translation termination sites in *Tetrahymena thermophila*". *J. Eukaryot. Microbiol.,* 46 (3): 239–47.

Yamauchi MM, Miya MU and Nishida M. 2004. Use of a PCRbased approach for sequencing whole mitochondrial genomes of insects: two examples (cockroach and dragonfly) based on the method developed for decapod crustaceans. *Insect Mol Biol.,* 13:435–442.

Zehner R, Amendt J, Schütt S, Sauer J, Krettek R and Povolný D, 2004. Genetic identification of forensically important flesh flies (Diptera: sarcophagidae). *Int. J. Legal Med.*, 118, 245-247.

******