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

International Journal of Current Research
Vol. 9, Issue, 02, pp.47060-47072, February, 2017

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

ISSN: 0975-833X

RESEARCH ARTICLE

MOLECULAR PHYLOGENETICS OF GLOBAL AND INDIAN MOSQUITO SPECIES IN TERMS OF BIOINFORMATICS AND BIOSTATISTICS

***,1Divya Damodaran, 1Dr. Sudarsanam, D., 1Maria Infanshia Sajini, 1Vasantha Suppriya, 1Joel Singh, D. and 2Siddhardha Solosan**

1Department of Advanced Zoology and Biotechnology, Loyola College, University of Madras, Chennai
2Govt .PHC, Medavakkam, Chennai-600100

ARTICLE INFO

Article History:
Received 20th November, 2016
Received in revised form
14th December, 2016
Accepted 16th January, 2017
Published online 28th February, 2017

Key words:

DNA barcoding,
Bioinformatics,
Phylogenetic analysis,
Timetree.

ABSTRACT

Recent studies on DNA-based approaches show a promising trend in the rapid description of biodiversity. Traditional morphology-based assessments are timeconsuming and require specialists whose numbers are insufficient and dwindling. A DNA-based method called DNA barcoding has been proposed as a rapid means of cataloguing species. Hebert *et al.,* (2003) specifically suggest the employment of DNA sequences as taxon 'barcode's and propose that the mitochondrial gene cytochrome oxidase I (COI) serves as the core of a global bio identification system for animals. The unique contribution of DNA barcoding to mosquito taxonomy and systematic is a compressed timeline for the exploration and analysis of biodiversity. According to Knowlton and Weigt (1998) among the mitochondrial genes, cytochrome oxidase subunit 1 (COI) is to be the most conserved gene in the amino acid sequences and hence has distinct advantage for taxonomic studies. The morphological identification keys used currently for identification of mosquitoes are mainly related to imaginal and fourth instars only. Hence, difficult to identify other stages of development collected in the field, if not reared in the laboratory.

Citation: Divya Damodaran, Dr. Sudarsanam, D., Maria Infanshia Sajini, Vasantha Suppriya, Joel Singh, D. and Siddhardha Solosan, 2017. "Molecular phylogenetics of global and indian mosquito species in terms of bioinformatics and biostatistics", *International Journal of Current Research*, 9, (02), 47060-47072.

INTRODUCTION

Mosquitoes are nematocerid flies that belong to the insect order Diptera and family Culicidae. There are around 3500 different Culicidae species in the world (Harbach and Howard, 2007). Massive hatching of river flood land mosquitoes, like Aedes (*Ochlerotatus) sticticus* and *Aedescinereus (Ae. cinereus)*, cause nuisance for humans and animals every year. The main vectors are considered to be species from genus *Culex*-and *Culiseta* (Lundstrom *et al*., 1990), but recent research have shown that *Ae. Cinereus* maybe just as important for the virus distribution (Lundstrom *et al*., 2007; Turell *et al*., 1990).Reducing taxonomy, laden as it is with theory and knowledge, to a high-tech identification service industry denies its many levels of rigorous hypothesis-testing, from characters to species to clades, and impoverishes the wider information base that is crucial for describing biodiversity (Lipscomb *et al*., 2003; Wheeler 2004; Godfray and Knapp 2004; Scoble 2004; Wilson 2004). If the climate changes to rainier and warmer in northern Europe, it´s possible that vectors like *Anopheles* could start to carry *Plasmodium* species again (Jaenson *et al*., 1986). DNA barcoding, which relies on the genetic variation within a standardized regionof the mitochondrial cytochrome *c* oxidase subunit I (COI) gene. Based on patterns of nucleotide variation in the COI barcode region (single linkage cluster analysis followed by Markov clustering), operational taxonomic units (OTUs) are assigned in BOLD to individuals for analysing animal biodiversity based on a molecular operational taxonomic unit (MOTU) algorithm (Jones *et al*., 2011).

MATERIALS AND METHODS

Basic Descriptive Statistics

The minimal statistics that are used to describe the data are called basic descriptive statistics. They provide simple summaries about the sample and the measures. Together with simple graphics analysis, they form the basis of virtually every quantitative analysis of data. Quantitative descriptions can be presented by studying the descriptive statistics of the data. Mean or average is probably the first described statistical method done on the data. To compute mean one just need to add up all the values and divide by the number of values.

****Corresponding author:*** **Divya Damodaran,**
Department of Advanced Zoology and Biotechnology, Loyola College, University of Madras, Chennai

Median is the value exactly in the middle of the data. Mode is the most frequently occurring value in the data. When discrepancy of a value from the mean value is squared and the square values are summed up, it is called Sum of Squares (SS). Skewness is a measure of symmetry, or more precisely, the lack of symmetry. A distribution, or data set, is symmetric if it looks the same to the left and right of the center point. Kurtosis is a measure of whether the data are heavy-tailed or light-tailed relative to a normal distribution. That is, data sets with high kurtosis tend to have heavy tails, or outliers. Data sets with low kurtosis tend to have light tails, or lack of outliers. A uniform distribution would be the extreme case. The histogram is an effective graphical technique for showing both the skewness and kurtosis of data set. The kurtosis of a standard normal distribution is three. Significant skewness and kurtosis clearly indicate that data are not normal. If a data set exhibits significant skewness or kurtosis. In this case, one approach is to apply some type of transformation to try to make the data appear more normal. Taking the log or square root of a data set is often useful for data that exhibit moderate right skewness. Another approach is to use techniques based on distributions other than the normal.

Basic descriptive statistics in SAS

PROC MEANS is used in a variety of analytic, business intelligence, reporting and data management situations. PROC MEANS capabilities may be employed in “data cleansing” or “exploratory data analysis” tasks to determine if incorrect or “bad” values of analysis variables are contained in the data set that must be transformed or removed prior to further analysis. PROC MEANS is included in the BASE Module of SAS System Software.

```
PROC MEANS noprint DATA=Aedes SKEW KURTOSIS
MEAN STD T PROBT;
OUTPUT OUT=Aedes_mean_global
SKEW=SKEWNESS_global
KURTOSIS=KURTOSIS_global
MEAN=MEAN_global
std=STD_global
T=NORMAL_global
PROBT=PVALUE_global;
varDistance_global;
RUN;
```

DATA= - Specify data set to use
NOPRINT - Do not print output

VAR variable - specifies which numeric variables to use
OUTPUT OUT = datasetname – statistics will be output to a SAS data file

SKEW= - specifies the name of the column to be assigned for the Skew column produced in Proc Means procedure

KURTOSIS= - specifies the name of the column to be assigned for the Kurtosis column produced in Proc Means procedure

MEAN - Arithmetic average

STD - Standard Deviation

Line overlay plots

A line plot is a graphical display of data along a number line with symbols connected by a line. The symbols represent data points. The symbols can represent frequency. It is best to use a line plot when comparing fewer than 25 numbers. It is a quick, simple way to organize data. A line plot will have outliers. An outlier is a number that is much greater or much less than the other numbers in the data set. Outliers are usually represented without any data transformation. A line plot consists of a horizontal line which is the x-axis with equal intervals. It is important for a line to plot to have a title and a label of the x-axis to provide the reader an overview of what is being displayed. Also, line plots must have legends to explain what is being measured.

Line overlay plots in SAS

The GPLOT procedure plots the values of two or more variables on a set of coordinate axes (X and Y). The coordinates of each point on the plot correspond to two variable values in an observation of the input data set. The GPLOT procedure creates a temporary SAS data set that is used to generate an image map in an SVG file when you are sending output to the LISTING destination. (This option is not necessary when you are sending output to the HTML destination.) The drill-down URLs in the image map must be provided by variables in the input data set.

```
procgplot data=Aedes;
title"Aedes";
SYMBOL1 I=JOIN V=DOT WIDTH=1 HEIGHT=1
CV=GREEN CI=GREEN; /*Global*/
SYMBOL2 I=JOIN V=DOT WIDTH=1 HEIGHT=1
CV=BLUE CI=BLUE; /*Indian*/
SYMBOL3 I=JOIN V=DOT WIDTH=1 HEIGHT=1
CV=BROWN CI=BROWN; /*Kerala*/
PLOT
(distance_globaldistance_indiandistance_kerala)*distance
/ CAXIS=BROWN no legend CFRAME=LIGHTBLUE
overlay; run; quit;
```

DATA= - specifies the SAS dataset name from plots are to be done
TITLE “” – specifies title for the plot output
SYMBOL1, SYMBOL2, SYMBOL3 statements – specifies different output symbols and colours for different line plots
PLOT statement – specifies the list of numeric variables to be plotted in the overlay plot
OVERLAY option – specifies that many line plots should be shown in the same plot

RESULTS AND DISCUSSION

Phylogram of Global and Indian Mosquito Species

The evolutionary history was inferred using the Neighbor-Joining method. Divergence times for all branching points in the topology were calculated with the RelTime method using the branch lengths contained in the inferred tree. The analysis involved 15 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 500 positions in the final dataset. Evolutionary analyses were conducted in MEGA6.

The evolutionary history was inferred using the Neighbor-Joining method. Divergence times for all branching points in the topology were calculated with the RelTime method using the branch lengths contained in the inferred tree. The analysis involved 16 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 490 positions in the final dataset. Evolutionary analyses were conducted in MEGA6. The evolutionary history was inferred using the Neighbor-Joining method. Divergence times for all branching points in the topology were calculated with the RelTime method using the branch lengths contained in the inferred tree. The analysis involved 28 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 611 positions in the final dataset. Evolutionary analyses were conducted in MEGA6.

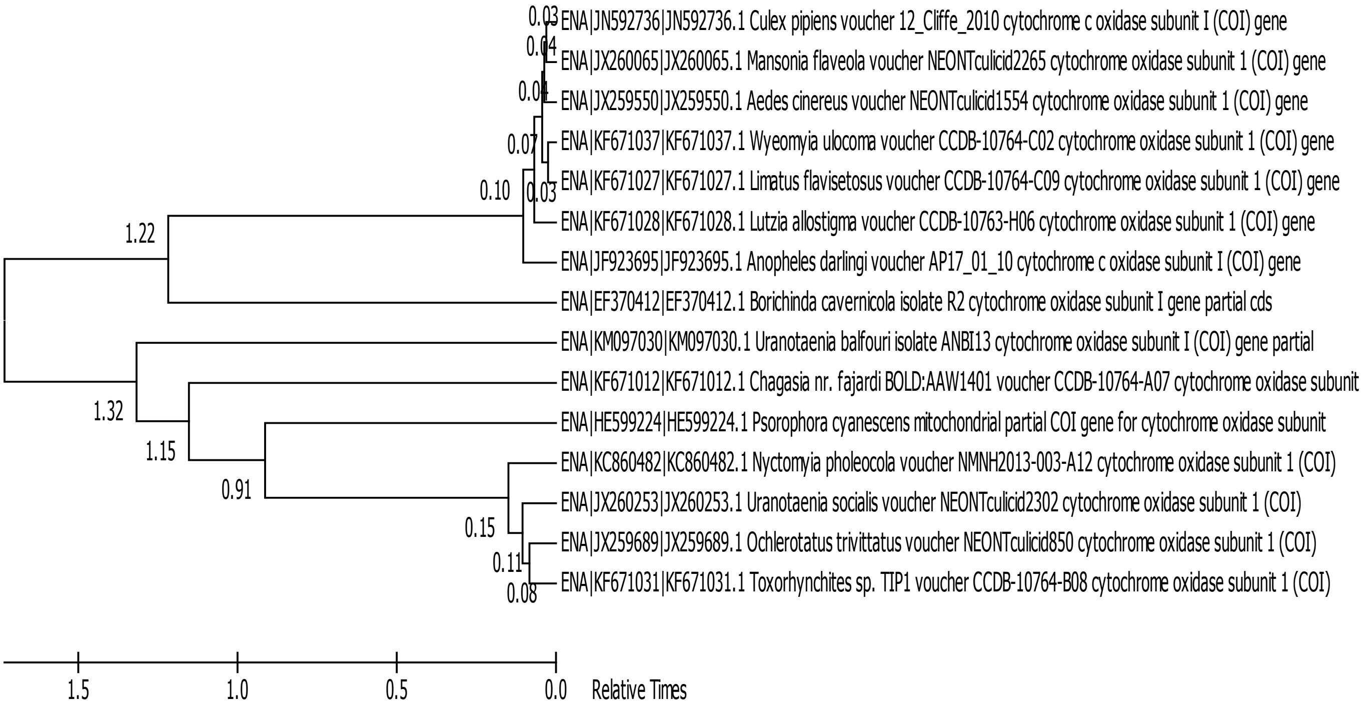


Fig. 1. Time tree for global mosquito species showing evolutionary relationship of Taxa

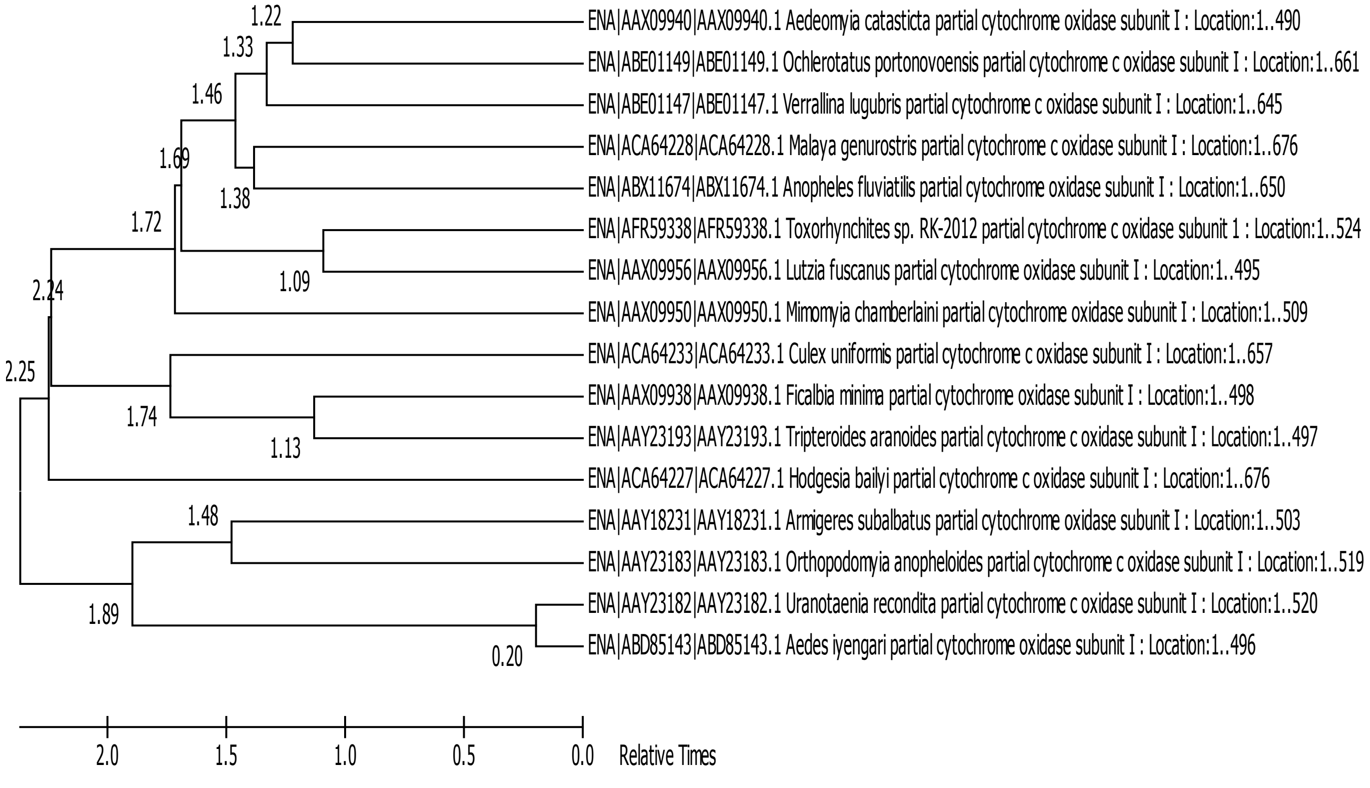


Fig. 2. Time tree for Indian mosquito species showing evolutionary relationship of Taxa

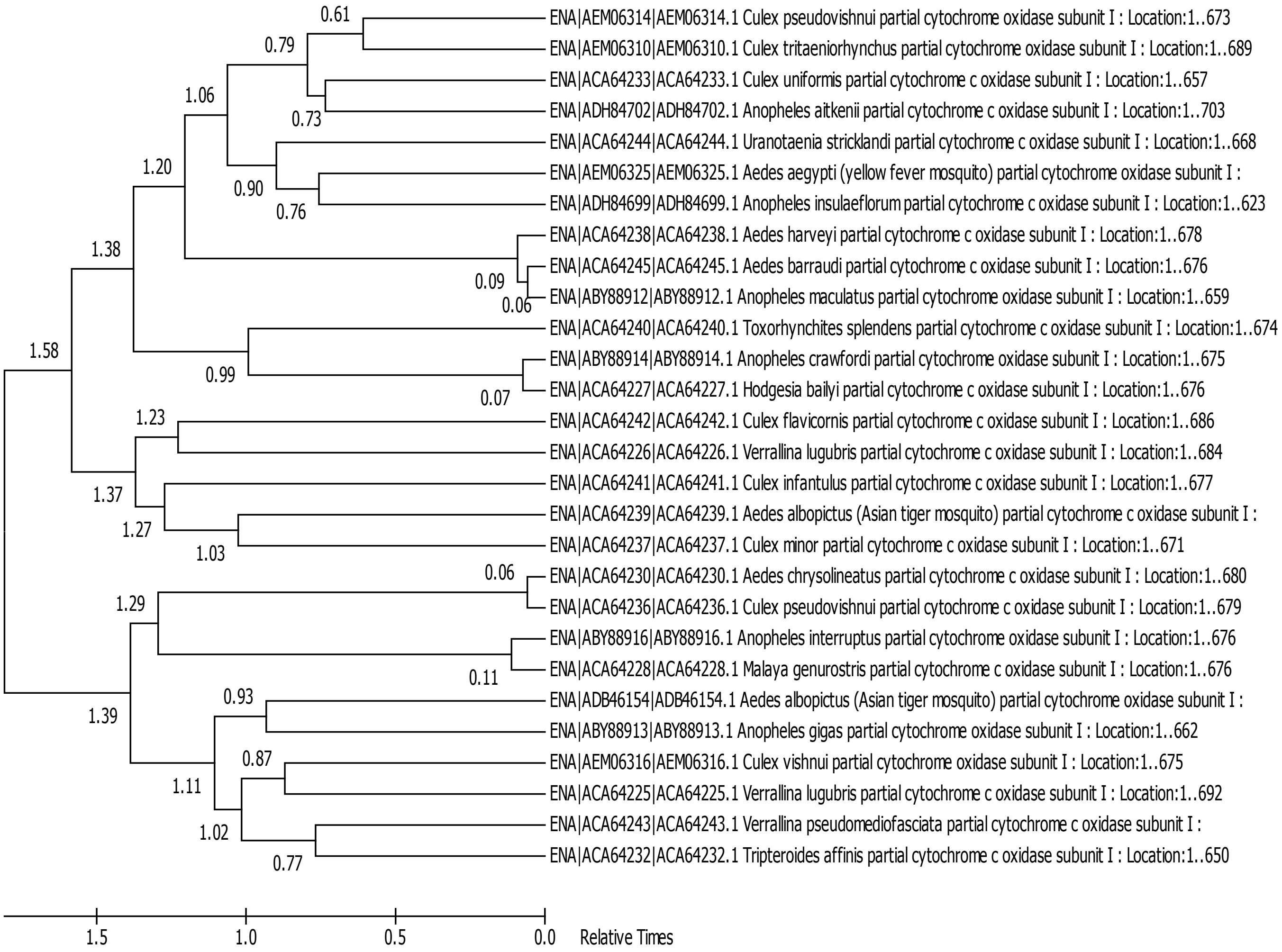


Fig.3. Time tree for Kerala mosquito species showing evolutionary relationship of Taxa

Time tree value analysis across genus

The following table shows the distribution of genus across the globe, in India and in Kerala:

Table 1. Distribution of genus across the globe, in India and the Indian state of Kerala

Global genus	Indiangenus	Kerala genus
Aedes	Aedeomyia	Aedes
Anopheles	Aedes	Culex
Borichinda	Anopheles	
Chagasia	Armigeres	
Culex	Ficalbia	
Limatus	Hodgesia	
Lutzia	Lutzia	
Mansonia	Malaya	
Nyctomyia	Mimomyia	
Ochlerotatus	Ochlerotatus	
Psorophora	Orthopodomyia	
Uranotaenia	Toxorhynchites	
Wyeomyia	Tripteroides	
	Uranotaenia	

Basic descriptive statistics of time tree values of all genus

The mean and standard deviation values for all the three groups are tabulated below.

The time tree values of global and Indian species are largely negatively skewed. The Kerala species which are distributed only in Aedes and Culex genus are negatively skewed.

Scatter plots - Comparing the time tree values of the three groups

Further, scatter plots were obtained for the time tree values of each genus. It should be noted here that the top hits of time tree values are taken for the analysis though the genus may actually be present in all the 3 groups – Global, Indian and Kerala. In the scatter plots below, the green line represents time tree values of global species, the blue line represents time tree values of Indian species and the brown line represents time tree values of Kerala species.

DISCUSSION

Average of time tree values for each species is obtained and tabulated as under.

A line plot is plotted across various species. There is lesser variation across the species suggesting that the mosquito species that are chosen for the analysis are closely related to each other. For example, the least average time tree value is 0.74 and the highest average time tree value is 1.45.

Table 2. **The mean and standard deviation values for all the three groups**

Genus	*No of species*	*MEAN_Global*	*STD_Global*	*MEAN_Indian*	*STD_Indian*	*MEAN_Kerala*	*STD_Kerala*
Aedeomyia	*14*			*1.942857143*	*0.381987285*		
Aedes	*84*	*0.9007692*	*0.688313*	*2.25*	*0*	*1.35028986*	*0.386211*
Anopheles	*13*	*0.9808333*	*0.653765*	*2.24*			
Armigeres	*11*			*2.11181818*	*0.253449*		
Borichinda	*6*	*1.2966667*	*0.118772*				
Chagasia	*3*	*1.35*	*0.173205*				
Culex	*50*	*0.8435714*	*0.695077*			*1.27*	*0.378599*
Ficalbia	*13*			*2.11923077*	*0.3282393*		
Hodgesia	*9*			*2.24333333*	*0.005*		
Limatus	*1*	*1.45*					
Lutzia	*9*	*1.095*	*0.71*	*2.024*	*0.3049262*		
Malaya	*4*			*2.03*	*0.433359*		
Mansonia	*8*	*1.08*	*0.619285*				
Mimomyia	*12*			*2.02666667*	*0.2707341*		
Nyctomyia	*5*	*1.13*	*0.563028*				
Ochlerotatus	*20*	*1.006*	*0.623435*	*1.977*	*0.3544025*		
Orthopodomyia	*6*			*2.13*	*0.1859032*		
Psorophora	*11*	*1.1936364*	*0.245613*				
Toxorhynchites	*7*			*1.92285714*	*0.4516162*		
Tripteroides	*8*			*2.24375*	*0.0051755*		
Uranotaenia	*19*	*1.219375*	*0.435515*	*1.56666667*	*1.1835681*		
Wyeomyia	*2*	*0.74*	*1.004092*				

Table 3. The skewness and kurtosis for the three groups

The skewness and kurtosis for the three groups are tabulated below.

Genus	Number of species	SKEWNESS _Global	KURTOSIS _Global	SKEWNESS _Indian	KURTOSIS _Indian	SKEWNESS _Kerala	KURTOSIS _Kerala
Aedeomyia	14			-0.6849835	-1.2016521		
Aedes	84	-0.5152402	-2.0415175			-2.2969239	5.2806279
Anopheles	13	-0.7796766	-1.6609453				
Armigeres	11			-1.9108551	3.24568033		
Borichinda	6	0.96824584	-1.875				
Chagasia	3	-1.7320509					
Culex	50	-0.304442	-2.2171646			-1.5878581	2.3230911
Ficalbia	13			-2.7884538	7.72430369		
Hodgesia	9			0.85714286	-1.7142857		
Limatus	1						
Lutzia	9	-2	4	-0.6077658	-3.332167		
Malaya	4			-1.9992919	3.99761546		
Mansonia	8	-1.3757992	-0.1048965				
Mimomyia	12			-0.3876615	-2.2621686		
Nyctomyia	5	-1.9677305	3.8648011				
Ochlerotatus	20	-0.9566333	-1.274727	-0.6491282	-1.7456221		
Orthopodomyia	6			-0.9682459	-1.875		
Psorophora	11	-0.1808625	-1.9789898				
Toxorhynchites	7			-1.1805478	0.5626596		
Tripteroides	8			0.64406119	-2.24		
Uranotaenia	19	-2.2932656	4.1715925	-1.7320509			
Wyeomyia	2						

Table 4. Average of time tree values for global species

Species	Average distance global
Aedes cinereus	0.901
Anopheles darlingi	0.981
Borichinda cavernicola	1.297
Chagasia nr fajardi	1.350
Culexpipiens	0.844
Limatusflavisetosus	1.450
Lutziaallostigma	1.095
Mansoniaflaveola	1.080
Nyctomyiapholeocola	1.130
Ochlerotatustrivittatus	1.006
Psorophoracyanescens	1.194
Uranotaeniabalfouri	1.219
Wyeomyiaulocoma	0.740

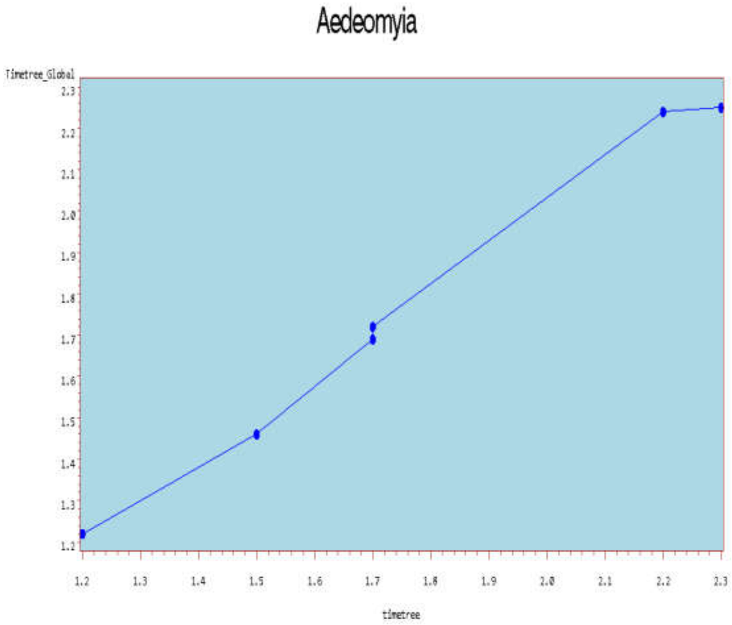


Fig. 4. Aedeomyia was found to be closer to other species among the Indian era

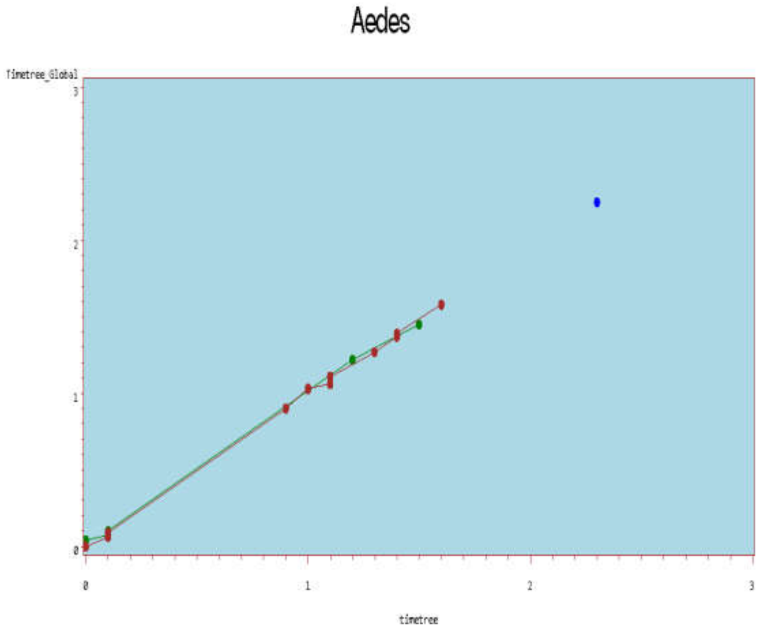


Fig. 5. Aedes species found across the globe, however it was prevalent in Kerala

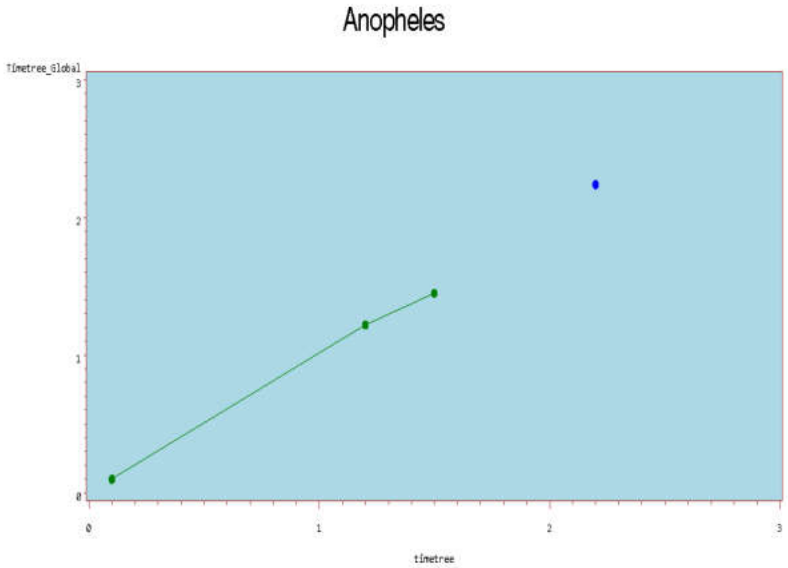


Fig. 6. The time tree values of Anopheles, globally

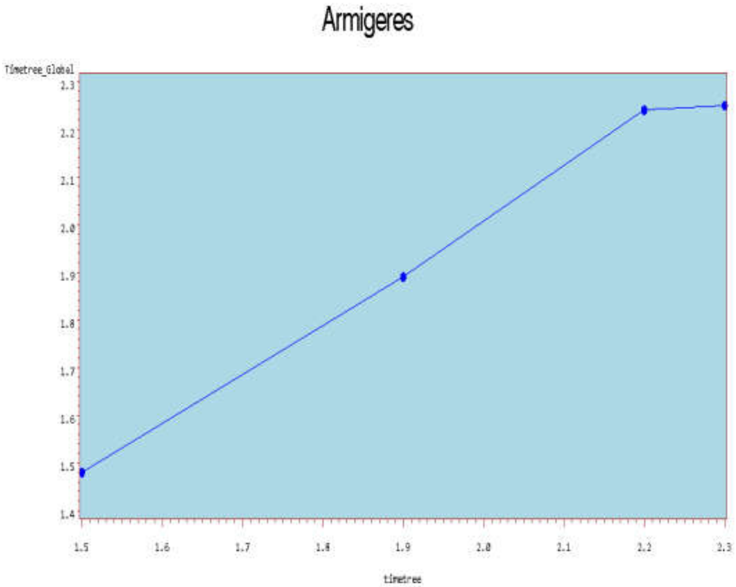


Fig. 7. The time tree values of Armigeres species, Indian context

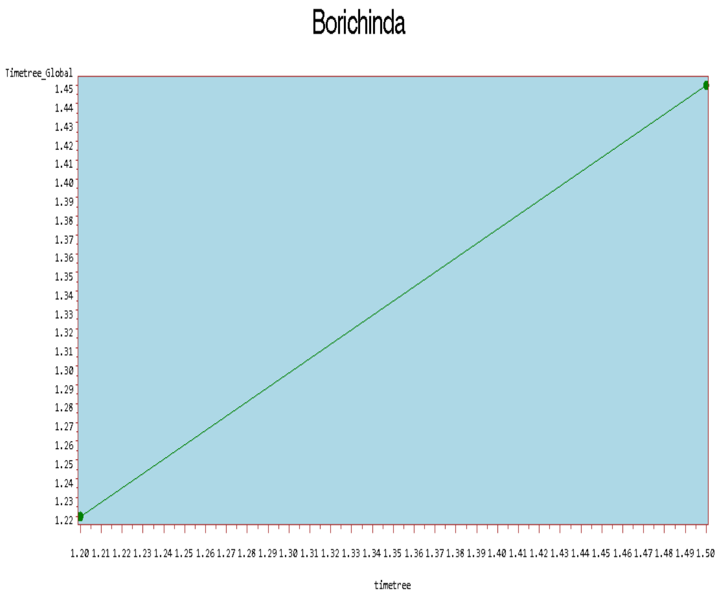


Fig. 8. Borichinda species found across the globe (but not in India)

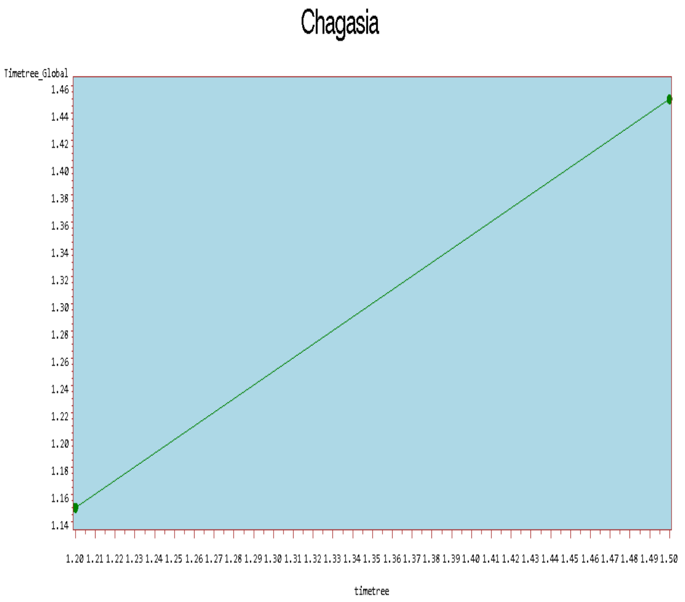


Fig. 9. Chagasia species are found across the globe (but not in India)

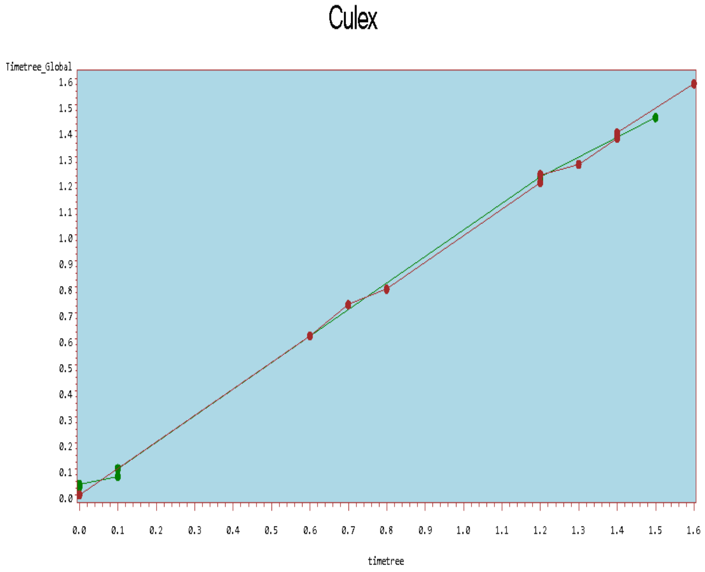


Fig. 10. Culex species are found across the globe however it was more prevalent in Kerala. This was evident from the time tree values being more in Kerala

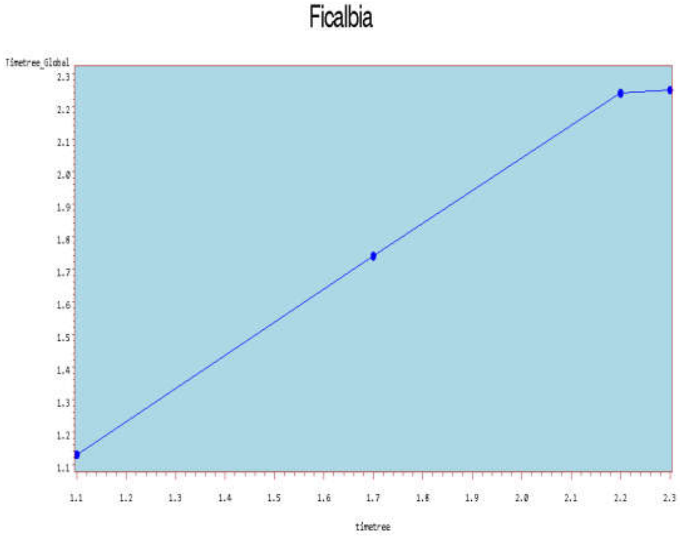


Fig. 11. Ficalbia species are closely related to Indian species

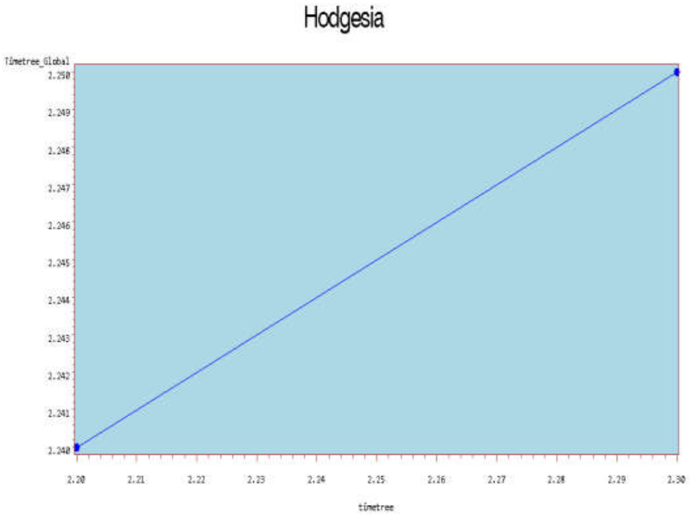


Fig. 12. Hodgesia species, found only in India

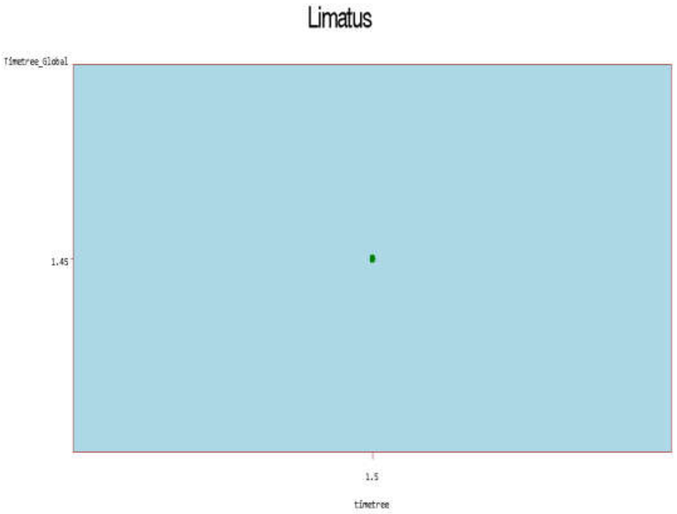


Fig. 13. Only one time tree value was obtained for the genus Limatus in the group of global mosquito species. (Limatus was not found to be related to any species in the Indian and Kerala species groups)

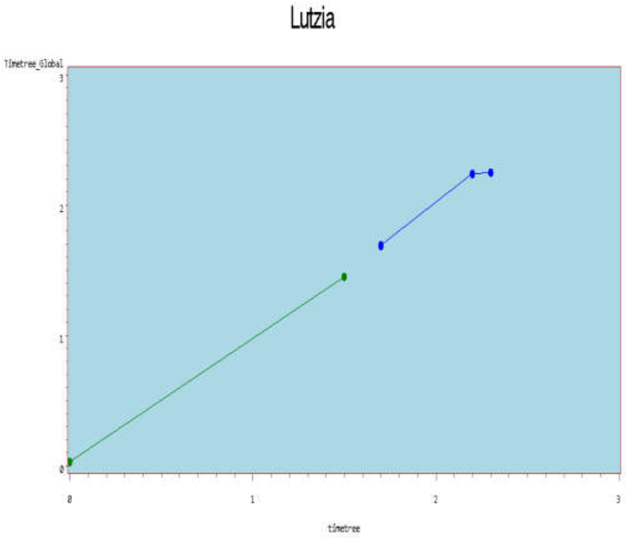


Fig. 14. Species of Lutzia were closely related to global and Indian species

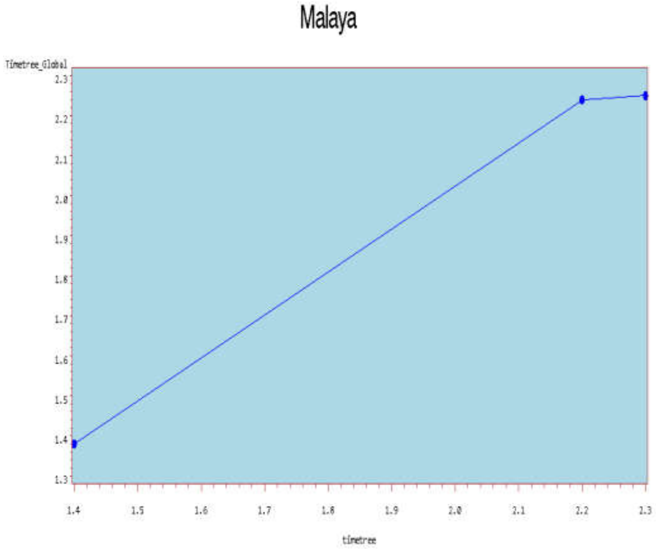


Fig. 15. Species of Malaya are present in the Indian species group with time tree value close to few other Indian species

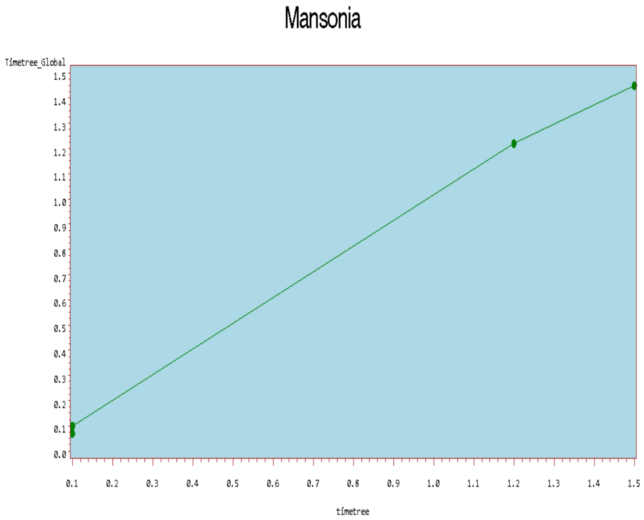


Fig. 16. Mansonia species is found among the global species

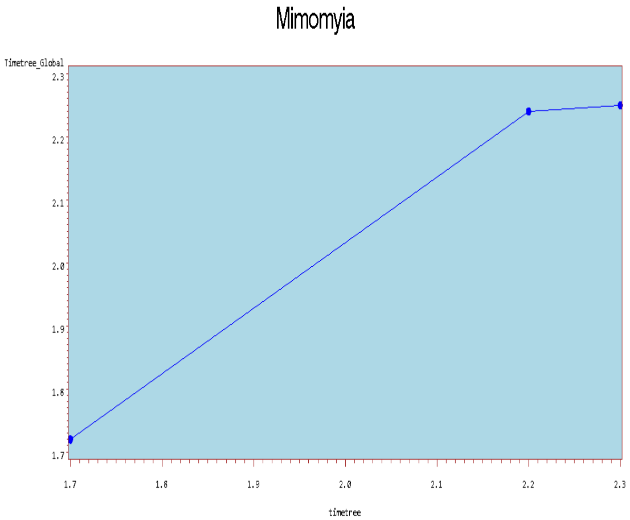


Fig. 17. Mimomyia species is found only in the India

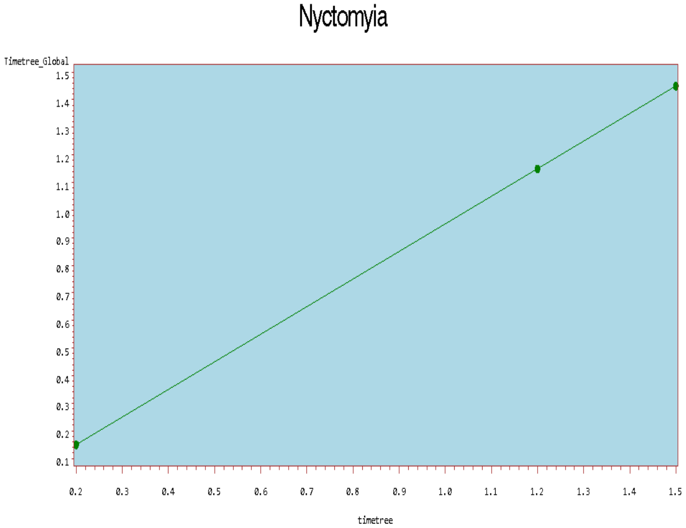


Fig. 18. Nyctomyiaspecies occur globally

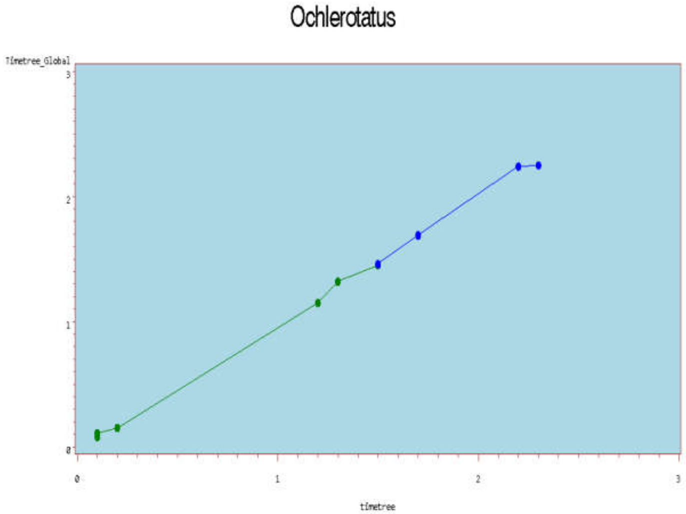


Fig. 19. Species of Ochlerotatus was found globally and India too

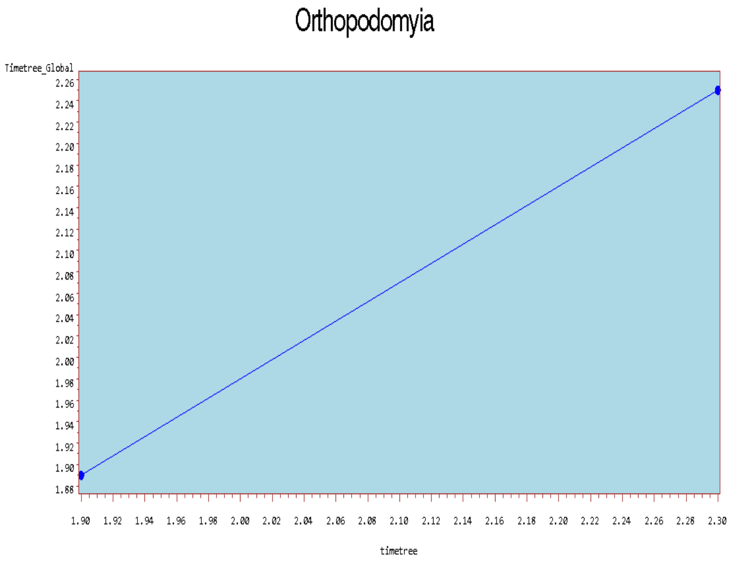


Fig. 20. Species of Orthopodomyiais found only in India

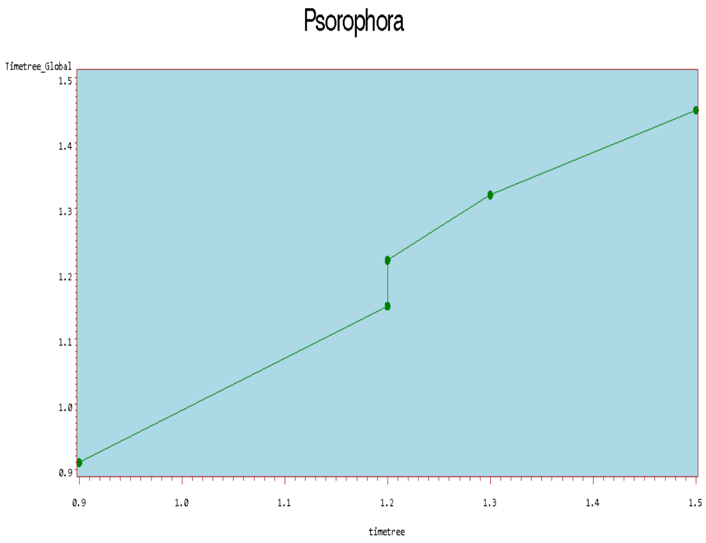


Fig. 21. Psorophora genus occurs globally

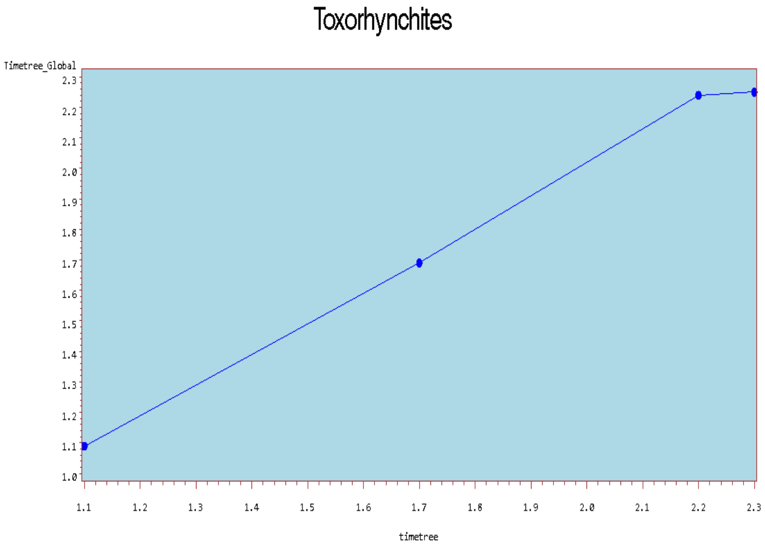


Fig. 22. Toxorhynchites genus occurs only in India

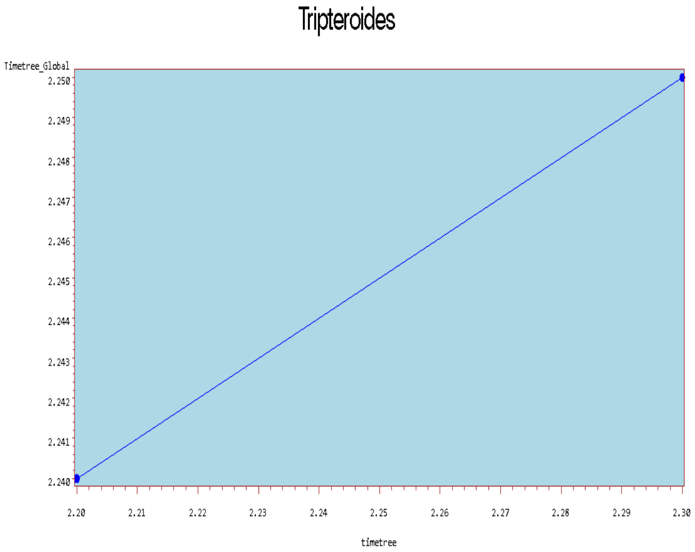
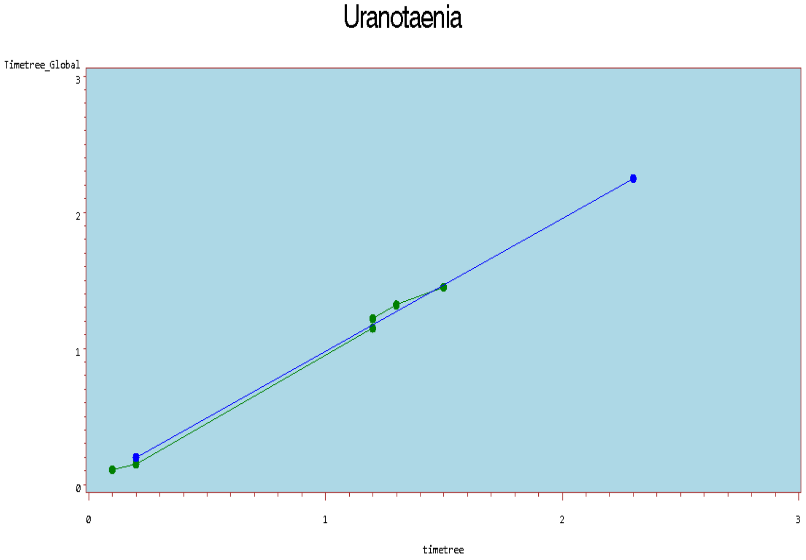


Fig. 23. Tripteroides genus occurs only in India



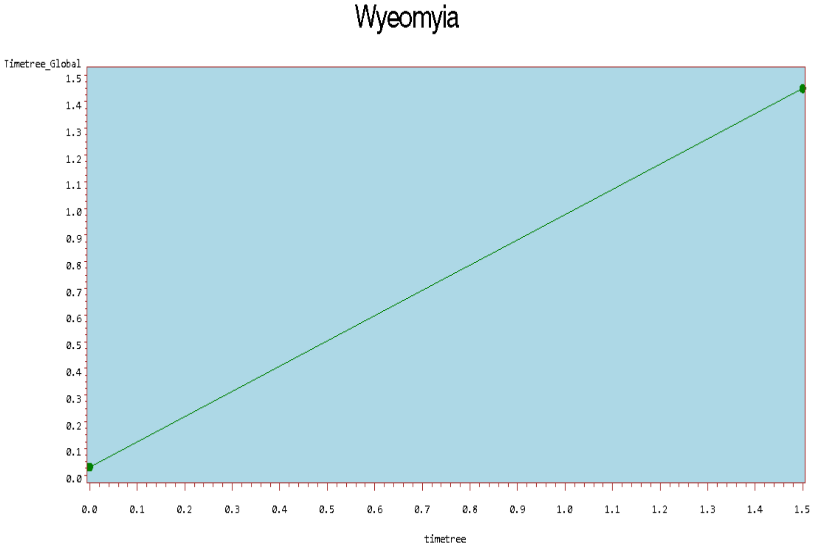
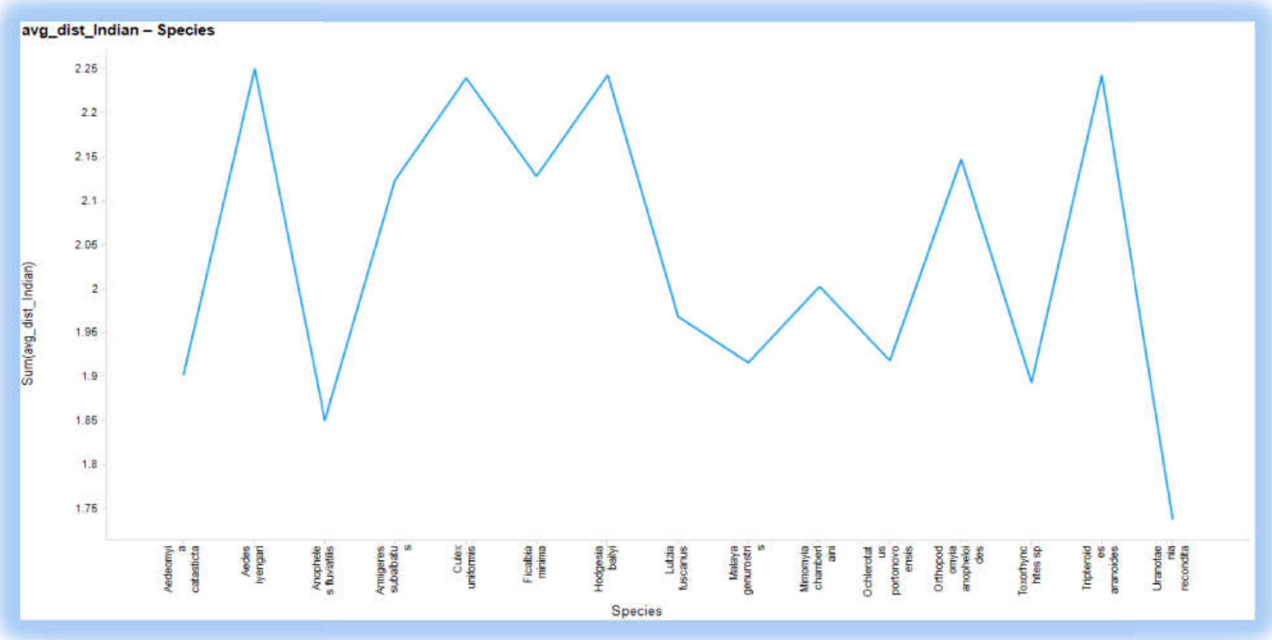
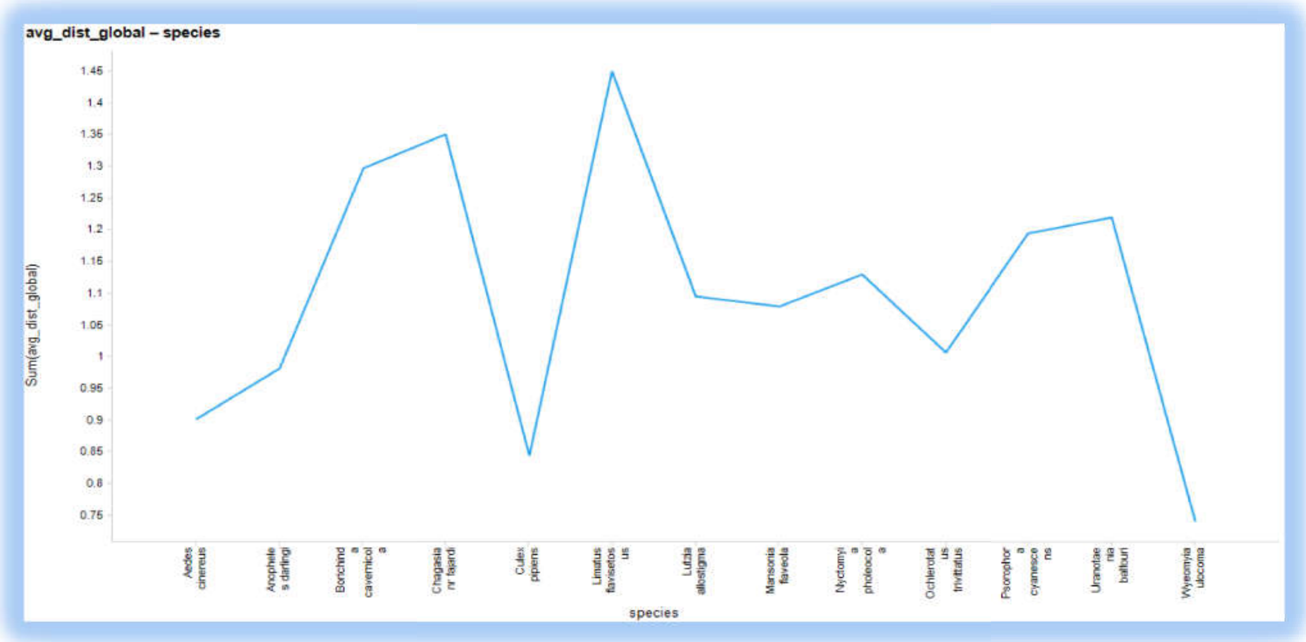


Fig. 24. Uranotaenia genus shows a wide range of time tree values in the global and Indian species group

Fig. 25. Wyeomyia shows close relatedness to only two species in the global category





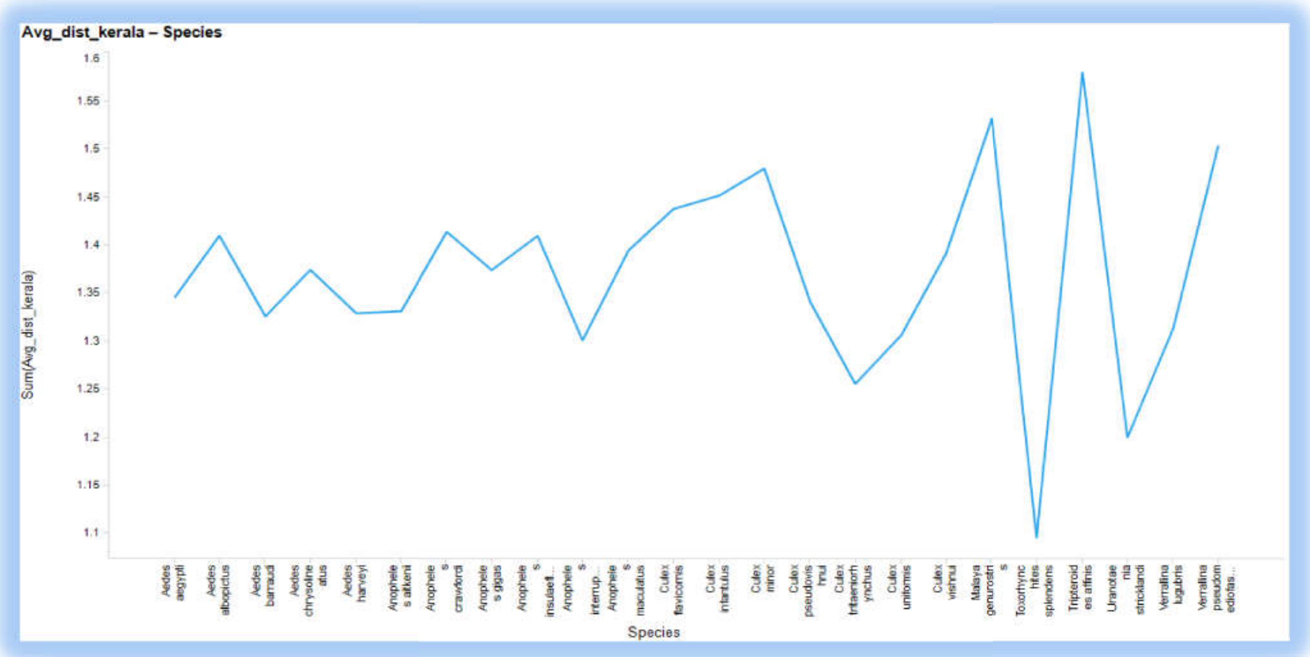


Table 5. Average of time tree values for Indian species

Species	Average distance Indian
Aedeomyiacatasticta	1.902
Aedesiyengari	2.250
Anopheles fluviatilis	1.850
Armigeressubalbatus	2.123
Culexuniformis	2.240
Ficalbia minima	2.128
Hodgesiabailyi	2.243
Lutziafuscanus	1.968
Malaya genurostris	1.916
Mimomyiachamberlaini	2.003
Ochlerotatusportonovoensis	1.918
Orthopodomyiaanopheloides	2.147
Toxorhynchitessp	1.894
Tripteroidesaranoides	2.243
Uranotaeniarecondita	1.738

Table 6. Average of time tree values for Kerala species

Species	Average distance Kerala
Aedesaegypti	1.345
Aedesalbopictus	1.410
Aedesbarraudi	1.325
Aedeschrysolineatus	1.374
Aedesharveyi	1.329
Anopheles aitkenii	1.331
Anopheles crawfordi	1.414
Anopheles gigas	1.375
Anopheles insulaeflorum	1.410
Anopheles interruptus	1.301
Anopheles maculatus	1.395
Culexflavicornis	1.437
Culexinfantulus	1.451
Culex minor	1.479
Culexpseudovishnui	1.341
Culextritaeniorhynchus	1.255
Culexuniformis	1.306
Culexvishnui	1.392
Malaya genurostris	1.533
Toxorhynchitessplendens	1.095
Tripteroidesaffinis	1.580
Uranotaeniastricklandi	1.200
Verrallinalugubris	1.314
Verrallinapseudomediofasciata	1.504

A line plot was plotted across various species. There is very less variation across Indian species suggesting that the mosquito species that are chosen for the analysis are more closely related to each other than the global or Kerala mosquito species. The least average time tree value 1.738 and the highest value is 2.25.

A line plot was plotted across various species. There is lesser variation across the species suggesting that the mosquito species that are chosen for the analysis are closely related to each other. For example, the average time tree value for almost all the species of Aedes ranges from 1.31 to 1.35.

Conflict of interests

The author declares no conflict of interests.

Author agreement: Material submitted is original; the author is in agreement to have the article published.

REFERENCES

Godfray HCJ. and Knapp S.2004. Introduction. Philos. *Trans. R. Soc. Lond. B Biol. Sci.No*., 359, 559–569.

Harbach RE. 2007. Classification within the cosmopolitan genus *Culex* (Diptera: Culicidae): the foundation for molecular systematics and phylogenetic research. *Acta Trop.,* 120,1–14. doi: 10.1016/j.actatropica.2011.06.005.

Hebert PD, Cywinska A. and Ball SL. 2003. Biological identifications through DNA barcodes. *Proc. R.Soc. Lond. B. Biol. Sci*., 270, 313-321.

Jaenson TGT, Lokki J. and Saura A. 1986. Anopheles (Diptera, Culicidae) and Malaria in Northern Europe, with Special Reference to Sweden. *J. Med. Entomol*., 23, 68-75.

Jones M, Ghoorah A. and Blaxter M. 2011. Turning DNA barcode sequences into annotated operational taxonomic units. *PLoS ONE,* 6: e19259.

Knowlton, N. and Weigt, L. A. 1998. New dates and new rates for divergence across the Isthmus of Panama. *Proc. Roy. Soc. Lond. B.,* 265, 2257-2263.

Lipscomb D, Platnick N. and Wheeler Q. 2003. The intellectual content of taxonomy: a comment on DNA taxonomy. *Trends Ecol. Evol*., 18, 65–66.

Lundstrom JO, Niklasson BO. and Francy DB. 1990. Swedish Culextorrentium and Culexpipiens (Dipitera: Culicidae) as experimental vectors of Ockelbo virus. *J. Med. Entomol*., 27, 561-563.

Lundstrom JO, Pfeffer M M S, PerssonVinnersten TZ. and Forsberg B. 2007. Sindbisvirusförekommerhosstickmyggor i Avestaområdet, vid Färnebofjärden, vid Hedesunda fjärdenochvid Storfjärden. (*Uppsala, Uppsala University*).

Scoble MJ. 2004. Unitary or unified taxonomy? Philos. *Trans. R.Soc. Lond. B Biol. Sci. No*. 359, 699–710.

Wheeler QD, Raven PH, Wilson EO. 2004. Taxonomy: impediment or expedient? *Science (Wash., D.C.)*.303, 285.

Wilson EO. 2004. Taxonomy as a fundamental discipline. Philos. *Trans. R. Soc. Lond. B Biol. Sci., No.* 359, 739

****