

# Revisiting the Evolutionary History of *Culex molestus* and *Culex pipiens*: Insights Beyond WWII Origin Narratives

Hemanth Kumar J<sup>1</sup>, Archana N<sup>2</sup>, Ashwini V Reddy<sup>2</sup>, Manohar G M<sup>3</sup>

<sup>1&2</sup> PG - M Sc In Biotechnology, Nrupathunga University, Bengaluru, Karnataka-01

<sup>3</sup>Researcher, halliLabs, Ragihalli, Bengaluru, Karnataka- 83

Orcid Id: <https://orcid.org/0009-0001-2824-1389>

## ABSTRACT

*Culex* is a genus of mosquitoes belonging to the family *Culicidae*. These species are the primary vectors for diseases like the West Nile virus, St. Louis encephalitis, lymphatic filariasis, and agricultural pests. The species are distributed all over the world with overlapping species distribution. To date, it is assumed that the *Culex molestus* ecotype originated and evolved in London underground tunnels during WWII, when people took shelter. We tested this hypothesis by constructing phylogenetic trees containing different ecotypes and species of *Culex*. We charted the ecological range of *C. pipiens* and *C. molestus* and investigated the evolutionary history of the *Culex* species using the CO1, IGS & ITS2 gene sequences. We find some indication that *C. pipiens* and *C. molestus* form a species complex at best. In all the phylogenetic trees we constructed, *C. pipiens* and *C. molestus* shared the same ancestor and showed long branch lengths indicative of a long evolutionary history. This indicates that *C. molestus* could not have evolved in London tunnels during WWII. *C. pipiens* in all the trees showed longer branch lengths as compared to *C. molestus*. Our study supports the concept that evolution selects from an already existing pool of genetic variants to evolve different ecotypes and species.

**Keywords; Culex, Phylogeny, World War II, Culex Pipiens, Culex Molestus, Evolution.**

## 1. Introduction

*Culex* is a genus of mosquitoes that includes over 1000 species worldwide. They are found in all regions of the world, except for Antarctica (3). Many species of *Culex* mosquitoes are important vectors of diseases such as West Nile virus, St. Louis encephalitis, and lymphatic filariasis (4). *Culex* mosquitoes are typically active at dusk and dawn, and they can breed in a variety of aquatic habitats, including stagnant pools, ponds, and even discarded containers (5). Some species of *Culex* mosquitoes are also important agricultural pests, causing damage to crops such as rice (6). *C. pipiens pipiens* (*C. pipiens*) and *C. pipiens molestus* (*C. molestus*) are two ecotypes that also differ in some morphological and behavioural and biological characteristics. Where *C. molestus* is autogenous, stenogamous, and anthropophilic, and *C. pipiens* is eurygamous, anautogenous, capable of diapause and mainly feeds on bird blood. *C. pipiens* is present both outdoors and indoors, however, *C. molestus* is present only indoors (13). *Culex molestus* can produce a limited number of eggs without the vertebrate blood meal. During the reproductive cycle, It requires a protein that is mainly present in the human blood (15).

*C. molestus* was first noticed during the time of world war II that is between 1940 and 1941 in Europe in subway tunnels in London (8). People used to frequently take shelter in the underground tunnels because of the bombing aboveground. These people suffered from a new type of mosquito that was aggressively biting humans. So it was thought that the bird-biting *Culex pipiens* may have evolved into *Culex molestus* adapted to bite humans and feed on their blood (9). In 1999 Byrne and Nicols conducted a study using allozymes, which revealed some evidence that the mosquitoes inhabiting the London underground were reproductively isolated from their counterparts residing aboveground (14). Furthermore, their study put forward the hypothesis that these underground mosquitoes underwent

evolutionary changes within their specific environment, suggesting an in-situ evolutionary process (14). However, there is some evidence that *Culex* mosquitoes with the type of behaviour exhibited by the *molestus* type may have existed way back in the 17th century in Egypt (13). At least in one of the sculptures of an Egyptian Pharaoh one leg is shown as swollen like in Elephantiasis disease indicating the prevalence of man-biting and disease-transmitting *Culex* mosquitoes (16). Two ecotypes are maintained as distinct ecotypes even though they can interbreed between them. Even when *C.pipiens* and *C.molestus* are present in the same environment like above ground habitat they still show behavioural differences. However, each of the ecotypes is confined to a certain region down the latitude though there is a region that harbours both ecotypes. The behaviours and some characters also gradually vary as a gradient within each ecotype down the latitude.

The phylogenetic tree provides evolutionary relationships among various biological species based on similarities and differences in their physical or genetic characteristics.

Based on the anecdotal shreds of evidence and observed cline in characters of each ecotype we hypothesised that *Culex pipiens molestus* may have shared a long history with *Culex pipiens pipiens* rather than the short believed duration from WWII. Here we constructed a phylogenetic tree for many species of *Culex* which includes *C.pipiens* and *C.molestus*, from different regions of the world to decipher the evolutionary relationship between the two ecotypes. Based on several phylogenetic trees we conclude that *C. molestus* share a long evolutionary history and have accumulated far greater genetic differences between the two ecotypes. We also propose that *Culex pipiens* may be a species complex.

## 2. Materials & Methods

**2.1 Sequence retrieval:** Nucleotide sequence from (HQ724614.1) CO1 gene of *Culex pipiens pipiens* from Tunisia mitochondrion complete genome. Nucleotide sequence from (U22120.1) ITS2 gene of *Culex pipiens pipiens*. Nucleotide sequence from (GU911333.1) IGS gene of *Culex pipiens pipiens*.

### 2.2 Retrieval of gene:

Retrieval Of CO1 gene: Accession number HQ724614.1 was used to query the database using Blastn. 451 sequences were retrieved with taxon ID 7174 that fulfilled a Criteria of query coverage of 50%. Also did a dataset of the sequences. the common sequences of the species which are common in all the 3 genes(IGS, CO1&ITS2)

Retrieval Of ITS2 gene: Accession number U22120 was used as a query to search the database using Blastn. 8 sequences were retrieved from the selected(common species in IGS &CO1) species of *Culex* with taxon ID 7174 that fulfilled a Criteria of query coverage of 50%.

Retrieval Of IGS gene: Accession number GU911333 was used as a query sequence to search the database using Blastn. 9 sequences were retrieved from the selected(common species in ITS2 /CO1) species of *Culex* with taxon ID 7174 that fulfilled a Criteria of query coverage of 50%.

**2.3 CD-HIT:** Removal of redundant sequences was done using CD hit in the Galaxy Pasteur tool (Version 4.8.1). (1) with a threshold frequency of 1.0 and a word length of 5. This reduced the number of sequences in the CO1 dataset.

**2.4 Location Retrieval:** For the sequences which remained in the MSA for those sequences the locations of the sequences were collected from GenBank data and represented in the map(Fig. 1)

### 2.5 Multiple sequence alignment(MSA):

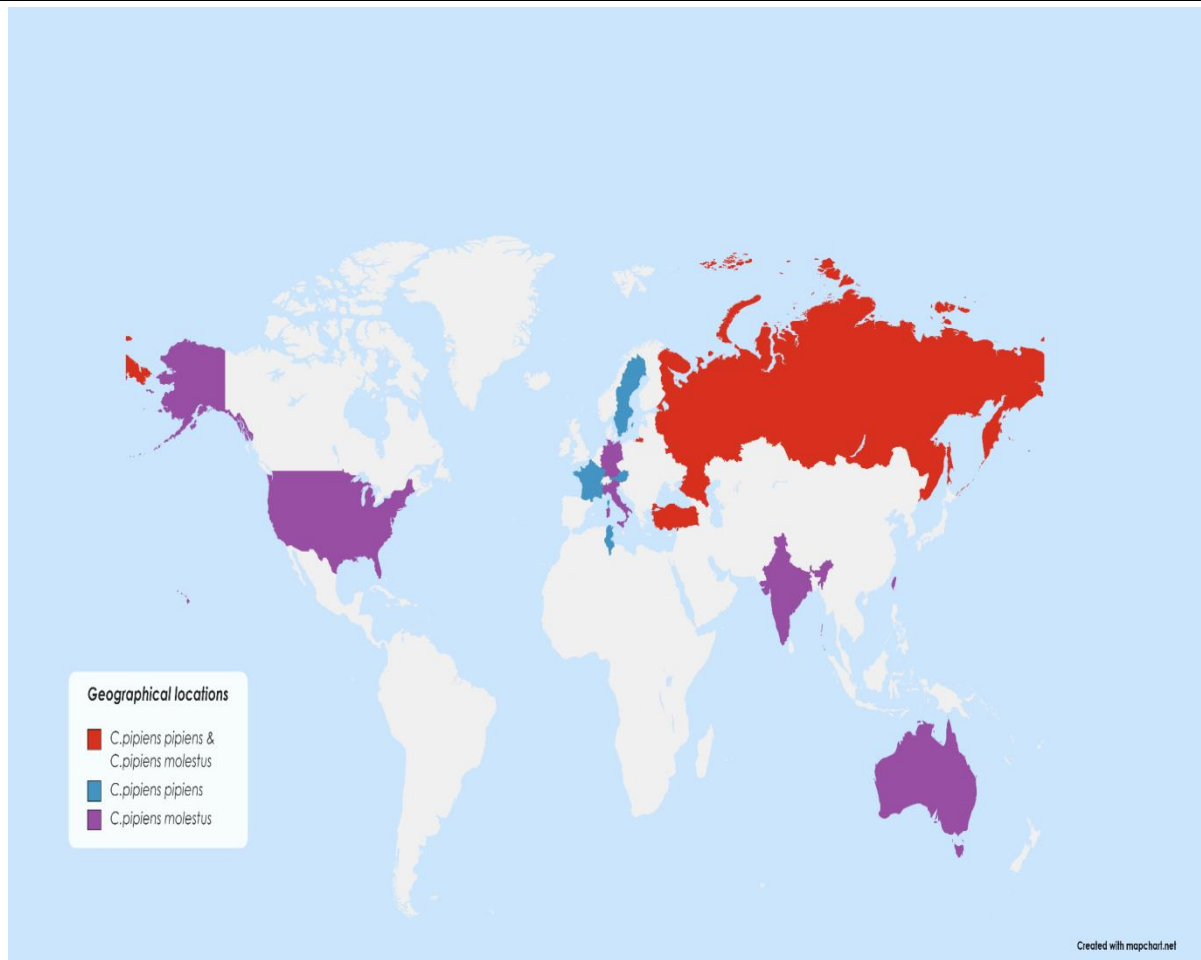
MSA of CO1: From the obtained 317 sequences multiple sequence alignment took the sequence length of 720 base pairs and removed the sequences which don't have a minimum sequence coverage of 70% have been removed. and sequences that don't have sequence similarity were removed.

MSA of ITS2: In the obtained 8 sequences the sequence length 458 base pairs did multiple sequence alignment.

MSA of IGS: In the obtained 9 sequences the sequence length 458 base pairs did multiple sequence alignment.

MSA of CO1: In the obtained 9 sequences the species present commonly in all 3 genes.





**Fig. 2:** Geographical distribution of *Culex pipiens pipiens*(blue), *Culex pipiens molestus* (purple) & both *Culex pipiens pipiens* and *Culex pipiens molestus* (red) together are shown. Map created using Mapchart(2).

We can see that there is a clear gradient in the distribution of the ecotypes of *C. pipiens pipiens* and *molestus*. This may be because of their differing adaptive phenotypes. For example, in Europe, *Culex pipiens pipiens* is confined to the more temperate northern region. While the *culex pipiens molestus* occupies the southern part of Europe and the Mediterranean.

We hypothesised that *Culex pipiens molestus* may have evolved in natural conditions dictated by the changing environments coming from north to south. And not in underground train stations during world war II when people took shelter to avoid bomb explosions on the ground. To delineate the evolutionary relationships we constructed different phylogenetic trees.

To find the evolutionary position of *molestus* and *pipiens* ecotypes of *Culex pipiens* concerning each other and with other species, we created phylogenetic trees. We retrieved 451 mitochondrial Cytochrome Oxidase 1 (CO1) sequences from the NCBI GenBank database using NCBI Blastn. These sequences were run in CD Hit Galaxy Pasteur (version:4.8.1). to remove redundant sequences which were 100% identical due to which some 134 sequences were removed (Table 1).

Multiple sequence alignment was done for the CO1 dataset (Tables 2 & 3) by Muscle alignment in MEGA (Molecular Evolutionary Genetics Analysis) version 11.0.13. We manually removed 50 sequences (Table 2). which include highly dissimilar sequences that were skewing the MSA and sequences which did not meet the criteria of a minimum length of 502bp (70 % of the original length) were removed. We obtained the MSA of the remaining 267 sequences (Fig. 3).



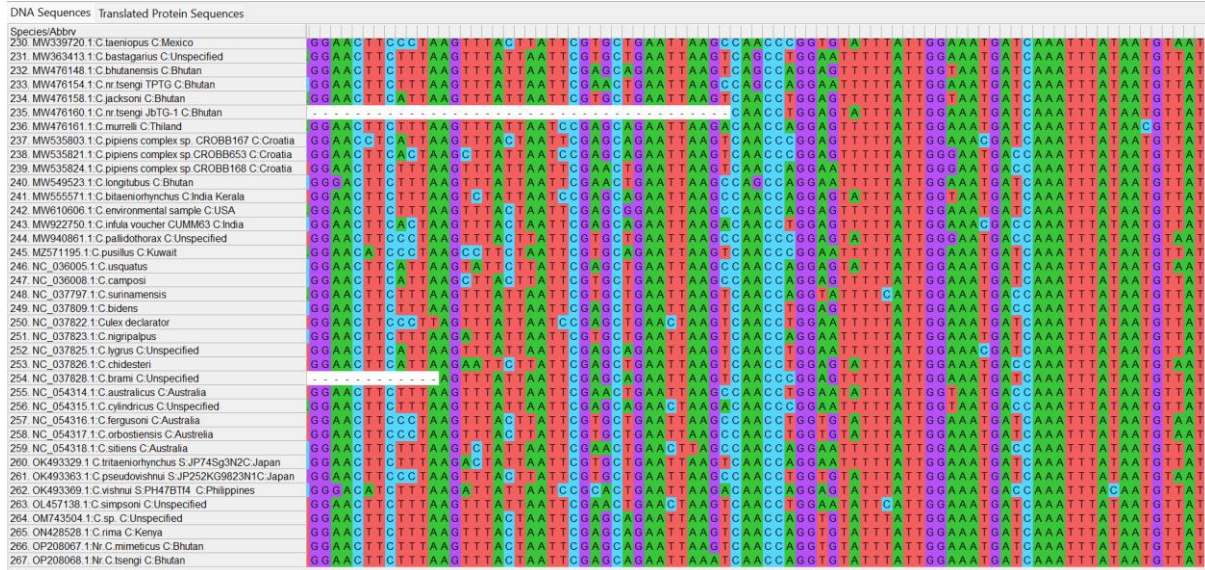


Fig. 3: Multiple Sequence alignment of CO1 gene dataset(Table 3).

Intergenic spacer sequences (IGS) sequences from 9 species (Table 5) were retrieved by Blastn using a *C.pipiens pipiens* sequence of AC.No GU911333 as a query. Those species which were common to the CO1 dataset ( ) were selected for MSA (Fig. 4).



Fig. 4: Multiple Sequence alignment of IGS gene dataset(Table 5).

Similarly, MSA was done on 9 species (Table 4) using CO1 sequences (Fig. 5).

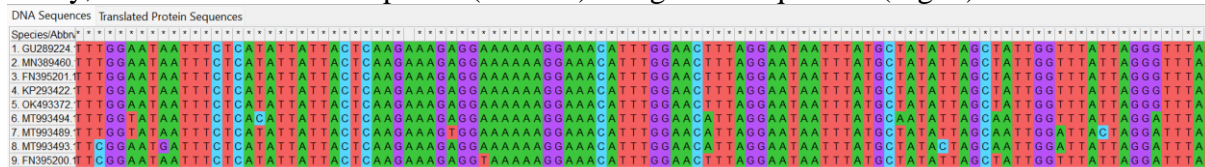


Fig. 5: Multiple Sequence alignment of CO1 gene dataset(Table 4).

And 8 species (Table 6) using ITS sequences retrieved by Blastn using a *Culex pipiens pipiens* sequence of Ac.No U22120. The sequences which were common to CO1 & IGS (Tables 4 & 5) datasets were selected for MSA (Fig. 6).



Fig. 6: Multiple Sequence alignment of ITS-2 gene dataset (Table 6).

All three genes (Tables 4, 5 & 6) were collated and MSA was done with the collated set of 9 *Culex* species (Fig. 7). Also for the collated genes of 5 *Culex* Species MSA was done (Fig. 8).



**Fig. 7:** Multiple Sequence alignment of Collated genes of 9 Culex species.



**Fig. 8:** Multiple Sequence alignment of Collated genes of 5 Culex species.

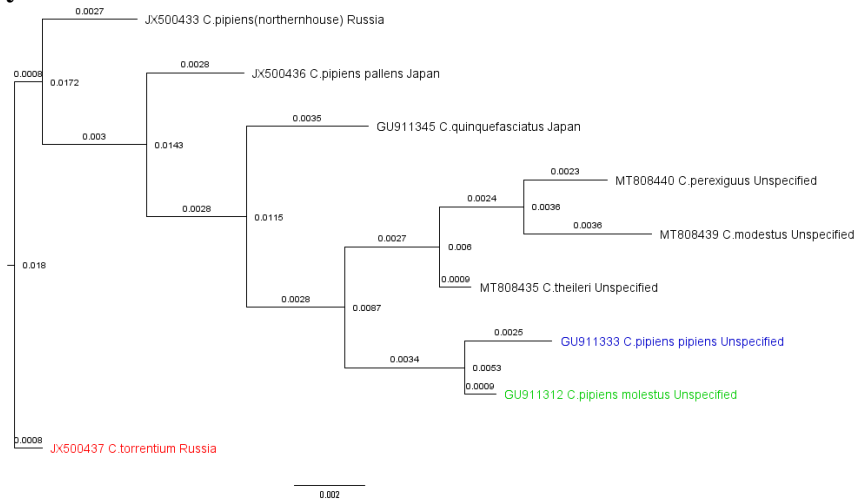
Pictures of all MSA are represented in the supplementary section.

**Partitionfinder:** To find the best model and partition the data we used partition finder 2. Cytochrome oxidase gene is a protein-coding gene we partitioned into three positions, each for one of the positions in the codon. This is because each of the codon positions evolves at different rates, especially the third base evolves at a different rate than the rest of the two because of 3rd base degeneracy. ITS2 and IGS were not partitioned according to the codon positions because they are non-protein coding genes. The best models as predicted by partition 2 for each of the gene and each of the positions in the codon is given in Table 7.

SL. No	MSA	Partitions	Best Model for Analysis
1	CO1 gene dataset (267 Sequences)	Gene1_pos1 (1-720\3) Gene1_pos2 (2-720\3) Gene1_pos3 (3-720\3)	GTR+I+G
2	CO1 gene dataset (9 sequences)	Gene1_pos1 (1-1101\3) Gene1_pos2 (2-1101\3) Gene1_pos3 (3-1101\3)	F81+I TIM+G TRN+I
3	ITS2 gene dataset (8 sequences)	Gene1 (458)	K81UF+I
4	IGS gene dataset (9 sequences)	Gene 1 (458)	TV MEF+I
5	All 3 genes combined dataset (9 sequences)	Gene1_pos1 (1-1101\3) Gene1_pos2 (2-1101\3) Gene1_pos3 (3-1101\3) Gene2 (1102-1559) Gene3 (1560-2017)	F81 K81UF+G TRN+G TVM+I+G K81UF+I
6	All 3 genes combined dataset (5 Sequences)	Gene1_pos1 (3-1101\3) Gene1_pos2 (3-1101\3) Gene1_pos3 (3-1101\3) Gene2 (1102-1559) Gene3 (1560-1960)	F81 HKY F81 K80+I+G K80+I+G

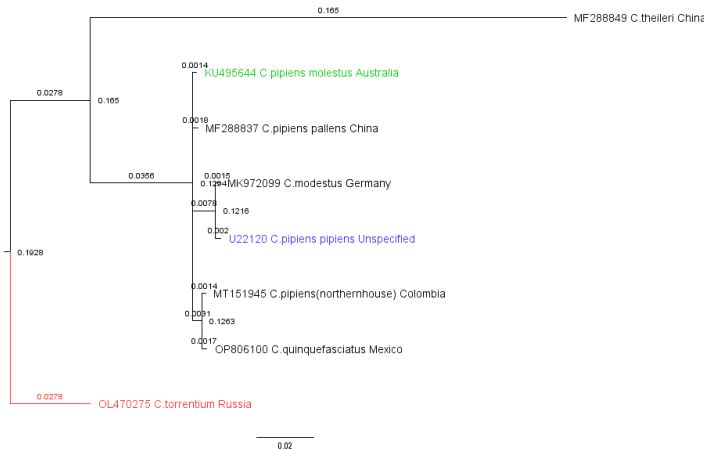
**Table 7:** Partition Test 2 analysis to find the best evolutionary model for Bayesian analysis

**Bayesian Analysis:**

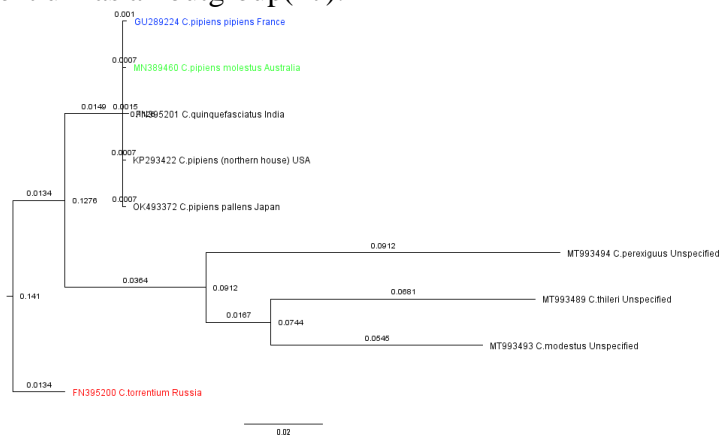


**Fig. 9:** Phylogenetic tree of *Culex* species created using Intergenic spacer(IGS) gene sequences & *C. torrentium* as an outgroup(17).

The IGS tree is showing that *C. pipiens pipiens* and *C. pipiens molestus* are sister ecotypes through *C. pipiens pipiens* showing a long branch length indicating accumulation of variation in the IGS sequence.



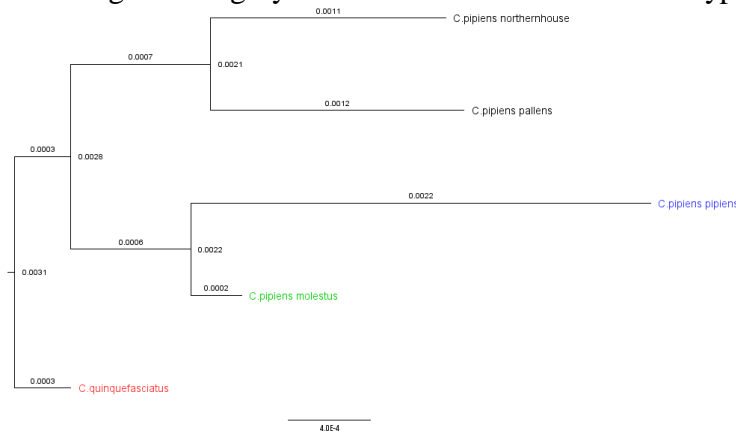
**Fig. 10:** Phylogenetic tree of *Culex* species was created using Internal transcribed spacer-2 (ITS-2) gene sequences *C. torrentium* as an outgroup(17).



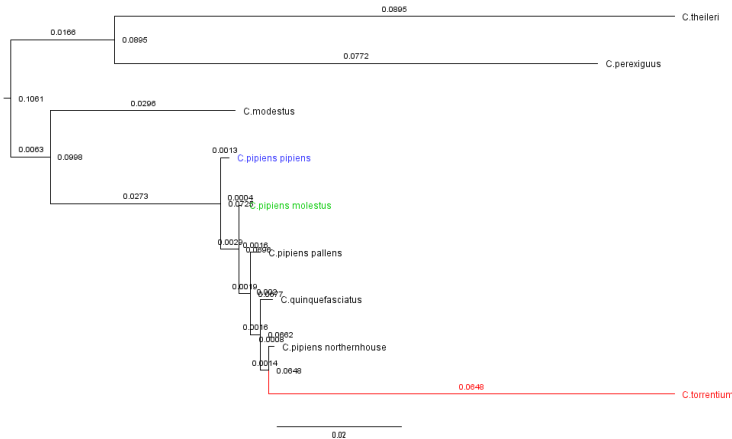
**Fig. 11:** Phylogenetic tree of *Culex* species created by using Cytochrome C Oxidase 1(CO1) gene sequences *C. torrentium* as out grouped

In this tree based on CO1 ( Fig. 11), we see that *C. pipiens pipiens* and *C. pipiens molestus* are sharing the same node and are more closely related than any of the other culex species. The branch lengths

of the pipiens clade are almost nil compared to the branches because we have included other species of *Culex* in the phylogenetic tree. This shows that all the different ecotypes (*pipiens*, *molestus*, *quinquefasciatus*, and *pallens*) are almost more closely related to each other than to any other species. It also implies that the CO1 gene is highly conserved between the two ecotypes.



**Fig. 12:** Phylogenetic tree of *Culex* species created by using All 3 (CO1, ITS&IGS,) gene sequences considering only the subspecies and ecotypes, *C. quinquefasciatus* is out grouped. The phylogenetic tree constructed using only the *Culex* ecotypes (Fig. 12) suggests that *C. pipiens pipiens* & *C. pipiens molestus* are more closely related than any of the other subspecies. This indicates that *C. pipiens* and *C. molestus* shared a most recent common ancestor.



**Fig. 13:** Phylogenetic tree of *Culex* species created using All 3 (CO1, ITS & IGS,) gene sequences considering subspecies and other species of *Culex*. Rooted by midpoint rooting. When we included other species in the collated gene tree (Fig. 27) (CO1, ITS & IGS) we found that the subspecies including *C. pipiens* and *C. molestus* were so closely clubbed with other subspecies that we can speculate that they form a species complex. Also, since each gene has a different origin concerning the location, it will be hard to interpret the collated tree. The three genes might have accumulated variation characteristics of the local regions.





**Fig. 14:** Phylogenetic tree of *Culex* species created using Cytochrome Oxidase 1 (CO1) gene sequences and *C.hortensis* as an outgroup(17).

A holistic tree created out of 267 sequences shows that *C pipiens pipiens* is not monophyletic and is seen spread across different clades. One sample ( FN395181) is seen clubbing with *C.pipiens molestus* (FN395171). since both are from the same geographical location ( Russia). The other *C pipiens* ecotypes are from different countries ( Russia, Sweden & France). It might be that *C.pipiens pipiens* ecotypes have varied highly among different geographical locations or that based on the morphological and behavioural characteristics some of the ecotypes are misidentified. We can see that the *Culex pipiens pipiens* of Europe ( Sweden and France) are in different clades while *Culex pipiens pipiens* and *molestus* of Russia are in one clade.

The long branch lengths between *C.pipiens* and *C. molestus* of Russia imply that they may have had a long time between them to evolve and accumulate variation sufficient to be called members of the *Culex pipiens* species complex. This result directly counters the assumption that *C.molestus* originated in London underground subways during world war II a few decades back. The time

between World War II and the present day is simply not sufficient to accumulate the variation that justifies the branch length found in our analysis.

#### 4.DISCUSSION

By constructing phylogenetic trees for the different *Culex* species, subspecies, and ecotypes using different genes, we observed there is a *Culex* species complex present. The evolutionary relationship between *C.pipiens* and *C.molestus* is deep in the phylogenetic tree (as observed in the Russian samples) indicating a shared evolutionary time much older than World War II. This supports earlier anecdotal evidence (as in the case of the Egyptian Pharaoh) that *Culex molestus* may have been in circulation with mankind thousands of years before. The branch lengths between *C. pipiens* and *C molestus* are longer than other branches between different species, showing that much faster variations and in many of the *C.pipiens* and *C.molestus* originate from the same clade and in the CO1 gene 267 species tree (fig 14) having more than one *C.pipiens pipiens* gene and they distribute in the different clades and it shows the species are much diversely distributed or the species are identified by their characters, they may have some characters but they are different in their genetic distribution, and in the IGS.

By analysing the phylogenetic tree we get to know that the evolutionary relationship is resembling that *C.pipiens pipiens* accumulating more variations over time as compared to *C.molestus* and also as we observed that in the phylogenetic tree of 267 species the species (FN395171 & FN395181) which are from Russia are showing huge genetic variations. These variations quantitatively are more than what is required to distinguish between two established species. So we can say that the *Culex molestus* might have evolved long back in WWII. Because accumulating such a huge variation between the two ecotypes needs much more time than the recent WWII.

As we observed the many species in the case of *C.pipiens pipiens* in the phylogenetic tree we saw the species which were identified as the *C.pipiens pipiens* are found in the different clades of the phylogenetic tree showing that they are non-monophyletic. It could be because some of the samples may be wrongly identified. Or *pipiens* may be a case of species complexity. A species complex is a group of genetically different species that appear morphologically similar. The environment gradient as we have seen in the *C.pipiens* and *C.molestus* as we observed in the map (fig.8) *C.pipiens* are seen mostly in the northern hemisphere. And *C.molestus* is seen mostly in southern regions though there is a region that overlaps between the two subspecies. The unique characteristics of *C. pipiens* are sharply distinct in the northern region. However, as we move south the characters become less unique to *C pipiens* and share some of the characteristics of *C molestus*. This gradient of characters agrees with our findings that *C molestus* was present much before and adapted with characters suitable to different environmental regions. This adaptability is a consequence of long-time evolution and contradicts the theory that *C molestus* originated in London tunnels.

The species' complex formation and also the non-monophyletic distribution of *C. pipiens* indicate that the species may have undergone hybridisation in regions where the two subspecies overlap as we see in Russia. ( Fig. 14).

As the data we used for constructing the phylogenetic tree is obtained from the NCBI GenBank data and the sampling was dictated by only what is available in public databases, this interpretation might be biased. A thorough sampling representing different regions in the north-south gradient could provide a more accurate picture of the evolutionary relationship between the two subspecies. A concatenated set of multiple loci from the same samples can also improve confidence in the phylogenetic tree.

**Significance of our study:** It helps the researchers to understand the drivers of species evolution in mosquitoes which can be translated to efficient control measures.

**CONCLUSION:** In conclusion, *C.pipiens* and *C.molestus* since sharing a long branch between them in the phylogenetic tree, we can say that they both share a long evolutionary history than the recent World War II constraints.

**Acknowledgement**

We deeply appreciate the help and mentoring from **Dr Meenakshi Iyer**, PhD and Postdoc from NCBS who is also the resource person from halliLabs, for regular inputs for this research.

## References

1. Li, W., Jaroszewski, L., & Godzik, A. (2001). Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics (Oxford, England)*, 17(3), 282–283. <https://doi.org/10.1093/bioinformatics/17.3.282>
2. Create your own Custom Map | MapChart. (n.d.). MapChart. <https://www.mapchart.net/index.html>
3. Harbach, R. E. (2018). The mosquitoes of the subgenus *Culex* in Southwestern Asia and Egypt (Diptera: Culicidae). *Zootaxa*, 4378(1), 1-76.
4. World Health Organization: WHO. (2020). Vector-borne diseases. [www.who.int](http://www.who.int). <https://www.who.int/news-room/fact-sheets/detail/vector-borne-diseases>.
5. Shaman, J., Day, J. F., & Komar, N. (2010). Hydrologic conditions describe West Nile virus risk in Colorado. *International journal of environmental research and public health*, 7(2), 494-508.
6. Zhao, T., & Xue, R. (2021). Integrated mosquito management in rice field in China. *Wetlands Ecology and Management*, 30(5), 963–973. <https://doi.org/10.1007/s11273-021-09840-6>
7. World Health Organization. (2019). Vector control: Methods for use by individuals and communities. World Health Organization.
8. Shute, P. G. (2009). CULEX MOLESTUS. *Transactions of the Royal Entomological Society of London*, 102(7), 380–382. <https://doi.org/10.1111/j.1365-2311.1951.tb00758.x>
9. Ayala, F. J. (1975). Genetic differentiation during the speciation process. *Evol. Biol.*, 8, 1-78.
10. Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., Calcott, B. (2016) PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular biology and evolution*. DOI: [dx.doi.org/10.1093/molbev/msw260](https://doi.org/10.1093/molbev/msw260)
11. Lanfear, R., Calcott, B., Ho, S. Y., & Guindon, S. (2012). PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular biology and evolution*, 29(6), 1695-1701.
12. Huelsenbeck, J. P. and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* 17:754-755. Ronquist, F. and J. P. Huelsenbeck. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572-1574
13. Haba, Y., & McBride, L. (2022). Origin and status of *Culex pipiens* mosquito ecotypes. *Current Biology*, 32(5), R237–R246. <https://doi.org/10.1016/j.cub.2022.01.062>
14. Zhao, T., & Xue, R. (2021). Integrated mosquito management in rice field in China. *Wetlands Ecology and Management*, 30(5), 963–973. <https://doi.org/10.1007/s11273-021-09840-6>
15. Roubaud, E. (1929). Autogenous Cycle of Winter Generations of *Culex pipiens* L. *Compte Rendu de l'Academie des Sciences*, 188(10).
16. Kenawy, M., & Abdel-Hamid, Y. (2015). Insects in ancient (Pharaonic) Egypt: a review of fauna, their mythological and religious significance and associated diseases. *Egyptian Academic Journal of Biological Sciences. A, Entomology*, 8(1), 15-32. doi: 10.21608/eajbsa.2015.12919.
17. Cox, F. E. G. (2002). History of Human Parasitology. *Clinical Microbiology Reviews*, 15(4), 595–612. <https://doi.org/10.1128/cmr.15.4.595-612.2002>
18. Da Silva, A. A., Machado, L. C., De Paula, M. B., Bernardo, C. S. S., De Moraes Bronzoni, R. V., De Melo Santos, M. a. V., & Wallau, G. L. (2020). Culicidae evolutionary history focusing on the Culicinae subfamily based on mitochondrial phylogenomics. *Scientific Reports*, 10(1). <https://doi.org/10.1038/s41598-020-74883-3>

## SUPPLEMENTARY SECTION:

[SUPPLEMENTASUPPLEMENTARY SECTION.pdf](#)

## URL



**For tables:** <https://drive.google.com/file/d/1zKVGe6l5VIE4pAzKgo4L5-q65U7MRtnW/view?usp=sharing>

**For Figures:**

[https://drive.google.com/drive/folders/18gMas91hfuwkB\\_z0EDt0xXyYZerykamM?usp=sharing](https://drive.google.com/drive/folders/18gMas91hfuwkB_z0EDt0xXyYZerykamM?usp=sharing)