



Stop Codon Usage of Three Gene Families in Selected Medical Insects, and their Phylogenetic Relationships

Titus, G.O., Ani, G.N., Animashaun, M., Abdulrahman-Giwa, Z.O. and Giwa, A.O.*
Department of Zoology and Environmental Biology, Lagos State University, Lagos, Nigeria

Abstract

Medical insects pose significant health risks, transmitting disease causing agents with significant morbidity and mortality globally. Understanding their genetic composition and evolutionary relationship is crucial to developing control techniques to combat them. This study aims to compare the stop codon usage of gustatory, odorant and kinase genes, and assess their use as markers for evaluating the phylogenetic relationship between selected medical insects. Fifteen genes from three gene families including gustatory receptors, odorant receptors, and kinases were studied in thirty-four insect species. Gene and transcript sequences were retrieved from Ensembl Metazoa database. The stop codon usage was obtained by identifying the stop codon present at the end of the transcript sequences. Multiple sequence alignment of the gene sequences was performed and phylogenetic trees were computed. There were variations in the stop codon usage across the different insect species. The odorant receptor and kinase genes grouped the insects into two major clusters. The stop codon usage highlighted the variation within and among the species. The phylogenetic analysis results supported existing insect classification. These approaches could be considered for phylogenetic analyses of other arthropod groups. Accurate measurement of stop codon usage will be important for understanding natural selection, the genetic variation of coding sequences of genes, and useful in design of more efficient expression constructs in gene-editing techniques for vector control.

Keywords: codon usage, medical insects, odorant receptors, phylogenetics, stop codon.

*Correspondence email:
abdulazeez.giwa@lasu.edu.ng

ORCID: 0000-0001-6536-3687

INTRODUCTION

Insects are important, abundant, and diverse group of organisms representing approximately 75% of all known animal species (Behura, 2015) making their ecological significance hard to overstate. The number of living insect species has been estimated to be between 2.5 and 3.7 million (Hamilton *et al.*, 2010) with this tremendous success due to their small size, exoskeleton, flight, high reproductive rate, and metamorphosis (Huis, 2014). Insects are also important because they are relatively cheap and easy to rear for research purposes (Adamski *et al.*, 2019) making them ideal models for many scientific studies (Mikulak *et al.*, 2018).

For the past century, scientists have tried to resolve the tree of life of organisms including insects using their phylogeny (Deline *et al.*, 2018). However, their high biodiversity and morphological disparity which contributed to their success as a group was a challenge. It made the reconstruction and understanding of their systematics and taxonomy using just morphological characteristics largely inadequate (Tihelka *et al.*, 2021). The production of molecular data (Genomes, transcriptomes) from sequencing technologies in the late 20th century provided a new and more suitable way called phylogenetics for the reconstruction of evolutionary history for different organisms including insects. Based on the introduction of molecular phylogenetics, the resolution of the tree of life of insect has improved during the past few decades (Simon *et al.*, 2012) such that the first molecular phylogeny of twelve insect species using a gene marker was published in 1989 (Wheeler, 1989).

Both phylogenetics and phylogenomics are areas of research which is carried out by using techniques and data from molecular biology to study evolution. The major difference is that phylogenetics involves using one or more gene markers while phylogenomics involves using the

whole genome to infer the phylogeny. They both involve the use of computational tools to align the molecular data (sequence alignment), and to generate species or gene trees from the aligned sequences (Katz and Grant, 2015). While sequence alignment is the most popular and used method in this field, there are other methods such as the comparison of DNA strings, random anchor, and gene content (Behura, 2015; Ceron-Romero *et al.*, 2019). With either of the above mentioned methods, the goal of the analysis is to infer phylogenetic relationships of organisms or to infer the functions of the sequences (Young and Gillung, 2019). Some popular tools used by the field for phylogeny analysis are Multiple Alignment using Fast Fourier Transform (MAFFT) (Kato *et al.*, 2002), Molecular Evolutionary Genetics Analysis (MEGA) (Tamura *et al.*, 2021), Minimap2 (Li, 2018), Phylogenetic Placement and Alignment Tool (PASTA) (Mirarab *et al.*, 2015) and ProbPPF (PHMM model optimized by particle swarm optimization) (Zhan *et al.*, 2019).

Some genes which can be used as markers because of their presence in various insect taxa are Odorant Receptors (ORs), Gustatory Receptors (GRs), and Kinases. The ORs are expressed as proteins found on the membrane of olfactory sensory neurons, and are involved in olfactory recognition (Butterwick *et al.*, 2018). They are broadly divided into the odorant receptor co-receptor (Orco) and ORx (the x being the chromosomal locus number) (Smart *et al.*, 2008). The GRs are genes which are expressed in taste tissues of insects (Isono and Morita, 2010). They are referred to as GRx and just like the ORx, the x represents the chromosomal locus number (Isono and Morita, 2010). Kinases are a large family of genes which has been conserved during evolution across all three domains of life (Manning *et al.*, 2002). They are also one of the largest gene families found in domain eukarya (Rauch *et al.*, 2011).

Another way that has been suggested for studying the evolutionary history of organisms because of the advent of molecular data is codon usage bias (Plotkin and Kudla, 2011). This bias is simply the non-uniform way in which the 64 three-nucleotide-code (codons) are used during protein synthesis (Behura and Severson, 2012). This bias causes specific codons to be used more than their alternatives which produce the same protein (Belinky *et al.*, 2018). Of the 64 genetic code (codons), 61 determines an amino acid and thus, the protein. These are known as the sense codons. The remaining three (UAA, UGA, and UAG in mRNA or TAA, TGA, and TAG in genomic sequence) do not produce an amino acid and are called stop codons (Chang *et al.*, 2019). Like synonymous codons, they too share the same “meaning” (Belinky *et al.*, 2018) and these principles are applicable in insects as in most organisms. Medical insects pose significant health risks, transmitting disease causing agents with significant morbidity and mortality globally. Understanding their genetic composition and evolutionary relationship is crucial to developing control techniques to combat them. The aim of this study is therefore to compare the stop codon usage of gustatory, odorant and kinase genes, and assess their use as markers for evaluating the phylogenetic relationship between selected medical insects.

MATERIALS AND METHODS

Sequence data collection

The first step of this phylogenetic analysis pipeline is the identification of a set of homologous sequences and collecting these sequences (Ceron-Romero *et al.*, 2019). Three gene families found in insects were selected including Gustatory receptors (GRs), Odorant receptors (ORs), and Kinases (Kn), because they have been highly conserved during the evolution of insects (Jones *et al.*, 2005; Ha and Smith, 2009; Ahier *et al.*, 2009). Fifteen genes made up of five each from the three gene

families were selected including Gustatory Receptor 2 (*GR2*), Gustatory Receptor 8 (*GR8*), Gustatory Receptor 10 (*GR10*), Gustatory Receptor 21 (*GR21*), Gustatory Receptor 22 (*GR22*), Odorant Receptor 2 (*OR2*), Odorant Receptor 3 (*OR3*), Odorant Receptor 10 (*OR10*), Odorant Receptor 30a (*OR30a*), Odorant Receptor 49b (*OR49b*), Serine kinase (*SK*), Cyclin Dependent Kinase (*CDK*), Mitogen Activated Protein Kinase (*MAPK*), Serine/threonine-protein kinase (*PK*), and Sphingosine kinase (*SphK*). The gene sequences were obtained from the EnsemblMetazoa database (Cunningham *et al.*, 2022). EnsemblMetazoa is a portal that acts as a single point of access for invertebrate metazoan species genome data. Each gene was searched in this database and the genomic and transcript(s) sequences of its orthologues in the appropriate species were downloaded in the FASTA format.

A total of 34 medical insect species belonging to 18 Genus, 14 Families, and 4 Orders were selected and used for this phylogenetic analysis. The insect species selected include; *Aedes aegypti*, *Aedes albopictus*, *Anopheles albimanus*, *Anopheles arabiensis*, *Anopheles atroparvus*, *Anopheles coluzzii*, *Anopheles culicifacies*, *Anopheles darlingi*, *Anopheles funestus*, *Anopheles gambiae*, *Anopheles melas*, *Anopheles sinensis*, *Anopheles stephensi*, *Anopheles quadriannulatus*, *Apis mellifera*, *Culex quinquefasciatus*, *Culicoides sonorensis*, *Drosophila melanogaster*, *Glossina austeni*, *Glossina fuscipes*, *Glossina morsitans*, *Glossina pallidipes*, *Glossina palpalis*, *Ixodes scapularis*, *Lucilia cuprina*, *Lutzomyia longipalpis*, *Megaselia scalaris*, *Musca domestica*, *Nasonia vitripennis*, *Phlebotomus papatasi*, *Pediculus humanus*, *Rhodnius prolixus*, *Stomoxys calcitrans* and *Teleopsis dalmanni*.

Stop codon usage

The stop codon usage was obtained by identifying the stop codon of each of the transcripts. This was done by identifying

_____ : Stop Codon Usage of Three Gene Families in Selected Medical Insects, and their Phylog 199
the stop codon present at the end of the transcript sequences using a custom python script.

Phylogenetic analysis

Multiple sequence alignment of the gene sequences was performed using MEGA software, Version 11 (MEGA11) (Tamura *et al.*, 2021). MEGA provides two methods to align sequence data namely Clustalw and Muscle, and the ClustalW method was selected for this work. Since the genomic sequence of each gene and not the transcript of the sequences were aligned, it was aligned by DNA and not by codons. Estimating a tree from the aligned sequence is the third step of this phylogenetic analysis. The evolutionary history was inferred by using the Maximum Likelihood method and Tamura-Nei model (Tamura and Nei, 1993). The bootstrap consensus tree inferred from 1000 replicates (Felsenstein, 1985) is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein, 1985). Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Joining and BioNJ algorithms to a matrix of pairwise distances estimated using the Tamura-Nei model, and then selecting the topology with superior log likelihood value. This analysis involved several nucleotide sequences for each gene and the analyses were conducted in MEGA11 (Tamura *et al.*, 2021). Presentation of the phylogenetic tree is the last step of this phylogenetic analysis. A circular tree was used to present the taxa relationship identified by a gene family as

it is better in comparing many organisms (Baum, 2008).

RESULTS

Stop codon usage

The preferred stop codon used by the various insect species for the fifteen genes studied are given in Tables 1 to 3 For *OR2*, there was no usage of the TAG stop codon. For *OR3*, TAG was the preferred stop codon in the *Anopheles spp* except *A. stephensi* which used TAA. For *OR49b*, the preferred stop codon was TAA except in *D. melanogaster* which used TGA (Table 1).

For the gustatory receptor gene, *GR22*, there was variation of stop codon usage even within the *Anopheles* group and the three stop codons was observed for the insect species studied. For *GR10*, TAG and TGA were the preferred stop codons with TAA only used by *R. prolixus*, *G. pallidipes* and *A. funestus* (Table 2). For the kinase gene, *SAK*, TAA and TGA were preferred to TAG stop codon. For *SPH*, TGA was the preferred codon in all of the *Glossina spp*.

This varied in *Anopheles* with all stop codons used in this group (Table 3). The preferred stop codon used by the various insect species during the expression of the three gene families are given below in Tables 1-3.

Phylogenetic relationship

The phylogenetic relationship of the 34 insects based on the three gene families; Odorant receptor, gustatory receptor and kinase gene families are shown in Figures 1-3. The five gustatory receptor genes grouped the insects into two major clusters (Figure 1) while the five odorant receptor genes grouped the insects into nine major clusters (Figure 2). The five kinase genes grouped the insects into two major clusters (Figure 3).

Table 1: Stop codon usage for odorant receptor genes in the studied insects

Species	OR2	OR3	OR10	OR30a	OR49b
<i>Aedes aegypti</i>	TGA	-	TGA	-	-
<i>Aedes albopictus</i>	TAA	-	TAA	-	-
<i>Anopheles albimanus</i>	TGA	-	TGA	-	-
<i>Anopheles arabiensis</i>	TGA	TAG	-	-	-
<i>Anopheles atroparvus</i>	-	-	TAG	-	-
<i>Anopheles coluzzii</i>	TAA	TAG	TGA	-	-
<i>Anopheles culicifacies</i>	TAA	-	-	-	-
<i>Anopheles darling</i>	TGA	-	-	-	-
<i>Anopheles funestus</i>	TAA	TAG	TAA	-	-
<i>Anopheles gambiae</i>	-	TAG	TGA	-	-
<i>Anopheles melas</i>	-	TAG	TAA	-	-
<i>Anopheles stephensi</i>	TAA	TAA	TAA	-	-
<i>Culex quinquefasciatus</i>	TGA	-	-	-	-
<i>Drosophila melanogaster</i>	-	-	-	TAA	TGA
<i>Glossina fuscipes</i>	-	-	-	TGA	TAA
<i>Lucilla cuprina</i>	-	-	-	TAA	TAA
<i>Lutzomyia longipalpis</i>	TGA	TAA	-	-	-
<i>Musca domestica</i>	-	-	-	TAA	TAA
<i>Pediculus humanus</i>	-	TAA	-	-	-
<i>Phlebotomus papatasi</i>	TGA	TAA	-	-	-
<i>Stomoxys calcitrans</i>	-	-	-	TGA	TAA
<i>Teleopsis dalmanni</i>	-	-	-	TAA	TAA

Table 2: Stop codon usage for gustatory receptor genes in the studied insects

Species	GR2	GR8	GR10	GR21	GR22
<i>Aedes aegypti</i>	TAA	-	TAG	-	TAG
<i>Anopheles albimanus</i>	TAA	TGA	-	-	TGA
<i>Anopheles arabiensis</i>	TGA	TAG	TAG	TAA	TAA
<i>Anopheles atroparvus</i>	TAG	-	-	-	-
<i>Anopheles coluzzii</i>	TAG	-	-	-	-
<i>Anopheles culicifacies</i>	TGA	-	-	-	-
<i>Anopheles funestus</i>	TAG	TGA	TAA	TGA	TAG
<i>Anopheles gambiae</i>	-	TAG	TAG	TAA	TAA
<i>Anopheles melas</i>	TGA	-	-	-	-
<i>Anopheles stephensi</i>	TAA	TAA	TAG	TGA	TAG
<i>Anopheles quadrimaculatus</i>	-	TAG	-	TAA	TAA
<i>Apis mellifera</i>	-	-	TGA	-	-
<i>Culex quinquefasciatus</i>	TAA	TGA	TGA	TAA	TGA
<i>Drosophila melanogaster</i>	TAA	TAA	TGA	TAA	TAA
<i>Glossina austeni</i>	TAA	TAG	TAG	-	-
<i>Glossina pallidipes</i>	-	TAA	TAA	-	-
<i>Ixodes scapularis</i>	-	-	TGA	-	TAG
<i>Lutzomyia longipalpis</i>	TAA	TAA	TAG	TAG	-
<i>Musca domestica</i>	TAG	-	-	-	-
<i>Nasonia vitripennis</i>	TAA	TAA	TAG	TAA	TGA
<i>Phlebotomus papatasi</i>	TGA	TGA	-	TGA	TAA
<i>Rhodnius prolixus</i>	TAA	TAA	TAA	TAA	TAA

Table 3: Stop codon usage for kinase genes in the studied insects

Species	CDK	MAPK	SAK	SER	SPH
<i>Aedes aegypti</i>	-	TAG	TGA	TGA	TAG
<i>Aedes albopictus</i>	-	TAA	TGA	TAA	TAA
<i>Anopheles albimanus</i>	-	TAG	TAA	TAA	-
<i>Anopheles arabiensis</i>	TAA	TAG	TGA	-	TGA
<i>Anopheles atroparvus</i>	-	TAG	-	-	-
<i>Anopheles coluzzii</i>	TAA	TAA	TGA	TAA	TGA
<i>Anopheles culicifacies</i>	-	TAA	-	TAG	-
<i>Anopheles funestus</i>	-	TAA	TGA	TAA	TAA
<i>Anopheles gambiae</i>	TAA	TAA	TGA	TGA	TAA
<i>Anopheles melas</i>	TAG	TAG		TGA	TAG
<i>Anopheles sinensis</i>	-	-	-	TAG	-
<i>Anopheles stephensi</i>	-	TAA	TGA	-	TAA
<i>Apis mellifera</i>	TAA	TAA	TAA	-	TGA
<i>Culex quinquefasciatus</i>	TGA	TGA	TAG	TAA	TGA
<i>Culicoides sonorensis</i>	-	TGA	TAA	-	TGA
<i>Drosophila melanogaster</i>	-	TAA	TAA	-	TAA
<i>Glossina austeni</i>	-	-	TGA	-	TGA
<i>Glossina fuscipes</i>	TGA	TAA	-	-	TGA
<i>Glossina morsitans</i>	-	TAA	TAG	-	TGA
<i>Glossina pallidipes</i>	-	-	TAA	-	TGA
<i>Glossina palpalis</i>	-	-	TGA	-	TGA
<i>Ixodes scapularis</i>	-	-	TAG	-	-
<i>Lucilla cuprina</i>	-	TAG	TAA	-	-
<i>Lutzomyia longipalpis</i>	TAA	-	TAA	TAG	TAA
<i>Megaselia scalaris</i>	-	-	-	TGA	-
<i>Musca domestica</i>	TAA	TAA	TAG	-	TGA
<i>Nasonia vitripennis</i>	-	-	TAA	-	-
<i>Phlebotomus papatasi</i>	-	-	TAG	-	TAA
<i>Rhodnius prolixus</i>	-	-	TAA	TAA	-
<i>Stomoxys calcitrans</i>	-	TAA	-	-	-

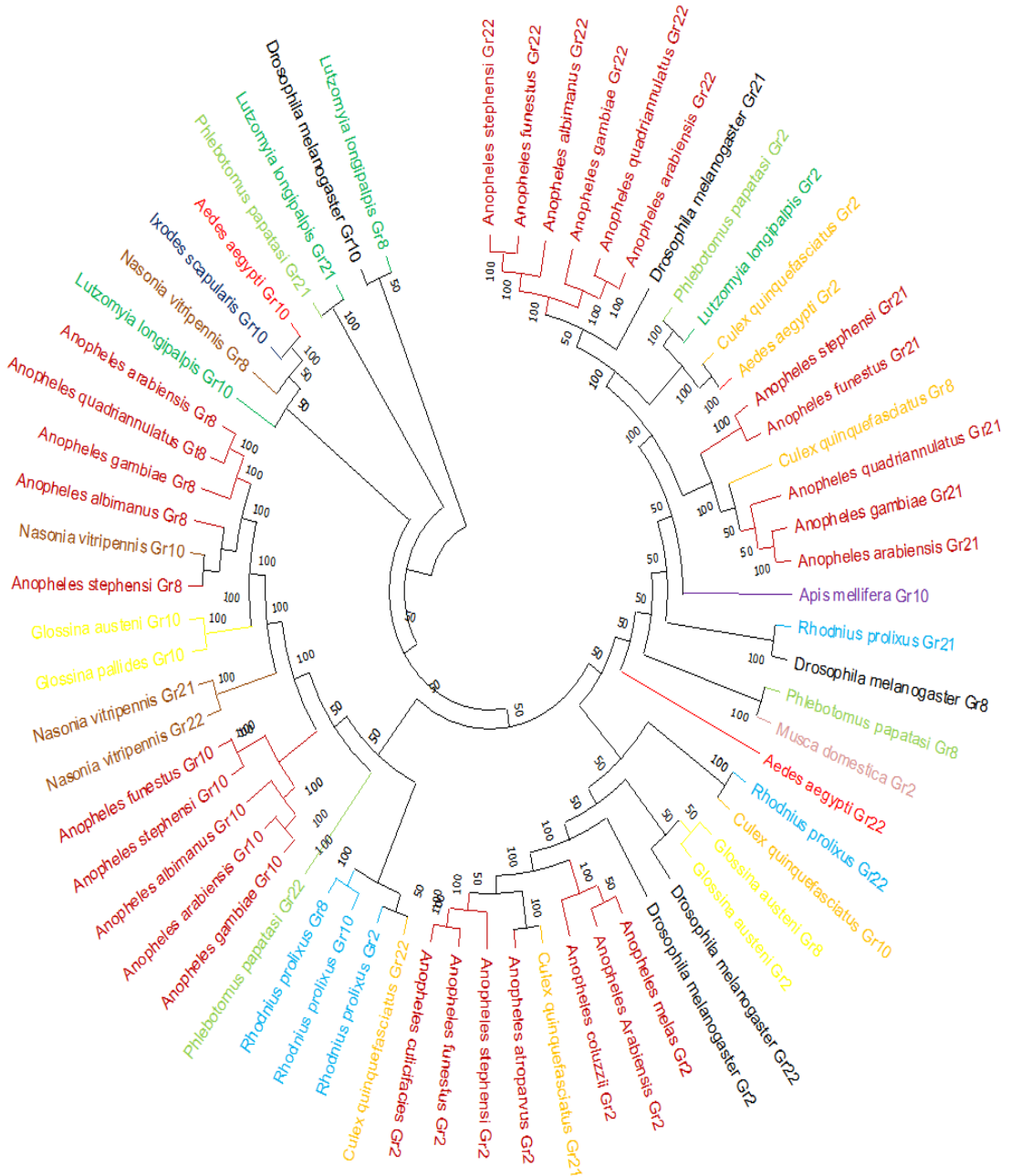


Figure 1: Phylogenetic relationship of the insects based on the five gustatory receptor genes

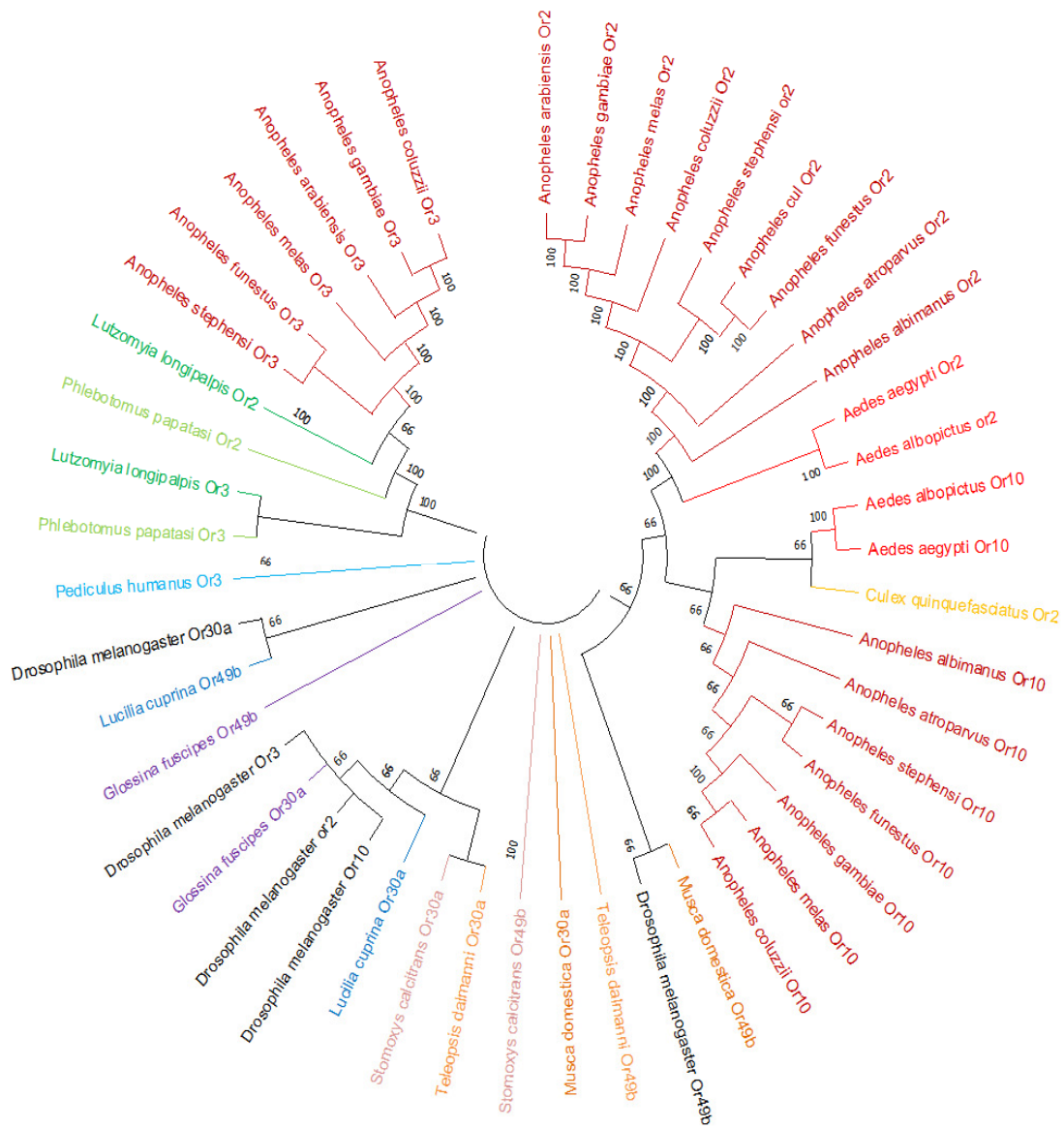


Figure 2: Phylogenetic relationship of the insects based on the five Odorant receptor genes

DISCUSSIONS

Gene orthology is central to organismic and evolutionary studies (Adipietro *et al.*, 2012). Orthologs as defined by Koonin, (2005) are "genes derived from a single ancestral gene that diverged since a speciation event; in contrast to paralogs, which are genes related via gene duplication". Recent advancement in sequencing technologies have made possible large scale studies of evolution of all known organisms using molecular data (Rognes *et al.*, 2016). Here, we used the orthology of three gene families to infer the phylogenetic relationship among various insect taxa. It can be seen that the current classification that places Anopheles, Aedes and Glossina into separate genus is correct. It can be inferred from the trees that the three genera are found in family Culicidae (mosquitoes), Aedes and Culex are more closely related. This is in line with existing classification that places both genera into subfamily Culicinae and the Anopheles genus into subfamily Anophelinae (Woodbridge and Edward, 2002). According to Scarpassa and Alencar (2013), Phlebotomus and Lutzomyia, both human leishmaniasis vectors are in subfamily Phlebotominae and this study's results generally placed them next to each other on the trees.

The stop codon usage results demonstrated variation within and among species. This is in line with what is known about codon usage bias which varies within and among species (Behura and Severson, 2012). The stop codon usage may also contribute to genome evolution. Factors such as GC content, expression level, codon position, gene length, RNA stability can influence codon usage bias within and among species (Behura and Severson, 2012). Particularly, the expression level of a gene may be correlated with the stop codon usage

bias, as the stop codon signals the end of a peptide during translation. Investigating the correlation between stop codon usage bias and expression levels is therefore necessary because optimal codons are selected for genes having high expression. Codon usage bias has been investigated in *Drosophila* (Vicario *et al.*, 2007), honeybee (Jorgensen *et al.*, 2007) and in mosquitoes (Behura and Severson, 2011). Accurate measurement of stop codon usage will be important for understanding natural selection and the genetic variation of coding sequences of genes.

CONCLUSION

In this study, we assessed the utility of fifteen genes belonging to three gene families; odorant receptors, gustatory receptors, and kinases, in constructing the phylogenetic relationship of several insect orders. We also used transcript information to assess the stop codon usage among the insect species. These analyses allowed elucidation of the interordinal phylogenetic relationships of the insects, and the results strongly supported existing classification of genus such as Anopheles, Aedes, Glossina, Lutzomyia, and Phlebotomus genera. Stop codon usage analysis demonstrated that insects closely related showed preference to the same stop codon and that most insects had a bias for the TAA, and TGA codon. This knowledge can be useful in the design of more efficient expression constructs in gene-editing technologies for vector control. The use of the gene sequences from these three gene families is an effective approach for resolving the relationships of groups of insects, and could be considered for the phylogenetic analyses of other arthropod groups.

REFERENCES

Adamski, Z., Bufo, S.A., Chowański, S., Falabella, P., Lubawy, J.,

- Marciniak, P., Pacholska-Bogalska, J., Salvia, R., Scrano, L., Słocińska, M. and Spochacz, M. (2019).** Beetles as model organisms in physiological, biomedical and environmental studies—a review. *Frontiers in Physiology* 10:431695.
- Adipietro, K.A., Mainland, J.D. and Matsunami, H. (2012).** Functional Evolution of Mammalian Odorant Receptors. *PLoS Genetics* 8 (7): e1002821.
- Ahier, A., Rondard, P., Gouignard, N., Khayath, N., Huang, S., Trolet, J., Donoghue, D.J., Gauthier, M., Pin, J.P. and Dissous C. (2009).** A new family of receptor tyrosine kinases with a venus flytrap binding domain in insects and other invertebrates activated by aminoacids. *PLoS One* 4(5): e5651.
- Baum, D. (2008).** Reading a Phylogenetic Tree: The Meaning of Monophyletic Groups. *Nature Education* 1(1): 190.
- Behura, S.K. and Severson, D.W. (2011).** Coadaptation of isoacceptor tRNA genes and codon usage bias for translation efficiency in *Aedes aegypti* and *Anopheles gambiae*. *Insect Molecular Biology* 20: 177–187.
- Behura, S.K. and Severson, D.W. (2012).** Codon usage bias; causative factors, quantification methods and genome-wide patterns: with emphasis on insect genomes. *Biological Reviews* 88(1): 49-61.
- Behura, S.K. (2015).** Insect phylogenomics. *Insect Molecular Biology* 24(4): 403–411.
- Belinky, F., Babenko, V.N., Rogozin, I.B. and Koonin, E.V. (2018).** Purifying and positive selection in the evolution of stop codons. *Scientific Reports* 8: 9260.
- Butterwick, J.A., del Marmol, J., Kim, K.H., Kahlson, M., Rogow, J., Walz, T. and Ruta, V. (2018).** Cryo-EM structure of the insect olfactory receptor Orco. *Nature* 560: 447–452.
- Cerón-Romero, M.A., Maurer-Alcalá, X.X., Grattepanche, J.D., Yan, Y., Fonseca, M.M. and Katz, L.A. (2019).** PhyloToL: A Taxon/Gene-Rich Phylogenomic Pipeline to Explore Genome Evolution of Diverse Eukaryotes. *Molecular Biology and Evolution* 36(8): 1831-1842.
- Chang, C.Y., Cheng, I.C., Chang, Y.C., Tsai, P.S., Lai, S.Y., Huang, Y.L., Jeng, C.R., Pang, V.F. and Chang, H.W. (2019).** Identification of neutralizing monoclonal antibodies targeting novel conformational epitopes of the porcine epidemic diarrhoea virus spike protein. *Scientific Reports* 9(1): 2529.
- Cunningham, F., Allen, J.E., Allen, J., Alvarez-Jarreta, J., Amode, M.R., Armean, I.M., Austine-Orimoloye, O., Azov, A.G., Barnes, I., Bennett, R. and Berry, A. (2022).** Ensembl 2022. *Nucleic Acids Research* 50(D1): D988-D995.
- Deline, B., Greenwood, J., Clark, J., Puttick, M., Peterson, K. and Donoghue, P. (2018).** Evolution of metazoan morphological disparity. *Proceedings of the National Academy of Science of the United States of America* 115(38): 8909-8918.
- Felsenstein, J. (1985).** Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39: 783-791.
- Ha, T.S. and Smith, D.P. (2009).** Odorant and pheromone receptors in insects. *Frontiers in Cellular Neuroscience* 3: 10.
- Hamilton, J., Basset, Y., Benke, K., Grimbacher, S., Miller, E.,**

- Novotny, V., Samuelson, A., Stork, E., Weiblen, D. and Yen, L. (2010).** Quantifying uncertainty in estimation of tropical arthropod species richness. *The American Naturalist* 176: 90-95.
- Huis, A. (2014).** *Edible Insects*. In: Meiselman, H. (eds). *Handbook of Eating and Drinking*. Springer, Cham. 768pp.
- Isono, K. and Morita, H. (2010).** Molecular and cellular designs of insect taste receptor system. *Frontiers in Cellular Neuroscience* 4: 20.
- Jones, W.D., Nguyen, T.A., Kloss, B., Lee, K.J. and Vosshall, L.B. (2005).** Functional conservation of an insect odorant receptor gene across 250 million years of evolution. *Current Biology* 15: R119–R121.
- Jorgensen, F.G., Schierup, M.H. and Clark, A.G. (2007).** Heterogeneity in regional GC content and differential usage of codons and amino acids in GC-poor and GC-rich regions of the genome of *Apis mellifera*. *Molecular Biology and Evolution* 24: 611–961.
- Katoh, K., Misawa, K., Kuma, K. and Miyata, T. (2002).** "MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform". *Nucleic Acids Research* 30(14): 3059–3066.
- Katz, A. and Grant, R. (2015).** Taxon-rich phylogenomic analyses resolve the eukaryotic tree of life and reveal the power of subsampling by sites. *Systematic Biology* 64(3): 406–415.
- Koonin, E.V. (2005).** Orthologs, paralogs, and evolutionary genomics. *Annual Review of Genetics* 39: 309-338.
- Li H. (2018).** Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34(18): 3094–3100.
- Manning, G., Plowman, G.D., Hunter, T. and Sudarsanam, S. (2002).** Evolution of protein kinase signaling from yeast to man. *Trends in Biochemical Sciences* 27: 514-520.
- Mikulak, E., Gliniewicz, A., Przygodzka, M. and Solecka, J. (2018).** *Galleria mellonella* L. as model organism used in biomedical and other studies. *Przegląd Epidemiologiczny* 72: 57–73.
- Mirarab, S., Nguyen, N., Guo, S., Wang, L.S., Kim, J. and Warnow, T. (2015).** PASTA: Ultra Large Multiple Sequence Alignment for Nucleotide and Amino-Acid Sequences. *Journal of Computational Biology* 22(5): 377-386.
- Plotkin, J. and Kudla, G. (2011).** Synonymous but not the same: the causes and consequences of codon bias. *Nature Reviews Genetics* 12: 32–42.
- Rauch, J., Volinsky, N., Romano, D. and Kolch, W. (2011).** The secret life of kinases: functions beyond catalysis. *Cell Communication and Signaling* 9(1): 23.
- Rognes, T., Flouri, T., Nichols, B., Quince, C. and Mahé, F. (2016).** VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4: e2584.
- Scarpassa, V.M. and Alencar, R.B. (2013).** Molecular taxonomy of the two *Leishmania* vectors *Lutzomyia umbratilis* and *Lutzomyia anduzei* (Diptera: Psychodidae) from the Brazilian Amazon. *Parasites & Vectors* 6(1): 258.
- Simon, S., Narechania, A., DeSalle, R. and Hadrys, H. (2012).** Insect Phylogenomics: Exploring the Source of Incongruence Using New Transcriptomic Data. *Genome Biology and Evolution* 4(12): 1295–1309.

- Smart, R., Kiely, A., Beale, M., Vargas, E., Carraher, C., Kralicek, A.V., Christie, D.L., Chen, C., Newcomb, R.D. and Warr, C.G. (2008).** Drosophila odorant receptors are novel seven trans-membrane domain proteins that can signal independently of heterotrimeric G proteins. *Insect Biochemistry and Molecular Biology* 38: 70–80.
- Tamura, K., Stecher, G. and Kumar, S. (2021).** MEGA11: Molecular Evolutionary Genetics Analysis version 11. *Molecular Biology and Evolution* 38: 3022-3030.
- Tamura, K. and Nei, M. (1993).** Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* 10: 512-526.
- Tihelka, E., Cai, C., Giacomelli, M., Lozano-Fernandez, J., Rota-Stabelli, O., Huang, D., Engel, M.S., Donoghue, J. and Pisani, D. (2021).** The evolution of insect biodiversity. *Current Biology* 31(19): R1299–R1311.
- Vicario, S., Moriyama, E.N. and Powell, J.R. (2007).** Codon usage in twelve species of *Drosophila*. *BMC Evolutionary Biology* 7: 226.
- Wheeler, W. (1989).** *The systematics of insect ribosomal DNA*. In: Fernholm, B., Bremer, K. and Jornvall, H. (Eds.). *The Hierarchy of Life: Molecules and Morphology in Phylogenetic Analysis*. Elsevier. pp307-321.
- Woodbridge, F. and Edward, W. (2002).** *Mosquitoes (Culicidae)*. In: Mullen, G. and Durden, L. (eds.), *Medical and Veterinary Entomology*. Academic Press. pp203-262.
- Young, A.D. and Gillung, J.P. (2020).** Phylogenomics – principles, opportunities and pitfalls of big-data phylogenetics. *Systematic Entomology* 45: 225–247.
- Zhan, Q., Wang, N., Jin, S., Tan, R., Jiang, Q. and Wang, Y. (2019).** ProbPFP: a multiple sequence alignment algorithm combining hidden Markov model optimized by particle swarm optimization with partition function. *BMC Bioinformatics* 20: 573.

