Düzce University
Journal of Science & Technology

Analysis of Positive Selection Provides Insights into Lifestyle- and Lineage-Specific Patterns of Molecular Evolution in Insects

Mehmet DAYI[a],[*]

ª Forestry Vocational School, Duzce University, Düzce, TÜRKİYE
* correspondig author: mehmetdayi@duzce.edu.tr
DOI: 10.29130/dubited.955354

ABSTRACT

Insects are among the most divergent and most rapidly evolving species, which allow them to adapt to virtually all ecosystems. Successful adaptation requires overcome of challenging environmental conditions. The best-known molecular mechanism underlying successful adaptation is positive selection. This mechanism favors in species by gaining new beneficial mutations and transferring these beneficial mutations to new generations in populations via reproduction. In this study, a total of 12 insect species belonging to 6 orders and two morphogenesis groups were used to investigate positive adaptive selection in insects and their common ancestors using a total of 535 one-to-one single-copy ortholog genes. The highest number of the positively selected gene was found in *Onthaphagus taurus* and *Dendroctanus ponderosae*, and the lowest number of positively selected genes were found in a homopteran species, *Acyrthosiphon pisum.* The highest number of positively selected genes was detected in the common ancestor of the orders Lepidoptera and Diptera, followed by the node that separated Hymenoptera from a recent common ancestor of the orders Homoptera and Isoptera. Genes involved in the fundamental biological process digestion, oxidative reduction, transcription, and translation were among the core positively selected genes. Lifestyle and lineage-specific genes were found to be under positive selection.

Keywords: Insects, Genome, Molecular evolution

Pozitif Seçilim Analizi, Böceklerde Yaşam Tarzına ve Soya Özgü Moleküler Evrimin İzlerini Ortaya Çıkarmaktadır

Öz

Böcekler, en çeşitli ve hızlı evrim geçiren organizmalar arasındadır ve bu, böceklerin neredeyse tüm ekosistemlere uyum sağlamalarına izin vermektedir. Başarılı bir adaptasyon, zorlu çevre koşullarının üstesinden gelmeyi gerektirir. Başarılı adaptasyonun altında yatan bilinen en iyi moleküler mekanizma pozitif seçilimdir. Bu mekanizma, yeni faydalı mutasyonlar kazanarak ve bu faydalı mutasyonları üreme yoluyla popülasyonlarda yeni nesillere aktararak türlerin lehine olmaktadır. Bu çalışmada 6 takım ve iki başkalaşım grubuna ait toplam 12 böcek türü kullanılmıştır. Bu böceklerde ve ortak atalarında adaptif pozitif seçilim toplam 535 bire bir tek kopya ortolog genlerin kodlayan dizileri kullanılarak incelenmiştir. En fazla pozitif seçilime maruz kalmış gen sayısı *Onthaphagus taurus* ve *Dendroctanus ponderosae*'de, en düşük pozitif seçilime uğramış gen sayısı ise bir homopteran türü olan *Acyrthosiphon pisum*'da bulunmuştur. Soya dayalı analizlerde ise, en yüksek sayıda pozitif seçilime uğramış gen Lepidoptera ve Diptera takımlarının ortak atasında ve onları takiben Hymenoptera'yı Homoptera ve Isoptera takımlarının yakın zamandaki ortak atasından ayıran atada tespit edilmiştir. Sindirim, oksidatif indirgeme, transkripsiyon ve translasyon gibi temel biyolojik süreçte yer alan genler, pozitif olarak seçilen ortak genler arasındadır. Yaşam tarzı ve soya özgü genlerin pozitif seçilim altında olduğu bulunmuştur.

Anahtar Kelimeler: Böcekler, Genom, Moleküler evrim

I. INTRODUCTION

Insects are considered one of the most successful organisms adapting to virtually all ecosystems on Earth. More than one million insect species have been identified so far, and their actual species number is estimated between 2 and 10 million [1]. Insects are recognized one of the earliest land animals, and their existence on earth dates back to around 400 million years [1].

Insects have been able to spread to a wide range of habits and have gained excellent adaptation and survival skills [2], which require beneficial mutations, also known as the positive selection, on key genes which are associated with signal mechanisms [3], transcription and translation [4], digestion of various food materials, reproduction, and detoxification of chemicals [5] released by host defense mechanisms or pesticides [6].

Positive selection is a mechanism by which species gain beneficial mutations required to overcome challenging environmental conditions and adapt to new environments [7]. Positive selection is a kind of natural selection by which beneficial phenotypes and genotypes become dominant after gaining beneficial mutations over other phenotypes and genotypes, which cause a change of allele frequency over time in the advantage of beneficial phenotype and genotype. [8]. Therefore, positive selection is recognized as the driving force of adaptive evolution and is one of the most powerful molecular evolutionary mechanisms used by species to maintain their populations.

Divergence ratio (ω or omega) at nonsynonymous (d_N) and synonymous (d_S) substitution rates is widely used to measure selection in evolutionary genetic studies. $dN/dS < 1$, $dN/dS = 1$ and $dN/dS > 1$ refer to purifying selection, neutral evolution, and positive selection, respectively [9]. In evolutionary studies, this ratio is accepted as an accurate and robust measure of adaptive evolution that acts on protein-coding sequences.

Advancements in genome sequencing accelerated genome-wide studies. To date, over one hundred insect genomes have been sequenced [10]. Several genome-wide evolutionary studies have been performed to identify genes under positive selection, but these studies included only a limited number of insect orders [11]-[13]. Therefore, it still largely remains a comparative positive selection event in a wide range of insect orders.

In this study, a total of 12 insect species from 6 orders were investigated signatures of positive selection by applying branch-site evolutionary selection analysis. In the analysis a total of 535 one-to-one single-copy gene set with a corresponding phylogenetic tree was used.

II. MATERIAL AND METHOD

A. INSECT DATA SET

A total of 12 insect species were used in the present study (Table 1). These include 7 species from Coleoptera (*Anoplophora glapripennis, Agrilus planipennis, Tribolium castaneum, Leptinotarsa decemlineta, Nicrophorus vespilloides, Onthaphagus taurus and Dendroctanus ponderosae*) and 1 species from Hymenoptera (*Apis melifera*), 1 species from Hemiptera (*Acyrthosiphon pisum*), 1 species from Isoptera (*Zootermipsis nevadensis*), 1 species from Diptera (*Drosophila melanogaster*) and 1 species from Lepidoptera (*Danaus plexippus*). The protein data set of each species was retrieved from National Center for Biotechnology Information (NCBI).

Table 1. *Insects used in the present study.*

Order name	Family name	Species name
Coleoptera	Cerambycidae	*Anoplophora glapripennis*
	Buprestidae	*Agrilus planipennis*
	Curculionidae	*Tribolium castaneum*
	Curculionidae	*Dendroctanus ponderosae*
	Silphidae	*Nicrophorus vespilloides*
	Chrypsomelidae	*Leptinotarsa decemlineta*
	Scarabaidae	*Onthaphagus taurus*
Diptera	Drosophilidae	*Drosophila melanogaster*
Isoptera	Archotermopsidae	*Zootermipsis nevadensis*
Lepidoptera	Nymphalidae	*Danaus plexippus*
Hymenoptera	Apidae	*Apis melifera*
Homoptera	Aphididae	*Acyrthosiphon pisum*

B. PHYLOGENOMIC RELATIONSHIPS IN INSECT SPECIES

For phylogenetic tree construction, one-to-one ortholog genes were identified using OrthoFinder v2.2.6 [14] with default parameters. Protein sequences of resulting one-to-one ortholog genes (535 genes) were aligned using MAFFT v7.221 [15] and alignments were further trimmed to remove poorly aligned positions and regions using trimAL v1.4.rev13 [16] with the default parameters. Alignments passing the filtering step was concatenated into one super matrix file, and this file was supplied as input to RAxML v8.0.26 [17] to generate the phylogenetic tree with 1000 bootstrap values using partition model to set the best amino acid substitution model for each ortholog gene. The resulting phylogenetic tree was further visualized using the ETE3 suite [18].

C. TEST OF ADAPTIVE SELECTION THAT ACTS ON BRANCHES AND SPECIES

The ratio $\omega = d_N/d_S$ (the ratio of nonsynonymous (d_N) to synonymous (d_S) substitution rates) was used to analyze positive selection on nodes and species in the phylogenetic tree that was generated using the one-to-one ortholog protein set (a total of 525) described previously (see previous chapter). CDS (coding sequences) of the one-to-one ortholog set were aligned to their corresponding protein sequences using pal2aln v14 with parameters --nogap --nomismatch. The aligned CDS of the one-to-one ortholog set was used as input to test positive selection on each species, and nodes that separate insect orders by comparing the null (M0, one ratio) and alternative model (Model A) of Codeml implemented in the package PAML v4.9 [9]. To test positive selection for species-wide, each species was marked in the phylogenetic tree as the foreground branch and tested against to remaining species (background).

For node test each node was marked as foreground node and tested against the remaining nodes (background) for each model.For each test, alternative and null models were tested. Each model (the null and the alternative) was launched with three independent runs for each tested (foreground) species, and nodes and the best likelihood value of each run was kept calculating the log-likelihood value.

The alternative model likelihood value was contrasted to the null model likelihood value using a likehood-ratio test (LRT) in which log-likelihood ratios were compared to a chi-square distribution with 1 degree of freedom. The *P*-values of comparisons were corrected for multiple testing by applying FDR (false discovery rate) correction at 0.05 level using qvalue R package [19] in R v3.6.1 [20]. The significant one-to-one ortholog genes (<0.05 FDR value) were considered as positively selected genes.

D. FUNCTIONAL ANNOTATION OF POSITIVELY SELECTED GENES

To identify potential functions of positively selected genes, a multiple annotation approach was applied. Amino acid sequences of those genes were searched against InterProScan [21], UniProt [22], SwissProt [23], PFAM [24] and NCBI NR (non-redundant) databases. The resulting searches were used as input in Blast2GO v5 [25] annotation suite to cluster multiple annotation results into one final result to avoid overlap and assign Gene Ontology (GO) terms.

III. RESULTS

A. EVOLUTIONARY RELATIONSHIPS IN INSECTS

The phylogenetic relationships of insects are given in Figure 1. The phylogenetic tree separated insects into two clades. While Coleoptera, Diptera, and Lepidoptera perform a clade, Hymenoptera, Homoptera, and Isoptera orders perform another clade. The orders Diptera and Lepidoptera were clustered as sister clades together, and these clades shared a common ancestor with Coleoptera insects. In another clade, Isoptera and Homoptera were clustered as sister species and shared a common ancestor with Hymenoptera order.
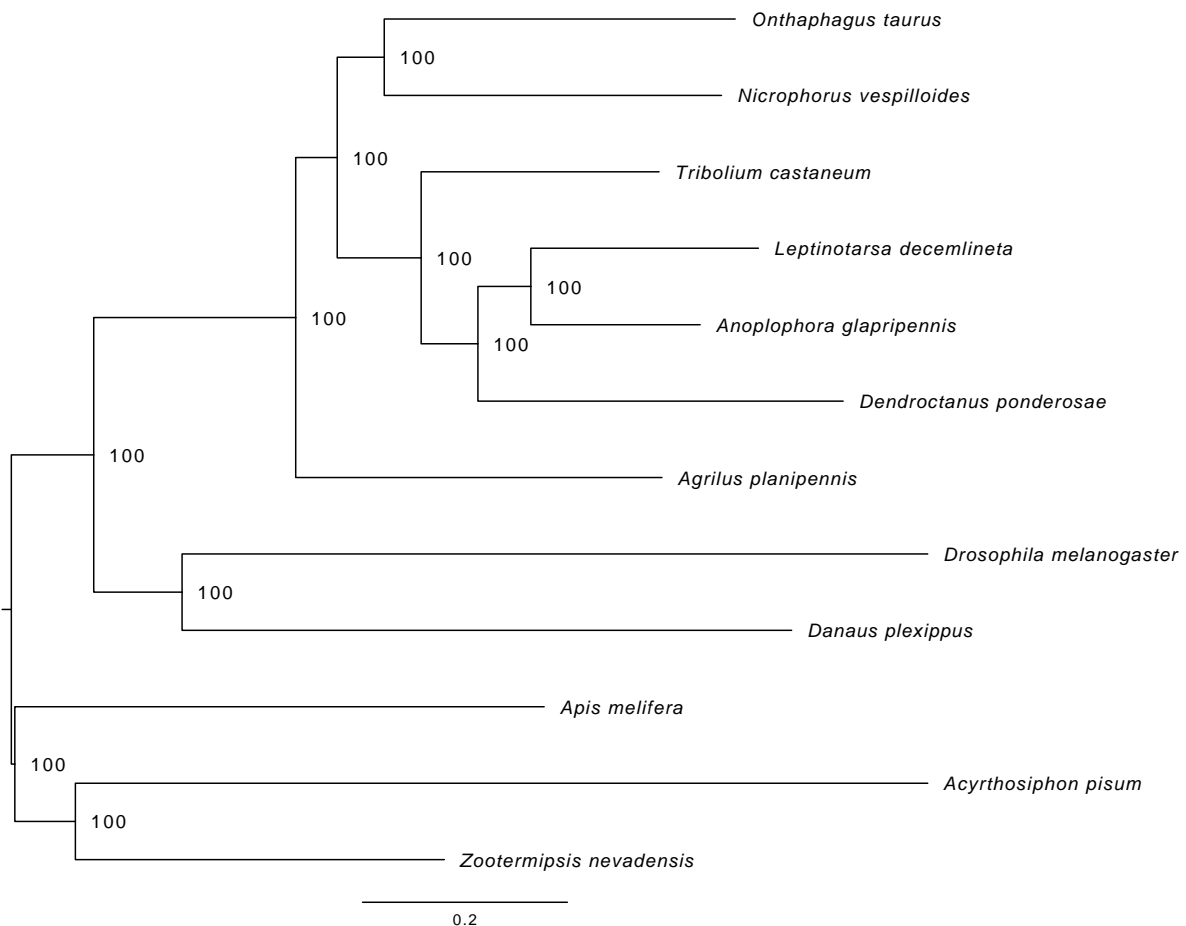

Figure 1. *Phylogenetic relationships among insect species.*

B. POSITIVE SELECTION IN INSECTS

A number of positively selected genes varied among species and branches tested. The highest number of positively selected genes were detected in *Onthaphagus taurus* (55 genes) and followed by *Dendroctanus ponderosae* (49 genes) (Table S1). The lowest number of positively selected genes were

767

detected in *Acyrthosiphon pisum* (19 genes) (Table S1). In tested nodes, the highest number of positively selected genes (92 genes) were detected in the node that separates the orders Coleoptera and from a recent common ancestor of the orders Lepidoptera and Diptera, followed by the node (73 genes) that separates the order Hymenoptera from a recent common ancestor of the orders Homoptera and Isoptera (Table S2).

A. FUNCTIONS OF POSITIVELY SELECTED GENES

In tested species, positively selected genes are mostly involved in binding (ATP binding, protein binding, and DNA binding), peptidase, transmembrane, ribosome, transferase activity, oxidation-reduction process, ATP binding, and translation (Figure 2). Functions of positively selected genes were found to be conserved in insect orders. For example, genes involved in tRNA synthesizing were observed in mostly coleopteran insects, while genes having signal-related functions (signal peptide production and signal process) were positively selected only in Hymenoptera and Diptera. In addition, genes associated with chitin-binding were positively selected in Lepidoptera only. Another example of order-specific positively selected genes is GTPase activity. Genes associated with GTPase activity were positively selected in Isoptera.
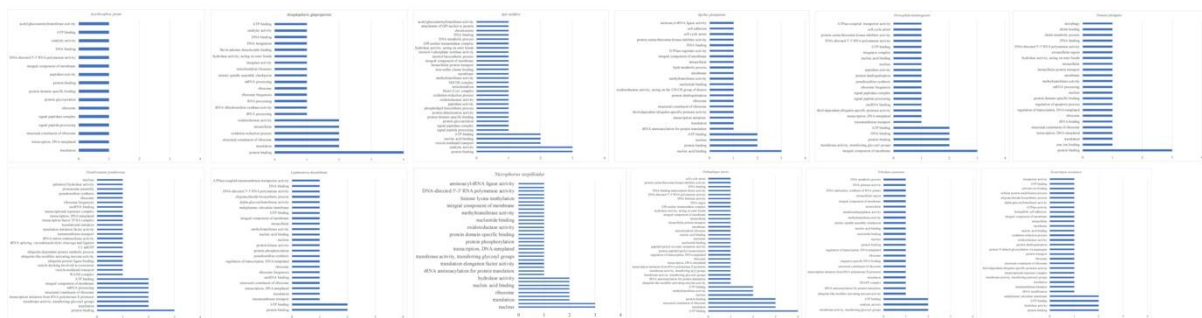

Figure 2. *Gene Ontology Enrichment results of positively selected genes in insects.*

On the other hand, species-specific positively selected genes varied among species. In *A. pisum*, genes encoding the signal activity process were found to expand, and genes having oxidoreductase activity were dominant in *A. glapripennis*. Signal peptide coding genes were detected as positively selected in *A. melifera* and *D. melanogaster*. Chitin process-associated genes were positively selected in *D. plexippus*. Protein synthesis associated genes (transcription, translation) were expanded positively in *D. ponderosae*. In *O. Taurus*, ATP binding was the most dominant function in positively selected genes.

In tested nodes, although it was found that positively selected genes are primarily involved in protein binding, ATP binding, translation, and structural constituent of ribosome, node-specific positively selected gene expansions were observed (Figure 3a, b,c).
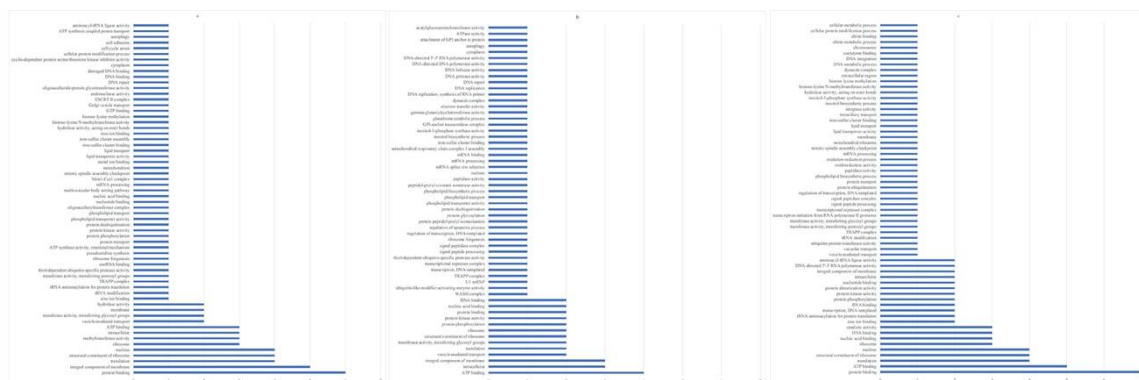

Figure 3. *Gene Ontology Enrichment results of positively selected genes in nodes.*

In addition, tRNA modification and protein post-process associated genes such as protein phosphorylation and protein kinase activity were found to under positive selection in the recent common ancestor of the orders Coleoptera, Diptera and Lepidoptera. It was found that number of positively selected genes that are associated with protein binding, integral component of membrane, translation, structural constituent of ribosome functions was higher in this node than other tested nodes (Figure 3a). Additionally, this node includes cyclin-dependent protein serine/threonine kinase inhibitor activity. In the node, the common ancestor of the orders Diptera and Lepidoptera, positively selected genes that are involved in the signal process (signal peptide processing and signal peptide complex), mRNA process (mRNA splice site selection, mRNA processing and mRNA binding) and phospholipid process (phospholipid transporter activity, phospholipid transport and phospholipid biosynthetic process) were found to be node-specific genes (Figure 3b). The node, common ancestor of the orders Hymenoptera, Homoptera, and Isoptera, was found to have positively selected genes involved in transferase activity and chitin metabolic process and chitin-binding function (Figure 3c).

IV. DISCUSSION

In the present study, a total of 12 insect species from 6 orders and common ancestors of these orders were examined to identify positively selected genes.

Insect species were found to share a common evolutionary mutation trend. It was observed that the positive selection mechanism acted on the fundamental genes of species. These genes include binding activities such as ATP binding, protein binding, and DNA binding. This function plays an essential role in gene activity by regulating their expression timing or expression level, which reflects that regulation activity of genes may have been exposed to a strong evolutionary force to switch on/off those genes based on the requirement of these species in challenging environments. Similar evolutionary trends were reported in ant [11] and social bees [12]. These studies also reported species-specific expansion or variation of positively selected gene functions. For example, in ant genomes, it was reported that mitochondrial genes were under positive selection [11].

Another common evolutionary trend shared by all insects was found in feeding mechanisms. Genes encoding peptidase, which is used to digest food molecules into small peptides [24], were found to be positively selected in all insects. Therefore, it can be said that insect genes associated with digestion activity were another mechanism to adapt to environments and target of evolutionary forces. Similar to these fundamental biological and metabolic processes, genes involved in the oxidation-reduction process are key factors in overcoming host defense, and pesticides to survive and maintain populations [5] were found to be under positive selection. These results suggest that insect species faced with common environmental challenges to adapt to new food sources and used similar mechanisms to overcome these challenges, such as regulating developmental process via modifying gene regulations by binding activity and mutations on digestion-associated genes such as peptidases. These results showed that genes involved in core metabolic and biological processes are under strong evolutionary pressure and are evolving more rapidly.

In addition to conserved adaptive evolution among insects, some species have been a target of different evolutionary forces, which allowed these insect species to gain beneficial mutations on different genes. For example, genes involved in the signal process were found to under positive selection in Diptera and Hymenoptera species. Positively selected genes encoding signal-related genes have been previously reported in several Hymenoptera and Diptera species as well [11]. In Hymenopteran species, signal transduction-associated genes have been reported as a target of positive selection [13]. Insects belonging to these orders mostly seek food sources, habitats, or mates by flying to distant environments. Therefore, signaling processes are a key factor for these species recognizing a wide range of signals and chemicals to survive and maintain their populations compared to insects in other orders.

Another example of species-specific positive selection was observed in *D. plexippus.* Genes involved in the chitin process were found to be positively selected in *D. plexippus.*

On the other hand, expansion of gene-specific positive selection was also found. Genes involved in protein synthesis (transcription, translation) in *D. ponderosae* and ATP binding in *O. taurus* were the most dominant function in positively selected genes. These results suggest that protein synthesize and ATP need of cell's during cellular respiration differ from other insects, and species-specific adaptation mechanism and need to overcome environmental challenges were greatly varied in these species compared to other species.

This study provided insights into core evolutionary mechanisms and lifestyle and lineage-specific molecular evolutionary footprints in insect genomes from various taxonomic groups and morphogenesis. It highlighted molecular evolution signatures targeting different biological and metabolic processes.

V. CONCLUSION

Insects are among the most divergent animals on Earth, and their genetic diversity and phenotypic plasticity allow them to invade and adapt to a wide range of ecosystems. Molecular mechanisms of this successful adaptation rely on their strong adaptive selection abilities that are forced by challenging environments. These are mediated by changes in nonsynonymous (d_N) ratio to synonymous (d_S) substitution rates. In this study, adaptive evolution was found to target on genes involved in fundamental biological process in all insects and nodes. Additionally, lifestyle and lineage specific positive selection were detected.

V. REFERENCES

[1]   G. Zhang, H. Wang, J. Shi, X.Wang, H.Zheng, GK. Wong, T. Clark, W. Wang, J. Wang, L. Kang, "Identification and characterization of insect-specific proteins by genome data analysis," *BMC Genomics,* vol. 8, p. 93, 2007.

[2]   M.W. Gaunt, M.A. Miles, "An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks," Molecular *Biology and Evolution,* vol. 19, no. 5, pp. 748-61, 2002.

[3]   C. Bleuven, C. R. Landry, "Molecular and cellular bases of adaptation to a changing environment in microorganisms," Proceedings *of the Royal Society B:Biological Sciences,* vol. 283, no. 1841, 2016.

[4]   J. H. Laity, B. M. Lee, P. E. Wright, "Zinc finger proteins: new insights into structural and functional diversity," Current *opinion in structural biology,* vol. 11, no. 1, pp. 39-46, 2001.

[5]   R. Feyereisen, "Insect P450 enzymes," Annual *Review of Entomology,* vol. 44, pp. 507-33, 1999.

[6]   N. Liu, T. Li, Y. Wang, S. Liu, "G-Protein Coupled Receptors (GPCRs) in Insects-A Potential Target for New Insecticide Development," *Molecules,* vol. 26, no. 10, 2021.

[7]   H. Weigand, F. Leese, "Detecting signatures of positive selection in non-model species using genomic data," *Zoological Journal of the Linnean Society,* vol. 184, no. 2, pp. 528-583, 2018.

[8]   M. Molles, A. Sher, "Ecology: Concepts and Applications", 8e, Graw Hill 2019.

[9]   Z. Yang, "PAML 4: phylogenetic analysis by maximum likelihood," *Molecular biology and evolution,* vol. 24, no. 8, pp. 1586-1591, 2007.

[10]   F. Li, M.Li, K.He, C. Huang, Y. Zhou, Z. Li, J.R. Walters, " Insect genomes: progress and

challenges," Insect *Molecular Biology,* vol. 28, no. 6, pp. 739-758, 2019.

[11]   J. Roux, E. Privman, S. Moretti, J. T. Daub, M. Robinson-Rechavi, L. Keller, "Patterns of positive selection in seven ant genomes," *Molecular biology and evolution,* vol. 31, no. 7, pp. 1661-85, 2014.

[12]   K. M. Kapheim *et al.*, "Social evolution. Genomic signatures of evolutionary transitions from solitary to group living," *Science,* vol. 348, no. 6239, pp. 1139-43, 2015.

[13]   B. A. Harpur, C.F. Kent, D.Molodtsova, J.M.D. Lebon, A.S. Alqarni, A.A. Owayss, A.Zayed, "Population genomics of the honey bee reveals strong signatures of positive selection on worker traits," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 111, no. 7, pp. 2614-9, 2014.

[14]   D. M. Emms, S. Kelly, "OrthoFinder: phylogenetic orthology inference for comparative genomics," *Genome Biology,* vol. 20, no. 1, pp. 238, 2019.

[15]   K. Katoh, D. M. Standley, "MAFFT: iterative refinement and additional methods," *Methods in Molecular Biology,* vol. 1079, pp. 131-46, 2014.

[16]   S. Capella-Gutierrez, J. M. Silla-Martinez, T. Gabaldon, "trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses," Bioinformatics, vol. 25, no. 15, pp. 1972-3, 2009.

[17]   A. Stamatakis, "RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies," Bioinformatics, vol. 30, no. 9, pp. 1312-3, 2014.

[18]   J. Huerta-Cepas, J. Dopazo, T. Gabaldon, "ETE: a python Environment for Tree Exploration," BMC *Bioinformatics,* vol. 11, p. 24, 2010.

[19]   A. Dabney, J. D. Storey, G. Warnes, "qvalue: Q-value estimation for false discovery rate control," Rv.*2.22,* vol. 1, no. 0, 2010.

[20]   R. C. Team, "R: A language and environment for statistical computing," 2013.

[21]   P. Jones, D. Binns, H.Y Chang, M. Frase, W.Li, C. McAnulla, H.McWilliam, J.Maslen, A.Mitchell, G. Nuka, S.Pesseat, A.F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S.Y. Yong, R.Lopez, S. Hunter, " InterProScan 5: genome-scale protein function classification," *Bioinformatics,* vol. 30, no. 9, pp. 1236-40, 2014.

[22]   R. Apweiler, A. Bairoch, C.H. Wu, W.C. Baerker, B. Boeckmann, S.Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, M.J. Martin, D.A. Natale, C. O'Donovan, N. Redaschi, L.SL. Yeh, " UniProt: the Universal Protein knowledgebase," Nucleic *Acids Research,* vol. 32, no. 47, pp. D115-9, 2004.

[23]   B. Boeckmann, A. Bairoch, R. Apweiler, M.C. Blatter, A. Estreicher, E. Gasteiger, M.J. Martin, K. Michoud, C. O'Donovan, I. Phan, S. Pilbout, M. Schneider, "The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003," Nucleic *Acids Research,* vol. 31, no. 1, pp. 365 370, 2003.

[24]   R.D. Finn, J. Tate, J. Mistry, P.C. Coggill, S.J. Sammut, H.R. Hotz, G. Ceric, K. Forslud, S.R. Eddy, E.L.L. Sonnhammer, A. Bateman, "The Pfam protein families database," Nucleic *Acids Research,* vol. 32, no. 40, pp. D138-41, 2004.

[25]   A. Conesa, S. Gotz, "Blast2GO: A comprehensive suite for functional analysis in plant genomics," International *Journal of Plant Genomics,* vol. 2008, pp. 619832, 2008.