Türk. entomol. derg., 2023, 47 (3): 307-316
DOI: http://dx.doi.org/10.16970/entoted.1337346

ISSN 1010-6960
E-ISSN 2536-491X

Original article (Orijinal araştırma)

Identification of odorant receptor protein genes in two Cephid stem borers (Hymenoptera: Cephidae) by high-throughput sequencing[1]

İki Cephid sap arısında (Hymenoptera: Cephidae) koku reseptör protein genlerinin yüksek verimli dizileme ile tanımlanması

Mahir BUDAK[2*]

Abstract

Insects are well adapted organisms to the terrestrial life on Earth. The evolution of the odorant receptor family is one of the causes underpinning this remarkable adaptation. Odorant receptors (ORs) sense aromas in the environment and cause the insect to respond. The ability of phytophagous insects to detect odor signals from their hosts is crucial for mating, oviposition, and feeding. The family of odorant receptor genes in Cephidae, pest on some economically important plants, is little understood. Bioinformatic tools were used to analyze the genomic data of the two pest species, *Syrista parreyssii* (Spinola, 1843) (Hymenoptera: Cephidae) (a rose pest) and *Pachycephus smyrnensis* J.P.E.F. Stein, 1876 (Hymenoptera: Cephidae), (a poppy pest), to determine their odorant receptors. The whole genome sequencing of *P. smyrnensis* collected in Sivas in 2020 was performed by next generation sequencing and short reads of *S. parreyssii* genome were obtained from previous studies. Following bioinformatic analyses, 67 and 82 putative odorant receptor genes were identified and annotated for *P. smyrnensis* and *S. parreyssii*, respectively. The ORs of these two species were found to be organized as repetitive genes in five separate clusters. No species-specific OR genes were identified in any of the investigated species. As a result, it was hypothesized that host specificity was acquired through the combined effect of multiple ORs.

Keywords: Cephini, next-generation sequencing, odorant receptors, repetitive genes, Sawflies

Öz

Böcekler, Dünya'daki karasal yaşama iyi uyum sağlamış organizmalardır. Koku reseptör ailesinin evrimi, bu olağanüstü adaptasyonun altında yatan nedenlerden biridir. Koku reseptörleri (OR'ler) çevredeki aromaları algılar ve böceğin tepki vermesine neden olur. Fitofag böceklerin konukçularından gelen koku sinyallerini algılama yeteneği çiftleşme, yumurtlama ve beslenme için çok önemlidir. Ekonomik açıdan önemli bazı bitkilerde zararlı olan Cephidae'deki koku reseptör genleri ailesi çok az anlaşılmıştır. Biyoinformatik araçlar, iki zararlı türün, *Syrista parreyssi* (Spinola, 1843) (Hymenoptera: Cephidae) (bir gül zararlısı) ve *Pachycephus smyrnensis* J.P.E.F. Stein, 1876 (Hymenoptera: Cephidae), (bir haşhaş zararlısı), koku reseptörlerini belirlemek amacıyla genomik verilerini analiz etmek için kullanılmıştır. Sivas'ta 2020 yılında toplanan *P. smyrnensis* tüm genom dizilemesi yeni nesil dizileme ile yapılmış ve *S. parreyssii* genomuna ait kısa okumalar ise önceki çalışmalardan elde edilmiştir. Analizler sonucunda *P. smyrnensis*'ten 67 olası koku reseptörü geni ve *S. parreyssii*'den 82 olası koku reseptörü geni tanımlandı ve açıklandı. Bu iki türün OR'lerinin beş ayrı kümede tekrarlayan genler olarak organize olduğu bulunmuştur. İncelenen türlerin hiçbirinde türe özgü OR genleri tespit edilmemiştir. Sonuç olarak, konakçı özgüllüğünün birden fazla OR'nin birleşik etkisi yoluyla kazanıldığı varsayılmıştır.

Anahtar sözcükler: Cephini, yeni nesil dizileme, koku almaçları, tekrarlayan genler, Testereli arılar

[1] This study was supported by Sivas Cumhuriyet University, Scientific Research Unit, Sivas, Türkiye, Grant Project No: F-508.
[2] Sivas Cumhuriyet University, Faculty of Science, Department of Molecular Biology and Genetics, 58140, Sivas, Türkiye
* Corresponding author (Sorumlu yazar) e-mail: mbudak@cumhuriyet.edu.tr
Received (Alınış): 03.08.2023 Accepted (Kabul ediliş): 30.09.2023 Published Online (Çevrimiçi Yayın Tarihi): 05.10.2023

Introduction

The Cephidae family is a member of the Cephoidea (Hymenoptera) superfamily. Fossil forms belonging to the Cephidae family were encountered in the Lower Cretaceous period (Gauld & Bolton, 1988). Cephidae contains approximately 160 species. The family is divided into three subfamilies: Athetocephinae, Australcephinae, and Cephinae. Cephinae is the most diverse and widespread subfamily among them. More than 40 of the species are distributed in Europe and the Mediterranean Region (Benson, 1951; Smith & Shinohara, 2002; Smith & Schmidt, 2009; Taeger et al., 2018). So far, twenty-eight species belonging to this family have been reported from Anatolia (Budak, 2012). All members of the Cephinae subfamily continue their life cycles as phytophagous in different plant groups. Several members of the Cephinae subfamily are responsible for substantial product losses on commercially relevant plants. *Pachycephus smyrnensis* J.P.E.F. Stein, 1876 (Hymenoptera: Cephidae), for example, causes harm to poppy plants (*Papaver somniferum* L., 1753 (Ranunculales: Papaveraceae)), whilst *Syrista parreyssi* (Spinola, 1843) (Hymenoptera: Cephidae) causes damage to roses by laying eggs on stems (Giray, 1985; Demirözer et al., 2011).

Pesticides are often used to control these insects; however, the chemicals are unsuccessful since the larvae develop in the stem (Altınayar, 1975; Li et al., 2015). Both to reduce product loss and to safeguard the environment, more environmentally friendly and effective pest control solutions must be developed. Insects use chemical sensors to identify host plants, which may be valuable for insect control (Venthur & Zhou, 2018). The chemosensory gene families are important members of insect genomes, which are involved in olfactory and gustatory functions (Sánchez-Gracia et al., 2009). Perception of chemicals in the environment enables insects to react. Therefore, the behaviors such as finding mates, hosts, food, oviposition sites and avoiding harmful situations are stimulated (Haverkamp et al., 2018). The olfactory system detects and recognizes many different volatile signal molecules belonging to various chemical classes (Keller & Vosshall, 2016). These molecules are detected by two gene families, odorant receptors (ORs), and ionotropic receptors (IRs). While ORs only detect volatile molecules, IRs can distinguish different types of molecules (Rimal & Lee, 2018). Research to date has shown that OR proteins cause stimulus-specific responses and that receptor structure provides a basis for detecting and identifying a wide range of chemicals (Hallem & Carlson, 2004; Hallem et al., 2004; Andersson et al., 2015).

The first insect ORs were identified in the fruit fly *Drosophila melanogaster* Meigen, 1830 (Diptera: Drosophilidae) at the end of the last century using a bioinformatics-based approach combined with the sequencing of an antenna cDNA library (Vosshall et al., 1999; Leal, 2013), and soon after, the data obtained by completing the genome sequence was expanded to complement 60 genes encoding 62 receptors (Vosshall et al., 2000; Robertson et al., 2003; Robertson, 2018). Since then, significant progress has been made in elucidating the roles of this ecologically essential gene family in insect biology. Based on a model of genetic evolution in *D. melanogaster*, Robertson et al. (2003) argued that the OR family may have evolved from the much larger (ancient) GR family, with the influence of terrestrial environment in insects, confirming the hypothesis of Scott et al. (2001). The origins of the GR family have been dates back to early animals, but ORs have to be a more recent gene family because they are not present in non-insect arthropods (Arthropods) (Robertson, 2015; Saina et al., 2015; Eyun et al., 2017). The appearance and spread of ORs among insects could be key to the insects' success (Yan et al., 2020). Also, it may have cleared the way for very distinct life strategies to develop over time.

Adapting to different hosts requires the development of specific odour receptors. Two large gene families, the odor-binding proteins (OBPs) and ORs, form the molecular basis of insect odor recognition (Hansson & Stensmyr, 2011). Great attempts have been made to reveal the evolutionary dynamics of the odorant receptor gene family since the initial identification of odorant receptors in insects (Clyne et al., 1999). Researchers have observed that the amount of OR genes and their diversity between insects can differ depending on the species (Missbach et al., 2014; Andersson et al., 2015). No correlation was found

between OR number and genome size. Genome size varies even between closely related Hymenoptera species but is more commonly associated with the insect's life history (Ardila-Garcia et al., 2010; Niu et al., 2022). Insect olfactory receptors function as heteromultimers consisting of at least one ligand-specific OR and the coreceptor Orco (Wicher et al., 2008). While the olfactory-related gene orco is substantially conserved among insect linages, other olfactory-related genes exhibit little sequence similarity, even among members of the Hymenoptera (Gress et al., 2013; Robertson et al., 2018). The number of OR genes vary from ten to six hundred in various insect genomes (Kirkness et al., 2010; McKenzie & Kronauer, 2018). Recent advances in the sequencing of genomes and transcriptomes have enabled the identification of putative OR genes of various insect species. Genes encoding ORs that are co-expressed in flies (*Drosophila*) and mosquitoes (*Anopheles*) are clustered together in the genome, as proven by transcriptome and genome studies involving OR genes (Ray et al., 2007). Then, expression analyses of clustered OR genes uncovered polycistronic RNA (Karner et al., 2015).

In this study, odorant receptor genes in two cephid pest species, *P. smyrnensis* and *S. parreyssii* were annotated by using next-generation sequencing and bioinformatics techniques. The gene annotation tools and were used to identify the sequences of odorant receptor gene families from *de novo* assembly of whole genome reads. Putative odorant receptor genes were then checked manually. On the basis of analyzed species, a gene family tree was constructed to infer the relationships between ORs across species. The main purpose of this study was to gain a better understanding of the evolutionary dynamics of the ORs in the Cephini subfamily.

Materials and Methods

The male specimens of *P. smyrnensis* were collected in Sivas (Türkiye), on flower of *Papaver rhoeas* L., 1753 (Ranunculales: Papaveraceae) in June 2020. Simple salting out protocol was used to recover whole genomic DNA from the specimen's hind leg (Miller et al., 1988). To verify the size distribution (>20 kbp), the isolated DNA was examined on a 0.8% agarose gel using a 1 kb extension ladder (Invitrogen). The libraries were constructed with Illumina DNA Prep with tagmentation and a 150bp pairwise short reads sequenced by Novagen (Beijing, China) by using HiSeq 2500 (Illumina, U.S.A.) Sequencing Platform. BioProject accession codes PRJNA816475 can be used to obtain short reads in GenBank and the SRA databases.

Genome assembly and annotation

The short reads produced in this study and other reads available in the SRA database (*P. smyrnensis*; SRX12828967, 164.3M paired reads and *S. parreyssii*; SRX12142183, 169.4M paired reads) were used for genome assembly analyses. The raw data was cleaned from adapter sequences and low-quality reads using fastp v 0.20.1. (Chen et al., 2018). Scaffolds construction was done using SPAdes v3.13.1 (Bankevich et al., 2012) with “-k 55, 87, 109, 121 --careful --cov-cutoff auto” options on TRUBA (TUBITAK ULAKBIM, High Performance and Grid Computing Center). Genome size (hence, GS) was estimated by Jellyfish version 1.1.11 (Marçais & Kingsford, 2011) and summary statistics for assemblies were calculated by a custom python script. The assembly quality was evaluated using BUSCO v.5.4.6 (Benchmarking Universal Single-Copy Orthologs) (Waterhouse et al., 2018) with the hymenoptera odb10 reference set.

Genome assembly of *Cephus cinctus* Norton, 1872 (Hymenoptera: Cephidae) (Acc.no: GCA_000341935.1) was included to analyses to test accuracy of genome annotations. Contigs over 1000 bp belonging to each specimen were analysed with Augustus v3.3.3 (Stanke et al., 2008) to predict coding sequences and saved as “gff” file. The amino acid sequences of each annotated gene were extracted from the “gff” file using the getAnnoFasta.pl script available in the Augustus (Stanke et al., 2006) and saved as a “fasta” file. InterProScan5 v54-87.0 (Jones et al., 2014; Blum et al., 2021) software was used to annotate the coding sequences of predicted genes identified in the previous analyses. The results were saved as “tsv” file, then sequences of ORs were pulled and saved separate files. The tblastn search was utilized to verify the

discovered sequences by the automated pipelines. The obtained results are available at *https://doi.org/10.5281/zenodo.7540708*.

Genome assembly and annotation

The OR sequences annotated from transcriptome analysis of antenna in the study of Robertson et al. (2018) was also included in the final data sets. OR protein alignments were produced with Mafft v7.453 (Katoh & Standley, 2013) with L-INS-i algorithm to find more accurately conserved regions. Gaps were trimmed using TrimAl v1.2 (Capella-Gutiérrez et al., 2009) with "-gt 0.7 -cons 60" options to remove gaps in 30% or more of the sequences. Gene tree inference was performed using RaxML v8.2.12 (Stamatakis, 2014) under the JTT + G substitution model, which was previously determined to be the most accurate model for OR gene trees (Brand & Ramrez, 2017) for 10 separate ML searches and 1000 bootstrap replicates.

Results and Discussion

The assembly of the genome of *P. smyrnensis* was accomplished by two pairs of short reads (193,287,106 in total), one from this study and one from the GenBank database. To facilitate genome assembly, we employed the haploid male *P. smyrnensis* to obtain NGS data. As a result, complexity associated with length polymorphisms between haplotypes were averted. The short reads used for *S. parreyssii* (169,365,520 reads) genome assembly were downloaded from GenBank. In their respective assemblies, the genomes of *P. smyrnensis* and *S. parreyssii* were found to have 95.6 (complete and single-copy BUSCOs: 95.4%, complete and duplicated BUSCOs: 0.2%) and 96.6 (complete and single-copy BUSCOs: 96.5%, complete and duplicated BUSCOs:0.1%) complete single-copy and duplicated genes, respectively. BUSCO is a powerful tool to assess the robustness and quality of genomic data based on expected gene content. For this purpose, BUSCO uses datasets of single copy orthologs derived from OrthoDB (Manni et al., 2021). Statistics for genome assemblies of *P. smyrnensis* and *S. parreyssii* were given in Table 1. The assembly sizes are quite similar to the in-silico genome size estimates (160.8 for *S. parreyssii* and 226.2 Mb for *P. smyrnensis*) determined by k-mer-based genome-size assessments (158 for *S. parreyssii* and 210.1 Mb *P. smyrnensis*). The NG50 and LG50 values are comparable with those for other hymenopterans in GenBank (Table 1).

Table 1. *Pachycephus. smyrnensis*, *Syrista parreyssi*, and *Cephus cinctus* genome assembly statistics

	Pachycephus smyrnensis	*Syrista parreyssi*	*Cephus cinctus*
Number of scaffolds	313,296	73,263	1,975
Contig length (max)	1,380,577	4,473,258	4,355,184
Number of scaffolds (>1K)	6366	1427	1,975
N50	113,745	733,566	622,163
L50	398	58	56
NG50	137,571	756,37	624,847
LG50	334	55	55
k-mer based estimated genome size	210.1M	158.0M	-
Genome size	226.2M	160.8 M	162.2M
k-mer based single copy region	149.1M	138.4M	-

Considering that *C. cinctus* and *S. parreyssii* are more recently evolved clades in Cephidae than *P. smyrnensis* (Budak, 2012), a decrease in genome size (GS) can be suggested. However, this reduction may have resulted from the insect-host relationship because insect-symbiont relationships are also known to play a role in genomic integration (Wernegreen, 2012). The k-mer-based estimates of single-copy regions are very close for both species (*S. parreyssii* and *P. smyrnensis*; Table 1). Therefore, this decrease in GS

can be mostly attributed to the reduction of repetitive regions. There may be a correlation between insect lifestyle and genome size, but this correlation is not always easy to demonstrate (Gregory, 2004). While the lifestyles of all three species studied here are similar, their host plants preferences are different. Although GS seems to evolve in response to host change, more research is needed to study the genomes of host plants and insects to elucidate this connection. According to Chak et al. (2021), eusocial organisms have larger genomes with the high proportion of repetitive sequences. Although several studies have attempted to explain the relationship between GS and insect lifestyles (Johnston et al., 2004; Tsutsui et al., 2008; Ardila-Garcia et al., 2010), additional studies are needed to shed light on genome size differences.

Selection of candidate OR genes

Augustus analysis projected 33,972 and 42,904 genes for the *S. parreyssii* and *P. smyrnensis* genomes, respectively (Table 2). It is a fact that this approach probably overestimates the actual number of genes present in the genome due to false positives (Saari et al., 2017). Genome annotation of non-model organisms is more challenging, but it is obvious that this obstacle will be addressed with high-throughput sequencing and improvements in bioinformatics. To identify the predicted genes obtained from the Augustus analysis, Pfam filtering and HMMER profiling were performed against the InterPro database (Blum et al., 2021) with interproseq5 software. Putative ORs were extracted from the identified genes using custom linux scripts. A total of 82, 67 and 56 ORs were found for *S. parreyssii*, *P. smyrnensis* and *C. cinctus*, respectively. To test the validity of the found OR genes, the amino acid sequences of the genes were compared with the GenBank database using the tblastn algorithm. One candidate gene identified in the *P. smyrnensis* genome did not match any OR available in the database (see supplementary material). OR genes annotated through bioinformatic analyses are named in the order in which they were discovered by the Augustus analysis. The suffixes (Cc for *C. cinctus*, Ps for *P. smyrnensis*, and Sp for *S. parreyssii*) were appended to the genes to help identify which ORs are related with which species on the tree (Figure 1).

Table 2. Number of predicted genes and ORs identified from insect genomes

	Pachycephus smyrnensis	*Syrista parreyssi*	*Cephus cinctus*
Number of predicted genes	42904	33972	36242
Number of predicted Ors	67	82	56
Ors in tandem repeats	27	45	24
Orco	1	1	1

Ten of the 72 CcinORs (CcinOR20, CcinOR23, CcinOR24, CcinOR36, CcinOR38, CcinOR41, CcinOR42, CcinOR53, CcinOR59 and CcinOR72) identified by Robertson et al. (2018) were not grouped with any of the annotated genes in the branches of the dendrogram. Five were found to be more closely linked to *S. parreyssii* OR genes (CcinOR5, CcinOR7P, CcinOR16, CcinOR34, and CcinOR43), while four were found to be more closely related to *P. smyrnensis* OR genes (CcinOR3p, CcinOR15, CcinOR28 and CcinOR71P). The remaining CcinORs were grouped with genes annotated from the *C. cintus* genome. Each of the three species tested had one conserved Orco gene as expected (g14373_Cc, g3653_Sp and g2593_Ps). Even in absence of transcriptome data, gene annotation success can be considered satisfactory. It is also necessary to correctly identify the repeating genes within a cluster to gain a better understanding of the evolution of ORs. The presence of repetitive sequences in the genome complicates genome assembly with short reads. The better way to circumvent this challenge is to use long reads in combination with short reads in genome assembly analyzes (Treangen & Salzberg, 2011; Claros et al., 2012).

Nearly half of the annotated ORs were in tandem-array (see Table 2, Figure 1), which was consistent with previous research but did not confirm the findings of Robertson et al. (2018). The phylogenetic tree using amino acid sequences of the identified ORS from these species revealed that the five clades were composed of tandem array genes (Figure 1). The existence of ORs in the genome as a tandem-array supports the

hypothesis of gene gain through duplication (Andersson et al., 2015). All three species have tandem-array genes in these five clades, however the cluster3 is dominated by *S. parreyssii* and *C.cinctus* ORs. Because *S. parreyssii* and *C. cinctus* are younger species (Budak et al., 2011), it is possible that *P. smyrnensis*' solitary OR (g28439_Ps, an orthologue of CcinOR25) at the clade's base gave rise to a new gene cluster.

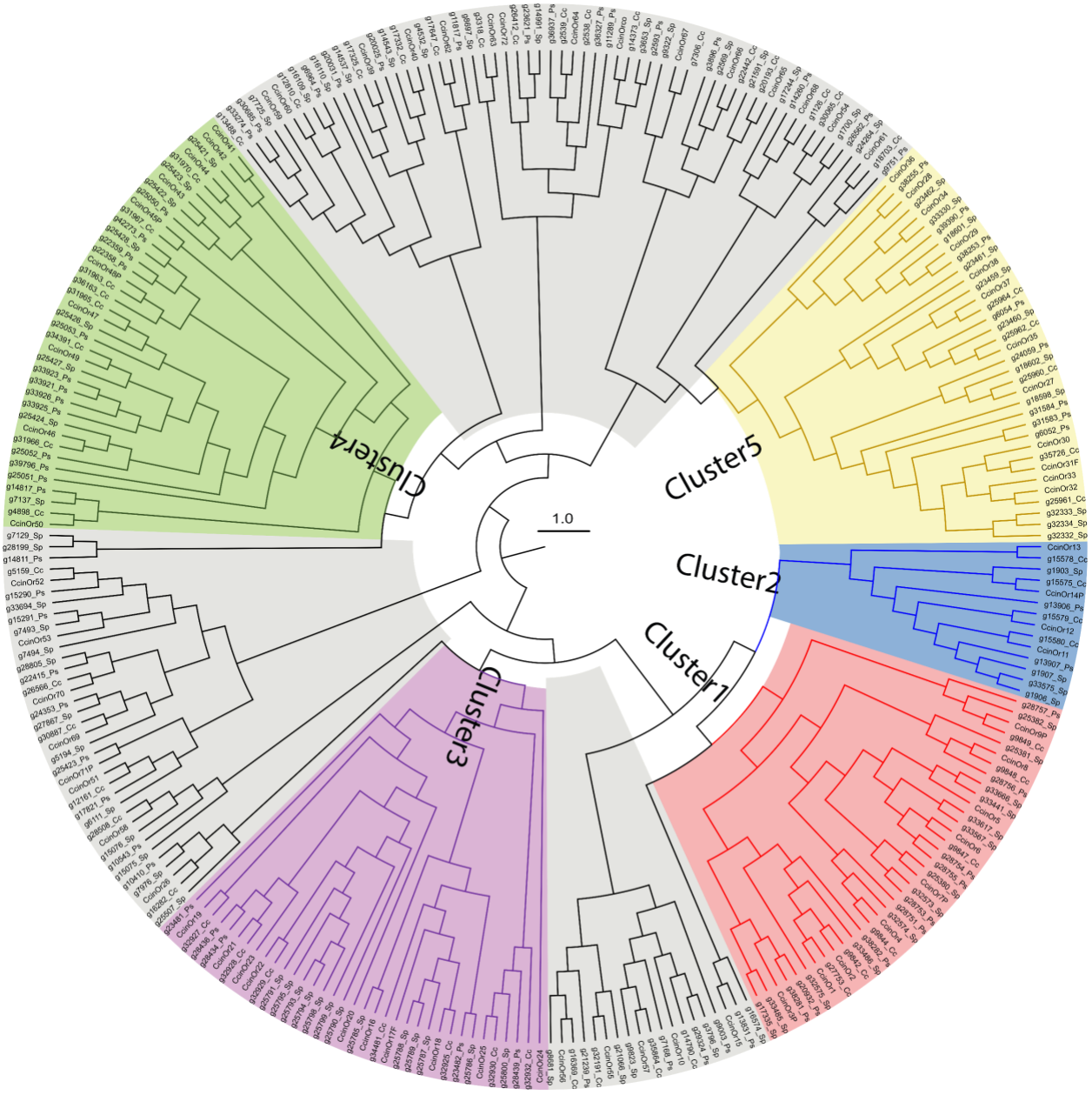

Figure 1. Phylogenetic relationships of the odorant receptor family in three cephid species. The branches are coloured by OR family clusters, red for cluster 1, blue for cluster 2, purple for cluster 3, green for cluster 4, and yellow for cluster 5. The scale bar indicates substitutions per site.

If ORs are the primary pathway for host recognition, it is reasonable to expect that organisms prefer different hosts will evolve specialized ORs for different odours. However, studies suggest that two strategies predominate for the processing and evaluation of odour signals. The first is combinatorial coding based on broadly tuned receptors: several different odours activate one type of receptor or, a particular odor can activate several types of receptors. The second is based on labelled lines: receptor types are narrowly

tuned for specific odours (Galizia, 2014; Wicher & Miazzi, 2021). There was no evidence of a species-specific OR gene family in any of the three species under investigation. It appears likely that it is more successful at host recognition when numerous receptors function in concert rather than when a single OR acts alone. The combined effect of mutations accumulated in different OR genes may have caused host shift in members of the Cephinae subfamily. On the other hand, it is likely that the plant generates a variety of odours, each of which is detected by different receptors, ensuring the insect's attraction to the host. Another notable finding is that while *S. parreyssii* has the fewest genes discovered as a result of gene annotations analysis, it has the highest number of ORs identified. This could also indicate that OR genes have evolved in concert within a clade, such as the observed evolutionary pattern in ribosomal genes. (Ganley & Kobayashi, 2007). Finally, it appears that there is no OR that is particular to host recognition among the ORs of these species. In order to better understand the host change mechanism in insects, it is important to examine a greater number of insect genomes as well as a greater number of gene interactions.

The functional characterization of insect ORs has been extensively studied in moths, flies, and mosquitoes, but has been neglected in other insect orders (Yuvaraj et al., 2021). The lack of functional data is a factor that severely limits our understanding of the molecular evolution of olfaction in symphyta. This challenge could be more easily addressed if more ORs for this group were identified and characterized. Studying the interactions between insect ORs and their ligands is important for understanding the molecular and functional evolution of insect OR families. This knowledge helps to find more effective OR agonists or antagonists for pest control. Future applications may lead to the development of more environmentally friendly strategies to replace chemicals in sawfly control.

Acknowledgements

This study is supported by the Scientific Research Project Fund of Sivas Cumhuriyet University under the project number F-508. The numerical calculations reported in this paper were partially performed at TÜBITAK ULAKBIM, High Performance and Grid Computing Center (TRUBA resources).

References

Altınayar, G., 1975. Ekin Sap Arıları (*Cephus pygmeus* (L.) ve *Trachelus tabidus* (F.) (Hymenoptera: Cephidae)'nin Konya Ilinde Biyo-Ekolojileri, Sebep Oldukları Ürün Kayıpları ve Savaş Yolları Üzerine Araştırmalar. Ankara Bölge Zirai Mücadele Araştırma Enstitüsü Yayınları, Araştırma Eserleri Serisi 36: 135 s (in Turkish).

Andersson, M. N., C. Löfstedt & R. D. Newcomb, 2015. Insect olfaction and the evolution of receptor tuning. Frontiers in Ecology and Evolution, 3: Article ID 53.

Ardila-Garcia, A. M., G. J. Umphrey & T. R. Gregory, 2010. An expansion of the genome size dataset for the insect order Hymenoptera, with a first test of parasitism and eusociality as possible constraints. Insect Molecular Biology, 19 (3): 337-346.

Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A.S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. Pham, A. D. Prjibelski, A. V. Pyshkin, A. V. Sirotkin, N. Vyahhi, G. Tesler, M. A. Alekseyev & P. A. Pevzner, 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology, 19 (5): 455-477.

Benson, R. B., 1951. Hymenoptera: 2. Symphyta. Section (a). Handbooks for the identification of British insects. Royal Entomological Society of London, London, Vol. VI, Part 2 (a): 50 pp.

Blum, M., H. Y. Chang, S. Chuguransky, T. Grego, S. Kandasaamy, A. Mitchell, G. Nuka, T. Paysan-Lafosse, M. Qureshi, S. Raj, L. Richardson, G. A. Salazar, L. Williams, P. Bork, A. Bridge, J. Gough, D. H. Haft, I. Letunic, A. Marchler-Bauer, H. Mi, D.A. Natale, M. Necci, C. A. Orengo, A. P. Pandurangan, C. Rivoire, C. J. A Sigrist, I. Sillitoe, N. Thanki, P. D. Thomas, S. C. E Tosatto, C. H. Wu, A. Bateman & R. D. Finn, 2021. The InterPro protein families and domains database: 20 years on. Nucleic Acids Research, 49 (D1): 344-354.

Brand, P. & S. R. Ramírez, 2017. The evolutionary dynamics of the odorant receptor gene family in corbiculate bees. Genome Biology and Evolution, 9 (8): 2023-2036.

Budak, M., 2012. Systematics, Biogeography and Phylogeny of Cephidae (Hymenoptera: Insecta) Species of Turkey. Cumhuriyet Üniversitesi, Fen Bilimleri Ensititüsü, (Unpublished) PhD Thesis, Sivas, 201 pp (in Turkish with abstract in English).

Budak, M., E.M. Korkmaz & H.H. Basibuyuk, 2011. A molecular phylogeny of the Cephinae (Hymenoptera, Cephidae) based on mtDNA COI gene: a test of traditional classification. ZooKeys, 130: 363-378.

Capella-Gutiérrez, S., J. M. Silla-Martínez & T. Gabaldón, 2009. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics, 25 (15): 1972-1973.

Chak, S. T. C., S. E. Harris, K. M. Hultgren, N. W. Jeffery & D. R. Rubenstein, 2021. Eusociality in snapping shrimps is associated with larger genomes and an accumulation of transposable elements. Proceedings of the National Academy of Sciences of the United States of America, 118 (24): e2025051118.

Chen, S., Y. Zhou, Y. Chen & J. Gu, 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics, 34 (17): i884-i890.

Claros, M. G., R. Bautista, D. Guerrero-Fernández, H. Benzerki, P. Seoane & N. Fernández-Pozo, 2012. Why assembling plant genome sequences is so challenging. Biology, 1 (2): 439-459.

Clyne, P. J., G. C. Warr, M. R. Freeman, D. Lessing, J. Kim & J. R. Carlson, 1999. A novel family of divergent deven-transmembrane proteins: candidate odorant receptors in *Drosophila*. Neuron, 22 (2): 327-338.

Demirözer, O., I. Karaca & Y. Karsavuran, 2011. Population fluctuations of some important pests and natural enemies found in oil-bearing rose (Rosa damascena Miller) production areas in Isparta province (Turkey). Turkish Journal of Entomology, 35 (4): 539-558.

Eyun, S., H. Y. Soh, M. Posavi, J. B. Munro, D. S. T. Hughes, S. C. Murali, J. Qu, S. Dugan, S.L. Lee, H. Chao, H. Dinh, Y. Han, H. Doddapaneni, K. C. Worley, D. M. Muzny, E. Park, J. C. Silva, R. A. Gibbs, S. Richards & C. E. Lee, 2017. Evolutionary history of chemosensory-related gene families across the Arthropoda. Molecular Biology and Evolution, 34 (8): 1838-1862.

Galizia, C. G., 2014. Olfactory coding in the insect brain: data and conjectures. European Journal of Neuroscience, 39 (11): 1784-1795.

Ganley, A. R. D. & T. Kobayashi, 2007. Highly efficient concerted evolution in the ribosomal DNA repeats: Total rDNA repeat variation revealed by whole-genome shotgun sequence data. Genome Research, 17 (2): 184-191.

Gauld, I. D. & B. Bolton, 1988. The Hymenoptera. British Museum (Natural History) & Oxford University Press, London, 332 pp.

Giray, H., 1985. A list of pests of popy (*Papaver somniferum* L.) in Turkey, with notes on the types of damage of the important species. Türkiye Bitki Koruma Dergisi (Turkish Journal of Entomology), 9 (2): 109-124 (in Turkish with abstract in English).

Gregory, T. R., 2004. Insertion-deletion biases and the evolution of genome size. Gene, 324 (1): 15-34.

Gress, J. C., H. M. Robertson, D.K. Weaver, M. Dlakić & K. W. Wanner, 2013. Odorant receptors of a primitive hymenopteran pest, the wheat stem sawfly. Insect Molecular Biology, 22 (6): 659-667.

Hallem, E. A. & J. R. Carlson, 2004. Coding of odors by a receptor repertoire. Cell, 125 (1): 143-160.

Hallem, E. A., M. G. Ho & J. R. Carlson, 2004. The molecular basis of odor coding in the Drosophila antenna. Cell, 117 (7): 965-979.

Hansson, B. S. & M. C. Stensmyr, 2011. Evolution of Insect Olfaction. Neuron, 72 (5): 698-711.

Haverkamp, A., B. S. Hansson & M. Knaden, 2018. Combinatorial codes and labeled lines: How insects use olfactory cues to find and judge food, mates, and oviposition sites in complex environments. Frontiers in Physiology, 9: Article ID 49.

Johnston, J. S., L. D. Ross, L. Beani, D. P. Hughes & J. Kathirithamby, 2004. Tiny genomes and endoreduplication in Strepsiptera. Insect Molecular Biology, 13 (6): 581-585.

Jones P., D. Binns, H. Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, S. Pesseat, A. F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S. Yong, R. Lopez & S. Hunter, 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics, 30 (9): 1236-1240.

Karner, T., I. Kellner, A. Schultze, H. Breer & J. Krieger, 2015. Co-expression of six tightly clustered odorant receptor genes in the antenna of the malaria mosquito *Anopheles gambiae*. Frontiers in Ecology and Evolution, 3: Article ID 26.

Katoh, K. & D. M. Standley, 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution, 30 (4): 772-780.

Keller, A. & L. B. Vosshall, 2016. Olfactory perception of chemically diverse molecules. BMC Neuroscience, 17: Article ID 55.

Kirkness, E. F., ,J. B. Haas, W. Sun, H. R. Braig, M. A. Perotti, J. M. Clark, S. H. Lee, H. M. Robertson, R. C. Kennedy, E. Elhaik, D. Gerlach, E. V. Kriventseva, C. G. Elsik, D. Graur, C. A. Hill, J. A. Veenstra, B. Walenz, J. M. C. Tubío, J. M. C. Ribeiro, J. Rozas, J. S. Johnston, J. T. Reese, A. Popadic, M. Tojo, D. Raoult, D. L. Reed, Y. Tomoyasu, E. Kraus, O. Mittapalli, V. M. Margam, H. Li, J. M. Meyer, R. M. Johnson, J. Romero-Severson, J. P. VanZee, D. Alvarez-Ponce, F. G. Vieira, M. Aguadé, S. Guirao-Rico, J. M. Anzola, K. S. Yoon, J. P. Strycharz, M. F. Unger, S. Christley, N. F. Lobo, M. J. Seufferheld, N. Wang, G. A. Dasch, C. J. Struchiner, G. Madey, L. I. Hannick, S. Bidwell, V. Joardar, E. Caler, R. Shao, S. C. Barker, S. Cameron, R. V. Bruggner, A. Regier, J. Johnson, L. Viswanathan, T. R. Utterback, G. G. Sutton, D. Lawson, R. M. Waterhouse, J. C. Venter, R. L. Strausberg, M. R. Berenbaum, F. H. Collins, E. M. Zdobnov & B. R. Pittendrigh, 2010. Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. Proceedings of the National Academy of Sciences, 107 (27): 12168-12173.

Leal, W. S., 2013. Odorant reception in insects: Roles of receptors, binding proteins, and degrading enzymes. Annual Review of Entomology, 58 (1): 373-391.

Li, H., R. Guan, H. Guo & X. Miao, 2015. New insights into an RNAi approach for plant defence against piercing-sucking and stem-borer insect pests. Plant, Cell & Environment, 38 (11): 2277-2285.

Marçais, G. & C. Kingsford, 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 27 (6): 764-770.

Manni, M., M. R. Berkeley, M. Seppey & E. M. Zdobnov, 2021. BUSCO: assessing genomic data quality and beyond. Current Protocols, 1 (12): e323.

McKenzie, S. K. & D. J. C. Kronauer, 2018. The genomic architecture and molecular evolution of ant odorant receptors. Genome Research, 28 (11): 1757-1765.

Miller, S. A., D. D. Dykes & H. F. Polesky, 1988. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Research, 16 (3): Article ID 1215.

Missbach, C., H. K. M. Dweck, H. Vogel, A. Vilcinskas, M. C. Stensmyr, B. S. Hansson & E. Grosse-Wilde, 2014. Evolution of insect olfactory receptors. eLife, 3: e02115.

Niu, G., M. Budak, E. M. Korkmaz, Ö. Doğan, A. Nel, S. Wan, C. Cai, C. Jouault, M. Li & M. Wei, 2022. Phylogenomic analyses of the Tenthredinoidea support the familial rank of Athaliidae (Insecta, Tenthredinoidea). Insects, 13 (10): Article ID 858.

Ray, A., W. G. van Naters, T. Shiraiwa & J. R. Carlson, 2007. Mechanisms of odor receptor gene choice in *Drosophila*. Neuron, 53 (3): 353-369.

Rimal, S. & Y. Lee, 2018. The multidimensional ionotropic receptors of *Drosophila melanogaster*. Insect Molecular Biology, 27 (1): 1-7.

Robertson, H. M., 2015. The insect chemoreceptor superfamily is ancient in animals. Chemical Senses, 40: 609-614.

Robertson, H. M., C. G. Warr & J. R. Carlson, 2003. Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. Proceedings of the National Academy of Sciences, 100 (2): 14537-14542.

Robertson, H. M., R. M. Waterhouse, K. K. O. Walden, L. Ruzzante, M. J. M. F. Reijnders, B. S. Coates, F. Legeai, J. C. Gress, S. Biyiklioglu, D. K. Weaver, K. W. Wanner & H. Budak, 2018. Genome sequence of the wheat stem sawfly, *Cephus cinctus*, representing an early-branching lineage of the Hymenoptera, illuminates evolution of hymenopteran chemoreceptors. Genome Biology and Evolution, 10 (11): 2997-3011.

Saari, T. W., A. L Schroeder, G. T. Ankley & D. L. Villeneuve, 2017. First-generation annotations for the fathead minnow (Pimephales promelas) genome. Environmental Toxicology and Chemistry, 36 (12): 3436-3442.

Saina, M., H. Busengdal, C. Sinigaglia, L. Petrone, P. Oliveri, F. Rentzsch & R. Benton, 2015. A cnidarian homologue of an insect gustatory receptor functions in developmental body patterning. Nature Communications, 6 (1): 6243.

Sánchez-Gracia, A., F. G. Vieira & J. Rozas, 2009. Molecular evolution of the major chemosensory gene families in insects. Heredity, 103 (3): 208-216.

Scott, K., R. Brady Jr., A. Cravchik, P. Morozov, A. Rzhetsky, C. Zuker & R. Axel, 2001. A Chemosensory Gene Family Encoding Candidate Gustatory and Olfactory Receptors in *Drosophila*. Cell, 104 (5): 661-673.

Smith, D. R. & A. Shinohara, 2002. A new genus and new species of Cephidae (Hymenoptera) from Sulawesi Utara, Indonesia. Proceedings of the Entomological Society of Washington, 104 (3): 624-628.

Smith, D. R. & S. Schmidt, 2009. A new subfamily, genus, and species of Cephidae (Hymenoptera) from Australia. Zootaxa, 2034 (1): 56-60.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics, 30 (9): 1312-1313.

Stanke, M., O. Keller, İ. Gunduz, A. Hayes, S. Waack & B. Morgenstern, 2006. AUGUSTUS: *ab initio* prediction of alternative transcripts. Nucleic Acids Research, 34 (2): W435-W439.

Stanke, M., M. Diekhans, R. Baertsch & D. Haussler, 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics, 24 (5): 637-644.

Taeger, A., A. D. Liston, M. Prous, E. K. Groll, T. Gehroldt & S. M. Blank, 2018. ECatSym: electronic world catalog of symphyta. (Web page: https://www.sdei.de/ecatsym) (Date accessed: June 2023).

Treangen, T. J. & S. L. Salzberg, 2011. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nature Reviews Genetics, 13 (1): 36-46.

Tsutsui, N. D., A. V. Suarez, J. C. Spagna & J. S. Johnston, 2008. The evolution of genome size in ants. BMC Evolutionary Biology, 8 (1): Article ID 64.

Venthur, H. & J. J. Zhou, 2018. Odorant receptors and odorant-binding proteins as insect pest control targets: A comparative analysis. Frontiers in Physiology, 9: Article ID 1163.

Vosshall, L. B., A. M. Wong & R. Axel, 2000. An olfactory sensory map in the fly brain. Cell, 102 (2): 147-159.

Vosshall, L. B., H. Amrein, P. S. Morozov, A. Rzhetsky & R. Axel, 1999. A spatial map of olfactory receptor expression in the Drosophila antenna. Cell, 96 (5): 725-36.

Waterhouse, R. M., M. Seppey, F. A. Simão, M. Manni, P. Ioannidis, G. Klioutchnikov, E. V. Kriventseva & E. M. Zdobnov, 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. Molecular Biology and Evolution, 35 (3): 543-548.

Wernegreen, J. J., 2012. Endosymbiosis. Current Biology, 22 (14): R555-R561.

Wicher, D. & F. Miazzi, 2021. Functional properties of insect olfactory receptors: ionotropic receptors and odorant receptors. Cell and Tissue Research, 383: 7-19.

Wicher, D., R. Schäfer, R. Bauernfeind, M. C. Stensmyr, R. Heller, S. H. Heinemann & B. S. Hansson, 2008. Drosophila odorant receptors are both ligand-gated and cyclic-nucleotide-activated cation channels. Nature, 452 (190): 1007-1011.

Yan, H., S. Jafari, G. Pask, X. Zhou, D. Reinberg & C. Desplan, 2020. Evolution, developmental expression and function of odorant receptors in insects. Journal of Experimental Biology, 223 (Suppl_1): jeb208215.

Yuvaraj, J. K., R. E. Roberts, Y. Sonntag, X. Q. Hou, E. Grosse-Wilde, A. Machara, D. Zhang, B. S. Hansson, U. Johanson, C. Löfstedt & M. N. Andersson, 2021. Putative ligand binding sites of two functionally characterized bark beetle odorant receptors. BMC Biology, 19: Article ID 16.