Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Characterization of the giant panda olfactory receptor repertoire reveals sensory evolution associated with bamboo diet specialization

LiWen Kang
Central South University of Forestry and Technology
Junjie Chen
Central South University of Forestry and Technology
mingsheng Hong
China West Normal University
Wei Wei
China West Normal University
Zuofu Xiang
Central South University of Forestry and Technology
Zejun Zhang
zhangzj@ioz.ac.cn

China West Normal University

Research Article

Keywords: Ailuropoda melanoleuca, olfactory receptor, multigene family, comparative genomics, bamboo dietary specialisation, plant volatiles, molecular docking

Posted Date: November 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11402347/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Olfactory perception mediates key behaviours including foraging and chemical communication in mammals. Olfactory receptors (ORs) constitute the largest G protein-coupled receptor multigene family responsible for detecting volatile odors. The giant panda (*Ailuropoda melanoleuca*, Carnivora: Ursidae) has evolved an extreme bamboo-specialized diet yet remains heavily dependent on olfactory cues for food selection and social signalling. Previous OR annotations for giant panda were partial, and expression profiles and potential ligand interactions remain poorly characterised.

Results

We combined OR gene identification from the giant panda genome, cross-carnivoran comparative genomic analyses, non-invasive nasal mucus transcriptome sequencing, and in silico molecular docking to characterize the giant panda OR repertoire. Using Genome2OR, we annotated 1397 OR genes (702 functional genes, 695 pseudogenes). Among eight sampled carnivoran species, the giant panda exhibited the highest OR pseudogene proportion, pointing to lineage-specific repertoire shifts linked to dietary specialisation. Most functional ORs belonged to class II and formed large chromosomal clusters. Transcriptomic profiling detected 11 expressed OR genes; OR554 and OR530 showed markedly high transcript abundance. Molecular docking simulations predicted favourable binding affinities of these two receptors to multiple bamboo-derived volatile compounds, nominating candidate receptor-ligand pairs for future experimental validation. Selective-pressure analyses revealed relaxed constraints on giant-panda OR orthologs compared with the other seven sampled carnivoran species.

Conclusions

Our combined genomic, expression, and docking results suggest that highly expressed giant-panda ORs may participate in recognising plant volatiles relevant to bamboo foraging and habitat sensing. This work provides new insights into the evolution, in-situ expression and putative functional properties of the giant-panda OR repertoire, advancing our understanding of olfactory adaptation in bamboo-specialised mammals.

1 Background

Chemoreception is an important way for animals to sense chemical signals in the environment, playing a key role in behaviors such as foraging, reproduction, territory recognition, and predator avoidance[1, 2]. In vertebrates, the olfactory system mainly relies on ORs[3], trace amine-associated receptors[4], vomeronasal type 1[5] and type 2 receptors[6], which are all seven-transmembrane G protein-coupled receptors (GPCRs)[7]. Among them, ORs form the largest family of GPCRs and are responsible for

detecting most volatile odors[8]. Odor molecules can bind to odorant-binding proteins and are transported to the ORs on the surface of olfactory sensory neurons, where they activate downstream signaling pathways to convert chemical signals into neural signals[9]. Different ORs can recognize different types of odor molecules, and a single odor molecule can activate multiple ORs[10]. This combinatorial coding allows animals to detect and distinguish a wide range of chemical signals in the environment.

OR genes encode seven-transmembrane GPCRs and usually lack introns in their coding regions, with a typical length of around 1 kb[11]. They are widely distributed in mammalian genomes. The number of OR genes varies greatly between species and is closely related to their ecological niche and reliance on the sense of smell[12]. For example, highly olfactory-dependent rodents have about 1,000 functional OR genes[13], whereas visually dominant primates have only about 400 functional OR genes, along with many OR pseudogenes[14]. This indicates that the OR gene family has undergone dynamic expansion and contraction during evolution, and these changes are closely linked to species' ecological adaptations[15, 16]. The proportion of OR pseudogenes can partly reflect the reduction of olfactory ability in a species[16], making the OR gene family an important model for studying sensory evolution and ecological adaptation.

The giant panda is a highly specialized bamboo-eating species in the order Carnivora[17]. During its long-term adaptation to a low-nutrient bamboo diet, it has developed a series of unique physiological and behavioral traits[18]. Despite its dietary specialization, the giant panda has a largely intact olfactory system and widely uses urine and perianal gland secretions for individual recognition, territory marking, and reproductive communication[19]. Hundreds of volatile compounds have been identified in its scent marks, including short-chain fatty acids, aldehydes, esters, and terpenoids[20]. With the release of high-quality genome data, studies on the giant panda OR gene family have made significant progress. Notably, multiple independent studies have annotated giant panda OR genes yet yielded inconsistent results. Liu et al. reported 611 functional OR genes in the giant panda[21], while Jian et al. identified 645 intact OR genes[22]. In 2023, Zhou et al. documented only 408 functional OR genes for this species[23]. Such obvious discrepancies exist in the counts of intact OR genes and the proportion of pseudogenes among these studies. However, previous studies have mainly focused on genome annotation and phylogenetic analysis, and functional studies of OR gene expression and receptor-ligand relationships are still limited.

Previous genomic investigations of giant-panda ORs have focused largely on gene inventory and phylogenetic classification. However, empirical expression profiles and putative receptor-ligand relationships remain insufficiently explored within a comparative carnivoran evolutionary framework. Here, we utilise the Genome2OR pipeline for comprehensive OR annotation across eight carnivoran taxa, combined with non-invasive nasal-mucus transcriptome sequencing and computational molecular-docking screening. We aim to characterise the giant-panda OR repertoire, examine lineage-specific pseudogenisation linked to dietary specialisation, and identify candidate ORs potentially responsive to

bamboo-derived volatiles. This study also provides an expanded genomic resource for future sensory-evolution research in carnivorans.

2 Results

2.1. Composition of OR Gene Repertoires

A total of 1,397 OR genes were identified in the giant panda genome, comprising 702 functional genes and 695 pseudogenes. The proportion of functional OR genes in the giant panda was 50.3%, the lowest among the eight species analyzed (Table 1), whereas the red panda showed the highest proportion of functional OR genes (72.8%). Notably, the American black bear possessed a substantially higher total number of OR genes and functional OR genes than the other seven species. In addition, the proportion of OR pseudogenes in the American black bear was relatively low (27.7%). These results indicate that the olfactory system of the American black bear may be well developed among ursid species. In contrast, the giant panda exhibited both the highest number and the highest proportion of OR pseudogenes among the eight species. In the other seven species analyzed here, the proportion of pseudogenes ranged from 27.3% to 41.3%. These pseudogenes have lost the ability to encode functional proteins during evolution. However, some OR pseudogenes may still be transcriptionally active, and their products could participate in the regulation of functional genes or pseudogenes. The complete list of predicted OR sequences is provided in Supplementary Table 2.

Table 1
Summary of the OR genes in eight species

Species	Functional OR Genes	OR Pseudogenes	Total OR Genes	Pseudogenes ratio
Giant panda (*Ailuropoda melanoleuca)*	702	695	1397	49.75%
American black bear (*Ursus americanus*)	1276	488	1764	27.66%
Asian black bear (*Ursus thibetanus*)	974	456	1430	31.89%
Brown bear (*Ursus arctos*)	946	405	1351	29.98%
Polar bear (*Ursus maritimus*)	1113	482	1595	30.22%
Spectacled bear (*Tremarctos ornatus*)	907	638	1545	41.29%
Sun bear(*Helarctos malayanus*)	813	536	1349	39.73%
Red panda(*Ailurus styani*)	793	297	1090	27.25%

Among the 702 functional OR genes identified in the giant panda, 498 genes were distributed across 17 chromosome pairs (Table 2). Notably, these genes were clustered on chromosomes: OR genes located within 1 Mb of each other were considered to belong to the same cluster. Chromosomes 8 (n = 134) and 16 (n = 149) contained the highest numbers of functional OR genes. On the remaining chromosomes, the number of functional OR genes ranged from 0 to 49 (Table 2). No OR genes were detected on chromosomes 9, 11, 14, or 19, possibly because some OR-containing genomic fragments have not yet been assembled at the chromosome level.

Table 2
Composition of functional OR genes for each chromosome of the giant panda.

ID	Location	Chromosome Size (bp)	Number of functional OR genes
Chr1	NC_048218.1	212,770,937	26
Chr2	NC_048219.1	199,809,881	9
Chr3	NC_048220.1	147,627,920	20
Chr4	NC_048221.1	144,794,249	49
Chr5	NC_048222.1	130,990,914	9
Chr6	NC_048223.1	131,587,389	12
Chr7	NC_048224.1	141,533,052	29
Chr8	NC_048225.1	129,245,720	134
Chr9	NC_048226.1	103,688,518	0
Chr10	NC_048227.1	110,580,333	4
Chr11	NC_048228.1	110,513,224	0
Chr12	NC_048229.1	81,781,225	5
Chr13	NC_048230.1	92,459,488	10
Chr14	NC_048231.1	106,650,358	0
Chr15	NC_048232.1	91,612,352	4
Chr16	NC_048233.1	91,341,929	149
Chr17	NC_048234.1	42,246,006	5
Chr18	NC_048235.1	38,124,791	7
Chr19	NC_048236.1	35,683,536	0
Chr20	NC_048237.1	30,944,335	9
ChrX	NC_048238.1	112,850,221	8

2.2 Classification of Olfactory Receptor Genes

Mammalian OR genes can be divided into two types: class I and class II, which generally form two distinct clades in phylogenetic trees and may have different evolutionary origins. Typically, only class I genes are found in the genomes of aquatic animals, whereas terrestrial animals possess both types, with most class I genes performing specific functions. In our phylogenetic tree, the blue branches represent class II ORs, and the green branches represent class I ORs (Fig. 1). Of the functional OR genes

identified in the giant panda, 597 were classified as class II and only 105 as class I, indicating a clear predominance of class II genes, a pattern commonly observed in most mammals.

Furthermore, clustering analysis of the OR gene family across eight species identified a total of 346 clusters. Among these, 167 clusters were shared across all eight species, indicating that they are evolutionarily conserved and may be responsible for detecting common odors, maintaining basic olfactory function. In the giant panda, four species-specific OR clusters were detected; the red panda had seven species-specific OR clusters, and the spectacled bear had three, while no species-specific OR clusters were found in the remaining species (Fig. 2).

2.3 Conserved Motif Patterns of OR Genes

Key features of OR genes include the absence of introns in the coding sequence (CDS), conserved protein sequences, and seven conserved transmembrane domains. To examine the conservation of OR protein sequences, the five most conserved motifs across eight species were identified using the MEME program and visualized as sequence logos (Fig. 3). In these logos, the height of each letter represents the conservation of that amino acid at the corresponding position, and the colors indicate different types of residues. The results indicate that, despite the diverse environments these species inhabit, the composition and positions of OR motifs are highly conserved, especially in regions corresponding to transmembrane domains and other functionally important residues. In addition, conserved N-linked glycosylation sites were detected in most functional OR genes, further indicating that OR proteins have highly similar structures and functions, which may be important for maintaining basic olfactory function in animals. Mammalian OR genes typically consist of a single exon, as their coding regions generally lack introns. However, untranslated exons may be present upstream of the coding region. Therefore, regardless of the number of transcript variants produced during expression, the translated protein sequences are usually identical.

2.4 Selective-pressure signatures of orthologous OR genes

We obtained one-to-one orthologous OR gene pairs across eight carnivoran species using OrthoFinder, and calculated pairwise Ka/Ks ratios to evaluate selective constraints acting on OR repertoires. The Ka/Ks distribution of orthologous OR genes was predominantly below 1, indicating that most OR orthologs have experienced strong purifying selection in carnivorans (Fig. 4). Notably, the Ka/Ks distribution of giant-panda-red-panda orthologous OR pairs was significantly higher compared with other ursid pairwise comparisons (Wilcoxon rank-sum test, $P < 0.05$).

2.5 Potential Odor Specificity of the OR Gene Family

Transcriptome analysis of nasal mucus samples from giant pandas at Chongqing Zoo yielded a total of 30 transcript sequences, among which 11 OR genes were detected as expressed (Table 3). OR554, OR530, and OR842 showed relatively high expression levels, with FPKM values ranging from 4.33 to

98.49, while the remaining OR genes exhibited lower expression levels, with FPKM values between 0.26 and 9.73 (Table 3).

Table 3
Expression profiles of olfactory receptor (OR) genes detected in nasal swab samples from two giant panda individuals.

OR gene symbol	FPKM		Number of assembled transcripts
	individual DD	individual HQZ	
OR554	98.49	8.96	5
OR530	62.26	4.33	13
OR842	13.62	24.95	1
OR418	9.73	0	2
OR380	1.32	0.1	1
OR525	1.06	0	1
OR649	0.85	0.26	1
OR763	0.8	0	1
OR38	0.45	0	1
OR231	0	0.59	2
OR469	0.35	0	1

Footnote: DD and HQZ are identifiers for two giant panda individuals. FPKM: fragments per kilobase of exon per million mapped fragments.

To further investigate the potential odor recognition properties of these OR genes, the protein sequences of the 11 expressed giant panda ORs were compared with human OR genes of previously characterized odor specificity. Three functional OR genes shared amino acid sequence identities exceeding 60% with known human ORs (Table 3). This high degree of sequence identity indicates that these OR genes may have similar odor recognition characteristics to their human counterparts.

Based on these results, the known odors of the matched human ORs, together with the bamboo-derived volatile compounds reported in the literature[24], were compiled into a set of 10 representative odor compounds. These 10 compounds were then docked against all 11 expressed giant panda OR proteins to predict potential receptor–ligand interactions.

The docking results revealed clear differences in predicted binding affinities among the different OR–odor pairs (Supplementary Table 1). Lower binding energy values indicate stronger predicted binding. Guaiacol generally showed higher binding energies with all 11 ORs, ranging from − 4.4 to − 6.1 kcal/mol, indicating relatively weak predicted interactions. In contrast, OR554, OR530, and OR38 exhibited lower

binding energies (≤ − 6 kcal/mol) with the remaining nine compounds (Fig. 5). However, the predicted binding pocket sizes differed: OR554 and OR530 had reasonably sized pockets ranging from 236 to 700 Å³, whereas OR38 had a relatively large predicted pocket of 1,582 Å³ (Supplementary Table 1). Among all interactions, ethyl 2,2-dimethylhydrocinnamal showed the lowest binding energy with OR554 (− 9.4 kcal/mol), indicating the strongest predicted binding, whereas cedrol showed relatively higher binding energies across most ORs, mainly ranging from − 6.3 to − 8.0 kcal/mol (Fig. 5).

To further investigate the molecular basis of odor recognition, representative docking complexes of OR554 and OR530 with plant-derived volatile compounds were visualized and analyzed (Fig. 6). The predicted structures showed that both ligands were located within the transmembrane cavity formed by multiple α-helices, consistent with the typical ligand-binding mode of class A GPCR ORs.

In the OR554–ethyl 2,2-dimethylhydrocinnamal complex, the ligand was deeply embedded within a compact hydrophobic pocket surrounded by several nonpolar and aromatic residues, including Phe75, Val106, Phe107, Val111, Leu161, Ile201, Val208, Phe253, Tyr261, and Tyr280. The aromatic residues Phe107, Phe253, Tyr261, and Tyr280 were positioned near the aromatic ring region of the ligand, indicating that hydrophobic and π-related interactions may contribute to ligand stabilization within the binding cavity. In addition, Asp204 and Ala205 were located close to the ligand and may participate in local structural stabilization.

Similarly, citral was predicted to bind within the transmembrane pocket of OR530, where the ligand interacted with multiple hydrophobic residues, including Tyr73, Phe105, Ile106, Leu160, Val182, Leu200, Val207 and Leu208. The aromatic residues Tyr73 and Phe105 are spatially proximal to the hydrophobic backbone of the ligand, implying that hydrophobic contacts and π-type stacking interactions collectively anchor the ligand and maintain its stable conformation inside the binding cavity. Compared with OR554, the binding cavity of OR530 appeared slightly larger and more open, potentially allowing accommodation of bulkier sesquiterpene compounds such as citral.

Overall, both OR554 and OR530 exhibited characteristic hydrophobic transmembrane binding pockets suitable for recognizing plant-derived volatile compounds, supporting their potential roles in bamboo-related odor perception in the giant panda.

3 Discussion

In this study, Genome2OR, a tool specifically optimized for olfactory receptor identification, was used to identify OR genes[25, 26]. Compared with previous studies that mainly relied on TBLASTN-based methods, Genome2OR could more efficiently identify a larger number of OR genes while maintaining sequence integrity and reducing false-positive predictions, thereby providing a more comprehensive gene set for subsequent functional analyses. However, because the giant panda genome still contains approximately 0.01 Gb of unresolved gaps, some OR genes may remain unidentified.

Notable numerical discrepancies in the count of functional OR genes of the giant panda have been reported across independent annotation studies, which stem from two core heterogeneous factors: distinct genome assembly versions, and inconsistent OR identification pipelines. Jian et al.[22] identified 645 functional ORs using an early fragmented scaffold-level giant panda genome and an in-house ORFAM pipeline with relaxed sequence filtering criteria, which retained numerous borderline candidate sequences. Liu et al.[21] recovered 611 functional ORs based on low-contiguity genome assemblies via a simple tBLASTn homology search without multi-layer structural validation. In contrast, Zhou et al.[23] adopted high-quality chromosome-level genome data and integrated a multi-step workflow combining tBLASTn alignment, Solar fragment assembly, GeneWise gene prediction and HMMER conserved domain screening; their strict filtering standard yielded only 408 functional OR genes. In the present study, we implemented the Genome2OR annotation pipeline and annotated 702 functional intact ORs for the giant panda. The larger number recovered in our work can be attributed to updated genome assembly resources, standardized full-length OR prediction algorithms, and optimized cutoff parameters that balance sensitivity and specificity compared to the three preceding studies. Pipeline-dependent variation is also evident across other surveys, which reported giant panda OR counts ranging from 639 to 1 695[16, 27, 28]. In addition, the 702 functional genes identified in this study were mapped to the chromosome level, with 498 functional genes (70.9%) successfully located on 17 chromosome pairs, whereas the previous study mapped only 186 functional genes (46.3%) to chromosomes[23]. These results indicate that the OR gene set obtained in this study has relatively high completeness and reliability, providing a more solid data foundation for subsequent analyses of OR gene evolution, expression patterns, and receptor–ligand interactions.

Olfaction is essential for vertebrates in behaviors such as foraging, social communication, and mating[29]. In giant pandas, the olfactory system plays a critical role in food selection, as pandas have relatively poor vision and rely mainly on their sense of smell to assess food quality, especially when foraging at night[30, 31, 32]. Previous studies have examined the expression of giant panda odorant-binding protein 3 (AimelOBP3) and found that AimelOBP3 preferentially binds plant-derived volatile compounds and long-chain aldehydes, with the highest affinities for several compounds present in bamboo leaves[33].

In this study, we characterized the OR gene repertoire of the giant panda and compared it with six other bear species within the family Ursidae as well as the red panda, a bamboo-specialized carnivore outside Ursidae. The results showed that the proportion of OR pseudogenes in the giant panda (49.8%) is higher than that in the other seven species analyzed here (27.3%–41.3%). The number of functional OR genes can reflect, to some extent, the selective pressure acting on the gene family during evolution. If OR genes become increasingly important over evolutionary time, they experience stronger selective pressure, which can lead to an increase in the number of functional genes, a phenomenon often referred to as birth in gene evolution. Conversely, death events occur when functional genes are disrupted and evolve into pseudogenes[34]. Extensive pseudogenization is a characteristic feature of the OR gene family.

The proportion of OR pseudogenes varies among species and is related to species-specific lifestyles. Our Ka/Ks analysis of one-to-one orthologous OR genes further supported this scenario. Although most carnivoran OR orthologs are constrained by purifying selection, the significantly elevated Ka/Ks ratios for orthologous OR gene pairs between the giant panda and the red panda indicate relaxed selective constraint on portions of the OR gene repertoire. Relaxed purifying selection could permit the accumulation of nonsynonymous mutations, which may ultimately accelerate pseudogenisation, consistent with the high pseudogene proportion observed in the giant panda OR set. Only a small number of OR orthologs showed signatures of positive selection, implying that adaptive amino-acid changes in ORs are relatively rare during bamboo-diet specialisation. For example, the platypus, an amphibious animal primarily living in aquatic environments, relies on electrosensory and tactile cues during foraging, and as a result, its class II OR genes have undergone extensive pseudogenization[35]. Similarly, the high pseudogene proportion in the giant panda may be due to its long-term specialization in bamboo forests, with a diet almost exclusively consisting of bamboo; this limited ecological exposure reduces sensitivity to other environmental odors, resulting in the pseudogenization of some functional OR genes during evolution. The red panda, which also feeds mainly on bamboo but retains a broader diet that includes fruits, insects, and other plant material and is more arboreal, showed a substantially higher proportion of functional OR genes (72.8%). This contrast indicates that the degree of dietary specialization, rather than bamboo feeding per se, may be a key determinant of OR pseudogenization in the two bamboo-eating carnivorans. It is also consistent with the broader mammalian pattern that OR family 5/8/9, which binds many plant-associated odors, retained relatively more functional copies after duplication in the giant panda[16].

Cluster analysis showed that the giant panda possesses four species-specific OR gene clusters among the eight species, indicating that it may have specialized olfactory capabilities adapted to bamboo forest habitats. In contrast, the other seven species analyzed here have more functional OR genes and lower proportions of pseudogenes, indicating a more developed olfactory system. These species inhabit more diverse environments and have access to a wider range of food sources, which likely increases the selective advantage of maintaining a sensitive sense of smell.

Transcriptome sequencing of giant panda nasal mucus samples identified 11 expressed OR genes, among which OR554 and OR530 showed markedly higher expression levels than the other OR genes. This indicates that these two OR genes may play more important roles in olfactory perception in the giant panda. Previous studies have shown that the expression levels of OR genes are often closely associated with functional requirements and environmental adaptation, and highly expressed ORs are usually involved in recognizing odor signals that are important for animal survival and behavior[36, 37]. Therefore, the high expression levels of OR554 and OR530 may be related to the demand for detecting plant-derived volatile compounds during the long-term adaptation of giant pandas to a bamboo-based diet.

Our molecular-docking simulations predicted comparatively low binding-energy values for OR554 and OR530 toward multiple volatile compounds isolated from bamboo-leaf tissues, including citral, safranal,

farnesol, β-ionone, and cedrol[24]. It should be emphasised that these outcomes represent purely computational predictions and are not equivalent to physiological ligand-recognition evidence. Even so, these outputs provide high-priority candidate pairs for subsequent heterologous-expression functional characterisation. Most of these volatile compounds are characterized by woody, citrus-like, floral, or fatty odors and are widely distributed in bamboo and other plant tissues, indicating that they may be associated with food selection, habitat recognition, and foraging behavior in giant pandas.

In addition, compared with OR554 and OR530, the other OR genes showed much lower expression levels in nasal mucus samples, indicating that these receptors may function only under specific physiological conditions, developmental stages, or particular odor stimulation conditions, and may also participate in recognizing other types of environmental odors or pheromone-related compounds. Mammalian OR gene families generally exhibit a combinatorial coding pattern, in which one receptor can recognize multiple odors and one odor can activate multiple receptors[38, 10]. Therefore, giant pandas may rely on the coordinated action of multiple OR receptors to perceive and distinguish complex plant-derived volatile compounds.

It is noteworthy that although the giant panda belongs to the order Carnivora, its long-term specialization toward feeding on bamboo may have resulted in the gradual degeneration of some OR genes associated with carnivorous feeding, whereas OR genes related to plant odor recognition may have been retained or even enhanced in expression. Elevated transcript abundance of OR554 and OR530 combined with their computationally-predicted binding preference for plant-origin volatiles indicates these paralogs could participate in plant-odor perception; definitive functional proof awaits experimental validation.

Several limitations of this study should be acknowledged. First, the transcriptome analysis was based on nasal mucus samples collected from captive giant pandas during routine health checks. Nasal mucus is likely to sample predominantly the superficial nasal lining, whereas many OR genes are expressed in deeper regions of the olfactory epithelium; consistent with this, only 30 transcript sequences corresponding to 11 OR genes were recovered, representing approximately 1.6% of the 702 functional OR genes identified in the genome. The expressed OR set described here is therefore an undersampling of the functional repertoire, and quantitative conclusions about the relative importance of individual receptors should be treated with caution. Second, the odor set used for molecular docking was selected on the basis of the volatile composition of bamboo and the known ligands of homologous human ORs; this targeted design may introduce bias, and the predicted binding energies are purely computational and require experimental validation (e.g., cell-based receptor assays). Third, because OR genes within a family share high sequence similarity, read assignment among closely related paralogs is not trivial; our expression estimates were generated using conservative mapping criteria, but some ambiguity cannot be fully excluded. Fourth, the giant panda genome still contains unresolved assembly gaps, so some OR genes may remain unidentified. Future studies using deeper sequencing of the olfactory epithelium, expanded odor sets including randomly selected control compounds, and functional validation will be needed to fully characterize the olfactory repertoire of the giant panda.

4 Conclusion

In this study, we systematically characterised the olfactory receptor multigene family of the giant panda using comparative genomics across eight carnivoran species, non-invasive nasal-mucus transcriptome sequencing, and in-silico molecular-docking screening. The Genome2OR pipeline identified 702 functional OR genes and 695 pseudogenes, yielding an expanded OR repertoire genomic resource compared with earlier giant panda reports. Cross-taxa comparison revealed that the giant panda harbours the highest proportion of OR pseudogenes in our carnivoran dataset. Phylogenetic and chromosomal-distribution analyses demonstrate giant-panda ORs are predominantly class II receptors organised in large genomic clusters, consistent with typical mammalian OR multigene-family architecture.

Despite extensive pseudogenisation across the genome, nasal-mucus transcriptomics detected a small set of transcribed OR loci among which OR554 and OR530 show markedly high expression. Molecular-docking simulations predict favourable binding between these two highly expressed receptors and multiple bamboo-derived volatile compounds; these interactions remain unvalidated by wet-lab functional experiments. Collectively, our findings illustrate adaptive sensory-genomic shifts combining gene loss via pseudogenisation and retention of specific expressed OR paralogs in the giant panda. This dataset provides valuable resources for future studies investigating mammalian OR evolution and dietary-adaptation mechanisms in carnivorans.

5 Methods

5.1 Genome Data Collection

The genome assemblies of the giant panda (*Ailuropoda melanoleuca*; GCF_002007445.2), the american black bear (*Ursus americanus*; GCF_020975775.1), the asiatic black bear (*Ursus thibetanus*; GCA_009660055.1), the brown bear (*Ursus arctos*; GCF_023065955.2), polar bear (*Ursus maritimus*; GCF_017311325.1), the spectacled bear (*Tremarctos ornatus*; GCA_028551375.1), the sun bear (*Helarctos malayanus*; GCA_028533245.1), and the red panda (*Ailurus fulgens*; GCA_002007465.1) were downloaded from the NCBI database (National Center for Biotechnology Information)[39]. The red panda, a bamboo-specialized carnivoran outside the family Ursidae, was included as an outgroup for comparative analyses.

5.2 Identification of OR genes

OR genes were identified from genome assemblies using the Genome2OR pipeline[25]. Genome2OR is an automated pipeline for annotating olfactory receptor genes from chordate genome assemblies, consisting of five integrated modules: nhmmer.py, FindOR.py, IdentifyFunc.py, Batch.py, and Iteration.py.

The annotation process begins with nhmmer.py, which employs clade-specific DNA profile hidden Markov models to scan the genome for OR loci. These profile HMMs were constructed by building DNA

multiple sequence alignments (MSAs) guided by protein MSAs, followed by model generation using hmmbuild. Given substantial sequence divergence of OR coding sequences across chordate evolutionary clades, separate profile HMMs were generated for seven major lineages: lancelets, jawless fish, jawed fish, amphibians, reptiles, birds, and mammals. Genome-wide scanning with nhmmer was performed under a stringent E-value cutoff of $1e^{-60}$[25]. Genomic regions matching OR-HMM signatures were retained as preliminary OR candidate loci.

The nhmmer output was subsequently processed by FindOR.py to extract putative OR coding sequences. For each genomic hit, fragments shorter than 1 200 bp were bidirectionally extended to reach at least 1 200 bp[25]. All potential start and stop codons were detected to define candidate coding regions. Open reading frame (ORF) assessment was performed by scanning all possible ORFs between start-stop codons on both forward and reverse strands. The longest ORF was selected as the primary candidate for further evaluation of sequence length, integrity and translational continuity. Sequences were preliminarily classified as pseudogene candidates if their longest potential coding region was shorter than 750 bp[25], exhibited frameshift mutations indicated by lengths not divisible by three nucleotides, or lacked an intact stop codon. Sequences passing this preliminary filtering were designated as potential functional genes and passed to IdentifyFunc.py for structural validation.

In the IdentifyFunc.py module, candidate sequences were combined with OR reference templates and aligned using MAFFT. Alignments were then segmented into transmembrane helices, intracellular loops, extracellular loops, and terminal regions. Four structural criteria were applied for final classification: fewer than ten matched conserved patterns, more than ten total amino-acid gaps within transmembrane domains, gaps disrupting more than two transmembrane helices, or abnormal N-terminal length. Sequences failing any one of these criteria were classified as pseudogenes based on their longest intact fragment. Sequences satisfying all structural checks were annotated as functional OR genes.

For large-scale datasets, Batch.py supported parallel annotation across multiple genomes, whereas Iteration.py implemented iterative annotation, especially suitable for non-mammalian clades. Under this iterative workflow, OR sequences recovered in the initial search round were used to update clade-specific profile HMMs; the updated models were then used in subsequent search cycles to capture divergent OR sequences that might be missed by the original HMM. The final output contained two groups: functional genes possessing intact coding sequences, complete open reading frames and conserved OR structural features; and pseudogenes carrying frameshift mutations, premature internal stop codons, or severe structural disruptions.

Compared with the OR gene identification method described by Niimura[40], Genome2OR achieved substantially improved performance. The Human Olfactory Data Explorer (HORDE) provides high-quality gold-standard OR gene annotations[41]. Our Genome2OR predictions showed high concordance with HORDE reference data: 92.3%–99.4% of identified OR sequences per species could be mapped to HORDE entries. Across all analysed species, the annotation accuracy of Genome2OR consistently outperformed the Niimura[40] workflow.

5.3 Phylogenetic tree construction

Protein sequences were aligned with the EINSi algorithm implemented in MAFFT using default parameters[42]. Phylogenetic trees were constructed to classify OR genes into class I and class II receptors and to assess their evolutionary relationships across the eight species, using the neighbor-joining (NJ) method with the Poisson correction for amino acid distances in MEGA 6[43, 44].

5.4 Chromosomal Distribution and Motif Analysis

The chromosomal locations of functional OR genes (but not pseudogenes) were visualized using TBtools[45] to better display their genomic organization. Conserved sequence motifs in the functional OR amino acid sequences were identified using the Multiple Expectation Maximization for Motif Elicitation (MEME) program[46]. Motif analysis was performed to evaluate the sequence conservation of functional OR proteins across species and to examine whether the seven-transmembrane structural fingerprints typical of ORs are preserved. The five most conserved motifs were selected, with lengths ranging from 5 to 50 amino acids. Potential N-glycosylation sites were predicted using the NetNGlyc server[47].

5.5 Transcriptome Analysis of Giant Panda Olfactory Receptor Genes

5.5.1 Sample Collection

Nasal mucus samples were collected from giant pandas at Chongqing Zoo during routine health checks. Nasal mucus, rather than tissue biopsy, was sampled because it can be obtained non-invasively without causing any discomfort to the animals[33]. Mucus contains cells shed from the nasal lining, and RNA recovered from nasal mucus has been used previously to examine gene expression in the nasal epithelium. Because nasal mucus can trap airborne particles, samples were collected with sterile swabs and immediately frozen, and RNA extraction was performed under sterile conditions. All procedures followed the regulations of Chongqing Zoo and relevant Chinese laws and standards for captive animals. All animal management protocols were approved by the Institutional Animal Care and Use Committee of China West Normal University (approval number 2026LLSC0093).

5.5.2 Transcriptome Sequencing

Nasal mucus samples were rapidly frozen in liquid nitrogen and stored at − 80°C. Total RNA was extracted using TRIzol Reagent (Invitrogen, USA). RNA purity and concentration were measured using a NanoDrop 2000 spectrophotometer (Thermo Scientific, USA), and RNA integrity was assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies, USA). Only samples with OD260/280 between 1.8–2.2 and RNA integrity number (RIN) > 7.0 were used for downstream experiments.

Qualified total RNA from each sample was used to construct cDNA libraries. Eukaryotic mRNA was enriched using Oligo(dT) magnetic beads, followed by random fragmentation into short fragments. First-strand and second-strand cDNA were synthesized using the fragmented mRNA as a template. After end

repair, A-tailing, and adapter ligation, libraries were purified with AMPure XP beads and PCR-amplified to generate the final libraries. Library quality was assessed using the Agilent 2100 Bioanalyzer, and paired-end sequencing was performed on an Illumina NovaSeq 6000 platform.

5.5.3 Transcriptome Data Quality Control and Mapping

Raw reads were filtered to remove adapter sequences, reads with a high proportion of N bases, and low-quality reads to obtain clean reads. Data quality was assessed using FastQC, and Q20, Q30, and GC content metrics were calculated. Clean reads were mapped to the reference genome (GCF_002007445.2) using HISAT2[48]. Because OR genes within a family are highly similar, reads were mapped allowing one reported alignment per read, and reads that could not be assigned uniquely to a single OR locus were excluded from OR-level expression quantification to avoid inflating the apparent expression of any paralog.

5.5.4 Gene Expression Quantification and Differential Expression Analysis

Gene-level read counts were calculated using featureCounts[49], and expression levels were normalized as FPKM (Fragments Per Kilobase of transcript per Million mapped reads). Differential expression analysis was performed using DESeq2[50]. Genes with $|\log_2 \text{fold change}| \geq 1$ and false discovery rate (FDR) < 0.05 were considered significantly differentially expressed.

5.6 Molecular Docking Analysis

The three-dimensional structures of all 11 expressed giant panda OR proteins were predicted using AlphaFold3 (https://alphafoldserver.com/)[51]. The structures of the odor molecules were downloaded from PubChem (https://pubchem.ncbi.nlm.nih.gov/)[52]. Molecular docking was performed using CBDock2 (http://183.56.231.194:8001/cbdock2/index.php) with the automatic blind docking option, following previously reported methods[53]. Docking results were visualized as heatmaps using GraphPad Prism (v8.0) software.

5.7 Selective-pressure analysis of orthologous OR genes

OrthoFinder was applied to identify one-to-one orthologous OR gene pairs across eight carnivoran species[54]. Corresponding CDS sequences of one-to-one orthologs were retrieved. Pairwise Ka (nonsynonymous substitution rate) and Ks (synonymous substitution rate) values were computed using Simple Ka/Ks Calculator implemented in TBtools[45]. The Wilcoxon rank-sum test was performed to compare the distribution of Ka/Ks ratios between species pairs.

Abbreviations

ORs
Olfactory receptors

GPCR
G protein-coupled receptor
Ka
Non-synonymous substitution rate
Ks
Synonymous substitution rate
MEME
Multiple Expectation Maximization for Motif Elicitation

Declarations

Ethics approval and consent to participate

Sample collection was performed as part of routine veterinary health examinations at Chongqing Zoo and adhered to the Chinese Regulations and Standards for Captive Animals. All sampling and experimental protocols were approved by the Institutional Animal Care and Use Committee of China West Normal University (approval number 2026LLSC0093). No wild giant pandas were captured or sampled in this study.

Consent for publication

Not applicable.

Availability of data and materials

Raw transcriptome sequencing reads generated in this study have been deposited in NCBI BioProject under accession PRJNA1478398. Supplementary tables containing the full OR gene annotation set are provided as additional files within this manuscript. All analysed public genome assemblies are downloadable from NCBI under the accession numbers listed in Methods section.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported by the National Natural Science Foundation of China (grant no. 32470516 and 31900337), and the National Key Research and Development Program of China, grant number 2023YFF1305400. We also thank the anonymous reviewers for their constructive comments.

Authors' contributions

KLW, XZF and ZZJ conceived and designed the study. KLW, CJJ, HMS and WW performed data analysis. KLW drafted the manuscript. All authors read, revised and approved the final manuscript.

Acknowledgements

We thank Chongqing Zoo for assistance in non-invasive sample collection from captive giant pandas.

References

1. Marquet N, Hubbard PC, Dasilva JP, et al. Chemicals released by male sea cucumber mediate aggregation and spawning behaviours. Sci Rep. 2018;8(1):239.
2. Motti CA, Bose U, Roberts RE, et al. Chemical ecology of chemosensation in Asteroidea: insights towards management strategies of pest species. J Chem Ecol. 2018;44(2):147-177.
3. Buck L, Axel R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. Cell. 1991;65(1):175-187.
4. Liberles SD, Buck LB. A second class of chemosensory receptors in the olfactory epithelium. Nature. 2006;442(7103):645-650.
5. Dulac C, Axel R. A novel family of genes encoding putative pheromone receptors in mammals. Cell. 1995;83(2):195-206.
6. Matsunami H, Buck LB. A multigene family encoding a diverse array of putative pheromone receptors in mammals. Cell. 1997;90(4):775-784.
7. Lindemann L, Hoener MC. A renaissance in trace amines inspired by a novel GPCR family. Trends Pharmacol Sci. 2005;26(5):274-281.
8. Kim C, Lee KK, Kang MS, et al. Artificial olfactory sensor technology that mimics the olfactory mechanism: a comprehensive review. Biomater Res. 2022;26(1):11.
9. Pourmorady AD, Lomvardas S. Olfactory receptor choice: a case study for gene regulation in a multienhancer system. Curr Opin Genet Dev. 2022;72:101-109.
10. Shirasu M, Yoshikawa K, Takai Y, et al. Olfactory receptor and neural pathway responsible for highly selective sensing of musk odors. Neuron. 2014;81(1):165-178.
11. Gilad Y. A comparison of the human and chimpanzee olfactory receptor gene repertoires. Genome Res. 2005;15(2):224-230.
12. Rospars JP, Gu YQ, Alexandre G, et al. Odour transduction in olfactory receptor neurons. Chin J Physiol. 2010;53(6):9.
13. Niimura Y, Nei M. Evolution of olfactory receptor genes in the human genome. Proc Natl Acad Sci U S A. 2003;100(21):12235-12240.
14. Niimura Y, Nei M. Extensive gains and losses of olfactory receptor genes in mammalian evolution. PLoS One. 2007;2(8):e708.
15. Dong D, He G, Zhang S, Zhang Z. Evolution of olfactory receptor genes in primates dominated by birth-and-death process. Genome Biol Evol. 2009;1:258-264.
16. Hughes GM, Boston ESM, Finarelli JA, Murphy WJ, Higgins DG, Teeling EC. The birth and death of olfactory receptor gene families in mammalian niche adaptation. Mol Biol Evol.

2018;35(6):1390-1399.
17. Zhang Z, Swaisgood R, Zhang S, Nordstrom LA, Wang H, Pan R, et al. Old-growth forest is what giant pandas really need. Biol Lett. 2011;7(3):403-406.
18. Wei F, Hu Y, Zhu L, Bruford MW, Zhan X, Zhang L. Black and white and read all over: the past, present and future of giant panda genetics. Mol Ecol. 2012;21(23):5660-5674.
19. Gilad O, Swaisgood RR, Owen MA, Zhou X. Giant pandas use odor cues to discriminate kin from nonkin. Curr Zool. 2016;62(3):333-336.
20. Hagey LR, MacDonald EA. Chemical cues identify gender and individuality in giant pandas (*Ailuropoda melanoleuca*). J Chem Ecol. 2003;29:1479-1488.
21. Liu A, He F, Shen L, Liu R, Wang Z, Zhou J. Convergent degeneration of olfactory receptor gene repertoires in marine mammals. BMC Genomics. 2019;20:977.
22. Jian Z, Li W, Zhang X, et al. Bioinformatics analysis of the giant panda olfactory receptor gene family. Sichuan Journal of Zoology. 2017;36(1):16.
23. Zhou C, Liu Y, Zhao G, et al. Comparative analysis of olfactory receptor repertoires sheds light on the diet adaptation of the bamboo-eating giant panda based on the chromosome-level genome. Animals. 2023;13(6):979.
24. Guo M, Hu ZQ, Strong PJ, et al. Evaluating the environmental health effect of bamboo-derived volatile organic compounds through analysis the metabolic indices of the disorder animal model. Biomed Environ Sci. 2015;28:595-605.
25. Han W, Wu Y, Zeng L, Zhao S. Building the Chordata olfactory receptor database using more than 400,000 receptors annotated by Genome2OR. Sci China Life Sci. 2022;65(12):2539-2551.
26. Han W, Bao S, Liu J, et al. The chordate olfactory receptor database. Protein Cell. 2025;16(4):286-295.
27. Christmas MJ, Kaplow IM, Genereux DP, Dong MX, Hughes GM, Li X, et al. Evolutionary constraint and innovation across hundreds of placental mammals. Science. 2023;380(6643):eabn3943.
28. Policarpo M, Baldwin MW, Casane D, Salzburger W. Diversity and evolution of the vertebrate chemoreceptor gene repertoire. Nat Commun. 2024;15(1):1421.
29. Zhou C, Liu Y, Zheng X, et al. Characterization of olfactory receptor repertoires provides insights into the high-altitude adaptation of the yak based on the chromosome-level genome. Int J Biol Macromol. 2022;209:220-230.
30. Kelling AS, Snyder RJ, Marr MJ, et al. Color vision in the giant panda (*Ailuropoda melanoleuca*). Learn Behav. 2006;34:154-161.
31. Hu J. The legend of the giant panda. Science Press; Beijing; 2016.
32. Huang XR, Li G, Zhang GQ, et al. Friend or foe? Using eye-tracking technology to investigate the visual discrimination ability of giant pandas. Curr Zool. 2024;70(4):430-439.
33. Zhu J, Arena S, Spinelli S, et al. Reverse chemical ecology: olfactory proteins from the giant panda and their interactions with putative pheromones and bamboo volatiles. Proc Natl Acad Sci U S A.

2017;114(46):E9802-E9810.
34. Rouquier S, Blancher A, Giorgi D. The olfactory receptor gene repertoire in primates and mouse: evidence for reduction of the functional fraction in primates. Proc Natl Acad Sci U S A. 2000;97(6):2870-2874.
35. Niimura Y, Nei M. Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods. Proc Natl Acad Sci U S A. 2005;102(17):6039-6044.
36. Kambere MB, Lane RP. Co-regulation of a large and rapidly evolving repertoire of odorant receptor genes. BMC Neurosci. 2007;8(Suppl 3):S2.
37. Nei M, Niimura Y, Nozawa M. The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. Nat Rev Genet. 2008;9(12):951-963.
38. Krautwurst D, Yau KW, Reed RR. Identification of ligands for olfactory receptors by functional expression of a receptor library. Cell. 1998;95(7):917-926.
39. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2016;44(D1):D7-D19.
40. Niimura Y. Identification of olfactory receptor genes from mammalian genome sequences. Methods Mol Biol. 2013;1003:39-49.
41. Glusman G, Yanai I, Rubin I, Lancet D. The complete human olfactory subgenome. Genome Res. 2001;11(5):685-702.
42. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772-780.
43. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 2013;30(12):2725-2729.
44. Takezaki N, Rzhetsky A, Nei M. Phylogenetic test of the molecular clock and linearized trees. Mol Biol Evol. 1995;12(5):823-833.
45. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data. Mol Plant. 2020;13(8):1194-1202.
46. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol. 1994;2:28-36.
47. Gupta R, Brunak S. Prediction of N-glycosylation sites in human proteins. Proc Int Conf Intell Syst Mol Biol. 2002;10:194-199.
48. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37(8):907-915.
49. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923-930.
50. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.

51. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature. 2024;630(8016):493-500.
52. Kim S, Chen J, Cheng T, et al. PubChem 2023 update. Nucleic Acids Res. 2023;51(D1):D1373-D1380.
53. Liu Y, Yang X, Gan J, Chen S, Xiao ZX, Cao Y. CBDock2: improved protein–ligand blind docking by integrating cavity detection, docking and homologous template fitting. Nucleic Acids Res. 2022;50(W1):W159-W164.
54. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biology. 2019;20:238.

Figures

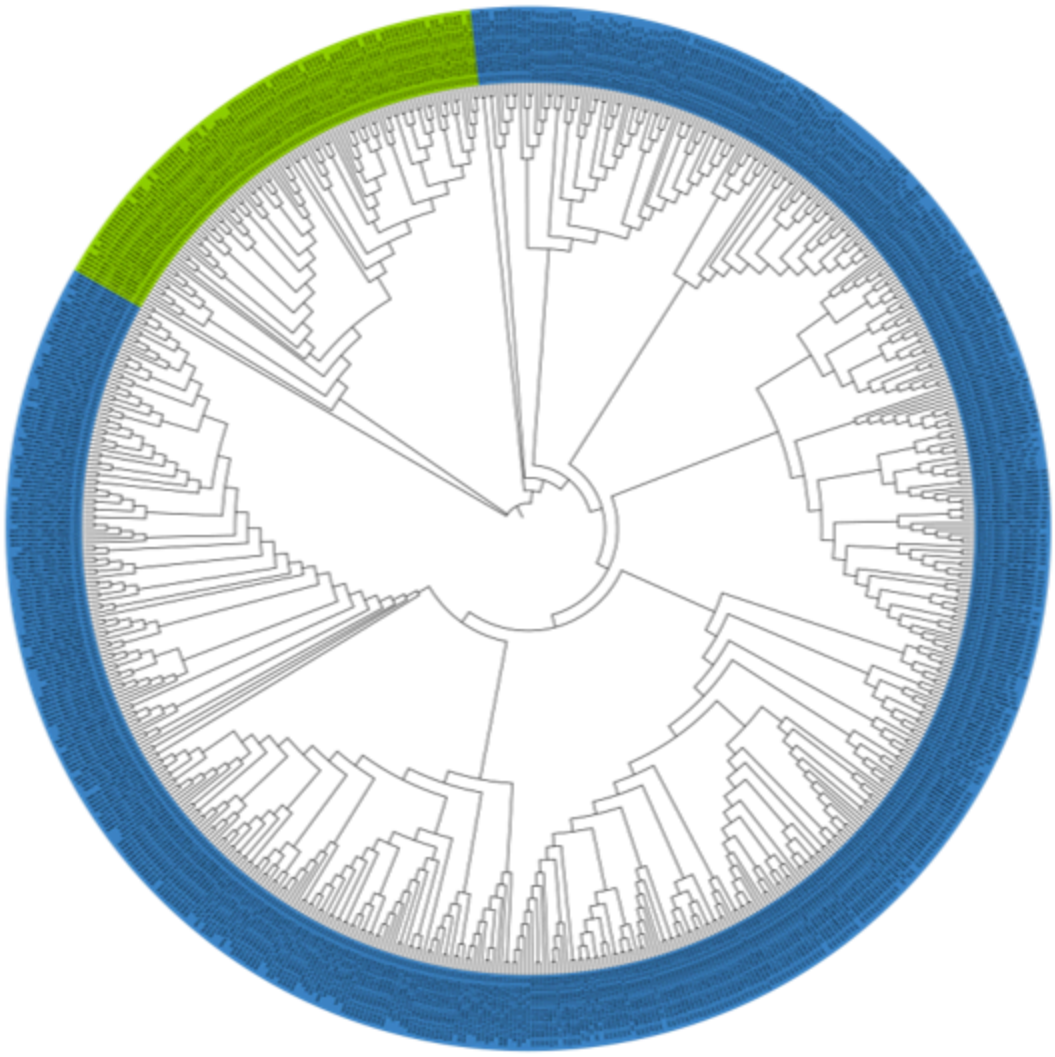


Figure 1

The phylogenetic tree of functional OR genes in the giant panda. The blue branch was Class II, and the green branch was Class I.

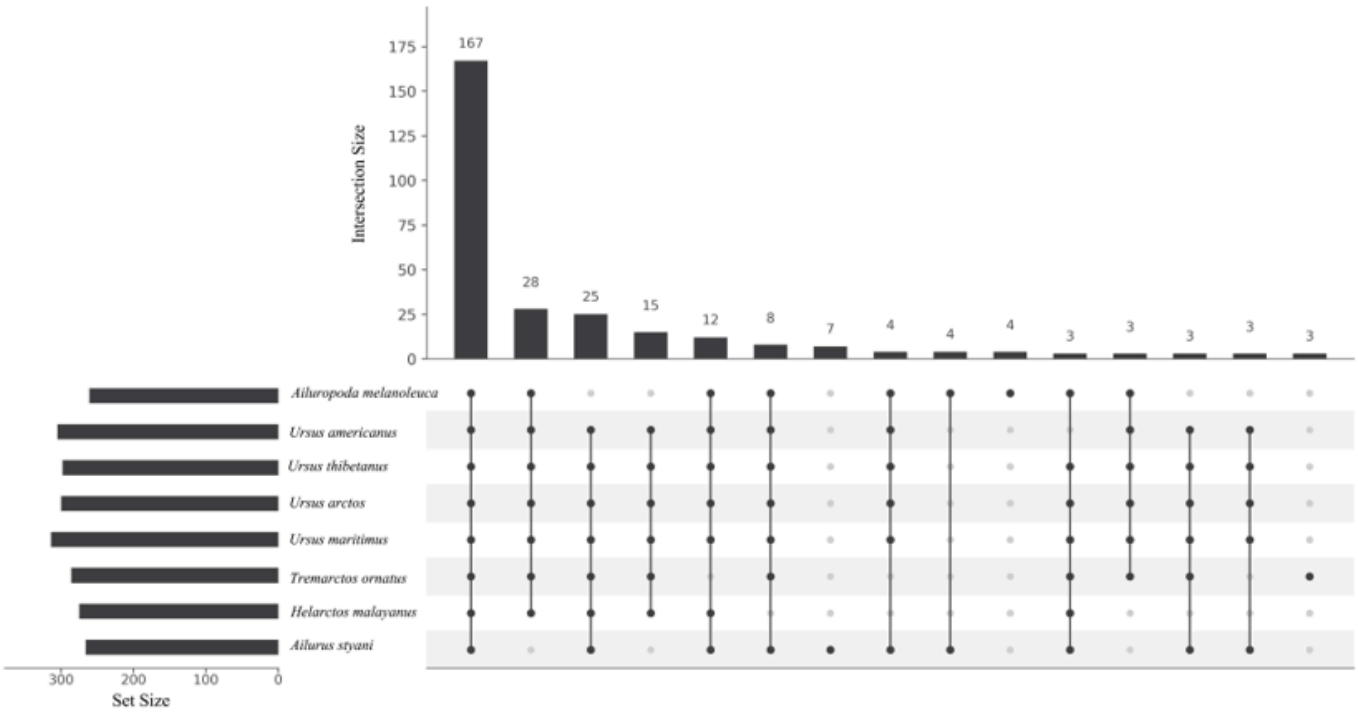


Figure 2

Comparison of OR gene clusters between eight species.

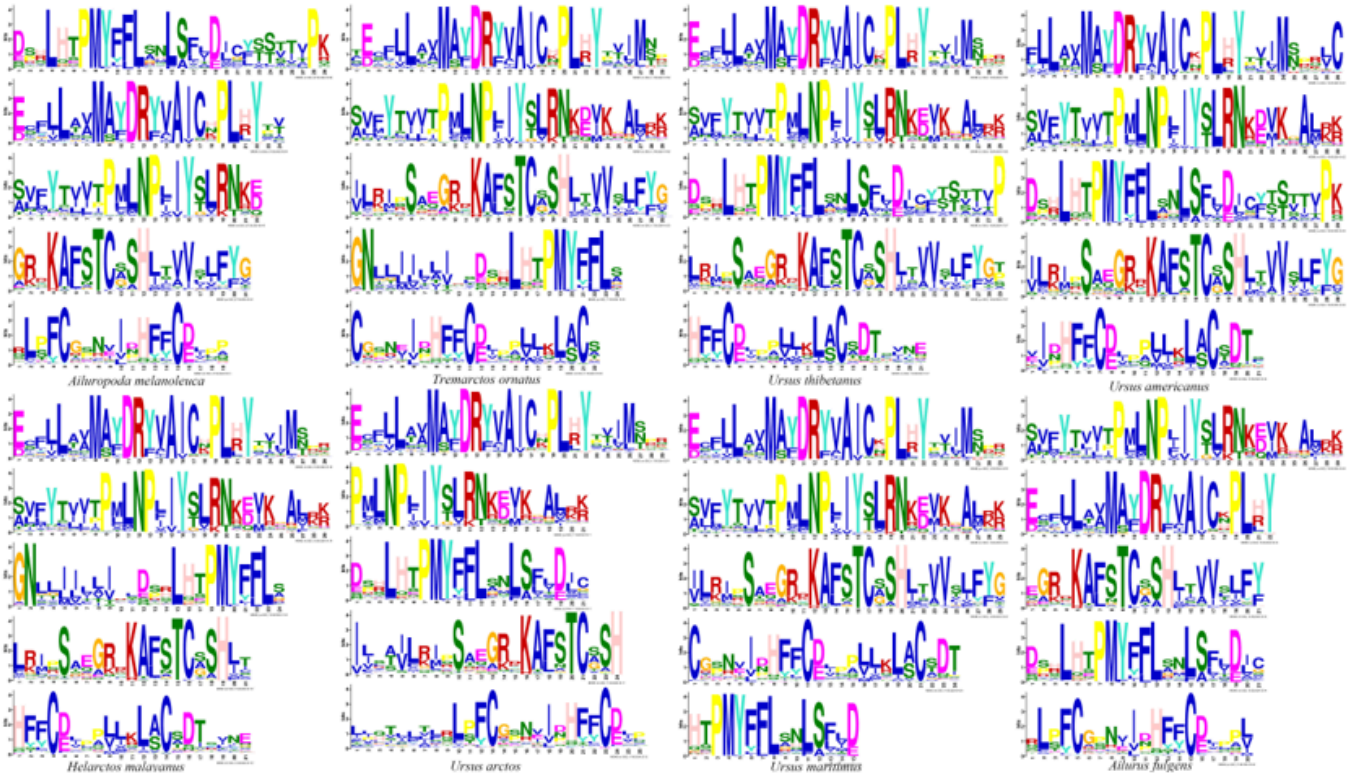


Figure 3

The five most conserved motifs of functional OR genes in eight species.

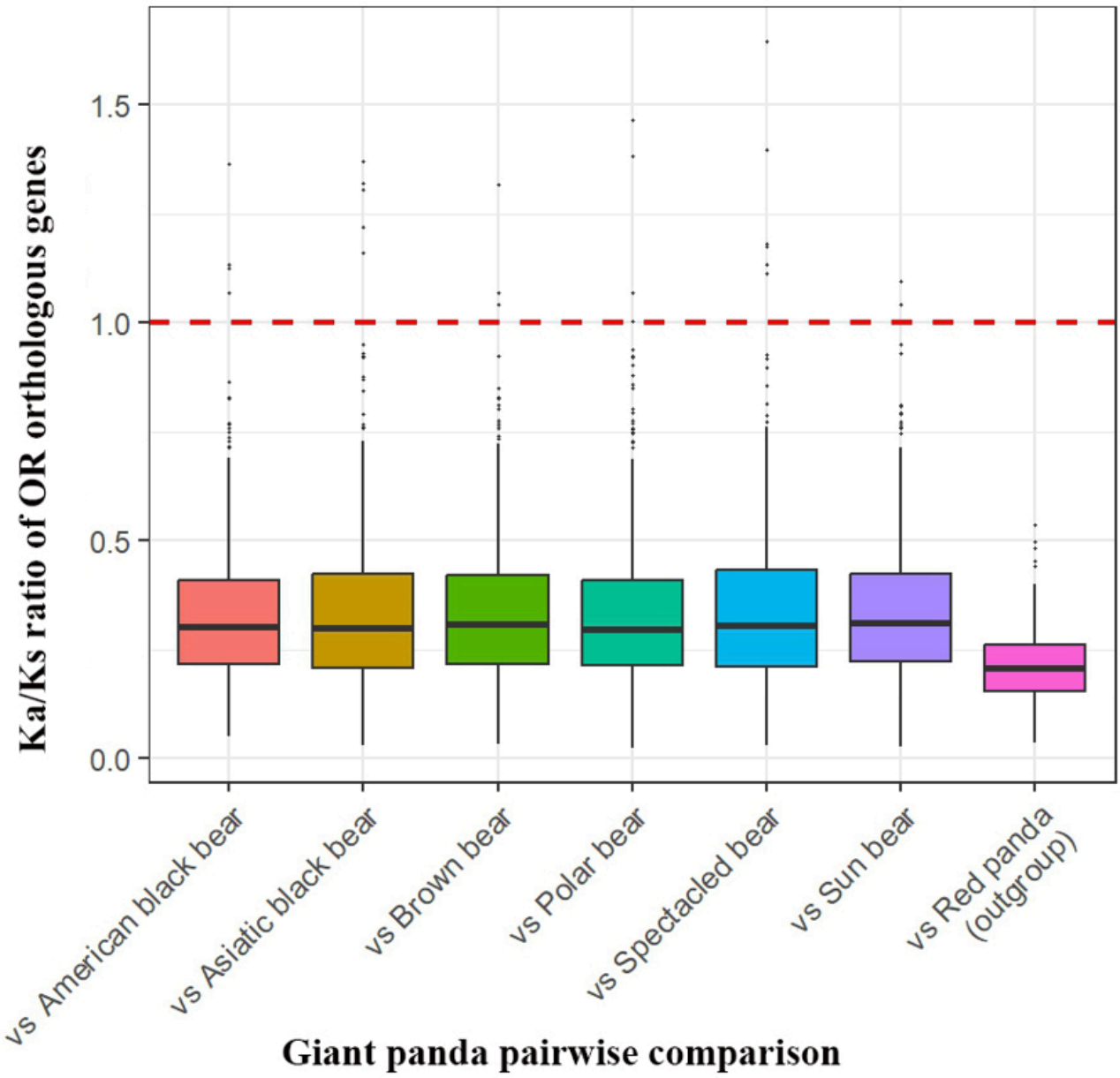


Figure 4

Pairwise Ka/Ks ratio distributions of one-to-one orthologous OR genes between giant panda and other carnivorans.

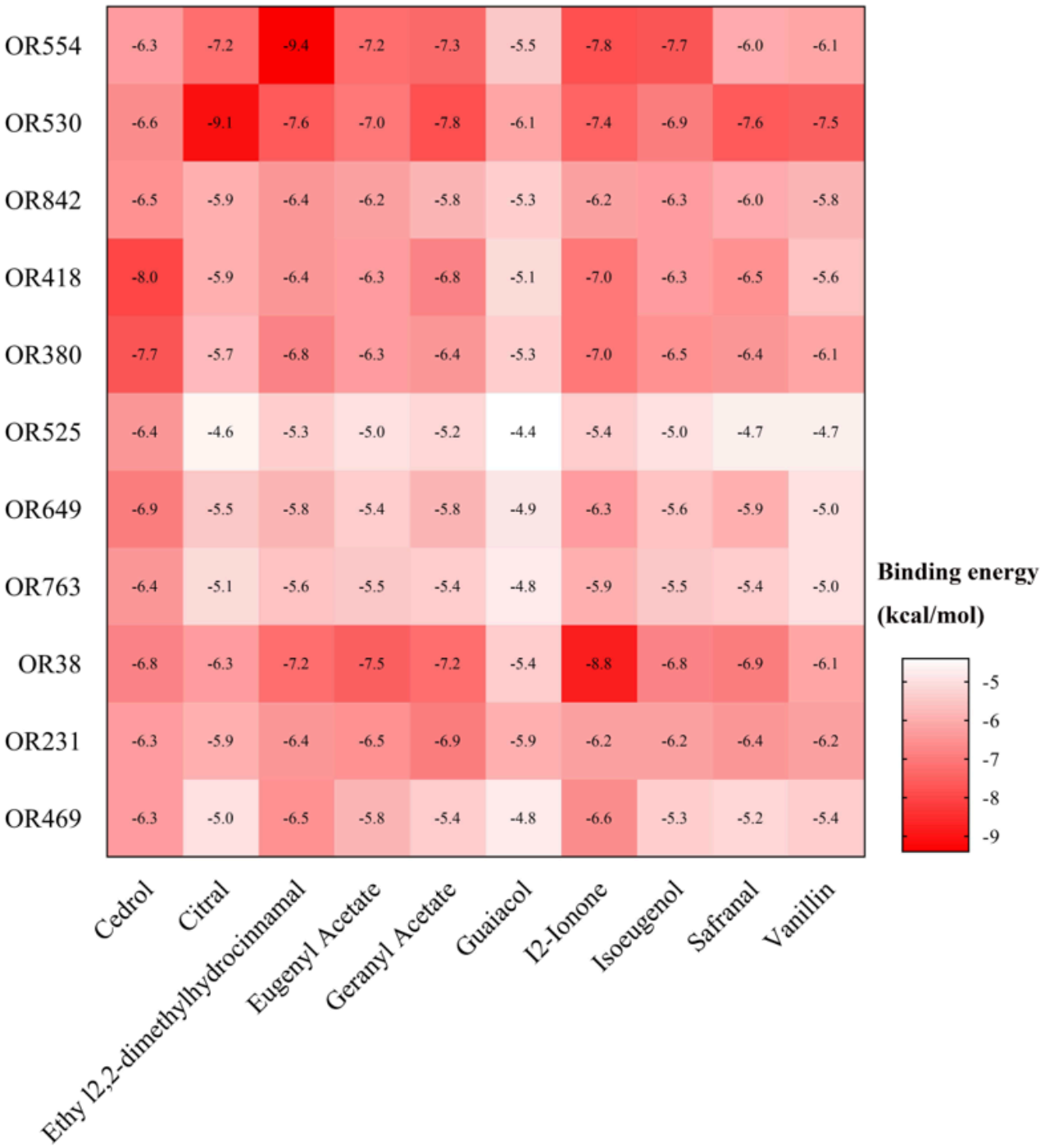


Figure 5

Heatmap of Predicted Binding Energies between OR Proteins and Plant Volatiles. Each cell represents the predicted binding energy (kcal/mol) for the interaction between an OR receptor (rows) and a volatile compound (columns). More negative values (darker red colour) correspond to stronger predicted receptor-ligand binding affinity.

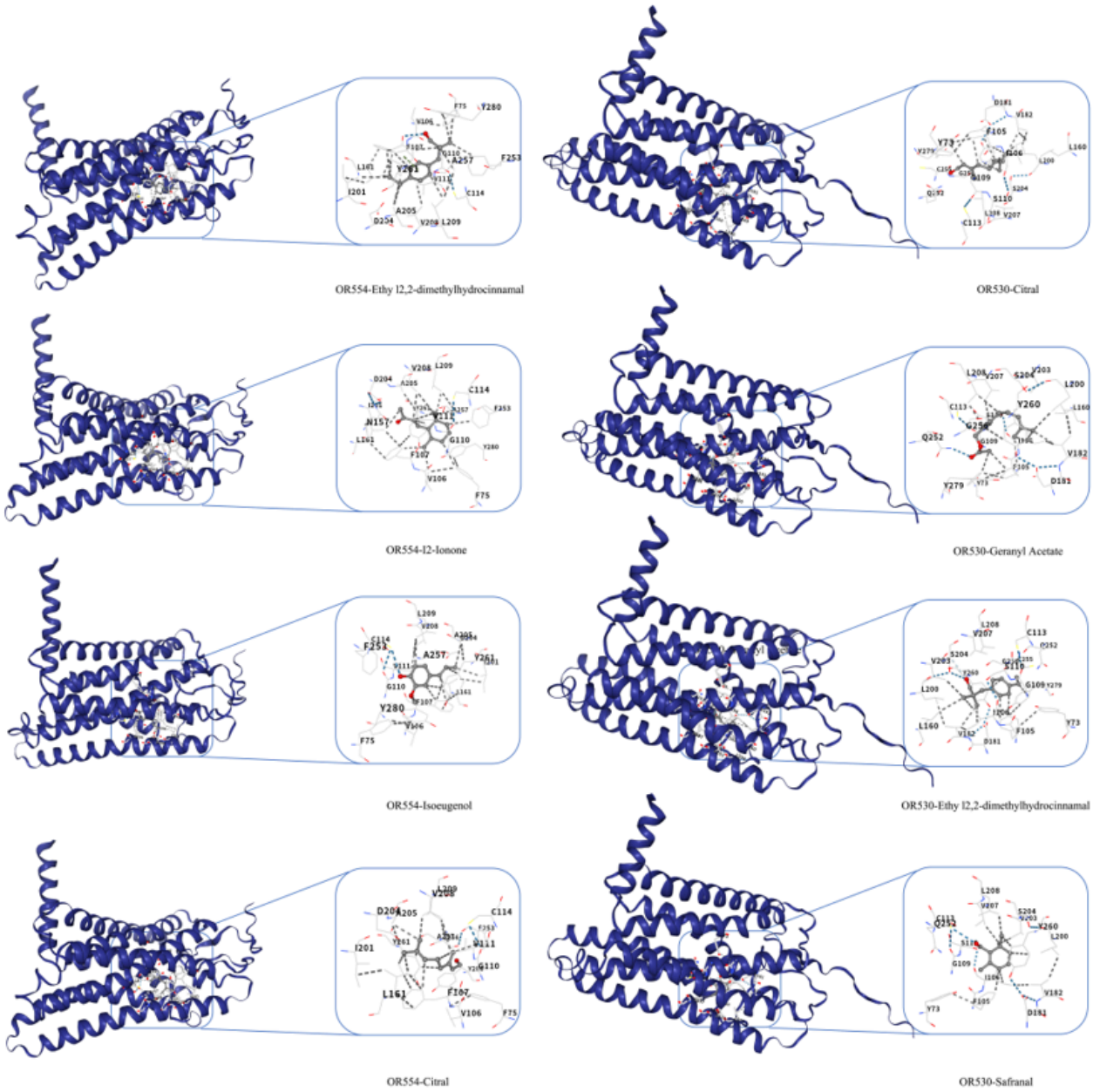


Figure 6

Predicted Binding Modes of OR554 and OR530 with Bamboo-Derived Volatiles

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable1.xlsx
- SupplementalTableS2.Olfactoryreceptorsequence.txt