BMC Genomics

RESEARCH Open Access

Lineage-specific expansion and positive selection of *NPC2* genes in the crab spider *Ebrechtella tricuspidata* and other arachnids

Tianlei Liu[1,2], Chunlei Cong[2], Jianxin Chen[2] and Daochao Jin[1*]

Abstract

Background Niemann-Pick type C2 (NPC2) proteins have been proposed as soluble carriers for semiochemicals in arthropods, serving as putative functional substitutes for the canonical odorant-binding proteins (OBPs) largely absent in arachnids. However, the evolutionary mechanisms driving NPC2 diversification in spiders—particularly in ambush predators relying heavily on chemical cues—remain unclear.

Results We identified a minimum expressed repertoire of 13 high-confidence *NPC2* genes in the crab spider *Ebrechtella tricuspidata* via comprehensive transcriptomic profiling. Comparative phylogenomics across 21 arthropod species revealed that while the generalist herbivore mite *Tetranychus urticae* underwent a significant repertoire expansion (+9 genes, $P < 0.001$), the *E. tricuspidata* lineage exhibited a modest numerical gain (+3 genes). Despite this, transcriptomic data revealed a striking functional dichotomy within the spider repertoire. The basal singleton *EtriNPC2-2* was highly enriched in the abdomen, suggesting a conserved role in visceral lipid transport. In contrast, the derived *Etri-B* subclade (*EtriNPC2-1*, *-3*) and the singleton *EtriNPC2-4* were predominantly expressed in chemosensory appendages (palps and legs). This spatial expression shift coincided with strong signatures of episodic positive selection acting on the *Etri-B* lineage. Structural modeling identified a key adaptive substitution (L33→Q31) at the pocket entrance of EtriNPC2-3, serving as a prime candidate for future functional validation regarding ligand access in the sensillar lymph.

Conclusions Our findings suggest a clear pattern of neofunctionalization, where gene duplication enabled spider *NPC2s* to diverge from an ancestral visceral function toward a specialized sensory role. This linkage between spatial expression shifts and adaptive sequence evolution provides essential molecular evidence for how ambush spiders evolved a complex chemosensory system in the absence of insect-type OBPs.

Keywords Niemann-Pick Type C2, *Ebrechtella tricuspidata*, Gene family expansion, Adaptive evolution, Positive selection, Chemosensation

*Correspondence:
Daochao Jin
dcjin@gzu.edu.cn
[1]Institute of Entomology, The Guizhou Provincial Key Laboratory for Plant Pest Management of Mountainous Region, The Scientific Observing and Experimental Station of Crop Pest in Guiyang, Guizhou University, Ministry of Agriculture, No. 2708 Huaxi Avenue, Guiyang 550025, China
[2]College of Agriculture, Anshun University, Anshun 561000, China

Introduction

Chemosensation plays a pivotal role in arthropod ecology, enabling precise detection of semiochemicals that guide foraging, mating, oviposition, and predator avoidance. In the peripheral olfactory pathway, hydrophobic odorants and pheromones must be solubilized and transported across the aqueous sensillar lymph to reach membrane-bound receptors [1, 2]. While odorant-binding proteins (OBPs) and chemosensory proteins (CSPs) dominate this carrier function in insects [3, 4], chelicerates present a markedly different molecular landscape. Classical insect-type OBPs are largely absent or highly divergent in spiders, mites, and ticks [5]. Consequently, these lineages must rely on alternative protein families to facilitate semiochemical transport.

Niemann-Pick type C2 (NPC2) proteins have emerged as strong candidates for this alternative carrier role. Initially described for intracellular cholesterol trafficking in vertebrates, NPC2 proteins are characterized by a conserved MD-2-related lipid-recognition (ML) domain, providing structural features highly suitable for binding diverse hydrophobic ligands. Recent genomic, transcriptomic, and functional studies have increasingly linked expanded NPC2 repertoires to chemosensory adaptation across arachnids. In predatory mites (*Neoseiulus barkeri* and *N. californicus*) and spider mites (*Tetranychus cinnabarinus*), *NPC2* genes are essential for olfactory-mediated behaviors and are rapidly upregulated by herbivore-induced plant volatiles (HIPVs) such as methyl salicylate (MeSA) and β-ionone. Recombinant mite NPC2 proteins demonstrate high-affinity binding to these volatiles (K_i values in the low micromolar range), and RNAi knockdown significantly impairs prey-location behavior [4, 6, 7]. Similar functional convergence occurs in ticks (*Rhipicephalus linnaei*, *Haemaphysalis longicornis*, and *Ixodes ricinus*): *NPC2* transcripts are highly enriched in the foreleg Haller's organ, localize to chemosensory sensilla, and readily bind volatile organic compounds and host-derived cues like ammonium hydroxide in vitro [8]. Together, these findings indicate that *NPC2* diversification makes up for the scarcity of canonical OBPs, facilitating effective chemoreception in lineages that heavily rely on lipid-based signals.

Recent comparative genomic work further highlights the evolutionary dynamism of the *NPC2* family. In the spider mite *Tetranychus cinnabarinus*, *NPC2* genes show significant expansion and respond to plant volatiles such as geranylacetone and farnesol, with functional binding validated by microscale thermophoresis and behavioral disruption upon RNAi [9]. Parallel expansions occur in other chelicerates, correlating with ecological specialization: generalist herbivores exhibit large repertoires to handle diverse plant volatiles, while ambush predators may favor refined ligand specificity for sparse prey-derived cues [10, 11]. Structural studies reinforce this functional shift. AlphaFold2 modeling of NPC2 proteins from predatory mites reveals a flexible β-barrel scaffold with a hydrophobic pocket that accommodates diverse HIPVs, while site-directed mutagenesis of key residues (e.g., in ligand-binding loops) abolishes binding affinity [4]. These advances indicate that NPC2 proteins not only solubilize ligands but may also modulate receptor activation or ligand release, extending beyond passive transport.

Despite this progress, critical gaps remain in our understanding of *NPC2* evolution and function in spiders. Most functional data derive from mites and ticks, where NPC2s have been linked to HIPV detection and host-seeking. In spiders, particularly ambush predators, the selective pressures shaping NPC2 repertoires are less clear. Crab spiders, such as *Ebrechtella tricuspidata* (Thomisidae), employ a sit-and-wait strategy on flowers, relying on precise discrimination of floral volatiles for camouflage and prey-derived cuticular hydrocarbons for prey assessment [12, 13]. This niche likely favors an expanded or refined NPC2 toolkit to detect low-abundance lipid signals amid complex floral backgrounds. However, systematic characterization of the *NPC2* repertoire, its phylogenetic placement, and signatures of adaptive evolution in such species has been limited.

In this study, we integrate hybrid long-read transcriptomics, systematic sequence curation, and molecular evolutionary analyses to characterize the *NPC2* gene family in the ambush predator *E. tricuspidata* and across representative arachnid lineages. Specifically, we aim to: (1) establish a high-confidence *NPC2* dataset using full-length Iso-Seq transcripts and unified quality control; (2) reconstruct the evolutionary history of the family and quantify lineage-specific changes in gene copy number through birth–death modeling; and (3) detect signatures of positive selection using branch-site and episodic selection models, and map selected sites onto predicted protein structures. By placing these findings in a comparative evolutionary framework, this work provides insights into how gene duplication, sequence divergence, and structural variation may contribute to chemosensory adaptation in Chelicerata.

Methods

Specimen collection and sampling

Laboratory colonies of *E. tricuspidata* were established from individuals collected in Anshun, Guizhou Province, China (26°15′N, 105°55′E). Spiders were reared under controlled conditions (25 ± 1 °C, 75 ± 5% RH, 16 L:8D photoperiod) and fed with *Drosophila melanogaster* ad libitum.

To construct a comprehensive transcriptomic atlas covering chemosensory tissues across different life stages, a

total of 53 RNA-seq libraries were sequenced. The sampling strategy was strictly stratified by developmental stage, sex, and tissue specificity. Adult (male and female) and sub-adult spiders were carefully dissected under a stereomicroscope into five distinct anatomical regions: (i) forelegs (legs I and II), (ii) hindlegs (legs III and IV), (iii) palps, (iv) abdomen, and (v) the remaining cephalothorax. Palps and forelegs were specifically targeted as they serve as the primary chemosensory appendages in spiders. For early developmental stages (juveniles), forelegs were specifically collected. Each specific biological replicate consisted of a pool of multiple individuals (e.g., up to 50 individuals per tissue sample) to ensure sufficient RNA yield and biological representativeness. The exact experimental metadata for all 53 libraries, including specific replicate counts per condition, are detailed in Supplementary Table S1.

Total RNA was extracted using the mirVana™ miRNA Isolation Kit (Ambion, Austin, TX, USA), following the manufacturer's protocol. RNA integrity was evaluated using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Only samples with an RNA Integrity Number (RIN) ≥ 7.0 were subjected to subsequent library construction.

Hybrid sequencing and de novo transcriptome assembly

Illumina sequencing and library construction: Library construction was performed using the VAHTS Universal V6 RNA-seq Library Prep Kit following the manufacturer's instructions. Briefly, mRNA was enriched using Oligo (dT) beads and fragmented. Double-stranded cDNA was synthesized, purified using Agencourt AMPure XP beads (Beckman Coulter), end-repaired, and adenylated. After adapter ligation and fragment size selection, libraries were amplified by PCR. Qualified libraries were sequenced on an Illumina NovaSeq 6000 platform (OE Biotech, Shanghai, China) to generate 150 bp paired-end reads.

PacBio SMRT sequencing: A pooled RNA sample (3 μg), combining equal amounts of total RNA from all 53 samples, was used for full-length transcriptome sequencing. A SMRTbell library was constructed using the SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences, Menlo Park, CA, USA) and sequenced on a PacBio Sequel II platform in Iso-Seq mode.

Raw Illumina reads were trimmed using Trimmomatic v0.39 [14] to remove adapters and low-quality bases (parameters: LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36). PacBio subreads were processed with SMRT Link software [15] to generate high-accuracy Circular Consensus Sequences (CCS). Full-length transcripts were corrected using the high-coverage Illumina reads and clustered at 95% identity using CD-HIT-EST v4.8.1 [16] to remove redundancy. Assembly completeness was assessed using BUSCO v5 [17]. Raw sequencing data are available in the NCBI SRA under BioProject PRJNA1301248.

Identification, re-annotation, and curation of NPC2 genes

Genomic assemblies, GFF annotations, and predicted proteomes for 18 arachnid species and two insect outgroups (*Cimex lectularius* and *Stomoxys calcitrans*) were retrieved from the NCBI GenBank database [18] (accessed October 2025; Supplementary Table S2). Including *E. tricuspidata*, a total of 21 species were analyzed. Assembly completeness was evaluated with BUSCO v5 lineage-specific reference datasets: *arachnida_odb10* for the 19 arachnid species (18 genome-based plus the *E. tricuspidata* transcriptome), *hemiptera_odb10* for *C. lectularius*, and *diptera_odb10* for *S. calcitrans*.

NPC2 homologs were initially identified using a combined approach: (1) HMMER v3.3 searches [19] against predicted proteomes using the MD-2-related lipid-recognition domain profile (PF02221) with an *E-value* < 10^{-5} [20], and (2) BLASTp searches [21] using 16 preliminary *E. tricuspidata NPC2* sequences as queries (*E-value* < 10 − 5). This initial screening yielded a total of 173 candidate sequences across the 21 species.

For *E. tricuspidata*, targeted re-annotation was performed to address potential N-terminal truncation artifacts prior to final filtering. Preliminary nucleotide sequences were mapped to the full-length Iso-Seq transcriptome and Unigene assemblies using exact sequence matching or containment matching. Candidate ORFs were refined from existing protein predictions and supplemented by six-frame translation of mapped transcripts, retaining only fragments ≥ 80 amino acids defined from the first methionine to the first stop codon.

To establish a high-confidence dataset, the 173 candidate ORFs were subjected to a rigorous multi-stage curation pipeline. First, distinct redundancy checks were applied to non-focal species (e.g., *Panonychus citri*, *Rhipicephalus microplus*), resulting in the removal of six sequences identified as obsolete or redundant based on updated annotation comparisons (retaining 167 sequences). Second, manual inspection of the *E. tricuspidata* candidates led to the exclusion of three sequences (*EtriNPC2-5*, *-12*, *and* – *16*) validated as short fragments lacking essential structural motifs (retaining 164 sequences). Finally, the remaining sequences underwent functional feature triangulation to verify biological plausibility using three independent predictors: SignalP 5.0 [22] for secretion signals, DeepTMHMM for transmembrane topology, and DeepLoc 2.0 [23] for subcellular localization. To be retained, candidates were required to exhibit consistent *NPC2* characteristics, specifically the presence of signal peptides, extracellular localization,

and a conserved cysteine pattern (≥ 4 residues). A strict domain integrity filter was further applied using HMMER (hmmscan against Pfam-A), which excluded nine additional sequences that failed to cover at least 70% of the MD-2-related lipid-recognition domain (PF02221). To recover full-length open reading frames (ORFs) and correct potential truncation artifacts, we performed six-frame stop-to-stop scanning of all full-length transcripts using a custom Python script. ORFs were defined from the first in-frame methionine to the first in-frame stop codon, and only those encoding ≥ 80 amino acids were retained. This process yielded a final core repertoire of 155 high-confidence sequences used for subsequent evolutionary analyses.

Phylogenetic analysis

Protein sequences were aligned with MAFFT v7.505 (L-INS-i algorithm) [24]. Corresponding coding sequences were back-translated to the protein alignment using PAL2NAL [25] to produce a codon-aligned dataset. Ambiguously aligned regions were trimmed with trimAl v1.5 [26] using the *automated1* heuristic method.

Phylogenetic trees were reconstructed with IQ-TREE 2 v2.1.4-beta [27] on the codon-partitioned CDS alignment (partitions by codon position). The best-fit substitution model for each partition was selected by ModelFinder [28] implemented in IQ-TREE 2. Branch support was assessed with 1,000 ultrafast bootstrap replicates (UFBoot2) [29] and the SH-aLRT test [30]. The codon-based tree was used as the primary phylogeny for downstream analyses.

Gene family expansion and contraction

Gene family dynamics of the *NPC2* family were analyzed with CAFE 5 v5.1.0 [31]. Copy numbers per species were provided as an integer count table for the single family. The species tree was rooted, made binary, and converted to an ultrametric tree (branch lengths in time units) using r8s v1.81 [32]. Divergence times were calibrated using secondary calibration points retrieved from TimeTree 5 [33] and previous arachnid phylogenomic studies.

CAFE 5 was run under the Base model with a single global birth–death rate (λ). Because maximum-likelihood estimation of λ lacks statistical power and is often unstable when applied to a single gene family, the global λ was fixed at 0.005. This value was selected based on empirically derived average turnover rates reported for rapidly evolving chemosensory gene families in arthropods [10, 34]. To ensure that our conclusions were not artifacts of this specific parameter choice, we performed a parameter sensitivity analysis by running CAFE 5 across a biologically plausible range of λ values (0.001, 0.003, 0.005, and 0.01). The identification of key significantly expanded or contracted branches (e.g., the expansion in *T. urticae*) remained highly consistent across these values. Final significant expansions and contractions ($P < 0.05$) under the representative λ = 0.005 were identified from branch-specific likelihood tests and copy-number change estimates (Δ = child − parent).

Selection pressure analysis

Codon alignments were analyzed with PAML v4.10.9 [35] and HyPhy v2.5 [36]. Gaps and ambiguous sites were removed (cleandata = 1). The F3 × 4 codon frequency model (CodonFreq = 2) was used in all runs.

PAML (codeml): Three nested models were compared: (1) one-ratio model (M0) estimating a single ω across the tree; (2) free-ratio model allowing independent ω per branch; (3) branch-site Model A on foreground branches. For each foreground branch, the null model fix $\omega = 1, \omega_2 = 1$ was compared against the alternative model fix $\omega = 1$,?$_2$estimated). Statistical significance was assessed using the likelihood-ratio test (LRT): $2?? = 2(?_1?? _0)$, where $?_1$ and $?_0$ are the log-likelihoods of the alternative and null models, respectively. The test statistic was assumed to follow a χ^2 distribution with 1 degree of freedom ($df = 1$). Positively selected sites were identified using Bayes Empirical Bayes (BEB) analysis with a posterior probability (*PP*) threshold of ≥ 0.95. Site indices were manually re-mapped to the original protein positions to account for the deletion of gaps during the *cleandata* = 1 filtering step.

HyPhy BUSTED [37] for gene-wide selection, aBSREL [38] for branch-specific selection, and MEME [39] for site-specific selection were performed. Sites detected by both PAML (BEB) and HyPhy (MEME) were considered high-confidence candidates of positive selection.

Structural modeling

Three-dimensional structures of representative mature NPC2 proteins (signal peptide removed) were predicted with AlphaFold 3 [40]. Models were visualized and positively selected sites were mapped using PyMOL (The PyMOL Molecular Graphics System, Version 3.1.4.1) [41].

Expression profiling and statistical analysis

To quantify tissue-specific expression, RNA-seq reads from 53 samples were mapped to the *E. tricuspidata* reference transcripts using Salmon v1.10. Gene expression levels were calculated as Fragments Per Kilobase of transcript per Million mapped reads (FPKM). To eliminate the statistical bias caused by variations in total normalized reads across different sequencing libraries and to enable robust cross-sample comparisons, FPKM values were subsequently converted to Transcripts Per Million (TPM). Specifically, the TPM for each gene was

calculated by dividing its FPKM value by the sum of all FPKM values within the same sample, and then multiplying by 10^6. This mathematical transformation ensures that the sum of TPM values in every sample is exactly one million, standardizing the proportional relationships and mitigating inter-library summation disparities. A total of 13 high-confidence *NPC2* genes were retained for analysis after excluding fragmentary sequences (*EtriNPC2-5*, *-12*, *-16*).

For statistical analysis, TPM values were log2 (TPM + 1) transformed to meet the assumptions of normality and homoscedasticity. Differences in expression levels across adult tissues (palps, legs I + II, legs III + IV, abdomen, and the remaining cephalothorax) were analyzed using a one-way analysis of variance (ANOVA) followed by Tukey's Honest Significant Difference (HSD) post-hoc test. All statistical analyses were performed in R v4.3.0 using the agricolae package [42]. Principal Component Analysis (PCA) was conducted on the transformed expression matrix using the factoextra package [43] to assess sample clustering and biological reproducibility.

Results

Identification and structural characterisation of NPC2 genes

Assembly completeness of the 20 genome-derived proteomes and the *E. tricuspidata* transcriptome was assessed with BUSCO v5 using lineage-appropriate reference datasets (*arachnida_odb10* for 19 arachnid species, *hemiptera_odb10* for *C. lectularius*, and *diptera_odb10* for *S. calcitrans*). For the 20 genome-based assemblies, complete BUSCO scores (C[S] + C[D]) exceeded 90% in most species (Supplementary Figure S1), indicating high assembly quality. For *E. tricuspidata*, BUSCO evaluation yielded a completeness of 59.0%. For a transcriptome-based dataset, which captures only expressed genes and is therefore inherently less complete than a genome assembly, this moderate completeness is expected. Importantly, the recovery of *NPC2* candidates in *E. tricuspidata* did not rely solely on BUSCO-assessed assembly completeness; all 16 initial candidates were independently validated against full-length Iso-Seq transcripts, ensuring that the *NPC2* repertoire was not underestimated due to transcriptome fragmentation.

A total of 173 *NPC2* candidate sequences were initially identified across the 21 species using HMMER (PF02221) and BLASTp searches. After the removal of six redundant or obsolete predictions from non-focal species, 167 candidates remained. These candidates were then subjected to the stepwise quality-control pipeline. A total of 12 sequences were flagged and excluded during this process: three were fragmentary *E. tricuspidata* candidates (*EtriNPC2-5*, *-12*, and *– 16*), and nine from other species failed to meet the structural thresholds (domain coverage ≥ 70% or conserved cysteines ≥ 4). After re-annotation using six-frame ORF scanning, all 13 retained *E. tricuspidata NPC2* genes were confirmed to encode full-length proteins with predicted signal peptides. The exclusion of these 12 flagged sequences yielded a final core dataset of 155 sequences. Within this high-confidence dataset, 144 sequences (92.9%) were classified as intact proteins with predicted signal peptides (qc_pass_full_sp), while 11 sequences (7.1%) were full-length but lacked a detectable signal peptide (qc_full_no_sp) (Supplementary Table S3; Supplementary Figure S2 and S3). The full list of identified sequences, including their physicochemical properties and signal peptide predictions, is available in Supplementary Table S4.

Variation in NPC2 gene copy number

NPC2 high-confidence core dataset size (155 sequences across 21 species) varied substantially, ranging from 2 genes in *Dermatophagoides pteronyssinus* to 20 in *Tetranychus urticae* (Fig. 1; Supplementary Table S3).

The Acari exhibited the greatest variation among arachnid orders. *T. urticae* possessed the largest repertoire (20 core genes), whereas the house dust mites *D. farinae* and *D. pteronyssinus* had only 3 and 2 core genes, respectively. Ticks showed intermediate to high counts: *Ixodes scapularis* (11), *Dermacentor andersoni* (9), *D. silvarum* (8), *Rhipicephalus sanguineus* (7), and *R. microplus* (6). Within the Araneae, copy numbers ranged from 6 (*Uloborus diversus*) to 13 (*E. tricuspidata*), with *Parasteatoda tepidariorum* (11), *Stegodyphus dumicola* (9), and *Argiope bruennichi* (7) at intermediate levels. The two insect outgroups, *Cimex lectularius* and *Stomoxys calcitrans*, contained 4 and 5 core genes, respectively. The scorpion *Centruroides sculpturatus* had 3 core genes.

Phylogenetic relationships of NPC2 genes

The phylogeny reveals that expansion of the *NPC2* repertoire in arachnids has largely occurred through multiple independent duplication events rather than single massive radiations. This pattern is evident in the two species with the largest repertoires.

In *E. tricuspidata* (13 genes), the sequences are distributed across three subclades and two singletons (Fig. 2). The *Etri-A* subclade comprises five members (*EtriNPC2-6*, *-7*, *-8*, *-9*, *and – 14*) and forms a well-supported monophyletic group exclusive to *E. tricuspidata* (MRCA UFBoot = 99; subtree = 5 tips, all *EtriNPC2*). The *Etri-B* subclade contains two members (*EtriNPC2-1* and *– 3*) and is likewise an *E. tricuspidata*-exclusive clade with maximum support (UFBoot = 100; subtree = 2 tips). The *Etri-E* subclade includes four members (*EtriNPC2-10*, *-11*, *-13*, and *– 15*) with moderate support (UFBoot = 65); unlike *Etri-A* and *Etri-B*, the *Etri-E* MRCA subtree encompasses 14 tips in total, of which

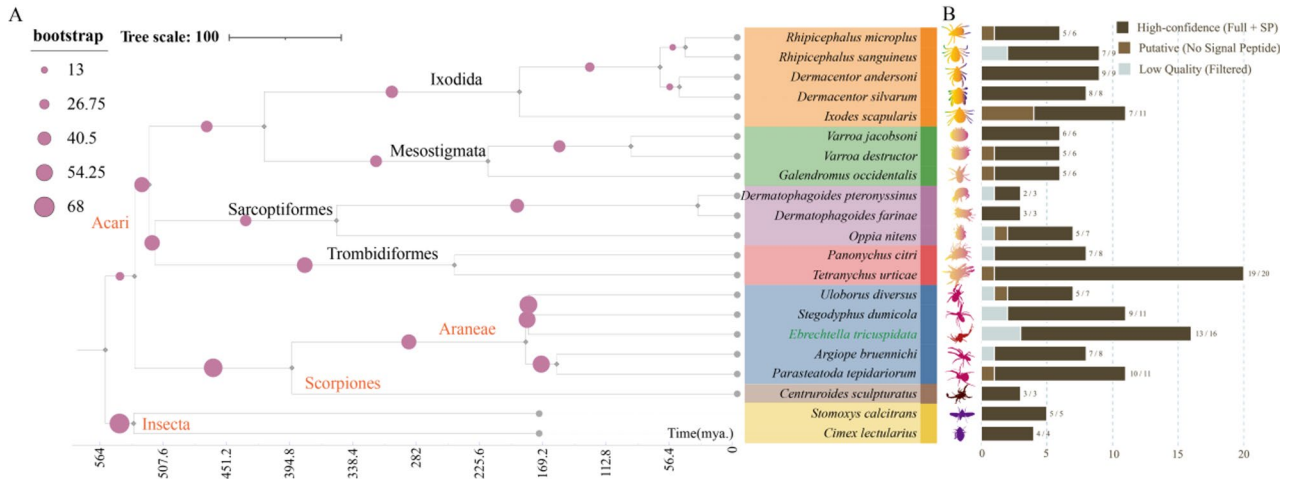

Fig. 1 Phylogenetic relationships and *NPC2* repertoire composition across 21 species. **A** Time-calibrated species tree with divergence times (Mya) derived from TimeTree 5. **B** Stacked horizontal bar charts showing the number of *NPC2* candidates per species, color-coded by quality-control outcome: high-confidence core dataset with predicted signal peptides (dark brown, retained for downstream analyses), full-length sequences lacking signal peptides (light brown), and fragmentary or low-cysteine sequences (gray, excluded). Numbers at bar ends indicate retained/total candidates

only 4 are *EtriNPC2* sequences, indicating that this subclade is nested within a larger mixed clade containing paralogs from other arachnid species. The remaining two sequences, *EtriNPC2-2* and *EtriNPC2-4*, are phylogenetically isolated singletons that do not group with any other *EtriNPC2* paralog. The dispersal of *E. tricuspidata* paralogs across five independent lineages indicates that the 13-gene repertoire was assembled through duplications originating from at least five distinct ancestral paralogs.

In *T. urticae* (20 genes), the mite sequences do not form a single monophyletic group. A subset of 8 sequences clusters together, while the remaining 12 are dispersed throughout the tree, intermingled with orthologs from other Acari species (Fig. 2).

Gene family expansion and contraction

Formal birth–death modeling with CAFE 5 identified three terminal branches with statistically significant copy-number changes ($P < 0.05$; Fig. 3; Supplementary Table S5). The strongest signal was a significant expansion on the *T. urticae* terminal branch, with a net gain of +9 genes relative to its reconstructed ancestral node (parent count = 11; $P = 1.84 \times 10^{-5}$). Two spider lineages showed significant contractions: *U. diversus* ($\Delta = -5$, parent count = 11; $P = 3.26 \times 10^{-3}$) and *A. bruennichi* ($\Delta = -3$, parent count = 10; $P = 0.048$). For *E. tricuspidata*,the analysis inferred a moderate net gain of +3 genes relative to its ancestral node (parent count = 10; $P = 0.086$). Crucially, because the *E. tricuspidata* dataset was derived from a transcriptome rather than a complete whole-genome assembly, the 13 identified sequences must be interpreted strictly as a minimum expressed repertoire. The CAFE 5 model fundamentally assumes complete genomic inputs; thus, any genes not expressed at the time of sampling are inherently mathematically penalized as evolutionary "losses." Therefore, identifying an expressed repertoire (13 genes) that already exceeds the inferred ancestral state (10 genes)—despite this strict transcriptomic penalty—serves as a highly conservative lower bound. This strongly implies an actual lineage-specific expansion trend in *E. tricuspidata*, even if the transcriptomic underestimation prevented the *P*-value from reaching formal significance. The remaining 17 terminal branches showed no significant departures from the background rate (all $P > 0.05$).

In *T. urticae*, 19 of 20 core genes (95.0%) possessed predicted signal peptides; in *E. tricuspidata*, all 13 core genes (100%) possessed predicted signal peptides. These high proportions indicate that the copy-number increases in both lineages predominantly involved functional secreted paralogs rather than pseudogenes or non-secreted duplicates. Overall, these results support asymmetric *NPC2* family evolution, with pronounced expansion in *T. urticae* and selective contractions in specific spider lineages.

Molecular evolution and positive selection

Selective pressure heterogeneity

Global BUSTED analysis provided strong evidence of episodic diversifying selection acting on at least one site in at least one branch across the arachnid *NPC2* phylogeny ($P = 3.95 \times 10^{-10}$).

To dissect lineage-specific selective pressures, Branch-site Model A (PAML) was applied to the three *E. tricuspidata*-specific subclades defined on the CDS tree (*Etri-A*, *Etri-B*, and *Etri-E*). For *Etri-A* and Etri-B, the MRCA branch was designated as the foreground; for *Etri-E*, which is nested within a mixed clade, individual tip branches were designated as the foreground. Pruned

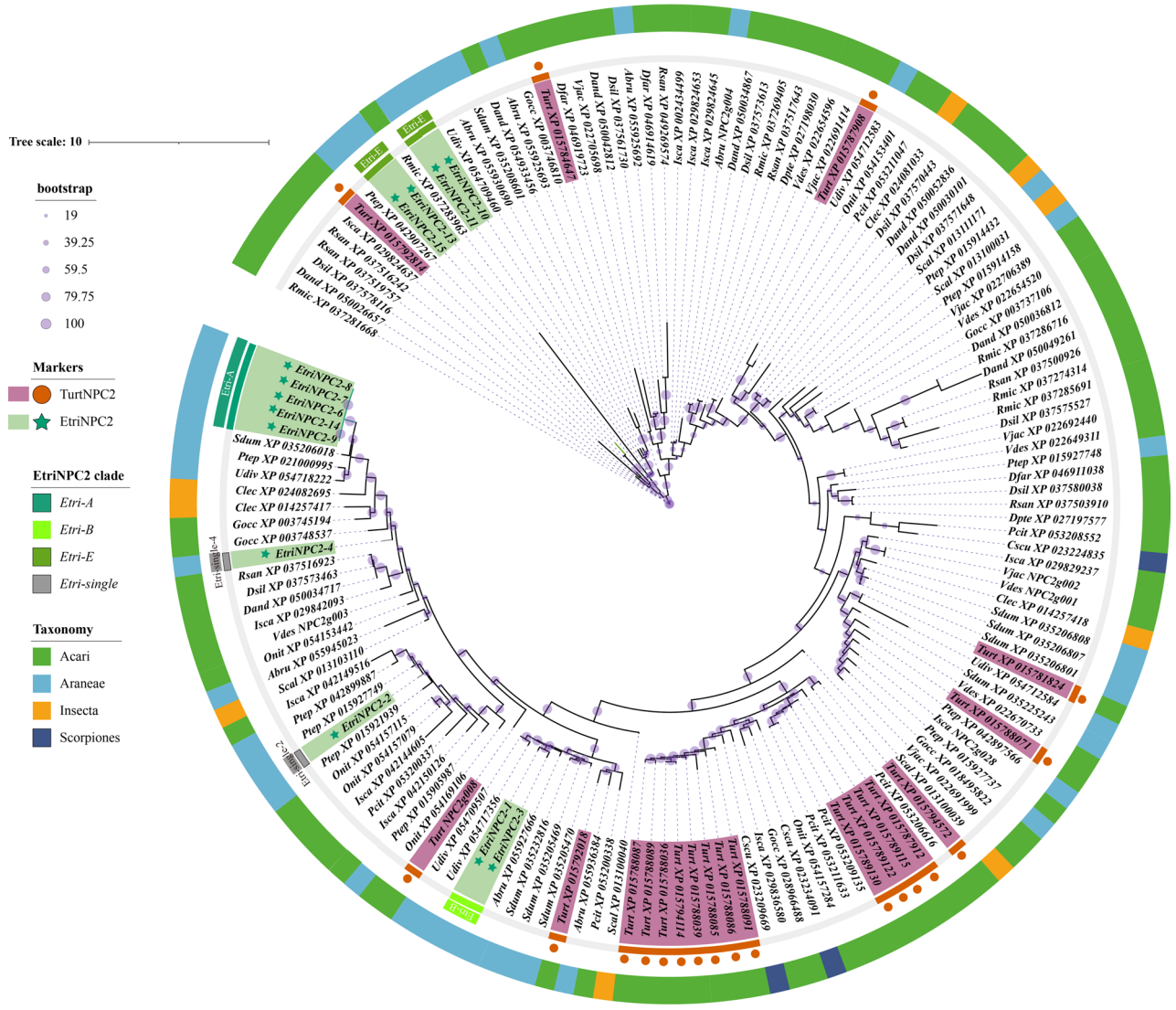

Fig. 2 Maximum-likelihood phylogeny of 155 high-confidence *NPC2* coding sequences. The circular tree was inferred from a codon-aligned CDS dataset (447 positions) using IQ-TREE 2 (codon-position partitioned, ModelFinder, 1,000 UFBoot + SH-aLRT replicates). Outer ring: taxonomy (Acari, blue; Araneae, red; Insecta, green; Scorpiones, purple). Inner color strips: *EtriNPC2* subclades (*Etri-A*, teal; *Etri-B*, orange; *Etri-E*, lime; singletons, gray). Green stars denote *EtriNPC2* sequences; orange circles denote *T. urticae* paralogs. Scale bar: 10 substitutions per site. Statistical support for internal branches was assessed using Ultrafast Bootstrap analysis. Purple circles at the internal nodes represent statistical support values, with the size of the circle proportional to the bootstrap percentage

subtree sizes ranged from 15 to 17 sequences, with 2–5 foreground branches per test (Supplementary Table S6).

The analyses uncovered significant heterogeneity in selective pressures across the three subclades, as indicated by Fig. 4 and detailed in Supplementary Table S6. The *Etri-B* subclade exhibited an exceptionally strong signal of episodic positive selection on its MRCA branch ($2\Delta\ell$ = 265.22, $P = 1.25 \times 10^{-59}$, BH $q = 3.75 \times 10^{-59}$; n~foreground ~ = 2), indicating intense adaptive pressure driving its divergence. The *Etri-A* subclade showed moderate but statistically significant evidence of selection ($2\Delta\ell$ = 11.35, $P = 7.54 \times 10^{-4}$, BH $q = 1.13 \times 10^{-3}$; n~foreground ~ = 5), whereas no support for positive selection was found in *Etri-E* ($2\Delta\ell$ = 0.0, $P = 1.0$, BH $q = 1.0$; n~foreground ~ = 4), suggesting that this lineage evolved under neutral drift or purifying selection. Consistent with these branch-site results, the free-ratio model (vs. M0) detected significant among-branch rate heterogeneity for the *Etri-A* ($2\Delta\ell$ = 51.89, $df = 30$, $P = 0.008$) and *Etri-B* ($2\Delta\ell$ = 50.38, $df = 26$, $P = 0.003$) subtrees, but not for *Etri-E* ($2\Delta\ell$ = 23.52, $df = 26$, $P = 0.604$).

Furthermore, we employed HyPhy aBSREL to evaluate selection at the individual branch level. aBSREL confirmed episodic selection on five individual branches within the *T. urticae* expansion (e.g., terminal branch

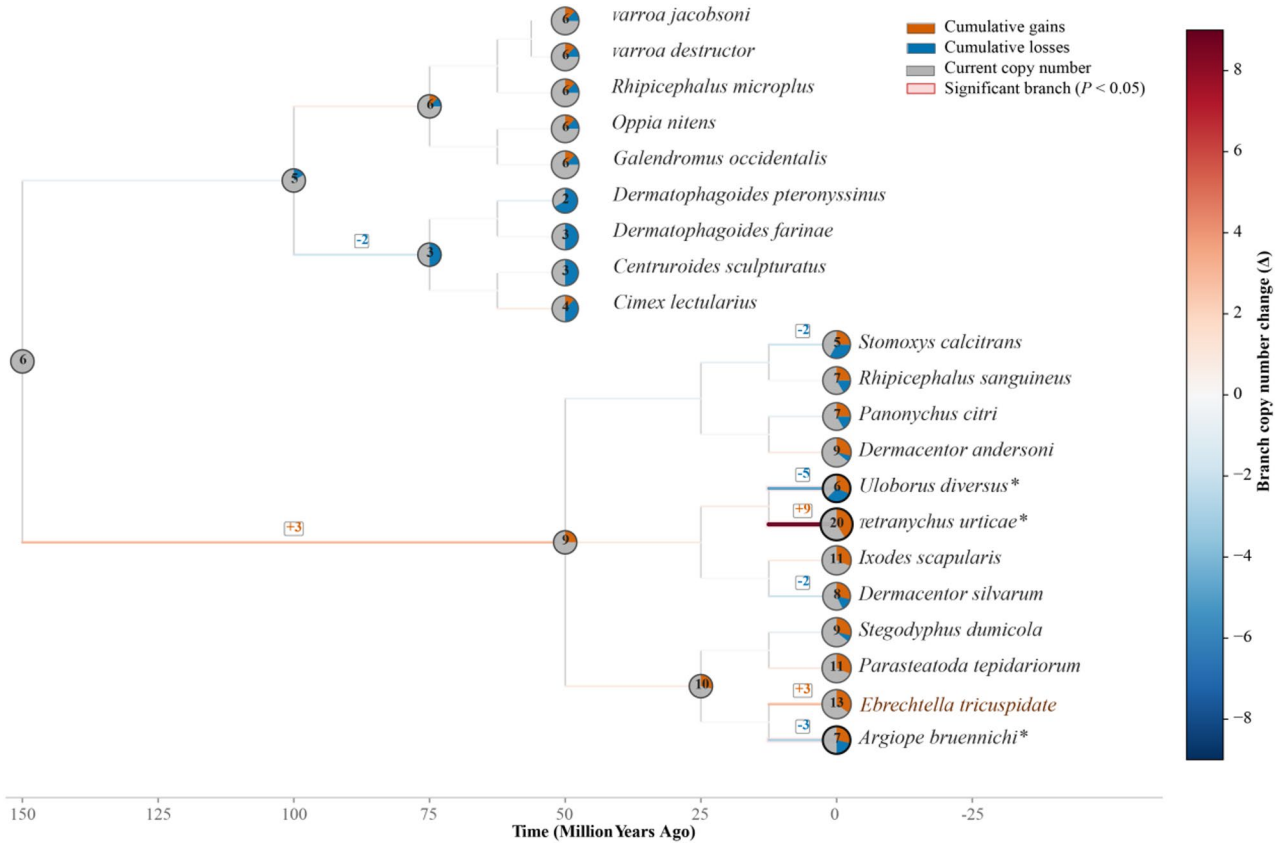

Fig. 3 *NPC2* gene-family expansion and contraction inferred by CAFE5 (pie-chart phylogram). The time-calibrated species tree is annotated with CAFE5-inferred *NPC2* copy-number changes. Branch colors indicate reconstructed copy-number change on each branch (Δcopy number = child − parent; color scale at right), and branch labels indicate large and/or significant changes. Pie charts are shown for terminal taxa and selected key internal nodes, summarizing lineage-level dynamics: cumulative gains (orange), cumulative losses (blue), and reconstructed current copy number (gray). The number at the center of each pie indicates the reconstructed copy number at that node. Asterisks (*) mark nodes with significant CAFE branch p-values ($p < 0.05$), and significant branches are highlighted

$P = 4.77 \times 10^{-14}$) but did not detect significant branches within the smaller *E. tricuspidata* subclades.

Identification and structural mapping of positively selected sites

Bayes Empirical Bayes (PAML) and MEME (HyPhy) identified overlapping high-confidence sites of positive selection (Fig. 5). In the *Etri-B* lineage, a high-confidence site at reference position 28 ($PP = 0.961$) corresponded to Gln31 (Q31) in the mature EtriNPC2-3 protein. An additional BEB site at reference position 2 (T5, $PP = 0.929$) was also identified. Two suggestive sites were detected in Etri-A (Glu38 (site 36), $PP = 0.928$; Ser40 (site 38), $PP = 0.941$).

Structural mapping onto the AlphaFold3-predicted model of EtriNPC2-3 revealed that Q31 is positioned on a flexible surface loop at the entrance to the hydrophobic ligand-binding pocket, while T5 is located in a more distal region (Fig. 6A). Superposition of the adaptive EtriNPC2-3 and the basal EtriNPC2-2 highlighted a distinct physicochemical shift at this site (L33 → Q31; Leu → Gln) accompanied by a noticeable loop deviation (Fig. 6B). It is important to note that AlphaFold provides static structural predictions rather than dynamic thermodynamic simulations. Consequently, rather than drawing definitive functional conclusions, these observations generate the hypothesis that the L→Q substitution might modulate ligand accessibility or alter surface polarity in the sensillar lymph. The Q31 site thus serves as a prime candidate for future experimental validation.

Tissue-specific expression patterns of NPC2 genes

The heatmap demonstrated distinct expression signatures across the *EtriNPC2* repertoire (Fig. 7A). PCA analysis of the 13 high-confidence genes further corroborated this, revealing clear clustering of samples by tissue type, where chemosensory appendages (palps and legs) formed a cluster distinct from the abdomen and the rest of the body (Fig. 7B).

Quantitative analysis based on TPM values normalized to the whole transcriptome revealed a striking functional dichotomy in expression magnitude and tissue specificity

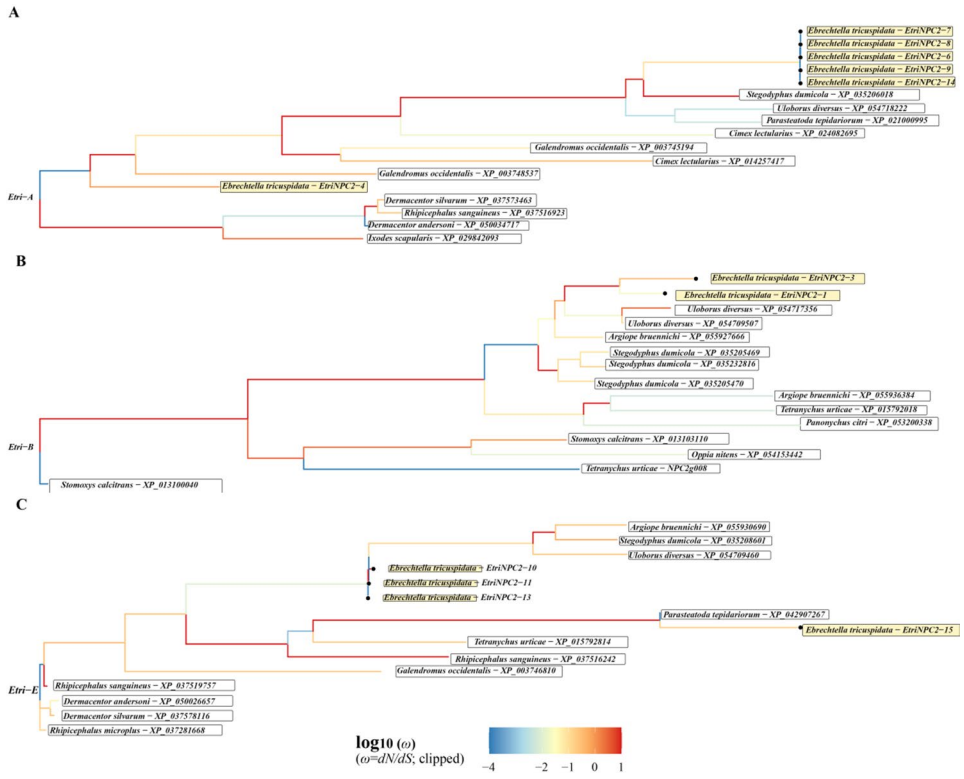

Fig. 4 Variation in selective pressures across *EtriNPC2* subclade phylogenies. Free-ratio model estimates of the non-synonymous to synonymous substitution ratio ($\omega = dN/dS$) mapped onto the phylogenies of (**A**) *Etri-A*, (**B**) *Etri-B*, and (**C**) *Etri-E* subfamilies. Branch colors represent ω values plotted on a log10 scale. Warmer colors (red/orange) indicate elevated ω ratios, suggesting relaxed selective constraints or positive selection; cooler colors (blue) indicate low ? ratios, reflecting strong purifying selection

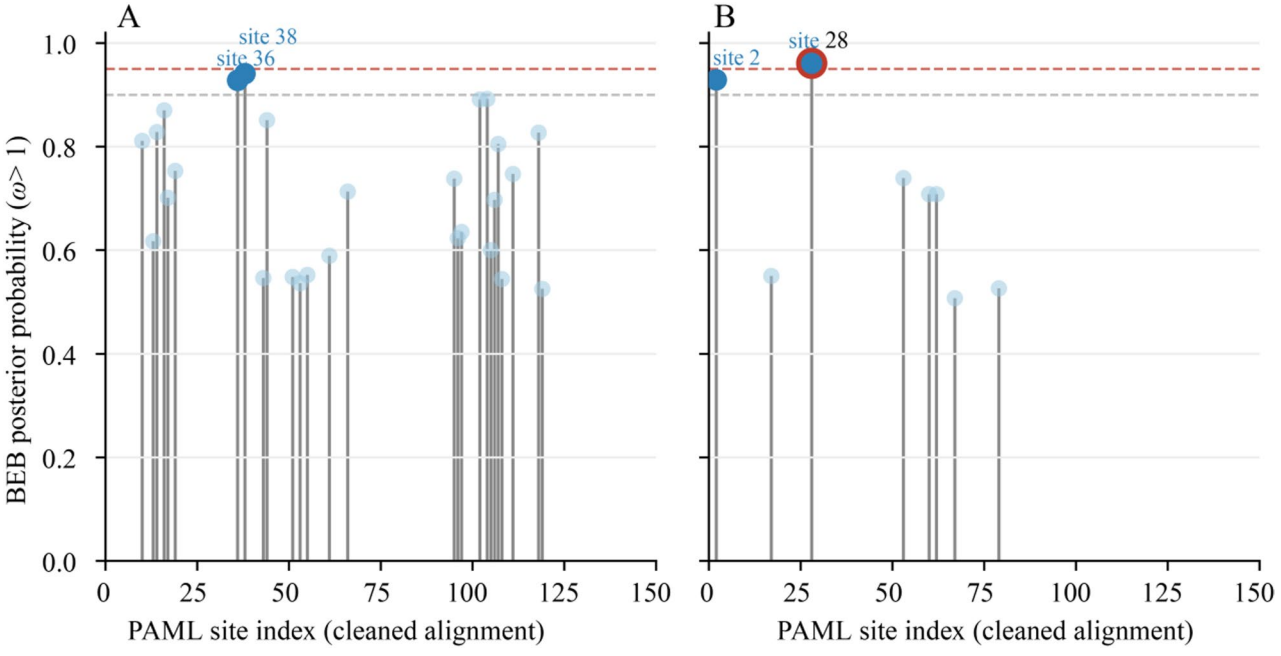

Fig. 5 BEB posterior probabilities of candidate adaptive sites in *Etri-A* and *Etri-B*. **A** Posterior probabilities for candidate sites in *Etri-A* mapped to alignment positions. **B** Posterior probabilities for candidate sites in *Etri-B* mapped to alignment positions. In both panels, labels indicate the reference (background) amino acid at each position, and the dashed line marks the significance threshold of PP = 0.95. For *Etri-B*, only one site (position 28; *PP* = 0.961) exceeds this threshold; the remaining three sites (*PP* = 0.928–0.941) are reported as suggestive candidates.

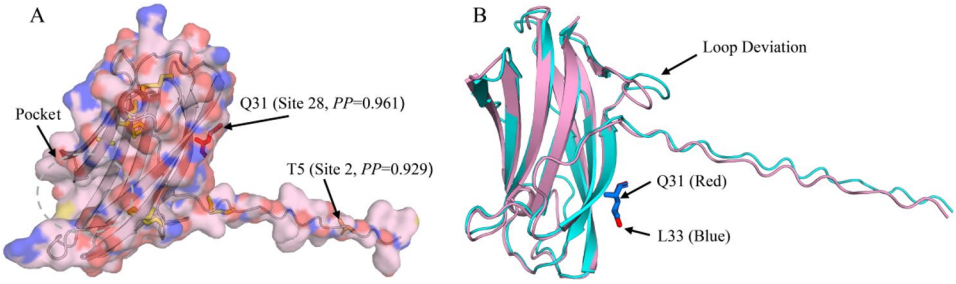

Fig. 6 Structural mapping of positively selected sites in EtriNPC2-3. **A** The high-confidence site Q31 (Site 28, *PP* = 0.961) is located on a flexible surface loop at the pocket entrance. An additional BEB site T5 (Site 2, *PP* = 0.929) is indicated distally. **B** Superposition of adaptive EtriNPC2-3 (Adaptive/Derived) and basal EtriNPC2-2 (Basal/Ancestral) showing the substitution at position 31 (Q31 vs. L33) and loop deviation (Supplementary Table S6)

(Fig. 7C). The genes could be broadly categorized into appendage-biased (sensory) and abdomen-biased (visceral) profiles. The derived singleton *EtriNPC2-4* and the *Etri-B* subclade members (*EtriNPC2-1*, *EtriNPC2-3*) exhibited a marked enrichment in the chemosensory appendages. *EtriNPC2-4* was the most abundant transcript in the sensory tissues, peaking in the pedipalps (7841.1 ± 1443.5 TPM) and legs (Legs III + IV: 5735.1 ± 758.6 TPM; Legs I + II: 5401.9 ± 709.8 TPM), which was significantly higher than its expression in the abdomen (512.7 ± 88.7 TPM; $P < 0.05$). Similarly, the positively selected *EtriNPC2-3* showed elevated mean expression in the forelegs (2340.3 ± 1473.5 TPM) and palps (1823.3 ± 1035.9 TPM) compared to the abdomen (161.5 ± 138.1 TPM), although the high variance in appendage samples suggests dynamic regulation. *EtriNPC2-1* also displayed a sensory-biased pattern, with expression levels in palps (770.5 ± 471.9 TPM) and forelegs (714.9 ± 392.2 TPM) approximately 15-fold higher than in the abdomen (47.5 ± 45.0 TPM).

In sharp contrast, the basal singleton *EtriNPC2-2* and the majority of *Etri-A* (*EtriNPC2-6*, *-7*, *-8*, *-9*) and *Etri-E* (*EtriNPC2-11*, *-14*, *-15*) paralogs were predominantly expressed in the abdomen. Notably, *EtriNPC2-2* exhibited the highest overall expression level among the entire family, reaching 11605.5 ± 2004.5 TPM in the abdomen, while its expression in the chemosensory palps was negligible (32.2 ± 13.0 TPM; $P < 0.05$). *EtriNPC2-6* also showed robust abdominal expression (6737.1 ± 802.3 TPM) compared to the appendages (< 600 TPM). This massive enrichment in internal tissues strongly implies a conserved ancestral role in visceral lipid metabolism, distinct from the neofunctionalized sensory roles observed in the derived lineages.

Discussion

The Niemann-Pick type C2 (NPC2) gene family, originally identified for intracellular cholesterol trafficking in vertebrates, has emerged as a major soluble carrier component in arthropod chemosensation. In chelicerates, which generally lack the canonical insect-type OBPs, NPC2 proteins are hypothesized to serve as the primary extracellular transporters of hydrophobic semiochemicals [1, 8]. Although OBP-like sequences have been reported in some spiders, NPC2 proteins appear to play a dominant role in chemosensory tissues, particularly in those lineages where classical OBPs are scarce or absent [10, 44]. This functional shift is supported by enriched expression in chemosensory appendages, such as Haller's organ in ticks, pedipalps and legs in spiders, and forelegs in predatory mites [4, 8, 9, 44].

Our study reveals pronounced variation in *NPC2* repertoire size across arachnids (2–20 core genes per species), with the most marked dynamics observed in the Acari and Araneae. CAFE 5 analysis identified statistically significant net gains on the *T. urticae* terminal branch (+ 9 genes) and a moderate, though non-significant, net gain along the *E. tricuspidata* lineage (+ 3 genes), consistent with the birth-and-death evolutionary model typical of chemosensory gene families [10, 45]. The global birth–death rate ($\lambda = 0.005$) is comparable to turnover rates reported for other arthropod chemosensory gene families [46]. The variation in repertoire size likely reflects differences in ecological complexity, host or prey range, and reliance on chemical communication. The absence of significant expansions in most spider lineages (e.g., *Argiope bruennichi*, which showed a contraction) suggests that a baseline repertoire of approximately 6–10 *NPC2* genes is sufficient for general chemosensory function, with expansions occurring under specific selective pressures associated with ecological specialization.

The largest *NPC2* repertoire (20 core genes) occurs in the extreme generalist herbivore *T. urticae*, whose broad host range (> 1,100 plant species) requires the discrimination of diverse plant volatiles and the coordination of detoxification pathways [9, 11]. An enlarged *NPC2* repertoire may enhance the capacity to bind and transport diverse plant-derived semiochemicals. Although our aBSREL analysis detected episodic positive selection on specific branches within the *T. urticae* expansion (e.g., terminal branch $P < 10^{-13}$), the signal was less pervasive than in the spider lineage, suggesting that for generalist

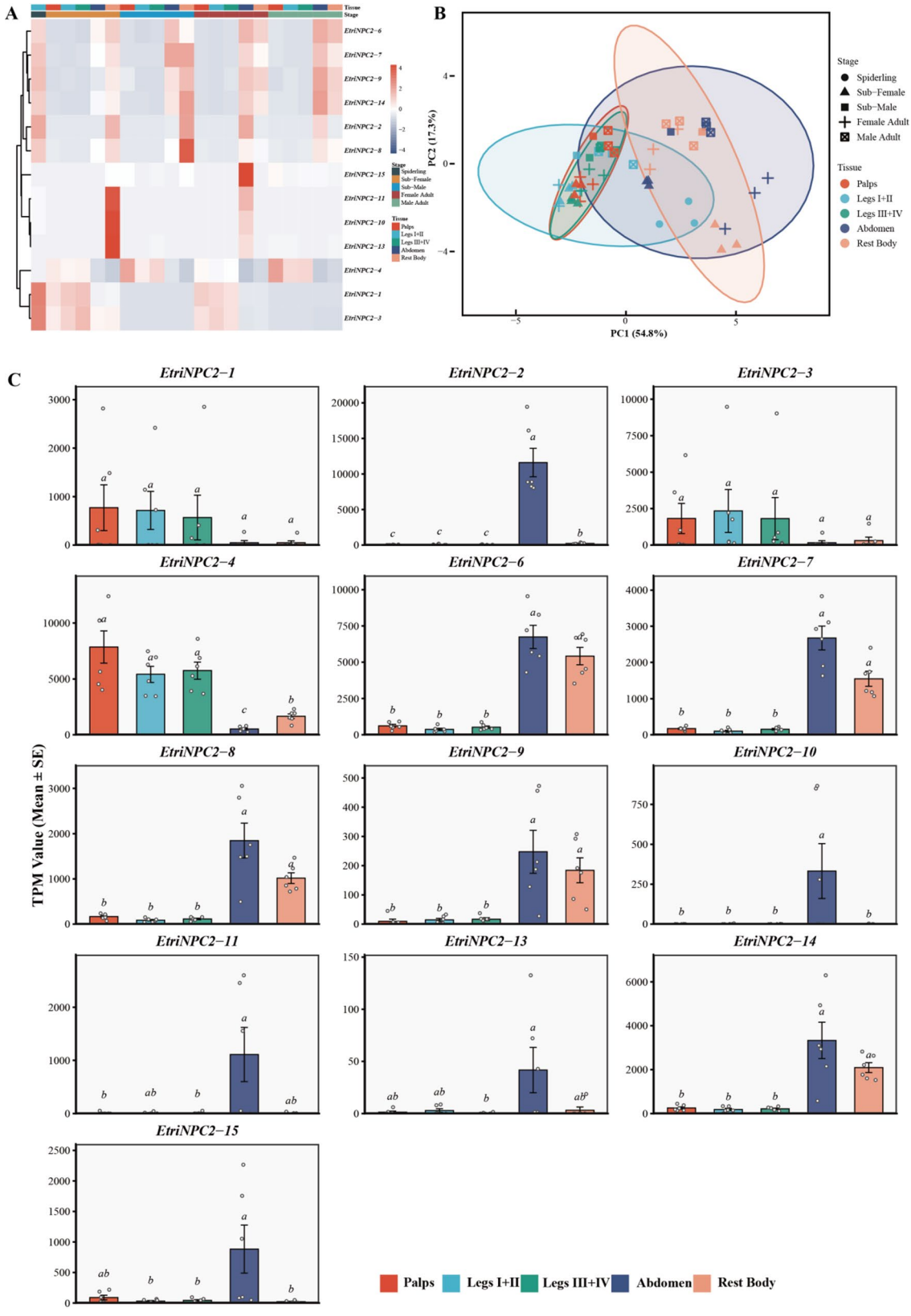

Fig. 7 (See legend on next page.)

(See figure on previous page.)

Fig. 7 Spatiotemporal expression landscape and tissue specificity of the *E. tricuspidata NPC2* gene family. **A** Hierarchical clustering heatmap showing the relative expression profiles of the 13 *EtriNPC2* genes across developmental stages and adult tissues. Expression values (TPM) were normalized by row Z-score; the color gradient (red to blue) represents high to low relative abundance. Top annotation bars indicate sample group metadata. **B** Principal Component Analysis (PCA) of the transcriptomic samples based on overall *NPC2* expression. Each dot represents an independent biological replicate (see Supplementary Table S1 for exact sample sizes). Ellipses denote 95% confidence intervals for each group. **C** Quantitative analysis of *NPC2* expression in adult tissues. Data represent mean TPM ± SD ($n = 6$; pooled from 3 males and 3 females). Individual dots overlaid on the bars indicate original data points. Statistically significant differences among tissues are indicated by different lowercase letters ($P < 0.05$, one-way ANOVA with Tukey's HSD post-hoc test)

herbivores, repertoire expansion and functional redundancy might play a more significant role compared to sequence-level refinement. Conversely, the ambush predator *E. tricuspidata* harbors 13 high-confidence secreted *NPC2* paralogs (out of 16 initial candidates), indicating a moderate net increase relative to the arachnid ancestor. The evolutionary trajectory of this repertoire appears closely tied to the spider's foraging strategy. Unlike web-building spiders (such as the orb-weaver *A. bruennichi*) that rely heavily on silk-borne vibrations to detect intercepted prey or active wandering spiders that constantly sample new environments, in contrast, ambush predators exhibit a strict "sit-and-wait" strategy. *E. tricuspidata* must integrate complex chemical information without the aid of a web, utilizing floral volatiles for hunting site selection and prey-derived cuticular hydrocarbons for prey recognition against complex floral backgrounds [12, 47]. In this specific ecological context, refined ligand specificity and optimized protein may be heavily prioritized over mere repertoire size. These contrasting patterns suggest that ecological specialization may shape distinct evolutionary trajectories in *NPC2* repertoires.

A methodological contribution of this study was the systematic re-annotation of *E. tricuspidata NPC2* candidates using full-length Iso-Seq transcripts combined with Illumina short reads and six-frame ORF scanning. This workflow recovered canonical N-terminal signal peptides in previously truncated sequences, confirming that 13 of the 16 identified transcripts encode intact secreted proteins. The three excluded candidates lacked signal peptides or contained frameshifts, likely representing assembly artifacts. Such curation is important because the presence of a functional signal peptide distinguishes extracellular carriers from intracellular lysosomal *NPC2* orthologs and directly supports the chemosensory carrier hypothesis [1, 48].

Phylogenetic reconstruction on the codon-aligned CDS dataset showed that arachnid *NPC2* diversity arose predominantly through lineage-specific duplications rather than the retention of ancient orthologs. The *EtriNPC2* genes clustered into three divergent subclades (*Etri-A*, *Etri-B*, *Etri-E*). Global BUSTED analysis detected episodic diversifying selection across the family. Branch-site Model A revealed intense positive selection on the ancestral branch of the *Etri-B* subclade and moderate support on *Etri-A*, while *Etri-E* showed no evidence of positive selection. The detection of strong positive selection on the *Etri-B* subclade, coupled with its likely expression in chemosensory tissues, is consistent with recent findings in predatory mites, where *NPC2* genes showing the strongest upregulation by HIPVs also exhibit signatures of positive selection [4].

Structural mapping of high-confidence adaptive sites onto AlphaFold3 models of EtriNPC2-3 revealed that positive selection primarily targeted solvent-exposed surface regions rather than the ligand-binding pocket interior. The high-confidence *Etri-B*-specific site Q31 (reference position 28, BEB $PP = 0.961$) lies on a flexible loop near the pocket entrance. This substitution represents a physicochemical shift from a hydrophobic or basic residue (such as Leu33 in the basal *E. tricuspidata* paralog EtriNPC2-2, or Val/Lys in other ancestral arachnid orthologs) to a polar glutamine (Q31 in EtriNPC2-3). Our structural mapping highlights the potential importance of surface-exposed residues, such as Q31, in the evolutionary divergence of *Etri-B* paralogs. However, because our structural insights rely on static AlphaFold predictions, we cannot definitively conclude that these substitutions alter dynamic structural flexibility or in vivo solubility. Instead, our findings show that lineage - specific positive selection acts on pocket -entrance loops to fine -tune semiochemical uptake. Future biochemical assays and molecular dynamics (MD) simulations are required to validate how these specific mutations affect thermodynamic stability and ligand-binding kinetics.

Transcriptional profiling further reveals that this adaptive sequence evolution is tightly coupled with spatial expression shifts, providing a functional context for the observed structural changes. Our data demonstrate a profound functional dichotomy within the *E. tricuspidata NPC2* repertoire. The basal singleton EtriNPC2-2, which lacks signatures of positive selection and retains the ancestral hydrophobic residue (Leu33), is predominantly expressed in the abdomen. The massive expression level of ~ 11,600 TPM in the abdomen (versus < 50 TPM in palps) strongly implies that the ancestral function of the spider NPC2 family involves intracellular cholesterol trafficking or visceral lipid transport related to digestion and reproduction. This massive abdominal enrichment strongly implies that the ancestral function of the spider *NPC2* family involves intracellular cholesterol trafficking or visceral lipid transport related to digestion and reproduction [1, 44]. In sharp contrast, members of the derived *Etri-B* subclade (*EtriNPC2-1*, *-3*) and the

singleton *EtriNPC2-4* have been exclusively recruited to the sensory appendages. *EtriNPC2-4* dominates the sensory repertoire with expression levels reaching ~ 7,800 TPM in the pedipalps, while the adaptive *EtriNPC2-3* also maintains high mean expression in the forelegs (~ 2,340 TPM). This spatial transition from the lipophilic environment of the visceral abdomen to the aqueous sensillar lymph of the appendages likely imposed novel selective pressures, driving the neofunctionalization of these proteins. We hypothesize that the adaptive Leu→Gln substitution (Q31) at the pocket entrance of *EtriNPC2-3* represents a specific physicochemical adaptation to this new fluid environment, optimizing protein solubility or facilitating interactions with chemosensory receptors [3, 10]. Ecologically, this robust appendage-specific expression aligns with the "sit-and-wait" foraging strategy of *E. tricuspidata*, which relies on sensilla on extended forelegs and palps to detect floral volatiles and prey cues without the aid of a web [47].

Several limitations should be noted, our repertoire estimates and tissue-specific patterns are based on transcriptomic data, which captures only expressed genes under the sampled conditions. CAFE 5 analysis assumes complete genomic input; using a transcriptome means that unexpressed or lowly expressed paralogs may be mathematically treated as evolutionary losses, potentially underestimating the true genomic repertoire. Therefore, the inferred net gain of + 3 genes in *E. tricuspidata* represents the minimum expressed repertoire rather than a definitive genomic event. Further validation via genome sequencing, in situ hybridization, or immunohistochemistry would be required to confirm the complete gene repertoire and spatial expression patterns. The absence of a chromosome-level genome for *E. tricuspidata* precludes confirmation of exact gene copy numbers and genomic clustering. Future work should prioritize the functional validation of these adaptive candidates. Ligand-binding assays comparing adaptive paralogs (e.g., EtriNPC2-3) with basal paralogs (e.g., EtriNPC2-2) using prey-derived volatiles would test whether sites such as Q31 alter binding affinity. Site-directed mutagenesis, such as the generation of Q31L revertants, can be instrumental in elucidating the functional contribution of individual surface residues to secretion efficiency or receptor interaction. These structural hypotheses warrant further validation via recombinant protein expression and fluorescence binding assays.

In summary, the *NPC2* family illustrates how gene duplication, positive selection, and spatial expression shifts interact to shape chemosensory repertoires in chelicerates. For *E. tricuspidata*, the partitioning of *NPC2*s into specialized sensory carriers in the appendages (*Etri-B*, *EtriNPC2-4*) and conserved internal lipid transporters in the abdomen suggests a targeted adaptive strategy for detecting sparse semiochemicals. These findings advance our understanding of arachnid chemosensory biology and highlight the role of non-canonical olfactory proteins in compensating for the absence of classical OBPs.

Conclusions

This study elucidates the molecular evolution and functional divergence of the *NPC2* gene family in the crab spider *E. tricuspidata*. Unlike the large-scale repertoire expansion observed in the generalist herbivore mite *T. urticae*, the *NPC2* repertoire in *E. tricuspidata* exhibits a moderate expansion characterized by profound functional partitioning. We provide robust transcriptomic evidence for neofunctionalization following gene duplication: the retention of the basal singleton *EtriNPC2-2* in the abdomen supports an ancestral intracellular lipid-transport function, whereas the derived *Etri-B* subclade has been recruited to sensory appendages (palps and legs) and exhibits strong signatures of adaptive evolution. This spatial expression shift coincides with positive selection acting on critical surface residues. Specifically, the Q31 substitution identified in EtriNPC2-3 serves as a high-value candidate for future functional validation to decipher its role in modulating protein solubility or ligand interactions within the aqueous sensillar lymph. Collectively, these findings underscore that gene repertoire size alone is not the sole determinant of chemosensory adaptation. Rather, the targeted recruitment of NPC2 variants to the peripheral sensory system, coupled with adaptive sequence refinement, provides a key molecular mechanism by which ambush spiders compensate for the absence of canonical insect OBPs.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-026-12756-1.

Supplementary Material 1.

Supplementary Material 2.

Acknowledgements
Not applicable.

Authors' contributions
TL: Conceptualization, methodology, software, formal analysis, investigation, writing - original draft. CC: Software, writing - original draft, funding acquisition. JC: Software, writing - original draft. DJ: Writing - review & editing, supervision, funding acquisition. All authors reviewed and approved the final manuscript.

Funding
This research was funded by the Scientific Research Platform of the Education Department of Guizhou Province, grant number Qianjiaoji [2022] 052.

Data availability
The raw transcriptome sequencing reads for ***E. tricuspidata*** generated in this study have been deposited in the NCBI Sequence Read Archive (SRA) under BioProject accession number PRJNA1301248. The genomic datasets for the 18 arachnid and 2 insect outgroup species analyzed in this study are publicly

available, with specific accession numbers listed in Supplementary Table S2. The final curated dataset of 155 NPC2 gene sequences (including nucleotide coding sequences and translated proteins), multiple sequence alignments, and phylogenetic tree files are provided in the Supplementary Data Archive. Detailed statistical results for gene family evolution and selection pressure analyses are provided in Supplementary Tables S3–S6. All other data generated or analyzed during this study are included in this published article and its supplementary information files.

Declarations

Ethics approval and consent to participate

Not applicable. The study involves spiders (*Ebrechtella tricuspidata*) and other arachnids, which are invertebrates and are exempt from ethical approval requirements according to local legislation and institutional guidelines.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 8 August 2025 / Accepted: 12 March 2026
Published online: 28 March 2026

References

1. Pelosi P, Iovinella I, Felicioli A, Dani FR. Soluble proteins of chemical communication: an overview across arthropods. Front Physiol. 2014;5. https://doi.org/10.3389/fphys.2014.00320.
2. Leal WS. Odorant reception in insects: roles of receptors, binding proteins, and degrading enzymes. Annu Rev Entomol. 2013;58:373–91. https://doi.org/10.1146/annurev-ento-120811-153635.
3. Pelosi P, Iovinella I, Zhu J, Wang G, Dani FR. Beyond chemoreception: diverse tasks of soluble olfactory proteins in insects. Biol Rev. 2018;93:184–200. https://doi.org/10.1111/brv.12339.
4. Yang H, Niu T, Ma R, Xiang D, Liu H, Chen H, et al. Functional characterization of Niemann-Pick proteins type C2 as odorant carriers in *Neoseiulus barkeri* olfactory perception. Insect Sci. 2025. https://doi.org/10.1111/1744-7917.70143.
5. Eyun S, Soh H, Posavi M, Munro J, Hughes D, Murali S, et al. Evolutionary history of chemosensory-related gene families across the Arthropoda. Mol Biol Evol. 2017;34:1838–62. https://doi.org/10.1093/molbev/msx147.
6. Song C, Wei J, Chen D, Yang X, Jiang C, Li Q. Niemann-Pick type C2 proteins mediate the perception of herbivore-induced plant volatiles in the predatory mite *Neoseiulus californicus*. J Agric Food Chem. 2025;73:31115–29. https://doi.org/10.1021/acs.jafc.5c10030.
7. Xiao F, Zhu J, Fu D, Li Q, Wang Z, Zhang H, et al. Expression and functional analysis of Niemann-Pick type C2 genes in *Neoseiulus californicus*. Entomol Res. 2025;55. https://doi.org/10.1111/1748-5967.70055.
8. Iovinella I, Ban L, Song L, Pelosi P, Dani FR. Proteomic analysis of castor bean tick *Ixodes ricinus*: a focus on chemosensory organs. Insect Biochem Mol Biol. 2016;78:58–68. https://doi.org/10.1016/j.ibmb.2016.09.004.
9. Wang Y, Wang C, Liu W, Huang Q, Xiao W. Niemann-Pick C2 proteins play crucial role in perception of plant volatiles in *Tetranychus cinnabarinus*. Pest Manag Sci. 2025;81:7511–20. https://doi.org/10.1002/ps.8839.
10. Vizueta J, Rozas J, Sánchez-Gracia A. Comparative genomics reveals thousands of novel chemosensory genes and massive changes in chemoreceptor repertories across chelicerates. Genome Biol Evol. 2018;10:1221–36. https://doi.org/10.1093/gbe/evy081.
11. Ngoc PCT, Greenhalgh R, Dermauw W, Rombauts S, Bajda S, Zhurov V, et al. Complex evolutionary dynamics of massively expanded chemosensory receptor families in an extreme generalist chelicerate herbivore. Genome Biol Evol. 2016;8:3323–39. https://doi.org/10.1093/gbe/evw249.
12. Su Q, Qi L, Yun Y, Zhang W, Peng Y. Visual preference of flower-visiting crab spiders *Ebrechtella tricuspidata* for host flowers. Ecol Entomol. 2020;45:626–34. https://doi.org/10.1111/een.12835.
13. Li Z, Zhong R, Yu L, Zhang H, Zhao Y, Peng Y. Exploiting floral signals: olfactory crypsis and visual attraction in crab spider predatory strategies. Entomol Exp Appl. 2024;172:1072–81. https://doi.org/10.1111/eea.13504.
14. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20. https://doi.org/10.1093/bioinformatics/btu170.
15. Wenger AM, Peluso P, Rowell WJ, Chang P-C, Hall RJ, Concepcion GT, et al. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. Nat Biotechnol. 2019;37:1155–62. https://doi.org/10.1038/s41587-019-0217-9.
16. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28:3150–2. https://doi.org/10.1093/bioinformatics/bts565.
17. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31:3210–2. https://doi.org/10.1093/bioinformatics/btv351.
18. Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2022;50:D20–6. https://doi.org/10.1093/nar/gkab1112.
19. Eddy SR. Accelerated profile HMM searches. PLoS Comput Biol. 2011;7:e1002195. https://doi.org/10.1371/journal.pcbi.1002195.
20. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: the protein families database in 2021. Nucleic Acids Res. 2021;49:D412–9. https://doi.org/10.1093/nar/gkaa913.
21. Altschul S. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389–402. https://doi.org/10.1093/nar/25.17.3389.
22. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol. 2019;37:420–3. https://doi.org/10.1038/s41587-019-0036-z.
23. Thumuluri V, Almagro Armenteros JJ, Johansen AR, Nielsen H, Winther O. DeepLoc 2.0: multi-label subcellular localization prediction using protein language models. Nucleic Acids Res. 2022;50:W228–34. https://doi.org/10.1093/nar/gkac278.
24. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80. https://doi.org/10.1093/molbev/mst010.
25. Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 2006;34:W609–12. https://doi.org/10.1093/nar/gkl315.
26. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25:1972–3. https://doi.org/10.1093/bioinformatics/btp348.
27. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37:1530–4. https://doi.org/10.1093/molbev/msaa015.
28. Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14:587–9. https://doi.org/10.1038/nmeth.4285.
29. Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol. 2018;35:518–22. https://doi.org/10.1093/molbev/msx281.
30. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010;59:307–21. https://doi.org/10.1093/sysbio/syq010.
31. Mendes FK, Vanderpool D, Fulton B, Hahn MW. CAFE 5 models variation in evolutionary rates among gene families. Bioinformatics. 2021;36:5516–8. https://doi.org/10.1093/bioinformatics/btaa1022.
32. Sanderson MJ. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics. 2003;19:301–2. https://doi.org/10.1093/bioinformatics/19.2.301.
33. Kumar S, Suleski M, Craig JM, Kasprowicz AE, Sanderford M, Li M, et al. TimeTree 5: an expanded resource for species divergence times. Mol Biol Evol. 2022;39:msac174. https://doi.org/10.1093/molbev/msac174.
34. Vieira F, Rozas J. Comparative genomics of the odorant-binding and chemosensory protein gene families across the Arthropoda: origin and evolutionary history of the chemosensory system. Genome Biol Evol. 2011;3:476–90. https://doi.org/10.1093/gbe/evr033.
35. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24:1586–91. https://doi.org/10.1093/molbev/msm088.

36. Kosakovsky Pond SL, Poon AFY, Velazquez R, Weaver S, Hepler NL, Murrell B, et al. HyPhy 2.5—a customizable platform for evolutionary hypothesis testing using phylogenies. Mol Biol Evol. 2020;37:295–9. https://doi.org/10.1093/molbev/msz197.
37. Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S, Aylward A, et al. Gene-wide identification of episodic selection. Mol Biol Evol. 2015;32:1365–71. https://doi.org/10.1093/molbev/msv035.
38. Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. Mol Biol Evol. 2015;32:1342–53. https://doi.org/10.1093/molbev/msv022.
39. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. Detecting individual sites subject to episodic diversifying selection. PLoS Genet. 2012;8:e1002764. https://doi.org/10.1371/journal.pgen.1002764.
40. 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;630:493–500. https://doi.org/10.1038/s41586-024-07487-w.
41. Schrödinger LLC. The PyMOL Molecular Graphics System, Version 1.8. 2015. http://www.pymol.org
42. de Mendiburu F. agricolae: statistical procedures for agricultural research. R package version 1.3-5. 2021. https://CRAN.R-project.org/package=agricolae
43. Kassambara A, Mundt F. factoextra: extract and visualize the results of multivariate data analyses. R package version 1.0.7. 2020. https://CRAN.R-project.org/package=factoextra
44. Xiu C, Xiao Y, Zhang S, Bao H, Liu Z, Zhang Y. Niemann-Pick proteins type C2 are identified as olfactory related genes of *Pardosa pseudoannulata* by transcriptome and expression profile analysis. Comp Biochem Physiol Part D Genomics Proteom. 2019;29:320–9. https://doi.org/10.1016/j.cbd.2019.01.004.
45. Nei M, Niimura Y, Nozawa M. The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. Nat Rev Genet. 2008;9:951–63. https://doi.org/10.1038/nrg2480.
46. Sánchez-Gracia A, Vieira FG, Rozas J. Molecular evolution of the major chemosensory gene families in insects. Heredity. 2009;103:208–16. https://doi.org/10.1038/hdy.2009.55.
47. Dodson GN, Lang PL, Jones RN, Versprille AN. Specificity of attraction to floral chemistry in *Misumenoides formosipes* crab spiders. J Arachnol. 2013;41:36–42. https://doi.org/10.1636/Hi11-94.1.
48. Cui Y, Wang J, Liu Q, Li D, Zhang W, Liu X, et al. Identification and expression of potential olfactory-related genes related to Niemann-Pick C2 protein and ionotropic receptors in *Haemaphysalis longicornis*. Exp Appl Acarol. 2022;87:337–50. https://doi.org/10.1007/s10493-022-00729-4.

Publisher's Note