Comparative Genomic Insights into the Longhorned Tick, *Haemaphysalis Longicornis*

Rui-Ling Zhang (✉ zhangrl_06@126.com)

Shandong First Medical University

Qian Zhang

Shandong First Medical University

Zhong Zhang

Shandong First Medical University

Short Report

Keywords: Ixodidae, genome annotation, comparative genomics, gene family evolution, detoxification

Posted Date: August 4th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-49223/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: The longhorned tick, *Haemaphysalis longicornis* Neumann, is widely distributed across temperate regions. It can parasitize terrestrial vertebrates, including birds and a large number of mammals. They are a concern in human and animal health notably for their potential to transmit infectious agents.

Methods: Genome survey was investigated using GenomeScope v1.0.0 with a maximum k-mer coverage cutoff of 1,000. Non-redundant assembly was polished with Illumina short reads using two rounds of NextPolish v1.1.0. Genome completeness was assessed using BUSCO v3.0.2 pipeline analyses against arthropod gene set (n = 1, 066). Ab initio predictions were generated using BRAKER v2.1.5. Transcriptomic reads were mapped to the genome with HISAT2 v2.2.0 and assembled with StringTie v2.1.2. Gene functions were assigned against UniProtKB database using Diamond v0.9.24. Orthogroups of 16 Chelicerata species were inferred using OrthoFinder v2.3.8 and gene family evolution was estimated using CAFÉ v4.2.1. Gene families related to digestion and detoxification, i.e. cytochrome P450 (CYP), carboxyl/cholinesterase (CCE), glutathione-S-transferase (GST), ATP-binding cassette (ABC) transporter were annotated by searching in the genome assembly.

Results: The final genome assembly has a size of 3.12 Gb, a scaffold N50 of 1.09 Mb, and captured 92.4% of the BUSCO gene set (n=1,066). Genome architecture pattern of the longhorned tick resembles another tick, *Ixodes scapularis* (Say), particularly in large size, highly repetitive DNA (~65%) and protein-coding genes (21,550). We also identified 5,601 non-coding RNAs with a high ratio of tRNAs (4,271). Gene family evolution revealed 350 rapidly evolving gene families. Combining function enrichment analyses of gene ontology (GO) and KEGG pathway, 255 families experiencing significant expansions mainly involves in cuticle synthesis, digestion and detoxification.

Conclusions: The new genome assembly, annotation and comparative genomic analyses provide a valuable resource for insights into parasitic life mode of the longhorned tick.

Introduction

Haemaphysalis longicornis (Acari: Ixodidae) is obligate hematophagous ectoparasites of vertebrates, which is widely distributed in China, Korea, Japan, Australia, New Zealand and Pacific Islands [1–5]. Hosts of this tick include a wide variety of mammals such as cattle, sheep, dogs, and birds. As one of the three-host ticks, *H. longicornis* take a single blood meal at larvae, nymphs and adult stage, respectively. In addition to causing anemia, weight loss, slow growth, and even the release of toxin that can cause death of hosts, *H. longicornis* also can transmit various zoonotic pathogens during the blood feeding. It has been reported that longhorned tick acts as vector of *Babesia ovate*, *Anaplasma bovis*, *Theileria orientalis*, *Coxiella burnetii*, *Ehrlichia chaffeensis*, *Rickettsia japonica*, as well as causal agent of the emerging infectious disease— severe fever with thrombocytopenia syndrome (SFTS) [6–11].

Despite the medical importance, there are major gaps in the knowledge of tick biology and methods to control ticks and the diseases they transmit are currently limited [12]. Genomic information will greatly contribute to the studies of vectors and its interactions with pathogens and hosts. Large genome sizes but small body sizes bring great difficulties in genome sequencing, assembly and annotation, and downstream applications. The present draft genome assembly of *H. longicornis* reached a size of 7.36 Gb with a very high ratio of redundancy, indicated by the 79% complete and duplicated BUSCO (Benchmarking Universal Single-Copy Orthologs) gene set; in addition, only repetitive elements were simply reported without any information of protein-coding genes and other comparative analyses [13]. These limits severely hinder our understanding of the biology of the longhorned tick.

We present a new version of the longhorned tick genome assembly by removing redundant sequences, annotate the repeats, non-coding RNAs (ncRNAs) and protein-coding genes, and compare gene family evolution across the main Chelicerata lineages, particularly several families related to dietary detoxification.

Material And Methods

Genome assembly

The available assembly (GCA_008122185.1) and raw PacBio (SRR9226158) and Illumina (SRR9226159) sequencing data were used for subsequent analyses. Pooled adults were collected for transcriptome sequencing to assist genome annotation. RNA extraction, library preparation and sequencing were carried out at Berry Genomics (Beijing, China). Genome survey was investigated using GenomeScope v1.0.0 [14] with a maximum k-mer coverage cutoff of 1,000. The k-mer frequencies were estimated with 21-mers using khist.sh (one of the BBTools v38.49 modules, [17]). Redundant heterozygous regions were removed with an identity cutoff of 60 using two rounds of Purge_dups v1.0.0 [16]. Non-redundant assembly was polished with Illumina short reads using two rounds of NextPolish v1.1.0 [17]. Genome completeness was assessed using BUSCO v3.0.2 pipeline [18] analyses against arthropod gene set (n = 1,066).

Genome annotation

We generate a custom repeat library by combining Dfam_3.0 [19], RepBase-20181026 databases [20], and *de novo* species specific library, which was constructed using RepeatModeler v2.0.1 [21]. RepeatMasker v4.0.9 [22] was used to mask repeats in the genome assembly upon the custom library. Non-coding RNAs were identified using Infernal v1.1.2 [23] and tRNAscan-SE v2.0 [24].

We employed MAKER v2.31.10 pipeline [25] to conduct gene prediction by integrating *ab initio*, transcriptome- and protein homology-based evidence. *Ab initio* predictions were generated using BRAKER v2.1.5 [26], which employed Augustus v3.3.2 [27] and GeneMark-ES/ET/EP v4.48_3.60_lic [28] based on transcriptomic data and OrthoDB [29]. Transcriptomic reads were mapped to the genome with HISAT2 v2.2.0 [30] and assembled with StringTie v2.1.2 [31]. Protein sequences of *Drosophila melanogaster*, *I.*

scapularis, *Tetranychus urticae*, *Stegodyphus mimosarum*, *Strigamia maritima*, *Daphnia pulex* and *Varroa destructor* were downloaded from the NCBI or Ensembl supporting the protein homology. We assigned gene functions using Diamond v0.9.24 [32] against UniProtKB database with the parameter ‘–more-sensitive -e 1e-5’. We also annotated protein domains, Gene Ontology (GO) and pathways (KEGG, Reactome) using eggNOG-mapper v2.0 [33] against eggNOG v5.0 database [34] and InterProScan 5.41-78.0 [35] against Pfam [36], PANTHER [37], Gene3D [38], Superfamily [39], and CDD [40] databases.

Gene family evolution

We inferred orthogroups of 16 Chelicerata species using OrthoFinder v2.3.8 [41]: *Centruroides sculpturatus*, *Dermatophagoides pteronyssinus*, *Galendromus occidentalis*, *H. longicornis*, *I. scapularis*, *S. mimosarum*, *Tachypleus tridentatus*, *Tetranychus urticae*, *V. destructor*. Phylogenetic tree was reconstructed using IQ-TREE v1.6.10 [42] with the partitioned strategy (‘-m MFP -mset LG -msub nuclear -rclusterf 10’) based on single-copy orthologs inferred from OrthoFinder. Protein sequences were aligned using MAFFT v7.394 [43] with the L-INS-I strategy, trimmed using trimAl v1.4.1 [44] with the heuristic method ‘automated1’, and concatenated using FASconCAT-G v1.04 [45]. Resulting tree was dated using r8s (Sanderson 2003) with calibration information extracted from the PDB database (<https://www.paleobiodb.org/navigator/>). Gene family evolution was estimated using CAFÉ v4.2.1 [46]. GO and KEGG function enrichment for those significantly expanded families were analyzed using R package clusterProfiler v3.10.1 [47].

Annotation of gene families associated to detoxification

We annotated gene families related to digestion and detoxification, i.e. cytochrome P450 (CYP), carboxyl/cholinesterase (CCE), glutathione-S-transferase (GST), ATP-binding cassette (ABC) transporter by searching in the genome assembly and automatically predicted gene models. Acari proteins were downloaded from NCBI RefSeq database as the reference. We conducted TBLASTN- and BLASTP-like searches using MMseq2 Release 8 [48] with four rounds of iterative searches. Candidate target genes were aligned to reference protein sequences to manually check intron/exon boundaries. The resulting proteins were examined using HMMER3 search in the Pfam database to confirm specific domains and using BLASTP in the non-redundant GenBank database to verify the classification.

Results And Discussion

Genome assembly and annotation

We estimated a genome size of ~ 3.09 Gb, a heterozygosity rate of 3.54–3.65% and a repetitive content of 65%. The left and the right peaks implied that the genome may have high levels of heterozygosity and repetition (FigureS1). The final assembly had 6,490 scaffolds/7,059 contigs, a total length of 3.12 Gb and scaffold/contig N50 length of 1.09/1.05 Mb. Assembled size was very close to the estimated one.

Genome completeness assessment against arthropod dataset (n = 1,066) (Table 2) revealed our assembled version covered 92.4% complete, 8.9% complete and duplicated, 1.1% fragmented, and 6.5% missing BUSCO genes.

A total of 64.81% (2.02 Gb) repetitive elements were identified in the longhorned tick genome. The top five abundant repeat categories were Unclassified (21.10%), LTR (18.37%), LINE (18.31%), DNA elements (2.99%), and SINE (2.23%) (Table S1). We identified 5,601 ncRNAs: 191 rRNAs, 65 miRNAs, 820 small nuclear RNAs (snRNAs), 2 long non-coding RNAs (lncRNAs), 8 small RNAs (sRNAs), 4,271 tRNAs (22 isoforms), and 244 other ncRNAs (Table S2). SnRNAs were classified into 771 spliceosomal RNAs (U1, U2, U4, U5, U6, U11), three minor spliceosomal RNAs (U12, U4atac, U6atac), and six H/ACA box and 40 C/D box snoRNAs. We predicted 21,550 protein-coding genes with the mean length of genes, exons and introns as 22,231.43, 262.16 and 3,451.43 bp, respectively, and 91.8% BUSCO completeness (Table 1). InterproScan and eggNOG functional annotations assigned protein domains of 17,963 (83.35%) genes, 14,464 GO terms, 10,903 KEGG ko terms, 4,148 Enzyme Codes, 6,788 KEGG and 4,083 Reactome pathways, and 15,557 COG categories, respectively.

Table 1
Genome assembly and annotation statistics of
the longhorned tick

Elements	Current version
Genome assembly	
Assembly size (Mb)	3,117.76
Number of scaffolds	6,490
Number of contigs	7,059
Longest scaffold (Mb)	8.69
Longest contig (Mb)	8.69
N50 scaffolds length (Mb)	1.09
N50 contig length (Mb)	1.05
GC (%)	47.50
Gaps (%)	0.003
BUSCO completeness (%)	92.4
Gene annotation	
Protein-coding genes	21,550
Mean protein length (aa)	437.15
Mean gene length (bp)	22,231.43
Exons per gene	7.02
Exon (%)	0.013
Mean exon length	262.16
Intron (%)	14.09
Mean intron length	3,451.43
BUSCO completeness (%)	91.8

Gene family evolution

A total of 162,773 (88.24%) genes were assigned into 18,611 gene families; among them, 3,482 families were shared by all nine species and 470 are single-copy ones; 143 families and 1,075 orthologs are common to four Parasitiformes species (Figure 1a, Table S3). In *H. longicornis*, 19,211 (91.91%) genes

were clustered into 9,639 gene families; 850 families and 4,008 genes were species-specific. Phylogenetic tree clustered *H. longicornis* and *I. scapularis*, both representing Ixodidae originated from early Cretaceous (Figure 1a). It indicated that the emergence of parasitic ixodids may be related to the pervasive reptiles, birds, mammals in Cretaceous.

CAFÉ identified 350 rapidly evolving gene families, 255 and 95 of them experienced significant expansions and contractions, respectively (Figure 1a). The largest expanded families were shown in Figure 1b and Table S4. Many of them are related to dietary digestion and detoxification, cuticle synthesis, such as ABC transporter, Cytochrome P450, carboxylesterase, tick histamine binding protein, insect cuticle protein, putative secreted salivary gland peptide, juvenile hormone acid O-methyltransferase, secretin family etc. GO (Figure 1c) and KEGG (Figure 1d) enrichment further confirmed it, involving various biological progress or pathways, for example, ABC transporters, insect hormone biosynthesis, fatty acid elongation and biosynthesis, fat digestion and absorption, and ovarian steroidogenesis (Figure 1d). KEGG pathways of toxoplasmosis and amoebiasis may relate to parasitic life of the longhorned tick. These findings adapted to the parasitic life are very similar to another ixodid tick *I. scapularis* [49].

Gene families related to detoxification

Digestion and detoxification function are important for parasitic progress unique to ticks, particularly feeding of blood meal, haemoglobin digestion, haem detoxification and prolonged off-host survival. We compared four dietary detoxification-related gene families in three tick and mite species. ABC, P450 and CCE families showed large expansions in the longhorned tick genome (Table 2). Expansions of ABCs occurred in the ABCA and ABCC subfamilies, which includes five and three large (≥ 5 orthologs) clusters on the phylogenetic tree (Fig. 2a). ABCA transporters function in lipid transport and resistance in insects [50–52]. ABCC transporters, also known as multidrug resistance proteins (MRPs), are known to be involved in ion transport, signal transduction, and toxin secretion [53]. Major P450 expansions of two tick species were discovered in clan 3 and clan 4 (Table 2), which had three clan 3 and three clan 4 large clusters on the tree for *H. longicornis* (Fig. 2b). P450 clan 3 and clan 4 may be linked to xenobiotic metabolism, insecticide resistance, odorant or pheromone metabolism [54]. GST expansions of two ticks mainly occurred in Mu class (Figure S2), which may activate in drug metabolism, particularly in the detoxification of reactive oxygen species (cyclised o-quinones) produced via oxidative metabolism of catecholamines [55, 56]. CCE in *H. longicornis* showed extreme expansions (Table 2) in neuro/developmental class although classifications of some members were unclear in Chelicerata [57]. These large expansions in the longhorned tick genome greatly enhance abilities in xenobiotic metabolism and insecticide resistance, and thus are considered to contribute to the parasitic adaptation.

Table 2
Comparison of digestion and detoxification gene families among three Acari species

Families	Clan/Group	<i>H. longicornis</i>	<i>I. scapularis</i>	<i>Tetranychus urticae</i>
P450	Clan 2	50	65	45
	Clan 3	99	100	10
	Clan 4	45	33	25
	Mitochondrial	10	7	5
	Total	204	205	85
GST	Delta	18	17	19
	Kappa	2	1	1
	Mu	17	23	10
	Sigma	1	1	0
	Zeta	1	7	1
	Omega	5	5	2
	Microsomal	5	1	0
	Total	49	55	33
CCE	Total	155	93	71
ABC	A	82	25	8
	B	6	5	2
	C	57	38	37
	D	4	2	2
	E	1	1	1
	F	3	3	4
	G	19	13	33
	Total	174	87	87

Conclusions

The genome research of *H. longicornis* will facilitate a number of applications including understanding its unique biology and enable the identification of unique tick genes and physiological processes such as blood-meal digestion, haemoglobin digestion, haem detoxification and prolonged off-host survival that could be exploited for tick control. Meanwhile, genome data could offers insights into the interactions of

ticks and pathogen and contribute to the designing novel strategies for blocking pathogen transmission and promote the development of tick vaccines.

Declarations

Ethical Approval and Consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

Raw transcriptomic data and genome assembly have been deposited at CNGBdb under the accessions CNX0199354 and CNA0014631, respectively, corresponding to the project CNP0001175 and sample CNS0254300. Genome annotations are available at the Figshare under the link [doi:10.6084/m9.figshare.12662036](https://doi.org/10.6084/m9.figshare.12662036).

Competing interests

The authors declare that they have no competing interests.

Funding

This research was supported by National Natural Sciences Foundation of China (No. 81871686); and Development Plan Project of Shandong province science and technology (No. 2017GSF221017).

Author Contributions

L.-R.Z. and Z.-Z designed the study. L.-R.Z. and Q.-Z collected the samples. L.-R.Z. and Z.-Z performed the analyses and wrote the paper. All authors edited and approved the final manuscript.

Acknowledgments

We gratefully acknowledge all researchers who submitted the related genome data of ticks and other 14 Chelicerata species.

References

1. Mediannikov O, Davoust B, Cabre O, Rolain JM, Raoult D. Bartonellae in animals and vectors in New Caledonia. *Comp Immunol Microbiol Infect Dis*. 2011;34(6):497–501. doi: 10.1016/j.cimid.2011.09.002.

2. Park SW, Song BG, Shin EH, Yun SM, Han MG, Park MY, et al. Prevalence of severe fever with thrombocytopenia syndrome virus in *Haemaphysalis longicornis* ticks in South Korea. *Ticks Tick Borne Dis.* 2014;5(6):975–977. doi: 10.1016/j.ttbdis.2014.07.020.
3. Hammer JF, Emery D, Bogema DR, Jenkins C. Detection of *Theileria orientalis* genotypes in *Haemaphysalis longicornis* ticks from southern Australia. *Parasit Vectors.* 2015;8(1):229.doi: 10.1186/s13071-015-0839-9.
4. Heath A. Biology, ecology and distribution of the tick, *Haemaphysalis longicornis* Neumann (Acari: Ixodidae) in New Zealand. *N Z Vet J.* 2016;64(1):10–20. doi: 10.1080/00480169.2015.1035769.
5. Tateno M, Sunahara A, Nakanishi N, Izawa M, Matsuo T, Setoguchi A, et al. Molecular survey of arthropod-borne pathogens in ticks obtained from Japanese wildcats. *Ticks Tick Borne Dis.* 2015;6(3):281–289. doi: 10.1016/j.ttbdis.2015.01.009.
6. Mahara F. Japanese spotted fever: report of 31 cases and review of the literature. *Emerg Infect Dis.* 1997;3:105–111. doi: 10.3201/eid0302.970203.
7. Jongejan F, Uilenberg G. The global importance of ticks. *Parasitology.* 2004;129:S3–S14. doi: 10.1017/s0031182004005967.
8. Yu XJ, Liang MF, Zhang SY, Liu Y, Li JD, Sun YL, Zhang LH, et al. Fever with thrombocytopenia associated with a novel bunyavirus in China. *N Engl J Med.* 2011;364(16):1523–1532. doi: 10.1056/NEJMoa1010095.
9. Kim KH, Yi J, Kim G, Choi SJ, Jun KI, Kim NH, et al. Severe fever with thrombocytopenia syndrome, South Korea, 2012. *Emerg Infect Dis.* 2013;19 (11):1892–1894. doi: 10.3201/eid2411.170756.
10. Matsuo T, Okura N, Kakuda H, Yano Y. Reproduction in a Metastriata tick, *Haemaphysalis longicornis*(Acari: Ixodidae). *J Acarol Soc Japan.* 2013;22:1–23.
11. Takahashi T, Maeda K, Suzuki T, Ishido A, Shigeoka T, Tominaga T, et al. The first identification and retrospective study of severe fever with thrombocytopenia syndrome in Japan. *J Infect Dis.* 2014;209:816–27. doi: 10.1093/infdis/jit603.
12. Hill CA, Wikel SK. The Ixodes scapularis Genome Project: an opportunity for advancing tick research. *Trends Parasitol.* 2005;21(4):151-153. doi: 10.1016/j.pt.2005.02.004.
13. Guerrero FD, Bendele KG, Ghaffari N, Guhlin J, Gedye KR, Lawrence KE, et al. The Pacific Biosciences de novo assembled genome dataset from a parthenogenetic New Zealand wild population of the longhorned tick, *Haemaphysalis longicornis* Neumann, 1901. *Data Brief.* 2019;27:104602. doi: 10.1016/j.dib.2019.104602.
14. Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, et al. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics.* 2017;33(14):2202–2204. doi: [10.1093/bioinformatics/btx153](https://doi.org/10.1093/bioinformatics/btx153)
15. Bushnell B. BBtools. 2014; Retrieved from <https://sourceforge.net/projects/bbmap/>
16. Guan et al. 2020. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics.* 36(9):2896–2898. doi: 10.1093/bioinformatics/btaa025.

17. Hu Df, McCarthy SA, Wood J, Howe K, Wang YD, Durbin R, et al. NextPolish: a fast and efficient genome polishing tool for long read assembly. *Bioinformatics*. 2020;36(7):2253–2255. doi: 10.1093/bioinformatics/btz891.
18. Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, et al. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol*. 2018;35:543–548. doi: 10.1093/molbev/msx319
19. Hubley R, Finn RD, Clements J, Eddy SR, Jones TA, Bao WD, et al. The Dfam database of repetitive DNA families. *Nucleic Acids Res*. 2016;44(D1):D81–89. doi: 10.1093/nar/gkv1272.
20. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA*. 2015;6:11. doi.org/10.1186/s13100-015-0041-9.
21. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. *PNAS*. 2020;117 (17):9451–9457. doi: 10.1073/pnas.1921046117.
22. Smit AFA, Hubley R, Green P. RepeatMasker Open-4.0. 2013-2015. Retrieved from <http://www.repeatmasker.org>
23. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*. 2013;29:2933–2935. doi: [org/10.1093/bioinformatics/btt509](https://doi.org/10.1093/bioinformatics/btt509).
24. Chan PP, Lowe TM. tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. *Methods Mol Biol*. 2019;1962:1–14. doi: [org/10.1007/978-1-4939-9173-0_1](https://doi.org/10.1007/978-1-4939-9173-0_1).
25. Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*. 2011;12:491. doi: [org/10.1186/1471-2105-12-491](https://doi.org/10.1186/1471-2105-12-491).
26. Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics*. 2016;32:767–769. doi: [org/10.1093/bioinformatics/btv661](https://doi.org/10.1093/bioinformatics/btv661).
27. Stanke M, Steinkamp R, Waack S, Morgenstern B. AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res*. 2004;32:W309–W312. doi: [org/10.1093/nar/gkh379](https://doi.org/10.1093/nar/gkh379).
28. Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M. Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Res*. 2005;33:6494–6506. doi: [org/10.1093/nar/gki937](https://doi.org/10.1093/nar/gki937).
29. Kriventseva EV, Kuznetsov D, Tegenfeldt F, Manni M, Dias R, Simão FA, et al. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res*. 2019;47(D1):D807-D811. doi: 10.1093/nar/gky1053.
30. 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. doi: 10.1038/s41587-019-0201-4.

31. Kovaka SK, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, Pertea M, et al. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* 2019;20(1):278. doi: 10.1186/s13059-019-1910-1.
32. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nature Methods.* 2015;12:59–60. doi: org/10.1038/nmeth.3176.
33. Huerta-Cepas JH, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol Biol Evol.* 2017;34(8):2115–2122. doi: 10.1093/molbev/msx148.
34. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 2019;47(D1):D309–D314. doi: 10.1093/nar/gky1085.
35. Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, et al. InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res.* 2017;45(D1):D190–199. doi: org/10.1093/nar/gkw1107.
36. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* 2019;47 (D1):D427–432. doi: org/10.1093/nar/gky995
37. Mi H, Muruganujan A, Ebert D, Huang X, Thoas PD. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* 2019;47:D419–D426. doi:org/10.1093/nar/gky1038.
38. Lewis TE, Sillitoe I, Dawson N, Lam SD, Clarke T, Lee D, et al. Gene3D: extensive prediction of globular domains in proteins. *Nucleic Acids Res.* 2018;46:D435-D439. doi: org/10.1093/nar/gkx1187.
39. Wilson D, Pethica R, Zhou Y, Talbot C, Vogel C, Madera M, et al. SUPERFAMILY—sophisticated comparative genomics, data mining, visualization and phylogeny. *Nucleic Acids Res.* 2009;37:D380–D386. doi: org/10.1093/nar/gkn762.
40. Marchler-Bauer A, Bo Y, Han LY, He J, Lanczycki CJ, Lu S, et al. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* 2017;45:D200–D203. doi:org/10.1093/nar/gkw1129.
41. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 2019;20 (1):238. doi: org/10.1186/s13059-019-1832-y
42. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Mol Biol Evol.* 2015;32(1):268–74. doi: 10.1093/molbev/msu300.
43. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol.* 2013;30(4):772–80. doi: 10.1093/molbev/mst010
44. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics.* 2009;25:1972–1973. doi: 10.1093/bioinformatics/btp348.

45. Kück P, Longo GC. FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Front Zool.* 2014;11(1):81. doi: 10.1186/s12983-014-0081-x.
46. Han MV, Thomas GW, Lugo-Martinez J, Hahn MW. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol Biol Evol.* 2013;30(8):1987–1997. doi: 10.1093/molbev/mst100.
47. Yu G, Wang L, Han Y, He Q. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012;16(5):284–287. doi: 10.1089/omi.2011.0118.
48. Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol.* 2017;35:1026–1028. doi: 10.1038/nbt.3988
49. Gulia-Nuss M, Nuss AB, Meyer JM, Sonenshine DE, Roe RM, Waterhouse RM, et al. Genomic insights into the *Ixodes scapularis* tick vector of Lyme disease. *Nat Commun.* 2016;7:10507. doi: 10.1038/ncomms10507.
50. Zhao ZW, Zera AJ. Differential lipid biosynthesis underlies a tradeoff between reproduction and flight capability in a wing-polymorphic cricket. *Proc Natl Acad Sci U S A.* 2002;99(26):16829–16834. doi: 10.1073/pnas.262533999.
51. Zera AJ, Zhao ZW. Effect of a juvenile hormone analogue on lipid metabolism in a wing-polymorphic cricket: implications for the endocrine-biochemical bases of life-history trade-offs. *Physiol Biochem Zool.* 2004;77(2):255–266. doi: 10.1086/383500.
52. Tay WT, Mahon RJ, Heckel DG, Walsh TK, Downes S, James WJ, et al. Insect Resistance to *Bacillus thuringiensis* Toxin Cry2Ab Is Conferred by Mutations in an ABC Transporter Subfamily A Protein. *PLoS Genet.* 2015;11(11):e1005534. doi: 10.1371/journal.pgen.1005534.
53. Dean M, Hamon Y, Chimini G. The human ATP-binding cassette (ABC) transporter superfamily. *J Lipid Res.* 2001;42(7):1007–17.
54. Feyereisen R. Insect cytochrome P450. In Gilbert LI, Iatrou K, Gill SS, editor. *Comprehensive molecular insect science. Vol 4, Biochemistry and molecular biology.* Oxford: Elsevier. 2005;p. 1–77.
55. Takahashi Y, Campbell EA, Hirata Y, Takayama T, Listowsky I. A basis for differentiating among the multiple human Mu-glutathione S-transferases and molecular cloning of brain GSTM5. *J Biol Chem.* 1993;268:8893–8898.
56. Hansson LO, Bolton-Grob R, Massoud T, Mannervik B. Evolution of differential substrate specificities in Mu class glutathione transferases probed by DNA shuffling. *J Mol Biol.* 1999;287:265–276. doi: 10.1006/jmbi.1999.2607.
57. Grbić M, Van Leeuwen T, Clark RM, Rombauts S, Rouzé P, Grbić V, et al. The genome of *Tetranychus urticae* reveals herbivorous pest adaptations. *Nature.* 2011;479(7374):487–492. doi: 10.1038/nature10640.

Figures

Figure 1

Phylogeny, orthologs, gene family evolution and function enrichment. (a) dating tree with node values representing the number of expanded, contracted and rapidly evolving families. '1:1:1' represents single-copy genes, 'N:N:N' as multi-copy genes, 'Parasitiformes' as common orthologs unique to Parasitiformes, 'Other' as unclassified orthologs, 'Unassigned' as orthologs which cannot be assigned into any known gene families. (b) the top twenty significantly expanded families. (c) GO enrichment of significantly

expanded gene families with the top 20 categories shown; (d) KEGG enrichment of significantly expanded gene families with the top 20 categories shown.

Figure 2

Phylogenetic tree and expansion clusters of ABCs (a) and P450s (b) in the longhorned tick. Red, *Haemaphysalis longicornis*; green, *Ixodes scapularis*; blue, *Tetranychus urticae*.

Supplementary Files

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

- [supplementarytables.xlsx](#)
- [Fig.S2.pdf](#)
- [Fig.S1.pdf](#)