The genome of horseshoe crab Tachypleus tridentatus reveals its evolutionary scenario and well-developed innate immunity

Yan Zhou
Fudan University
Yuan Liang
Fudan University
Qing Yan
Chinese National Human Genome Center at Shanghai
Liang Zhang
Chinese National Human Genome Center at Shanghai
Dianbao Chen
Minjiang University
Lingwei Ruan
Ministry of Natural Resources of the People's Republic of China
Yuan Kong
Minjiang University
Hong Shi
Ministry of Natural Resources of the People's Republic of China
Mingliang Chen
Ministry of Natural Resources of the People's Republic of China
Jianming Chen (chenjianming@tio.org.cn)
https://orcid.org/0000-0003-0400-3092

Research article

Keywords: Tachypleus tridentatus, immunity, pattern recognition receptor, coagulation

Posted Date: May 1st, 2019

DOI: https://doi.org/10.21203/rs.2.9427/v1

Version of Record: A version of this preprint was published on February 10th, 2020. See the published version at https://doi.org/10.1186/s12864-020-6488-1.

Abstract

Background Horseshoe crabs are ancient marine arthropods with a long evolutionary history. Results Here, we describe the 2.06 Gbp genome assembly of Tachypleus tridentatus with predicted 24,222 protein-coding genes. Comparative genomics show that T. tridentatus and the Atlantic horseshoe crab Limulus polyphemus had the most orthologs shared in only these two species, including genes involved in immune related Jak/STAT signaling pathway. Divergence time dating results show that the last common ancestor of Asian horseshoe crabs (including T. tridentatus) and L. polyphemus appeared at about 130 Mya (121-141) and the split of two Asian horseshoe crabs was dated to about 63 Mya (57-69). Hox gene analysis suggests two clusters in both horseshoe crab assemblies. Surprisingly, selective analysis of immune related gene families revealed high expansion of conservatively presented pattern recognition receptors. Genes involved in IMD and Jak/STAT signaling transduction pathways also exhibited a certain degree of expansion in both genomes. Intact coagulation cascade related genes were presented in T. tridentatus genome with a higher number of coagulation factor genes. Moreover, most antibacterial peptides reported have been identified in T. tridentatus with their potentially effective antimicrobial sites. Conclusions The well-developed innate immunity of T.tridentatus may affect the quality of the adaptive properties with regard to complicated marine environments.

Background

Horseshoe crabs are marine arthropods, representing an ancient family with an evolutionary history record extending back approximately 450 million years (1). For their static morphology and their position in the arthropod family tree, they are therefore labeled “living fossils” for a long time (2). *Tachypleus tridentatus* (Leach, 1819), one of extant horseshoe crab species, is mainly distributed from coastal Southeast China to western Japan and in a few islands in Southeast Asia (3). Similar to other invertebrates, *T. tridentatus* lacks an acquired immune system, relying entirely on its innate immune system and has developed multiple defense systems. These defense systems include hemolymph coagulation, phenoloxidase activation, cell agglutination, release of antibacterial substances, active oxygen formation and phagocytosis (4-8), which together constitute the innate immune system of *T. tridentatus*. Previous studies have been devoted to investigate important signaling pathways and gene families from other arthropods such as insects, crustaceans and myriapods, revealing extensive conservation and functional diversity in innate immune components across arthropods (9, 10). A genome survey of horseshoe crabs, including *T. tridentatus,* has been denoted before (11), focusing on homeobox gene components, recovering ~ 1.5 Gbp of the assembly. Currently, the immune molecular mechanisms under how horseshoe crabs achieve distinguishing "self" and "non-self" antigenic epitopes, also known as pathogen-associated molecular patterns (PAMPs) still remain to be explored.

Here, we present the analysis of genome sequence of *T. tridentatus*, together with comparative genomic and divergence time analyses of other available Chelicerata genomes to date, including previously released *Limulus polyphemus* assembly (12). Particular attention was paid to gene families related to assessing the genomic and phenotypic changes of horseshoe crabs, and exploring immune signaling

pathways, coagulation factors and antimicrobial peptides which may contribute to their robust and effective innate immunity for self-defense in marine environments with enormous number of invading pathogens and may have an important implication for this species continuing to this day.

Results

General genome features

The genomic DNA isolated from *T. tridentatus* were sequenced to 124× coverage and assembled into 2.06 Gb of genome. The final draft assembly consists of 143,932 scaffolds with N50 scaffold size of 165 kb, among which the longest scaffold size is 5.28 Mb and the shortest is 1 kb. The GC content of the genome is 32.03% (Table 1). A total of 24,222 protein-coding genes were conservatively predicted in the *T. tridentatus* genome. The average exon length and intron length predicted for the assembly are 333 bp and 3,792 bp. A total of 88.25% of the predicted genes were assigned and annotated by comparing to the NCBI non-redundant database (NR), KEGG database and Interpro database.

The screening of repeat content from RepeatMasker analysis based on similarity alignments identified 20.29 Mb and 20.33 Mb in genomes of *T. tridentatus* and *L. polyphemus*, representing 0.99 % and 1.11 % of the genome size, respectively. Most of the identified repeat sequences were simple repeats (0.77 % in *T. tridentatus* and 0.94 % in *L. polyphemus*). This suggests that the repeat sequences from horseshoe crabs have a great difference from the existing homologous repeats, indicating that the evolutionary distance of horseshoe crabs is far from the current arthropods. For the estimate of repeat sequences which are more difficult to detect in both assemblies, RepeatModeler was used to predict possibly existing but unidentified repeats, and then RepeatMasker was employed to mask these sequences. On the basis of this, repeat elements totalled 34.83 % in *T. tridentatus* and 34.22 % in *L. polyphemus*, including a proportion of 13.26 % and 10.30 % of transposable elements in *T. tridentatus* and *L. polyphemus*. Long interspersed elements (LINEs) (6.21 % in *T. tridentatus* and 3.43 % in *L. polyphemus*), LTR elements (1.72 % in *T. tridentatus* and 3.72 % in *L. polyphemus*), and DNA elements (5.33 % in *T. tridentatus* and 3.15 % in *L. polyphemus*) were detected in both assemblies. The detailed categorization of remaining unclassified repeats may require more time and experts who are professional in the field of repeat sequences.

Assembly validation

The completeness of the *T. tridentatus* genome assembly was assessed using the transcriptome data of general sample of *T. tridentatus*. It was found that 99.04% of the transcriptome contigs were aligned to the assembly scaffolds, with an e-value cut-off of e-30. To further confirm the completeness of the predicted genes, the commonly used genome assembly validation pipeline BUSCO gene mapping method with 1066 BUSCO Arthropoda gene sets were utilized. The predicted genes of *T. tridentatus* reveals 98.7% conserved proteins of homologous species with 1,052 BUSCOs (76.6% complete single-copy BUSCOs, 10.8% complete duplicated BUSCOs and 11.3% fragmented BUSCOs). Only 1.3% of the benchmarked universal single-copy orthologous groups of arthropod genes were missing in the assembly. This

suggested a remarkable completeness of genome assembly and predicted gene repertoire of *T. tridentatus*, containing most of the evolutionarily conserved core genes.

Phylogeny analysis and divergence time dating

Two assemblies of *L. polyphemus* have been documented before (12, 13), one of which was selected to perform comparative genomics according to a relatively higher assembly level. The OrthMCL calculation by merging the genes of *T. tridentatus* and *L. polyphemus* resulted in a total of 12,116 orthologous groups. Of these, 10,968 orthologs contained genes found in both two horseshoe crab genomes with 15,905 *T. tridentatus* and 20,390 *L. polyphemus* genes included, plus almost 6,880 shared genes were single-copy. Functional enrichment analysis showed that these shared genes involved in several important pathways (P-value < 0.05), such as metabolic pathways (pyruvate, glycerolipid, amino sugar, nucleotide sugar and so on), ribosome biogenesis and DNA replication. The analysis also identified 1,418 protein-coding genes presented only in *T. tridentatus*. There were 1,956 genes found specific to *L. polyphemus*.

In order to place the *T. tridentatus* within the most current understanding of the evolution of Chelicerata species, comparative genomic analysis was performed using coding gene repertoire of recently published Arthropod Chelicerata species close to *T.tridentatus* (Figure 1b). Comparing orthologous results of *T.tridentatus* and 12 other Chelicerata and one Myriapoda outgroup revealed in 14,479 orthologous groups containing genes in more than or equal to two different species, among which 1,993 shared groups were identified common to the 14 species sampled, with 111 single-copy orthologs. Those single-copy genes were enriched in KEGG pathways such as ribosome, oxidative phosphorylation, proteasome, metabolic pathways, and carbon metabolism. In addition, *T. tridentatus* and *L. polyphemus* had the most orthologs shared in only two species (2,720 (22.2%) and 2,648 (21.5%)). Pathway enrichment of these genes showed significant enrichment (P-value < 0.01) in neuroactive ligand-receptor interaction, FoxO signaling pathway and AGE-RAGE signaling pathway in diabetic complications. The latter two KEGG pathways include the important signaling pathway Jak/STAT related to innate immunity in arthropods. Concerning species specific gene families, 1,124 genes were found unique to *T. tridentatus* compared with other Chelicerata arthropods. *C. sculpturatus* had the most (7,328) expanded species unique genes, followed with 6,247 *N. clavipes* specific gene families. In contrast, only 161 genes were unique to *T. mercedesae*. The numbers of species specific genes of *T. tridentatus* and *L. polyphemus* showed in the middle position with 1,124 and 857, respectively.

Considerable progress of the phylogenetic relationship of Arthropoda has been made, particularly in the last five years, by employing transcriptomic and several existing genomic data. The recovery of Euchelicerata (Xiphosura plus Arachnida) with 62 single-copy nuclear protein-coding genes from 75 arthropod species only favoured Arachnida (scorpions, spiders, mites and ticks) with a relatively low bootstrap support (14). Conflicts of phylogeny remained that the monophyly of Arachnida was influenced by the evolving rates of genes in the dataset, including 3,644 orthologous groups of Chelicerata taxa into analysis (15). The phylogenetic tree in our analysis was rooted using the centipede *S. maritima* as the

outgroup (Figure 1a). Strong bootstrap support was obtained for spider, mite and tick clade forming a monophyletic group, respectively. *T. tridentatus* and *L. polyphemus* were grouped together, forming the Xiphosura clade, which might be explained by the more sharing orthologous groups in the two lineages.

Available mitochondrial coding sequences of Chelicerata species and the corresponding mitochondrial sequences recovered from the assembly of the genome and transcriptom of *T. tridentatus* were used to investigate the divergence time for horseshoe crabs with their close species. We set the separation time for common ancestor of Chelicerata species according to the fossil evidence reported before (16), the resulting divergence time estimate result for the 7 Chelicerata species showed that the last common ancestor of Asian horseshoe crabs (including *T. tridentatus*) and *L. polyphemus* was dated to 130 Mya (121-141) and the split of two Asian horseshoe crabs *T. tridentatus* and *C. rotundicauda* was dated to 63 Mya (57-69) (Figure 2), while the internal split of the *T. tridentatus* across the southern costal China to Korean Peninsula was dated to 12 Mya (11–14). Both species tree and time tree suggested that horseshoe crabs are closely related to scorpions and the split of scorpions from horseshoe crabs were dated to 440 Mya (412-468).

Two Hox gene clusters

The Hox genes, one highly conservative subclass of homeobox super-class genes, which has been extensively investigated, are usually distributed in clusters (17-20). This was demonstrated in a number of studies that Hox genes play an essential role in determination of anterior-posterior axis, organ and nervous system formation of the embryonic development (21-24) . Analysis of Hox gene family showed that *T. tridentatus* assembly contained 46 Hox genes, and 43 Hox genes are identified in *L. polyphemus* (Table S1). This is the most complete set of Hox genes we obtained based on homeobox domains from two available horseshoe crab assemblies. It was also found that most Hox genes were presented at least two representatives in both genomes, which was consistent with previous whole-genome duplication study in horseshoe crabs (11).

We further examined positions of identified Hox genes in two genomes and found two clusters of Hox1 and Hox4 adjacent distribution in *T. tridentatus* assembly. As to *L. polyphemus*, there were also one Hox clusters of adjacent Hox1 and Hox4 genes and one additional Hox1, Hox2 and Hox3 clusters. Other clusters such as adjacent Hox2 and Hox3 clusters, longer clusters of Hox4, Hox7, Ubx, AbdA and AbdB genes found in two assemblies could probably had connection to each two clusters mentioned above. Combining these, based on Hox gene positions on the assemblies, our analysis indicates a good agreement with previous study and suggests that there are highly possible two Hox gene clusters presented in horseshoe crabs if Hox genes were linearly arranged in clusters along the anterior-posterior axis like the ancestral arthropod Drosophila (25).

Expansion of crucial gene families of the innate immune signaling pathways in *T. tridentatus* and *L. polyphemus*

Horseshoe crabs are an independent ancient group with distributions in Southeast Asia and North America. There must be inherent reasons for their survival against complicated marine environmental change. Some clews could be found from the comparison of previously determined *L. polyphemus* and our presented *T. tridentatus* genomes. Among them, the immune system undoubtedly provides an important guarantee. However, the immune molecular mechanisms that horseshoe crabs utilize to specifically respond to potential pathogens were still unclear. Therefore, we manually searched the *T. tridentatus* and *L. polyphemus* genomes and *T. tridentatus* transcriptome for homologues of essential immune signaling related genes. Genes that are not found in the *T. tridentatus* genome are supplemented with the transcriptome data to restore as accurate numbers as possible. Although the draft genome has certain incompletely, the gene counts error should be small enough in respect of the predicted genes according to the BUSCO evaluation of the 98.7% conserved proteins of homologous species.

Immune-related genes can be broadly classified into pattern recognition receptors, signaling transduction pathways and effectors. Although *T. tridentatus* and *L. polyphemus* only have innate immune pathways, their receptor genes show expansion in large amounts, and key genes in the signaling transduction pathways also exhibit a certain degree of expansion (Figure 3). Pattern recognition receptors (PRRs) could recognize foreign pathogens and play a leading role in defensing PAMPs (26). Sequence homologues for two families of PRRs were extensively presented in both genomes with functional domains, including 42 fibrinogen-related proteins (FREPs) and 117 down syndrome cell adhesion molecules (DSCAMs) in *T. tridentatus*. The two abundantly expressed PRRs in *T. tridentatus* transcriptome suggest an effective ability to recognize a broad range of pathogens, which may be inductive to cope with a great diversity of invading microorganisms in marine environment. Pathogen recognition would further lead to activation of the signaling transduction and amplification of immune response, thus producing immune factors that are resistant to microbial activity. Major gene families that play important roles in signaling transduction of innate immunity in arthropods (such as Toll, IMD, JAK/STAT, and JNK pathways) (27-29) are present in genomes of *T. tridentatus* and *L. polyphemus*. Dredd genes of IMD and JNK genes of Jak/STAT pathway in *T. tridentatus* and *L. polyphemus* exhibit a certain degree of expansion. The phylogeny analysis for shared genes of horseshoe crabs with their evolutionary closely related species found that horseshoe crabs have the most unique gene orthologs shared in only two species, including the above-mentioned expanded gene families, predicting that the innate immunity of horseshoe crabs has strong signal transducting ability. For Dredds required for IMD signaling pathway, the phylogeny tree shows one branch including 7 corresponding genes identified in two horseshoe crabs and 1 gene in *C. sculpturatus*. Another branch encompasses 2 genes in *P. tepidariorum* (Figure 4a). For JNKs related to JAK/STAT pathway in the phylogeny, there were three branches consisting of a pair of corresponding genes identified in genomes of *T. tridentatus* and *L. polyphemus*, and one branch formed by a pair of genes in *C. sculpturatus* and *S. mimosarum* (Fig 4b).

Antimicrobial peptide diversity in *T. tridentatus*

Horseshoe crabs have developed a unique and effective host defense system and coagulation system to protect against invading pathogens. Various immune response factors have been isolated and purified

from the hemolymph of horseshoe crabs, which are involved in their host defense action (4, 30, 31) . Of these, an important component of the innate immune self defense system of these marine animals are antimicrobial peptides, which act as the important innate immune effectors (32). We searched the *T. tridentatus* genome for antimicrobial peptide genes and successfully identified most of antibacterial peptides which have been reported, including one anti-LPS, two tachyplesins and two big defensins (Fig 3).

The anti-LPS gene found in *T. tridentatus* genome contains the antimicrobial peptide (AMP) region between G23 to R83 with two conserved cysteine residues, including hydrophobic NH2-terminal and cationic residues clustered in its disulfide loop, which were supposed to act as an affinity site in combination with LPS (33, 34). Tachyplesin family are cationic peptides in constitutive expression comprised of 17-18 amino acids, which strongly inhibit the growth of both gram-negative and -positive bacteria, such as pathogenic microorganisms of marine bivalves *Bonamia ostreae*, *Perkinsus marinus* and Vibrio P1, and can also have strong inhibitory effects on the growth of fungi (32, 35). In this study, we succeeded in identifying two precursors of tachyplesins in *T. tridentatus*, each of which consists of 77 amino acids encompassing a putative signal peptide sequence, a mature tachyplesin peptide sequence, a C-terminal arginine followed by the amidation signal residues Gly-Lys-Arg and a peptide of 22 aa length in the C-terminal portion (36). In addition to this, we identified two protein precursors of big defensins in *T. tridentatus* genome, one of which is in a longer length of 118 amino acids, containing a hydrophobic N-terminal half and a cationic C-terminal half, which may be closely related to its biological activity for broad antimicrobial properties (37).

Intact coagulation cascades in *T. tridentatus*

Rapid coagulation of the hemolymph is of great significance for both host defense and hemostasis in horseshoe crabs (6, 38). The coagulation cascade of horseshoe crabs consists of four coagulation factors and one coagulogen. The coagulation factors include factor C, factor B, factor G and proclotting enzyme, all of which belong to the serine proteinase family (4, 30, 31) . Clotting factor C is an endotoxin-sensitive serine proteinase in blood cells of horseshoe crabs that can be activated by endotoxin to initiate the blood coagulation cascade system (39, 40). After this, the proclotting enzyme can be activated by activated factor B (41, 42). The clotting enzyme further activates coagulogen to form an insoluble coagulation protein gel (43, 44).

We found that *T. tridentatus* and *L. polyphemus* have all known coagulation related genes while other related species lack a part of the coagulation pathway (Table 2), indicating a wider diversity in coagulation factors and a relatively intact coagulation cascade presented in horseshoe crabs. Factor G, a heterodimer that is specifically activated by fungal cell wall component 1,3-β-D-glucan, is a special serine protease precursor and provides another starting point for the clotting reaction (45, 46). We identified 4 factor G sequences in our *T. tridentatus* genome and transcriptome assembly, including genes coding the alpha subunit and the beta subunit, respectively. However, we failed to identify any clotting factor G homologues in other Chelicerata species.

Discussion

The complete genome of *T. tridentatus*, the second sequenced horseshoe crab genome could provide Chelicerata clade with a high quality of publicly available sequence, and would have more hopeful prospects in eliminating the uncertainty in evolution relationship of Chelicerata. The genome size and predicted gene number of *T. tridentatus* assembly are close to the previous sequenced *L. polyphemus*, with a remarkably completeness validated by 98.7% BUSCOs and 99.04% transcriptome coverage.

Considering that previous phylogenetic studies only used transcriptomic data with multiple representation of one gene or obtained low bootstrap support for Arachnida, our phylogenetic tree using 111 single-copy orthologous groups of 13 Chelicerata species and 1 outgroup was unsuccessful to support Euchelicerata (Xiphosura plus Arachnida). Even so, relatively wider sampling range of species and more comprehensive information of previous studies showed that further experiments would be helpful for the exploration of Chelicerata taxa. We further investigated divergence time using mitochondrial coding sequences of 7 Chelicerata species and our analysis suggest that the diversification of the *Limulidae* and *Tachypleus tridentatus* lineages was congruent around 121-141 Mya and the two Asian horseshoe crabs *T. tridentatus* and *C. rotundicauda* lineages was congruent around 57-69 Mya. According to the continental drift theory, before the Triassic virtually all continents were joined to form the supercontinent Pangea, in the Triassic the breakup of Pangea commenced (47). Around 170–120 million years ago (MYA), Pangea broke up into two supercontinents: Laurasia and Gondwana (48), the subsequent lineage divergence within the reptiles (49), amphibians (50, 51), mammals (52) and even plants (53) matches the separation and fragmentation of Laurasia and Gondwana. The Laurasia was fragmenting at mid-Mesozoic (54), but until late-Cretaceousthe the Eurasia and North American plate was still joined together (55). The ancestor of horseshoe crabs (or their progenitor species) was reported probably originated in the Mesozoic waters of Europe (56, 57). After the final breakup of Eurasia and North American plate, the European land mass was formed as the shallow seas disappeared, the ancestors of the horseshoe crab migrated, one group to the west is found along the east coast of North America from Maine through south Florida and the Gulf of Mexico to the Yucatan Peninsula evolved into American species *L. polyphemus*. One group to the east through the Tethys is found along Asia from Japan to India evolved into: *T. tridentatus*, *T. gigas*, and *C. rotundicauda*. There is evidence show that India–Asia collision was underway in northern Pakistan at ca. 56–55 Mya (58). The diversity of Asian horseshoe crabs *T. tridentatus* and *C. rotundicauda* may be related to India–Asia collision.

The research about Hox genes suggest duplication of clusters from one common ancestor (59). Analysis of Hox gene family of both horseshoe crabs found that two Hox clusters distribution if Hox genes were linearly arranged in clusters along the anterior-posterior axis like the ancestral arthropod *Drosophila* (25). More intact Hox clusters in both assemblies may be recovered in both horseshoe crabs with more complete assemblies.

The native immune system of hemolymph in *T. tridentatus* has drawn much attention. As a member of arthropods, horseshoe crabs lack immunoglobulin-like substances in their hemolymph, and the immune

function is mainly performed by their blood cells called hemocytes or amebocytes, which undergo adhesion and deformation by protruding the cytoplasm and degranulation when stimulated by environmental pathogens (4, 30). Granular blood cells of *T. tridentatus* are filled with large and small granules storing most of the immune proteins or polypeptides. We searched the two horseshoe crab genomes to investigate the molecular basis of immune signaling pathways which have effect on specificity of foreign pathogen recognition and performance of releasing immune response substances. From the results we have obtained, two horseshoe crabs showed a high sequence homology in most immune signaling related gene families which have been studied in other arthropods before. Homologs of FREP and DSCAM families playing important role in pattern recognition receptors were extensively presented in two horseshoe crab genomes with corresponding functional domains. The two abundantly expressed PRRs in *T. tridentatus* transcriptome suggest an effective ability to recognize a broad range of pathogens, which may be inductive to cope with a great diversity of invading microorganisms in marine environment. In addition, two gene families of IMD and JAK/STAT pathway members exhibit a certain degree of expansion in *T. tridentatus* and *L. polyphemus*, predicting that the innate immunity of horseshoe crabs has strong signal reception and signal transduction ability.

Antimicrobial peptides refer to polypeptides having antibacterial activity, which are evolutionarily conservative and exist in all eukaryotes, forming an important part of the innate immune system. There is increasing evidence that antimicrobial peptides in horseshoe crabs not only possess broad-spectrum antimicrobial capabilities (35), but also have a strong resistibility to enveloped-virus, parasites and tumor cells (60-63) . We identified most antimicrobial peptides isolated in previous studies in our *T.tridentatus* assembly, including anti-LPS, tachyplesins, and big defensins. However, in the *L. polyphemus* genome, we failed to find any of the tachyplesin family genes, probably because these antimicrobial peptides are usually shorter and have a high degree of species specificity.

Hemolymph coagulation is a major component of the innate immune system in horseshoe crabs. Two biosensors for pathogen recognizing, factor C and factor G, are activated by the corresponding activator to initiate the coagulation cascade, leading to the formation of coagulin fibrils. This pathogen-activated coagulation response is not only important for hemostasis, but also critical for host defense (64). We found all four coagulation factors with a relatively high number of genes in *T. tridentatus* assembly, indicating a complete coagulation cascade in horseshoe crabs. Other Chelicerata species show a lack of factor G and coagulogen. It might be because of a limited annotation of the draft genome assemblies or there might exist other special biosensors in the process of their antifungal recognition. Although the more direct internal cause for the long-term evolutionary conservation and success of horseshoe crabs with abilities to maintain morphological stability in a fluctuating marine environment may be the combination of wide feeding spectrum, substantial saline tolerance and insensitive to temperature of horseshoe crabs (65, 66), there is a presumption about the features of hemolymph of horseshoe crabs may improve the quality of their adaptive strategy and increase their population survival rate (67, 68). Moreover, counterparts of the antimicrobial peptides in the hemolymph of horseshoe crabs have also been identified in other evolutionarily conservative representatives of animal populations (69-71) .

Conclusions

Our draft assembly of *T. tridentatus* was sequenced into 2.06Gb of genome, consists of 143,932 scaffolds with N50 scaffold size of 165 kb. A total of 24,222 protein-coding genes were conservatively predicted present in the *T. tridentatus* genome, revealing 98.7% completeness with 1,052 BUSCOs.

Further analysis of the *T. tridentatus* genome included phylogeny analysis and divergence time dating using newly published Chelicerata species, and selective analysis of Hox genes, innate immunity-related genes, cogulation factors and antimicrobial peptides. It was found that two horseshoe crabs, *T.tridentatus* and *L. Polyphemus* had the most orthologs shared in only two species, which were enriched in immune related Jak/STAT signaling pathway. Furthermore, at least two clusters of Hox genes were predicted in both assembly. As for innate immunity gene investigation, conservatively presented gene families for pattern recognition receptors FREPs and Dscams, signaling transduction pathway Dredd and JNK genes involved in IMD and JAK/STAT exhibit a certain degree of expansion in both horseshoe crab gemomes. Besides, according to our analysis, intact cogulation cascade related genes were identified in *T. tridentatus* genome with a higher number total gene counts. Beyond that, all antibacterial pepetides reported in previous studies have been identified in *T. tridentatus* genome with effectively antimicrobial sites. Apart from the above, there may remain other aspects which may be greatly conducive to the adaptive advantages in marine environments through the long evolutionary history of horseshoe crabs, which need to be further studied and established in future studies.

Methods

T. tridentatus specimens and DNA extraction

The *T. tridentatus* samples were collected from the aquatic market in Xiamen, Fujian province, China, and then subjected to separating the leg muscle tissue. The muscle tissue samples were preserved in liquid nitrogen then stored in -80℃. Total genomic DNA was extracted using E.Z.N.A.®Insect DNA Extraction Kit.

Genome sequencing and assembly

The Illumina TruSeq Nano DNA Library Prep Kit was used to construct the 400 bp and 800 bp paired-end sequencing libraries. The long mate-pair libraries featuring inserts with size of 3-kb were prepared using CHGC kit. Nextera MP Sample Prep Kit was used to build long mate-pair libraries with 8- and 12-kb insert sizes. Whole-genome sequencing was performed on the Illumina HiSeq 2500 and HiSeq X Ten platform generating 250 bp and 150 bp paired-end reads. Five libraries of nominal insert sizes 400 bp, 800 bp, 3-kb, 8- kb and 12- kb were sequenced at expected genome coverages of 60×, 30×, 7×, 7× and 20×, respectively. Reads were assembled into contigs using Velvet (72) by cutting the short reads into k-mer and establishing the de Bruijn table to correct and complete the contigs, which were further scaffolded and gap-filled using SSPACE-STANDARD (73) with long mate-pair reads (8- and 12-kb).

Automated annotation

The automatic gene prediction of *T. tridentatus* genome assembly were established using AUGUSTUS (Version 3.3) (74). The predicted genes were annotated by comparing to the NCBI non-redundant database (NR)KEGG database and Interpro database (http://www.ebi.ac.uk/interpro/) with an E-value threshold of 10−5. The three annotation results were combined as the annotation of those predicted genes. The benchmarking sets of universal single-copy orthologs (BUSCOs) (75) were used to assess the completeness of the predicted genes with 1066 Arthropoda datasets. The repeat content of *T. tridentatus* and *L. polyphemus* were first analyzed using RepeatMasker (version open-4.0.5) (A.F.A. Smit, R. Hubley & P. Green RepeatMasker at http://repeatmasker.org) with merostomata as query species and running with rmblastn (version 2.2.27+), RepBase (Update 20140131) and RM database (version 20140131). RepeatModeler (version open-1.0.11) was used to build repeat database which was further masked by RepeatMasker (version -open-4-0-8).

Transcriptome analysis

RNA-seq raw data of *T. tridentatus* were downloaded from NCBI SRA database (accession number SRX330201) (76). De novo assembly of the transcriptome was performed with Trinity (77) by default parameters. Protein coding sequences and longest ORFs of transcripts were predicted via Transdecoder.

Orthology and phylogeny analysis

The published genome assembly, coding sequences and protein sequences of the Atlantic horseshoe crab, *L. polyphemus*, submitted by Washington University, were downloaded from NCBI with RefSeq ID 2304488, accession GCF_000517525.1. Non-redundant protein sequences of *L. polyphemus* were selected by sorting the scaffold position of proteins and filtering out overlapped proteins. Protein sequences of other 11 Chelicerata species, including three Araneae *Nephila clavipes*, *Parasteatoda tepidariorum*, *Stegodyphus mimosarum*, seven Acari *Varroa jacobsoni*, *Tropilaelaps mercedesae*, *Sarcoptes scabiei*, *Metaseiulus occidentalis*, *Tetranychus urticae*, *Varroa destructor*, *Ixodes scapularis*, and one scorpione *Centruroides sculpturatus* were downloaded from GenBank database. Additional outgroup protein sequences of the centipede *Strigamia maritima* were downloaded from UniProt database. OrthoMCL (78) was used to perform orthologous gene family clustering and provide calculation results of the numbers of genes in each orthologous group. Selected single copy orthologs were subjected to multiple sequence alignments using ClustalW v2.0.12 and Maximum Likelihood phylogenetic trees were built by RAxML v8.2.12 (79) using the PROTGAMMAWAG model and 1000 bootstraps, with centipede as an out-group. The web server KOBAS 3.0 (80-82) was used for functional gene set enrichment of shared and species specific genes of *L. polyphemus* and *T. tridentatus* and the single-copy gene set of the 13 Chelicerata species and the centipede, with KEGG pathway and PANTHER databases.

Divergence time estimates

Intact 13 mitochondrial coding sequences of 7 species were downloaded from NCBI gene database and used to build the time tree, containing three horseshoe crabs *Tachypleus tridentatus*, *Carcinoscorpius rotundicauda* and *Limulus polyphemus*, one scorpione *Mesobuthus gibbosus* and three Araneae *Argiope bruennichi*, *Wadicosa fidelis* and *Tetragnatha nitens*. The 13 mitochondrial coding sequences of *T. tridentatus* collected in Korea were used as query for searching identical genes in our *T. tridentatus* genome and transcriptom assembly using blastn. Alignment sequences with identities of at least 97% were considered to be the counterparts of mitochondrial coding sequences in *T. tridentatus* identified in this study. Multiple sequence alignments of concatenated mitochondrial coding sequences of 7 speices were processed using ClustalW v2.0.12. BEAUti v2.5.1 (83-85) was used to generate the BEAST input XML files then the tree was dated using BEAST v.2.5.1. HKY substitution model with empirical frequencies and a strict clock model were used. Fossil information of all Chelicerata species were used to calibrate the tree with an normal distribution of 530 Mya and standard deviation 5 Mya (16). Chain length of 6,000,000 generations were run with sampling every 1,000 generations. Software Tracer was used for analyzing the output of BEAST log file. TreeAnnotator and FigTree were used for tree production and tree visualization.

Identification of Hox genes, immune pathway genes and coagulation factors

Protein sequences of Hox genes, essential immune signaling related genes and coagulation factors of species closely related to horseshoe crabs were downloaded from NCBI protein database and used as query sequences. Blastp was then used with e-value of 10-15 to search for homologues using genomes of *T.tridentatus* and *L. polyphemus* as the databases. Tblastn was used with e-value of 10-15 to search corresponding transcripts in the transcriptom of *T. tridentatus*. Transcripts found in the transcriptome were further compared to the genome of *T. tridentatus* by blastn with e-value of 10-5 to complement identified genes and to reduce the data omission in the genome. Putative genes were selected based on positive scores and alignment length percentages defined by dividing alignment length by query length, and then filtrated according to their annotation in NCBI non-redundant database (NR) and Interpro database.

Dredd and JNK gene phylogenetic analysis

MEGA (version 7.0) (86) was used to construct Neighbor-Joining tree with bootstrap 1,000 using putative protein sequences in default parameters. Domains of all selected genes annotated as IPR001309 and IPR000719 in Interpro database were used for multiple alignments, respectively. Dredd of *Stegodyphus mimosarum* and JNK of *Parasteatoda tepidariorum* were chosen as outgroups for two phylogenetic trees, respectively.

Identification of putative antimicrobial peptides

Because antimicrobial peptides (AMPs) are commonly short in length, with a wide variety and have large differences in structure and function within species, tblastn were used to search potential AMPs using *T.tridentatus* and *L. polyphemus* genome assemblies as the databases. Putative AMPs identified were

further used as query sequences to search for corresponding transcripts. The prediction of ORFs was performed using NCBI ORF finder (https://www.ncbi.nlm.nih.gov/orffinder/) and the annotation of those ORFs were followed via blastp with nr database.

List Of Abbreviations

PAMPs	Pathogen-Associated Molecular Patterns
LINEs	Long Interspersed Elements
LTR	Long Terminal Repeat
BUSCO	Benchmarking Universal Single-Copy Orthologs
FoxO	The Forkhead Box O
RAGE	Receptor for Advanced Glycation Endproducts
JAK	Janus kinase
STAT	Signal Transducer and Activator of Transcription Protein
Hox genes	Homeotic Genes
Ubx	Ultrabithorax
AbdA	Abdominal-A
AbdB	Abdominal-B
PRRs	Pattern Recognition Receptors
FREPs	Fibrinogen-Related Proteins
DSCAMs	Down Syndrome Cell Adhesion Molecules
IMD	Immune deficiency
JNK	c-Jun N-terminal Kinase
LPS	Lipopolysaccharide
AMP	Antimicrobial Peptide
ORF	Open Reading Frame

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and material

All data are available from the following databases as described. The Whole Genome Shotgun project has been deposited at GenBank under the accession RXZF00000000. The version described in this paper is version RXZF01000000. RNA-seq raw data of *T. tridentatus* were downloaded from NCBI SRA database (accession number SRX330201). The published genome assembly, coding sequences and protein sequences of the Atlantic horseshoe crab, *L. polyphemus*, submitted by Washington University, were downloaded from NCBI with RefSeq ID 2304488, accession GCF_000517525.1.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported in part by the National Key R&D Program of China (2017YFC1404504), the Major State Basic Research Development Program of China (973 Program) (2015CB755906), Program for Innovative Research Team in Science and Technology in Fujian Province University, and Beihai Pilot City Program for the National Innovative Development of the Marine Economy.

Authors' contributions

ZY, CML and CJM designed the study. CML, CDB, RLW, KY, HS collected horseshoe crab samples. CML, LY, YQ, and ZL carried out the experiments. ZY, LY, YQ analyzed the data. ZY, CML, CJM, LY and KY wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable

References

1. Rudkin DM, Young GA, Nowlan GSJP. The oldest horseshoe crab: a new xiphosurid from Late Ordovician Konservat-Lagerstätten deposits, Manitoba, Canada. 2008;51(1):1-9.

2. Eldredge N, Stanley SM. Living Fossils: Springer Verlag; 1984. 257–60 p.

3. Sekiguchi K. Biology of horseshoe crabs: International Specialized Book Service Incorporated; 1988.

4. Iwanaga S, Kawabata S-i, Muta TJTjob. New types of clotting factors and defense molecules found in horseshoe crab hemolymph: their structures and functions. 1998;123(1):1-15.

5. Tan NS, Ho B, Ding JLJTFJ. High-affinity LPS binding domain (s) in recombinant factor C of a horseshoe crab neutralizes LPS-induced lethality. 2000;14(7):859-70.

6. Muta T, Iwanaga SJCoii. The role of hemolymph coagulation in innate immunity. 1996;8(1):41-7.

7. Nellaiappan K, Sugumaran MJCB, Biochemistry PPB, Biology M. On the presence of prophenoloxidase in the hemolymph of the horseshoe crab, Limulus. 1996;113(1):163-8.

8. Pieters JJCoii. Evasion of host cell defense mechanisms by pathogenic bacteria. 2001;13(1):37-44.

9. Iwanaga S, Lee B-LJBR. Recent advances in the innate immunity of invertebrate animals. 2005;38(2):128-50.

10. Chipman AD, Ferrier DE, Brena C, Qu J, Hughes DS, Schröder R, et al. The first myriapod genome sequence reveals conservative arthropod gene content and genome organisation in the centipede Strigamia maritima. 2014;12(11):e1002005.

11. Kenny NJ, Chan KW, Nong W, Qu Z, Maeso I, Yip HY, et al. Ancestral whole-genome duplication in the marine chelicerate horseshoe crabs. 2016;116(2):190.

12. Battelle B-A, Ryan JF, Kempler KE, Saraf SR, Marten CE, Warren WC, et al. Opsin repertoire and expression patterns in horseshoe crabs: evidence from the genome of Limulus polyphemus (Arthropoda: Chelicerata). 2016;8(5):1571-89.

13. Simpson SD, Ramsdell JS, Watson III WH, Chabot CCJIjog. The draft genome and transcriptome of the Atlantic horseshoe crab, Limulus polyphemus. 2017;2017.

14. Regier JC, Shultz JW, Zwick A, Hussey A, Ball B, Wetzer R, et al. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. 2010;463(7284):1079.

15. Sharma PP, Kaluziak ST, Perez-Porro AR, Gonzalez VL, Hormiga G, Wheeler WC, et al. Phylogenomic interrogation of Arachnida reveals systemic conflicts in phylogenetic signal. 2014;31(11):2963-84.

16. Tanaka G, Hou X, Ma X, Edgecombe GD, Strausfeld NJJN. Chelicerate neural ground pattern in a Cambrian great appendage arthropod. 2013;502(7471):364.

17. Holland PWJWIRDB. Evolution of homeobox genes. 2013;2(1):31-45.

18. Hui JH, McDougall C, Monteiro AS, Holland PW, Arendt D, Balavoine G, et al. Extensive chordate and annelid macrosynteny reveals ancestral homeobox gene organization. 2011;29(1):157-65.

19. Gehring WJ, Affolter M, Burglin TJArob. Homeodomain proteins. 1994;63(1):487-526.

20. Ferrier DJEPktiEDB, . Cambridge University Press: Cambridge, UK. When is a Hox gene not a Hox gene? The importance of gene nomenclature. 2008:175-93.

21. Gehring WJHl. Homeotic genes, the homeobox, and the spatial organization of the embryo. 1985;81:153-72.

22. Gehring WJJGtthg. A history of the homeobox. 1994:1-10.

23. De Robertis EMJGtthg. The homeobox in cell differentiation and evolution. 1994.

24. Holland P, Hogan BJG, development. Expression of homeo box genes during mouse development: a review. 1988;2(7):773-82.

25. Hughes CL, Kaufman TCJE, development. Hox genes and the evolution of the arthropod body plan 1. 2002;4(6):459-99.

26. Akira S, Uematsu S, Takeuchi OJC. Pathogen recognition and innate immunity. 2006;124(4):783-801.

27. Tanji T, Ip YTJTii. Regulators of the Toll and Imd pathways in the Drosophila innate immune response. 2005;26(4):193-8.

28. Boutros M, Agaisse H, Perrimon NJDc. Sequential activation of signaling pathways during innate immune responses in Drosophila. 2002;3(5):711-22.

29. Li F, Xiang JJF, immunology s. Signaling pathways regulating innate immune responses in shrimp. 2013;34(4):973-80.

30. Kawabata S-i, Osaki T, Iwanaga S. Innate immunity in the horseshoe crab. Innate Immunity: Springer; 2003. p. 109-25.

31. Iwanaga S, Kawabata S-iJFB. Evolution and phylogeny of defense molecules associated with innate immunity in horseshoe crab. 1998;3:D973-D84.

32. Miyata T, Tokunaga F, Yoneya T, Yoshikawa K, Iwanaga S, Niwa M, et al. Antimicrobial peptides, isolated from horseshoe crab hemocytes, tachyplesin II, and polyphemusins I and II: chemical structures and biological activity. 1989;106(4):663-8.

33. Aketagawa J, Miyata T, Ohtsubo S, Nakamura T, Morita T, Hayashida H, et al. Primary structure of limulus anticoagulant anti-lipopolysaccharide factor. 1986;261(16):7357-65.

34. MUTA T, MIYATA T, TOKUNAGA F, NAKAMURA T, IWANAGA SJTJoB. Primary structure of anti-lipopolysaccharide factor from American horseshoe crab, Limulus polyphemus. 1987;101(6):1321-30.

35. Nakamura T, Furunaka H, Miyata T, Tokunaga F, Muta T, Iwanaga S, et al. Tachyplesin, a class of antimicrobial peptide from the hemocytes of the horseshoe crab (Tachypleus tridentatus). Isolation and chemical structure. 1988;263(32):16709-13.

36. Shigenaga T, Muta T, Toh Y, Tokunaga F, Iwanaga SJJoBC. Antimicrobial tachyplesin peptide precursor. cDNA cloning and cellular localization in the horseshoe crab (Tachypleus tridentatus). 1990;265(34):21350-4.

37. Saito T, Kawabata S-i, Shigenaga T, Takayenoki Y, Cho J, Nakajima H, et al. A novel big defensin identified in horseshoe crab hemocytes: isolation, amino acid sequence, and antibacterial activity. 1995;117(5):1131-7.

38. Muta T, Iwanaga S. Clotting and immune defense in Limulidae. Invertebrate Immunology: Springer; 1996. p. 154-89.

39. Nakamura T, MORITA T, IWANAGA SJEjob. Lipopolysaccharide-sensitive serine-protease zymogen (factor C) found in Limulus hemocytes: Isolation and characterization. 1986;154(3):511-21.

40. Nakamura T, Tokunaga F, Morita T, Iwanaga SJTJoB. Interaction between lipopolysaccharide and intracellular serine protease zymogen, factor C, from horseshoe crab (Tachypleus tridentatus) hemocytes. 1988;103(2):370-4.

41. Nakamura T, HORIUCHI T, MORITA T, IWANAGA SJTJoB. Purification and properties of intracellular clotting factor, factor B, from horseshoe crab (Tachypleus tridentatus) hemocytes. 1986;99(3):847-57.

42. Muta T, Oda T, Iwanaga SJJoBC. Horseshoe crab coagulation factor B. A unique serine protease zymogen activated by cleavage of an Ile-Ile bond. 1993;268(28):21384-8.

43. Nakamura S, MORITA T, HARADA-SUZUKI T, IWANAGA S, TAKAHASHI K, NIWA MJTJoB. A clotting enzyme associated with the hemolymph coagulation system of horseshoe crab (Tachypleus tridentatus): its purification and characterization. 1982;92(3):781-92.

44. Muta T, Hashimoto R, Miyata T, Nishimura H, Toh Y, Iwanaga SJJoBC. Proclotting enzyme from horseshoe crab hemocytes. cDNA cloning, disulfide locations, and subcellular localization. 1990;265(36):22426-33.

45. Morita T, Tanaka S, Nakamura T, Iwanaga SJFl. A new (1→ 3)-β-D-glucan-mediated coagulation pathway found in limulus amebocytes. 1981;129(2):318-21.

46. Muta T, Seki N, Takaki Y, Hashimoto R, Oda T, Iwanaga A, et al. Purified Horseshoe Crab Factor G RECONSTITUTION AND CHARACTERIZATION OF THE (13)-β-D-GLUCAN-SENSITIVE SERINE PROTEASE CASCADE. 1995;270(2):892-7.

47. Dietz RS, Holden JCJJoGR. Reconstruction of Pangaea: breakup and dispersion of continents, Permian to present. 1970;75(26):4939-56.

48. Pyron RAJSB. Biogeographic analysis reveals ancient continental vicariance and recent oceanic dispersal in amphibians. 2014;63(5):779-97.

49. Gamble T, Bauer AM, Greenbaum E, Jackman TRJJoB. Evidence for Gondwanan vicariance in an ancient clade of gecko lizards. 2008;35(1):88-104.

50. Roelants K, Bossuyt FJSB. Archaeobatrachian paraphyly and Pangaean diversification of crown-group frogs. 2005;54(1):111-26.

51. San Mauro D, Vences M, Alcobendas M, Zardoya R, Meyer AJTAN. Initial diversification of living amphibians predated the breakup of Pangaea. 2005;165(5):590-9.

52. Springer MS, Murphy WJ, Eizirik E, O'Brien SJJPotNAoS. Placental mammal diversification and the Cretaceous–Tertiary boundary. 2003;100(3):1056-61.

53. Mao K, Milne RI, Zhang L, Peng Y, Liu J, Thomas P, et al. Distribution of living Cupressaceae reflects the breakup of Pangea. 2012;109(20):7793-8.

54. Gardner JD, Nydam RL. “Mesozoic and Cenozoic lissamphibian and squamate assemblages of Laurasia”—introduction to the special issue. Springer; 2013.

55. Wicander R, Monroe JS. Historical geology : evolution of Earth and life through time: Thomson Brooks/Cole; 2007.

56. Shuster JCJPic, research b. A pictorial review of the natural history and ecology of the horseshoe crab Limulus polyphemus, with reference to other Limulidae. 1982;81:1-52.

57. Briggs DE, Moore RA, Shultz JW, Schweigert GJPotRSoLBBS. Mineralization of soft-part anatomy and invading microbes in the horseshoe crab Mesolimulus from the Upper Jurassic Lagerstätte of Nusplingen, Germany. 2005;272(1563):627-32.

58. Ding L, Qasim M, Jadoon IA, Khan MA, Xu Q, Cai F, et al. The India–Asia collision in north Pakistan: Insight from the U–Pb detrital zircon provenance of Cenozoic foreland basin. 2016;455:49-61.

59. Garcia-Fernàndez JJNRG. The genesis and evolution of homeobox gene clusters. 2005;6(12):881.

60. Morvan A, Iwanaga S, Comps M, Bachere EJJoip. In VitroActivity of the Limulus Antimicrobial Peptide Tachyplesin I on Marine Bivalve Pathogens. 1997;69(2):177-82.

61. Morimoto M, Mori H, Otake T, Ueba N, Kunita N, Niwa M, et al. Inhibitory effect of tachyplesin I on the proliferation of human immunodeficiency virus in vitro. 1991;37(3):206-11.

62. Murakami T, Niwa M, Tokunaga F, Miyata T, Iwanaga SJC. Direct virus inactivation of tachyplesin I and its isopeptides from horseshoe crab hemocytes. 1991;37(5):327-34.

63. Chen Y, Xu X, Hong S, Chen J, Liu N, Underhill CB, et al. RGD-Tachyplesin inhibits tumor growth. 2001;61(6):2434-8.

64. Kawabata SJNdiii. Clotting cascade and defense molecules found in the hemolymph of the horseshoe crab. 1996:255-83.

65. Ehlinger GS, Tankersley RA. Ecology of Horseshoe Crabs in Microtidal Lagoons2009.

66. Walls EA, Berkson J, Smith SAJRiFS. The horseshoe crab, Limulus polyphemus: 200 million years of existence, 100 years of study. 2002;10(1):39-73.

67. Mikkelsen T. The secret in the blue blood: Science Press Beijing; 1988.

68. Leibovitz L, Lewbart GJTAHCHUP, Cambridge. Diseases and symbionts: vulnerability despite tough shells. 2003:245-75.

69. Zasloff MJPotNAoS. Magainins, a class of antimicrobial peptides from Xenopus skin: isolation, characterization of two active forms, and partial cDNA sequence of a precursor. 1987;84(15):5449-53.

70. Moore KS, Wehrli S, Roder H, Rogers M, Forrest JN, McCrimmon D, et al. Squalamine: an aminosterol antibiotic from the shark. 1993;90(4):1354-8.

71. Merchant ME, Leger N, Jerkins E, Mills K, Pallansch MB, Paulman RL, et al. Broad spectrum antimicrobial activity of leukocyte extracts from the American alligator (Alligator mississippiensis). 2006;110(3-4):221-8.

72. Zerbino D, Birney EJGr. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. 2008:gr. 074492.107.

73. Boetzer M, Pirovano WJBb. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. 2014;15(1):211.

74. Stanke M, Tzvetkova A, Morgenstern BJGb. AUGUSTUS at EGASP: using EST, protein and genomic alignments for improved gene prediction in the human genome. 2006;7(1):S11.

75. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EMJB. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. 2015;31(19):3210-2.

76. Chen M, Wang C, Wang W, Ji G, Hu B, Du M, et al. De Novo Assembly and Characterization of Early Embryonic Transcriptome of the Horseshoe Crab Tachypleus tridentatus. 2016;11(1):e0145825.

77. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. 2013;8(8):1494.

78. Li L, Stoeckert CJ, Roos DSJGr. OrthoMCL: identification of ortholog groups for eukaryotic genomes. 2003;13(9):2178-89.

79. Stamatakis AJB. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. 2014;30(9):1312-3.

80. Wu J, Mao X, Cai T, Luo J, Wei LJNar. KOBAS server: a web-based platform for automated annotation and pathway identification. 2006;34(suppl_2):W720-W4.

81. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, et al. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. 2011;39(suppl_2):W316-W22.

82. Ai C, Kong LJJoG, Genomics. CGPS: A machine learning-based approach integrating multiple gene set analysis tools for better prioritization of biologically relevant pathways. 2018;45(9):489-504.

83. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, et al. BEAST 2: a software platform for Bayesian evolutionary analysis. 2014;10(4):e1003537.

84. Heled J, Drummond AJJSB. Calibrated tree priors for relaxed phylogenetics and divergence time estimation. 2011;61(1):138-49.

85. Hasegawa M, Kishino H, Yano T-aJJome. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. 1985;22(2):160-74.

86. Kumar S, Stecher G, Tamura KJMb, evolution. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. 2016;33(7):1870-4.

87. McTaggart SJ, Conlon C, Colbourne JK, Blaxter ML, Little TJJBG. The components of the Daphnia pulex immune system as revealed by complete genome sequencing. 2009;10(1):175.

Tables

Table 1 Summary of the *Tachypleus tridentatus* genome assembly and annotation statistic

Summary of the *Tachypleus tridentatus* genome assembly and annotation statistics.

Tachypleus tridentatus assembly statistics	
Assembly size (Gb)	2.06
Number of scaffolds	143,932
N50 scaffold length (kb)	165
Largest scaffold (kb)	5,278
Shortest scaffold (kb)	1
GC content	32.03%
Average exon length (bp)	333
Average intron length (bp)	3,792
Tachypleus tridentatus assembly annotation statistics	
Total number of genes	24,222
% BUSCOsa	87.4 [10.8], 11.3, 1.3

a of 1,066 arthropod BUSCOs Complete [Duplicated], Fragmented, Missing, in the assembly

Table 2 Coagulation Cascade genes in 2 horseshoe crabs, 1 scorpione, and 3 spiders.

	Horseshoe Crabs		Scorpiones		Spiders	
species	*T. tridentatus*	*L. Polyphemus*	*C. sculpturatus*	*N. clavipes*	*S. mimosarum*	*P. tepidariorum*
Factor C	2	4	5	4	6	4
Factor B	10	12	8	4	7	7
Factor G	4	1*	0	0	0	0
Proclotting Enzyme	9	11	7	5	5	5
Coagulogen	6	6	0	0	0	0
Total	30	33	20	13	18	16

* Identified in previous study

Figures

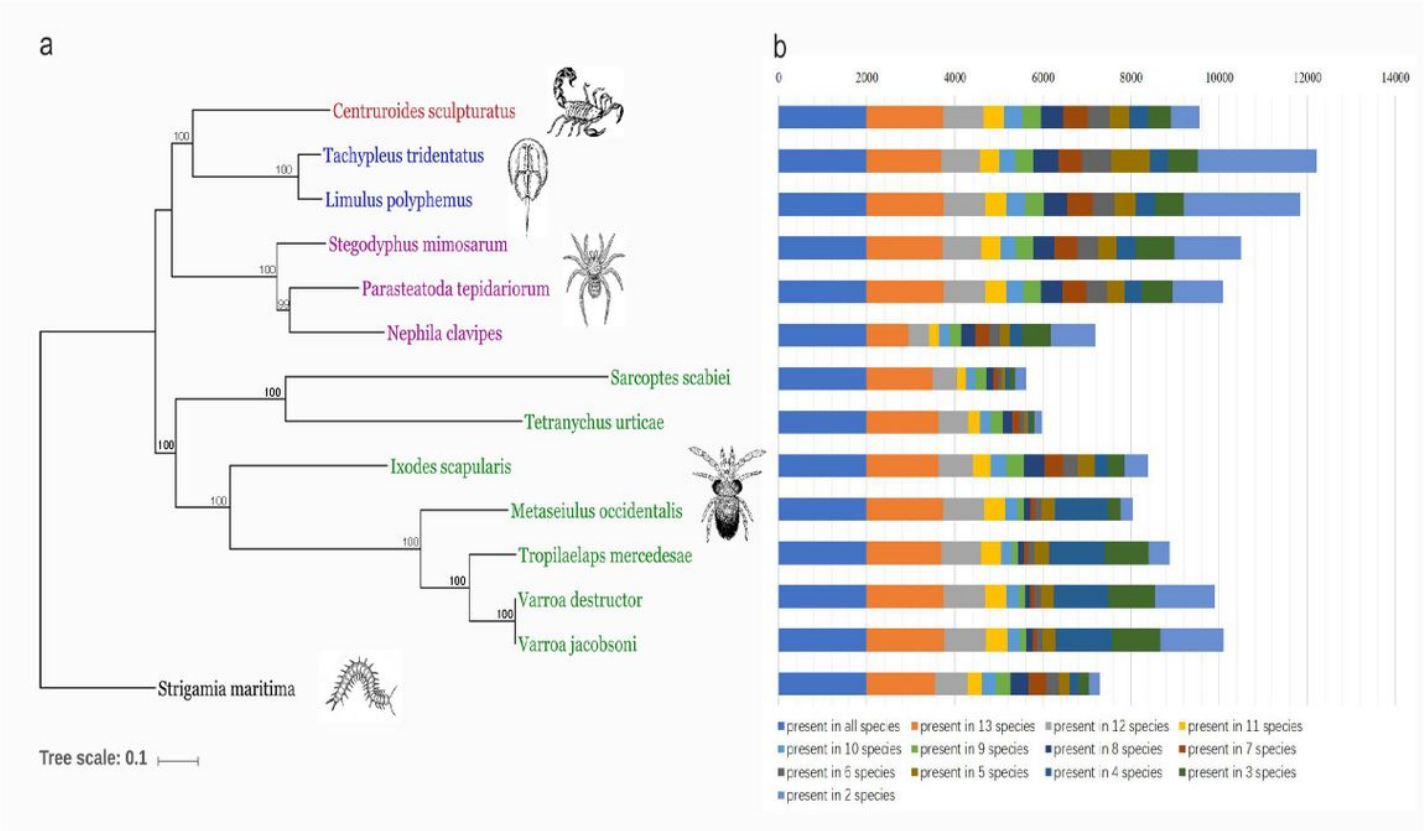

Figure 1

Comparative genomics. (a) Phylogenetic placement among T. tridentatus and other Chelicerata species. The phylogeny with 111 single-copy orthologous genes presented in all 14 species was built using RAxML. The tree was rooted with S. maritima. (b). Orthology comparsion among T. tridentatus and other Chelicerata species. There were 2,720 (22.2%) and 2,648 (21.5%) orthologs of T. tridentatus and L. polyphemus shared in only two species. C. sculpturatus had the most expanded species unique genes (7,328), followed with 6,247 N. clavipes specific genes. The number of species specific genes of T. tridentatus and L. polyphemus showed in the middle position with 1,124 and 857, respectively.

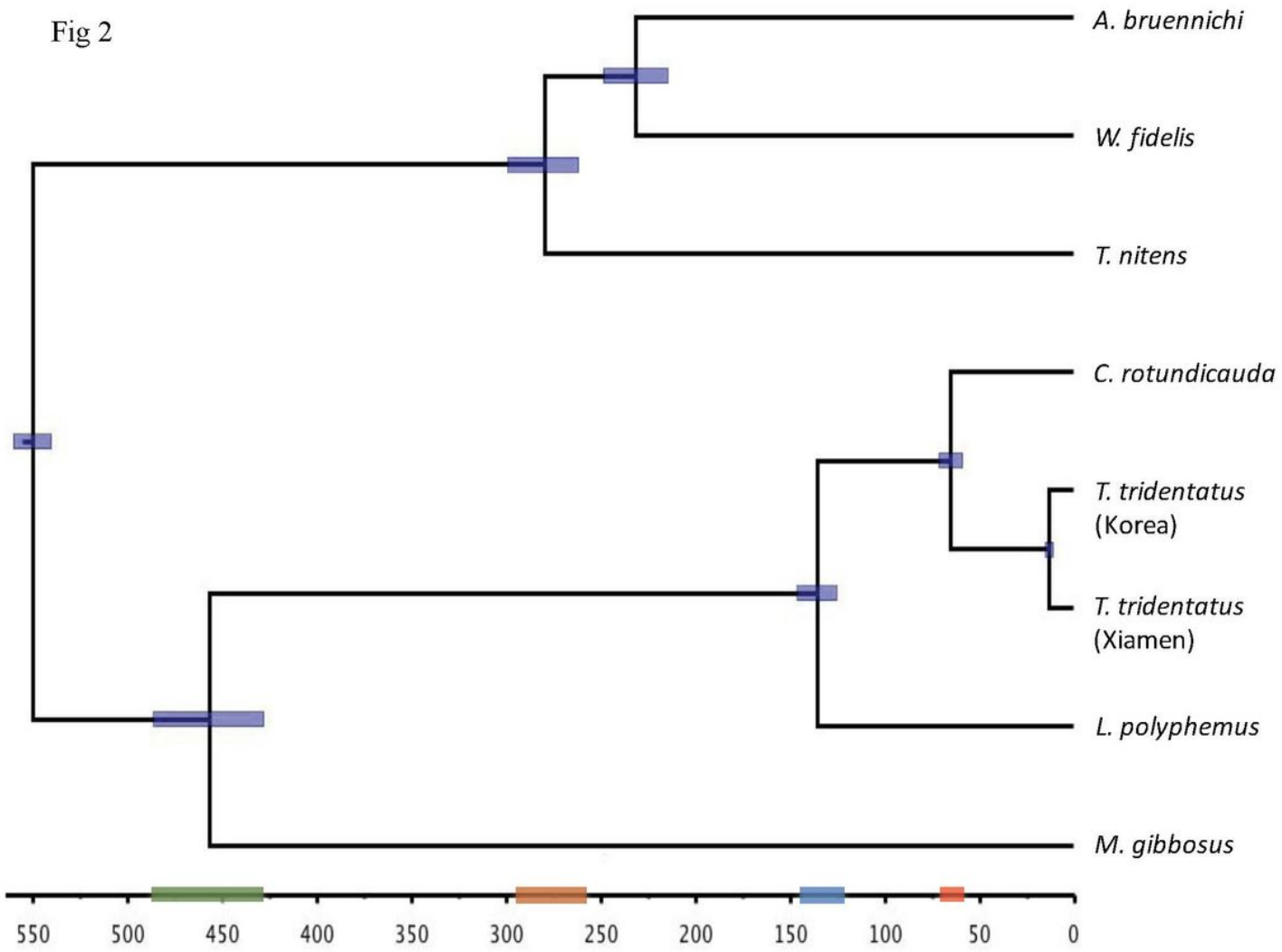

Figure 2

Bayesian maximum-clade-credibility tree based on the concatenated mitochondrial coding genes dataset in BEAST 2.5.1 with a strict clock, showing the estimated divergence time of Chelicerata species. Node shows the mean estimated divergence times in million years ago (MYA). Purple bars indicate 95% confidence levels. On the time axis, the green bar shows the divergence time for split of the scorpion from horseshoe crabs; the brown bar shows the inner split time of the three spiders; the blue bar shows the origin of the the last common ancestor of Asian horseshoe crabs (including T. tridentatus) and L. polyphemus; the red bar shows the inner split of C. rotundicauda and T. tridentatus.

	T. tridentatus (genome)	*T. tridentatus* (transcriptom)	*L. polyphemus*	*S. maritima*	*D. melanogaster*
Pattern recognition receptors					
PGRP	2	2	2	16	13
TEP like	23	2	24	4	6
FREP like	42	32	46	13	13
Dscam like	117	57	118	1	4
Galectin	5	2	5	-	-
CTL	27	25	27	-	-
Signaling and Transduction					
Toll pathway					
Toll like	18	8	17	36	9
spz like	8	2	13	1	6
Myd88	1	1	1	1	1
tube	1	1	1	0	1
pelle	3	1	1	1	1
cactus	1	1	1	1	1
dorsal	1	1	2	1	1
TRAF2	8	8	6	-	-
IMD pathway					
IMD	1	1	1	~1	1
Dredd	4	4	3	1	1
Tak1	3	3	3	1	1
Relish	6	6	4	2	1
IKK	1	1	1	-	-
Jak/stat pathway					
domeless	5	4	5	1	1
JAK (hop)	1	0	1	1	1
Stat92E	2	2	2	1	1
JNK (bsk)	3	3	3	1	1
Hem	1	1	1	1	1
Effectors					
Anti-LPS factor	1	-	~1	-	-
Tachyplesin	2	-	-	-	-
Big defensin	2	-	-	-	-

Figure 3

Presence of immune related gene families in T. tridentatus and L. polyphemus. Counts of immune related genes are shown for T. tridentatus, L. polyphemus, S. maritima (10) and D. melanogaster (87). The gene number counts according to results of blastp based on NR annotation and InterproScan from the genome of T. tridentatus and L. polyphemus and the transcriptome of T. tridentatus. Abbreviations: PGRP,

peptidoglycan recognition protein; TEP, thioester-containing protein; FREP, fibrinogen-related protein; CTL, C-type lectin.

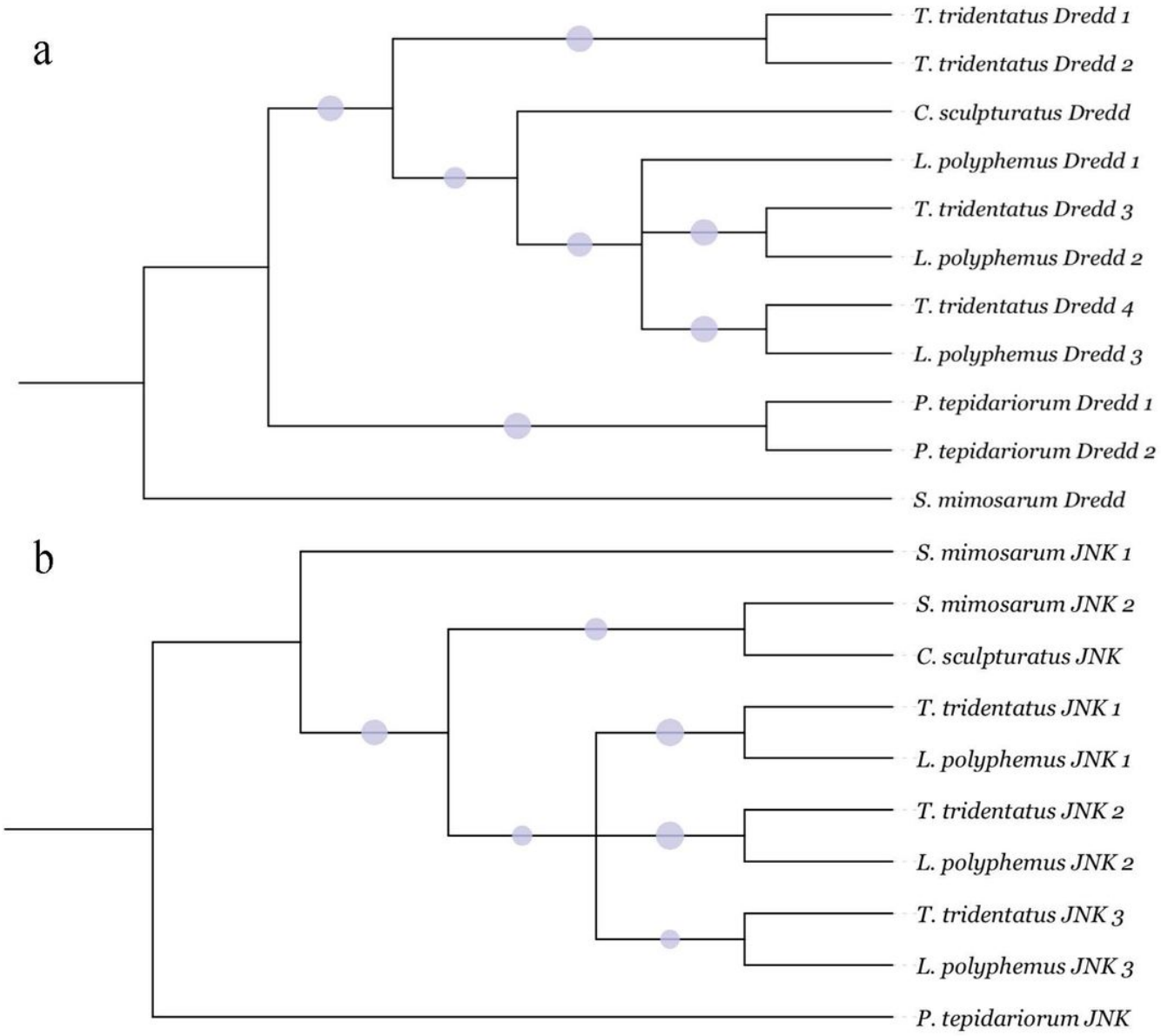

Figure 4

Phylogenetic analysis of immune signaling pathway related genes. (a) Phylogeny of Dredd genes involved in IMD and Jak/STAT signaling pathways among multiple Chelicerata species. (b) Phylogeny of JNK genes involved in IMD and Jak/STAT signaling pathways among multiple Chelicerata species The two Neighbor-Joining trees was constructed using MEGA with 1000 bootstrap, and rooted with S. mimosarum for Dredds and P. tepidariorum for JNKs, respectively.

Supplementary Files

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

- supplement1.doc