




## Article

# A Long-Read Genome Assembly of a Native Mite in China *Pyemotes zhonghuajia* Yu, Zhang & He (Prostigmata: Pyemotidae) Reveals Gene Expansion in Toxin-Related Gene Families

Yan-Fei Song<sup>1,2,†</sup> , Li-Chen Yu<sup>3,†</sup>, Mao-Fa Yang<sup>2,4</sup>, Shuai Ye<sup>2</sup>, Bin Yan<sup>2</sup>, Li-Tao Li<sup>3</sup>, Chen Wu<sup>5</sup>   
and Jian-Feng Liu<sup>1,2,\*</sup> 

- <sup>1</sup> State Key Laboratory Breeding Base of Green Pesticide and Agricultural Bioengineering, Key Laboratory of Green Pesticide and Agricultural Bioengineering, Ministry of Education, Guizhou University, Guiyang 550025, China
- <sup>2</sup> Institute of Entomology, Guizhou University, Guizhou Provincial Key Laboratory for Agricultural Pest Management of the Mountainous Region, Scientific Observing and Experiment Station of Crop Pest Guiyang, Ministry of Agriculture, Guiyang 550025, China
- <sup>3</sup> Changli Institute of Pomology, Hebei Academy of Agriculture and Forestry Sciences, Changli 066600, China
- <sup>4</sup> College of Tobacco Science, Guizhou University, Guiyang 550025, China
- <sup>5</sup> The New Zealand Institute for Plant and Food Research Limited, Auckland 1142, New Zealand
- \* Correspondence: [jfliu3@gzu.edu.cn](mailto:jfliu3@gzu.edu.cn)
- † These authors contributed equally to this work.

**Abstract:** *Pyemotes zhonghuajia* Yu, Zhang & He (Prostigmata: Pyemotidae), discovered in China, has been demonstrated as a high-efficient natural enemy in controlling many agricultural and forestry pests. This mite injects toxins into the host (eggs, larvae, pupae, and adults), resulting in its paralyzation and then gets nourishment for reproductive development. These toxins have been approved to be mammal-safe, which have the potential to be used as biocontrol pesticides. Toxin proteins have been identified from many insects, especially those from the orders Scorpions and Araneae, some of which are now widely used as efficient biocontrol pesticides. However, toxin proteins in mites are not yet understood. In this study, we assembled the genome of *P. zhonghuajia* using PacBio technology and then identified toxin-related genes that are likely to be responsible for the paralytic process of *P. zhonghuajia*. The genome assembly has a size of 71.943 Mb, including 20 contigs with a N50 length of 21.248 Mb and a BUSCO completeness ratio of 90.6% (n = 1367). These contigs were subsequently assigned to three chromosomes. There were 11,183 protein coding genes annotated, which were assessed with 91.2% BUSCO completeness (n = 1066). Neurotoxin and dermonecrotic toxin gene families were significantly expanded within the genus of *Pyemotes* and they also formed several gene clusters on the chromosomes. Most of the genes from these two families and all of the three agatoxin genes were shown with higher expression in the one-day-old mites compared to the seven-day-pregnant mites, supporting that the one-day-old mites cause paralyzation and even death of the host. The identification of these toxin proteins may provide insights into how to improve the parasitism efficiency of this mite, and the purification of these proteins may be used to develop new biological pesticides.

**Keywords:** *Pyemotes*; genome annotation; protein; toxin

**Key Contribution:** *Pyemotes zhonghuajia* is an important biocontrol agent against Isoptera, Homoptera, Hymenoptera, Lepidoptera, and Coleoptera pests. However, the genetic resources are largely lacking in the genus of *Pyemotes*; which has impeded our understanding of the parasitic and physiological mechanism of those venomous ectoparasitic mites at the molecular level. In this study, we have provided a genome assembly of *P. zhonghuajia* with a specific description of the toxin-related gene families. Our resources have provided opportunities for understanding the molecular mechanism underlying the use of toxins in this mite in order to improve its parasitism efficiency.



**Citation:** Song, Y.-F.; Yu, L.-C.; Yang, M.-F.; Ye, S.; Yan, B.; Li, L.-T.; Wu, C.; Liu, J.-F. A Long-Read Genome Assembly of a Native Mite in China *Pyemotes zhonghuajia* Yu, Zhang & He (Prostigmata: Pyemotidae) Reveals Gene Expansion in Toxin-Related Gene Families. *Toxins* **2022**, *14*, 571. <https://doi.org/10.3390/toxins14080571>

Received: 13 July 2022

Accepted: 17 August 2022

Published: 21 August 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

*Pyemotes zhonghuajia* Yu, Zhang & He (Prostigmata: Pyemotidae) is a native viviparous mite initially collected from *Sinoxylon japonicum* Lesne and *Phloeosinus hopehi* Schedl by Lichen Yu in the 1990s [1]. It distributes naturally in Shanxi, Xinjiang, Ningxia, Hebei, Tianjin, and Beijing, China [2]. *Pyemotes zhonghuajia* is a dominant efficient ectoparasitic mite and is regarded as an important natural enemy in controlling many agricultural and forest pests [3]. This mite can inject toxins through puncturing the intersegmental cuticle using its mouthpart to paralyze a large number of pests including Isoptera, Homoptera, Hymenoptera, Lepidoptera, and Coleoptera [4–7]. After the host is completely paralyzed, the mite finds an optimal position to settle down and then obtains nourishment for reproductive development [7]. The gland of *P. zhonghuajia* is located at the junction of the head and neck, and those toxins are produced in secretory cells in the gland. The toxins produced by the mite are highly efficient in paralysis. A one-day-old mite *P. zhonghuajia* female can kill a third *Spodoptera litura* instar larva heavier than 680,000 times its own weight [7–9]. One single *P. zhonghuajia* female can lead to over 50% mortality rate of the first to third instar larvae of *Mythimna separata* Walker and *Spodoptera frugiperda* (Smith) [3,6]. The toxins are also found to be safe to mammals, which have the potential to be used as biological pesticides [6]. Nowadays, *Pyemotes zhonghuajia* is mass-produced and commercialized to control *M. separata*, *Aphis citricola* van der Goot, *S. frugiperda* Smith, *Sinoxylon japonicum* (Motschulsky), *Monochamus alternatus* Hope, and *Zeuzera leuconotum* Bulter [3,4,6,10–13].

Toxin proteins from the order Scorpiones and Araneae have been widely studied, some of which are now used for pest control. The scorpion insect-specific neurotoxins *AaIT* and *LqhIT2* target the insect's voltage-gated sodium channels that bind to the host's motor nerve branches to play critical roles in electrical signaling in stimulating skeletal muscles [14]. Identified from the highly toxic scorpion *Androctonus australis* [15], *AaIT* is a single-chain polypeptide containing 70 amino acids and four disulfide bridges [16]. As it is safe to mammals, *AaIT* can be used for pest control [17,18]. *AaIT*-expressed baculovirus has shown to reduce the survival time of the pests coupled with a significantly enhanced infection efficiency of this virus [19]. For example, *AaIT*-expressed *Bombyx mori* nucleopolyhedrovirus (BmNPV) can reduce feeding damage from silkworms to the host, and introducing *AaIT* gene into entomopathogen *Beauveria bassiana* can enhance its virulence to mosquitoes [14]. *LqhIT2* is another protein known with the potential of pest control. It was isolated from the scorpion *Leiurus quinquestriatus hebraeus*, which comprises 61 amino acids with four disulphide bridges [17]. The feeding capacity of the rice leafroller (*Cnaphalocrocis medinalis* Guenee) was shown to be decreased in the *LqhIT2*-inserted transgenic rice compared to the wild type [17]. In Araneae, more than 800 toxins have been isolated and described [20]. Dermonecrotic toxins, well-characterized from the Brown spider (*Genus Loxosceles*), are biochemical constituents in spider crude venom, and it can also induce necrotic and dermonecrotic lesions on rabbits and mice [21]. Agatoxins are neurotoxins, identified from *Agelenopsis aperta* and classified into three classes ( $\alpha$ -,  $\mu$ -, and  $\omega$ - agatoxins), specifically targeting three classes of ion channels (voltage-activated calcium channels, transmitter-activated cation channels, and voltage-activated sodium channels), respectively [22]. The  $\alpha$ -agatoxins have an acylpolyamine structure and can induce immediate but reversible paralysis. The  $\mu$ -agatoxins do not have this structure and cause immediate but irreversible paralysis. The  $\omega$ - agatoxins divide into four types ( $\omega$ -Aga-1A,  $\omega$ -Aga-IIA, Type III  $\omega$ -Agatoxins, and Type IV  $\omega$ -Agatoxins), and inhibit voltage-activated calcium channels in nerve terminals. Some enzymes also play important roles in arachnid biotoxins. Identified from spider venoms, enzymes mainly serve two important functions: (1) lysing polymers in the extracellular matrix and (2) binding to the compounds in the membrane [23,24].

There are only a few studies on toxin proteins in mites. In straw itch mite (*Pyemotes tritici*), a low molecular weight protein TxP was identified and showed inducing a rapid, muscle-contracting paralysis [25]. The toxicity of TxP -1 has shown to be comparable to *AaIT* or even stronger [26]. The TxP proteins were found to be translated by a range of cDNAs with

variable length that are homologs to the insect-selective paralytic neurotoxin *tox34* [25]. It is notable that more than 18 recombinant baculoviruses engineered with *tox34* have been used for pest control [19,26]. In *P. zhonghuaia*, 12 *tox34* homologs were identified with sequence similarity ranging from 84.21% to 90.42% compared with those found in *P. tritici* [9]. However, other than these sequences, there is little genetic information about toxin-related proteins in the genus of *Pyemotes* that contains many venomous ectoparasitic mites.

Here, we sequenced the genome of *P. zhonghuaia* using Pacific Bioscience technology on the single-molecule real-time (SMRT) platform. The assembled genome was annotated with protein-coding genes, repeats, and non-coding RNAs (ncRNAs). We analyzed gene family evolution across Arachnida (1 Araneae, 1 Scorpiones, and 10 Acarina) with the main focus on the toxin-related genes.

## 2. Results and Discussion

### 2.1. Genome Assembly

A total of ~13 Gb and ~11 Gb Illumina short reads (150 bp) and PacBio long reads (average length of 9699.12 bp and N50 of 12.537 kb) were generated for genome assembly. The k-mer analysis based on short reads estimated genome size of 69.33 Mb comprising 7.09 Mb repetitive regions with genome heterozygosity of 0.04% (Supplementary Table S1). The *P. zhonghuaia* genome was assembled into 19 scaffolds containing 71.943 Mb with N50 of 21.248 Mb (Table 1), and 68.511 Mb was assigned to three pseudo-chromosomes. This assembled genome has a size comparable to the k-mer estimation and achieved a BUSCO complete gene ratio of 90.6% with duplicated and missing gene ratios of 1.3% and 7.9%, respectively. The mapping back rates from short and long reads as well as RNA-Seq data were 98.44%, 96.72%, and 92.84%, respectively. Compared with the other two mite genome assemblies containing thousands of scaffolds/contigs, our *P. zhonghuaia* genome assembly is much more continuous (Table 1). This genome assembly has a comparable size with the other two mite genomes.

**Table 1.** Genome Assembly and Annotation Statistics of *P. zhonghuaia*, *Tetranychus urticae*, and *Dermatophagoides pteromyssinus*.

Elements	<i>Pyemotes zhonghuaia</i>	<i>Stratiolaelaps scimitus</i>	<i>Tetranychus urticae</i>
Genome assembly			
Assembly size (Mb)	71.943	426.50	89.6
Number of scaffolds/contigs	19/20	158	-/2035
Longest scaffold/contig (Mb)	25.136/22.128	31.29	7/0.929
N50 scaffold/contig length (Mb)	22.128/21.248	7.66	10/120
GC (%)	25.02	45.85	-
Gaps (%)	0.00	0.00	-
BUSCO completeness (%)	90.6	93.1	-
Annotation			
Protein-coding genes	11,183	13,305	18,414
Mean protein length (aa)	480.97	500.59	-
Mean gene length (bp)	3243.42	7870.13	2652
Exons/introns per gene	3.59/2.45	6.24/-	3.82/-
Exon (%)	26.77	7.25	-
Mean exon length	475.53	372.35	178
Intron (%)	24.07	5.02	-
Mean intron length	626.12	1105.66	400
BUSCO completeness (%)	91.2	95.8	-

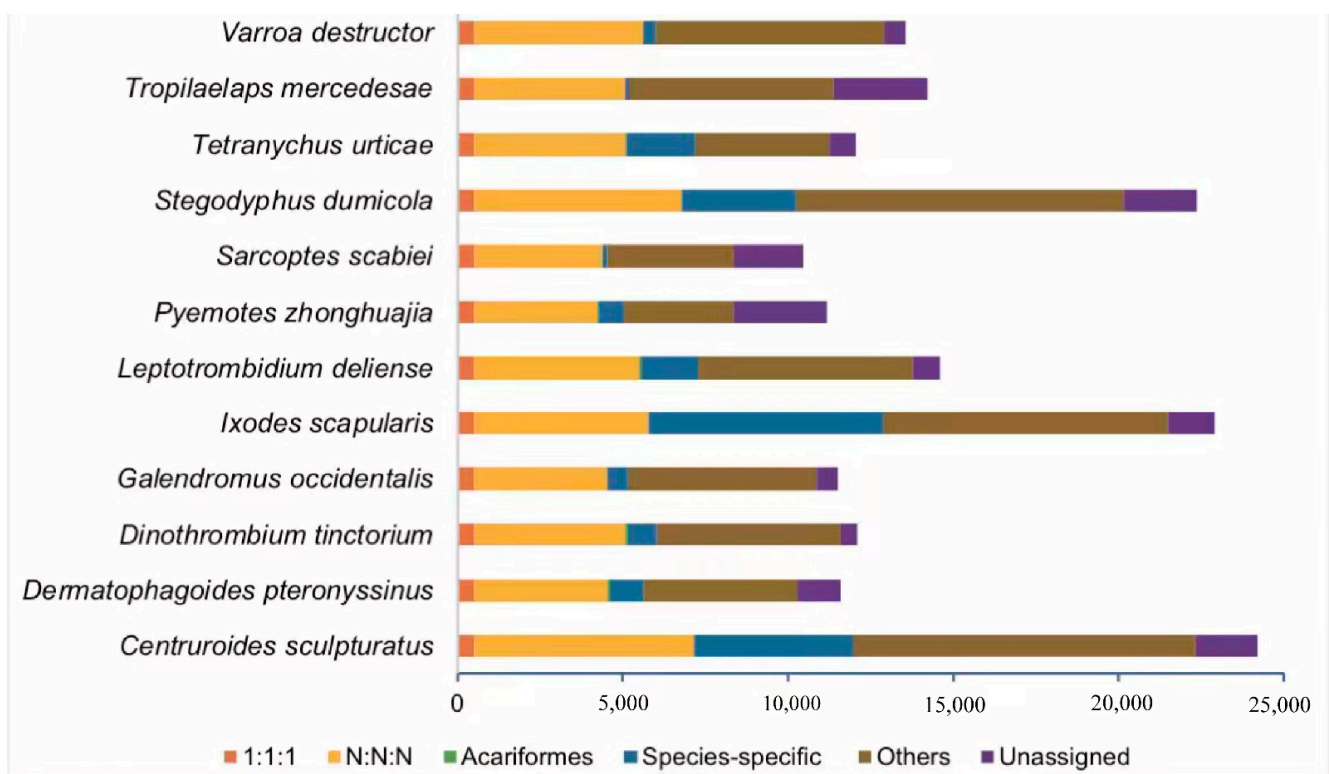
### 2.2. Genome Annotation

There was 11.4% of the assembly annotated as repetitive regions, which contained 77,050 repeats taking up ~8.2 Mb. The most abundant repeat class was LTR elements, taking up 3.62% of the assembly, followed by simple repeats (2.92%), unclassified repeats (1.98%), low complexity repeats (1.42%), and DNA elements (0.79%) (Supplementary Table S2). There were 11,183 gene models annotated with the average length of 3243.42 bp and 3.59 exons per gene. The average lengths of exon and intron were 475.53 bp and 626.12 bp, respectively. The BUSCO result showed 91.2% complete genes with duplicated and missing genes being 2.5% and 6.5%, showing that most of the annotated genes are likely to have complete lengths. It is noticed that the BUSCO complete gene ratio resulted from the

annotated gene set is slightly higher than the genome assembly. It is likely because BUSCO only employs AUGUSTUS for gene prediction and has less power to predict complete genes correctly compared with our gene annotation method that was supported by RNA-Seq data. [27]. There were 218 ncRNAs annotated in the genome assembly, including eight miRNA, 48 rRNA, nine snRNA, and 112 tRNA. The annotated snRNAs included five spliceosomal RNAs (U1, U2, U4, U5 and U6), one minor spliceosomal RNA (U6atac), and three C/D box snoRNAs (U3 and snoR38) (Supplementary Table S3). The tRNAs Supres and SelCys were absent in the annotation. This is the first non-coding RNA set annotated in a mite genome.

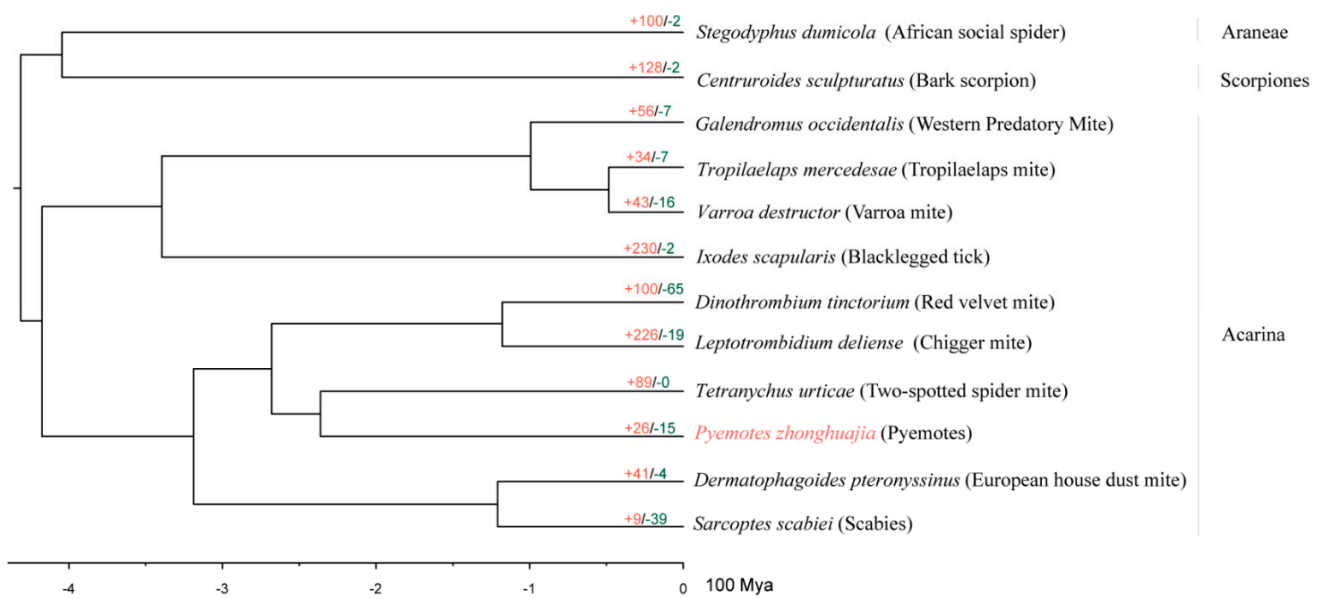
### 2.3. Species Phylogeny and Gene Family Evolution

There were 12 species including *P. zhonghuajia* selected for phylogenetic construction. In total, 180,720 genes were clustered into 17,407 gene families. In *P. zhonghuajia*, a total of 11,183 genes were analyzed, and there were 8388 genes assigned into 6224 gene families with species-specific gene families and genes being 178 and 761, respectively (Figure 1).



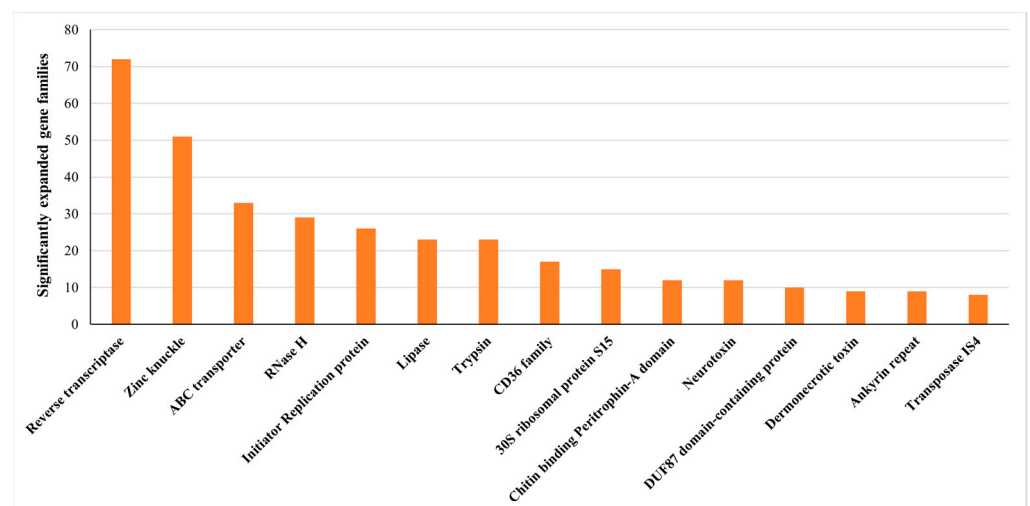
**Figure 1.** Histogram shows the number of genes assigned to different groups. The “1:1:1” and “N:N:N” groups represent single- and multi-copy genes found in all the species. The group “Acariformes” represents orthologs unique to Acariformes. The “Others” group indicates other orthologs which do not belong to any above-mentioned ortholog categories. The group “Unassigned” represent the orthologs which can’t be assigned to any orthogroups.

There were 527 single-copy genes found in all the species, 473 of which contained 143,759 amino acid sites were used to construct a phylogenetic tree. All node supports were 100/100 (SH-aLRT support /ultrafast bootstrap support). The phylogenetic tree shows that *P. zhonghuajia* is sister to the two-spotted spider mite (*T. urticae*) and they are clustered with the other two mite species: chigger mite (*Leptotrombidium deliense*) and red velvet mite (*Dinothrombium tinctorium*) (Figure 2). The phylogenetic tree was consistent with the published classifications [1] and our calculation indicated *P. zhonghuajia* together with *T. urticae* emerged during Triassic (223.43~247.66 Mya).



**Figure 2.** Dating tree with node values representing the number of expanded, contracted.

Compared with the closely related *T. urticae* that has over one thousand gene families identified as expanded families, *P. zhonghuaia* only has 621 (containing 1224 genes) gene families that were calculated as expanded families (Figures 2 and 3). These families are related to digestion, detoxification, and toxins, and those such as ABC transporter, Lipase, Trypsin, Dermonecrotic toxin, CD36 family, and neurotoxins (Supplementary Table S4). The families of dermonecrotic toxins and neurotoxins are mostly likely to participate in the parasitic mechanism [25,28–31]. There were 3927 genes lost in 3664 contracted families, including 15 gene families that were identified as significantly contracted families. These contracted families might be associated with viviparous reproduction in pyemotid mite; no free-living stages of larvae and nymph occur during the life cycle of *P. zhonghuaia* (Supplementary Table S5) [32].



**Figure 3.** Top fifteen significantly expanded families with gene numbers of the families shown above the bars.

The results from the enrichment analysis of GO terms and KEGG pathways also showed the expanded gene families belong to the categories of digestion, such as lipid metabolic process and phosphatidic acid biosynthetic process (Supplementary Figure S1A), and PPAR signaling pathway and Biosynthesis of unsaturated fatty acids (Supplementary Figure S1B).

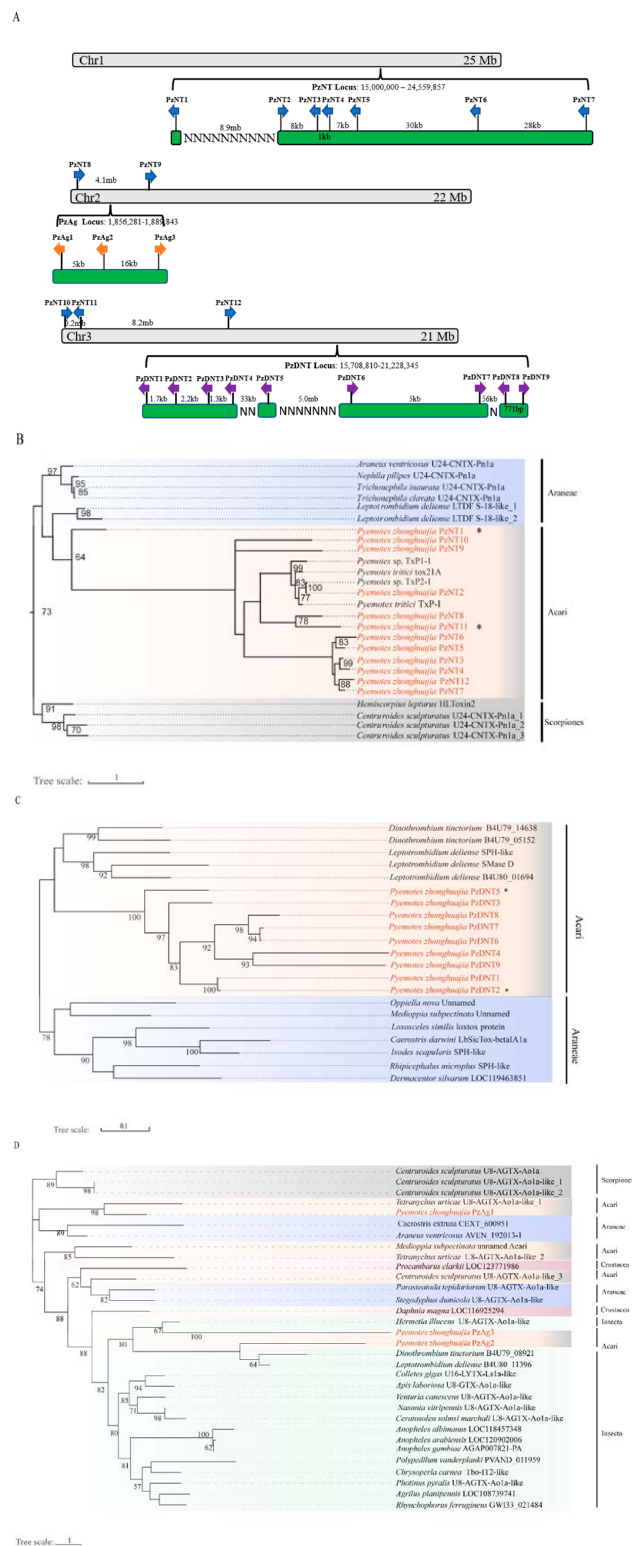
This suggests some of the gene family expansions are likely to be involved in the extensive feeding habits. The expanded chitin-binding families (Supplementary Figure S1A) may be related to intense enlargement of the parasitoid body during the reproductive period [33].

#### 2.4. Neurotoxin, Dermonecrotic Toxin and Agatoxin Genes

It is interesting that we found significant expansion of neurotoxin and dermonecrotic toxin gene families in the *P. zhonghuajia* genome. Neurotoxins in spiders were found as the main component of the venom that targets the prey's ion channels leading to paralysis or death [34]. The *P. zhonghuajia* neurotoxin gene family was expanded within the Acari clade (Figure 4B). The 12 neurotoxin genes were present on all the three chromosomes with a six-gene locus being located near the end of chromosome 1 (Figure 4A). *PzNT3* and *4* were located nearby and closely related on the phylogenetic tree with nearly 85% sequence similarity (Supplementary Table S6), suggesting they were recently duplicated under a tandem duplication event. Similar observations were from *PzNT5* and *6*. Located on the two different chromosomes, a recent duplication event might have also occurred between *PzNT7* and *12* as they were closely related on the phylogenetic tree with high sequence similarity. *PzNT2* was present as a homologous sequence to *P. tritici* TxP2-1 (80% sequence similarity) and they were closely related to *P. tritici* TxP-1 (encoded by *Tox34*) that was found playing a role in paralyzing and even killing insects [26,35]. Most neurotoxin genes (except *PzNT1* and *11*) have a higher number of reads aligned from RNA-Seq data obtained from the one-day-old mites compared with the seven-day-pregnant mites, indicating higher levels of gene expression (Supplementary Figure S2). *PzNT1* was located distantly from the six-gene locus on chromosome 1 and had shown to be distantly related to the rest of the neurotoxin genes on the phylogenetic tree. It is possible that there was a functional divergence of this gene compared with the rest of the family members.

Dermonecrotic toxins have been demonstrated to cause dermonecrotic lesions on the prey in spiders [21] and it is likely that the dermonecrotic toxins of *P. zhonghuajia* may also cause the dermonecrotic lesion of the host. In *P. zhonghuajia*, all nine dermonecrotic toxin genes were present roughly close to each other on the second half of chromosome 3 (Figure 4A). Our phylogenetic tree showed these genes formed a single clade with the gene family expansion occurring after Pyemotes separated from Leptotrombidium and Dinotrombidium (Figure 4C). *PzDNT1* and *2* were located next to each other and showed to be closely related on the phylogenetic tree with over 97% sequence similarity (Supplementary Table S8), suggesting a recent tandem gene duplication event. However, their expression divergence (higher expression of *PzDNT1* in the one-day-old mites compared with seven-day-pregnant mites and *PzDNT2* is opposite) suggests a functional difference. A recent duplication event was also shown between *PzDNT6* and *7*, both of which exhibited similar expression patterns in the two adult forms. Similar to *PzDNT2*, *PzDNT5* was also shown with higher expression level in the seven-day-pregnant mites compared with the one-day-old mites.

$\alpha$ -agatoxins was found to paralyze the prey in funnel web spiders [22]. There were three agatoxin genes identified in the *P. zhonghuajia* that formed a gene cluster on chromosome 2, which, with each other, has shown relatively distantly related on the phylogenetic tree (Figure 4D). *PzAg2* and *3* were separated from *PzAg1* at a very early stage during evolution. All three genes exhibited higher expression in the one-day-old mites compared with seven-day-pregnant mites, suggesting a role in paralyzing the hosts.



**Figure 4.** Distribution of toxin genes on the chromosomes and phylogenetic trees of the three toxin gene classes. (A) Distribution of toxin genes on the chromosomes. Blue: neurotoxin genes; Orange: agatoxin genes; Purple: dermonecrotic toxin genes. (B–D) Phylogenetic trees of neurotoxin genes, dermonecrotic toxin genes, and agatoxin genes. The “\*” indicates the gene was shown with higher expression in the seven-day-pregnant mites compared with the one-day-old mites. The accession numbers of all the sequences used in the phylogenies are listed in Supplementary Table S12.

### 3. Conclusions

Pyemotidae is a significant family with several species being natural enemies, such as *P. tritici* and *P. zhonghuaajia*. They can paralyze and even cause the death of the stored product insects and agriculture and forestry insects. Understanding the composition and function of toxin-related proteins of *P. zhonghuaajia* is crucial to improve its predation efficiency and it can also provide the knowledge for the potential of transferring the mite toxin-related genes into crop genomes for pest control. In this study, we have generated a chromosome-level genome assembly of *P. zhonghuaajia*, which is the first whole-genome assembly in Pyemotidae that provides an important genomic resource for the study of biocontrol potential and ecological importance. The two gene families encoding neurotoxins and dermonecrotic toxins were found with significant expansion within the genus of *Pyemotes*, which also formed several gene clusters on the chromosomes. It is possible that gene expansion provides a high dosage of toxin proteins that are released during parasitic process. Gene expansion might also result in a highly diverged population of toxins that enable the mite to have a wide range of hosts. Several recent gene duplication events that we observed from the two gene phylogenies indicate they may underlie key adaptive events in the evolution of *P. zhonghuaajia*. All the toxin-related genes including the three agatoxin genes were shown with expression in the adults, and most of them exhibited higher level of expression in the one-day-old mites compared with seven-day-pregnant mites. This matches our observations on the one-day-old mite forms paralyzing and killing the hosts such as *S. frugiperda*, *M. separata*, and *S. litura* [3,6,7]. Future research will focus on confirming the presence of the proteins encoded by these genes through proteomic studies and functional characterization of the proteins through protein purification approaches and feeding experiments in *P. zhonghuaajia*. Similar studies will also be conducted in other venomous ectoparasitic species from *Pyemotes*.

### 4. Materials and Methods

#### 4.1. Sample Collection and Sequencing

Colonies of *P. zhonghuaajia* were reared on mature larvae of *Sitotroga cerealella* (Oliver) (Lepidoptera: Gelechiidae) with wheat bran in a climate chamber at  $25 \pm 1$  °C with  $60 \pm 5\%$  relative humidity (RH) at Changli Institute of Pomology, Hebei Academy of Agriculture and Forestry Sciences. There were 2000 seven-day-pregnant mites used for Illumina whole-genome and PacBio sequencing, respectively. Genomic DNA was extracted using the QIAGEN DNeasy Blood & Tissue kit, which was then used to construct a 350 bp insert-size library using the Truseq DNA PCR-free kit for sequencing on the Illumina NovaSeq 6000 platform and a 15 kb insert-size library using the SMRTbell™ Template Prep Kit 2.0 for sequencing on the PacBio Sequel II platform. The whole-individual transcriptome was performed using RNA-Seq from 2000 one-day-old mites and 2000 seven-day-pregnant mites with three biological replicates for each group. Total RNA was extracted using the TRIzol™ Reagent kit and the RNA-Seq library was constructed using TruSeq RNA v2 kit. DNA/RNA extraction, library construction, and sequencing were performed at Berry Genomic (Beijing, China).

#### 4.2. Genome Assembly

Illumina raw reads were cleaned using two tools under BBTools v38.67 [36] with the following steps: (1) removing duplicated reads; (2) trimming low-quality reads; (3) removing poly-A/G/C tails; (4) filtering reads less than 15 bp; and (5) correcting reads based on overlapping ends between pairs. The tool Clumpify was used for step (1) and steps (2) to (5) were performed using BBDuk with parameters “qtrim = rl trimq = 20 minlen = 15 ecco = t maxns = 5 trimpolya = 10 trimpolyg = 10 trimpolyc = 10”. K-mer analysis based on Illumina short reads was performed using BBNorm (*k-mer*: 21) and the k-mer profile was visualized using the online version of Genomescope v2.0 [37] with parameters “-k 21 -p 2m 1000”. A preliminary PacBio long-read assembly was performed using Flye v2.7.1 [38] with parameters “-i 2-m 3000”. Purge\_Dups v1.0.0 [39] was used to remove allelic contigs



based on the read depth with a minimum alignment score of 70 after the long reads were mapped back to the assembly with Minimap2 (v2.17) [40]. Illumina short reads were used for two rounds of contig polishing performed by NextPolish (v1.1.0) [41] after mapping the reads back to the assembly using Minimap2. The contaminated contigs were assessed and removed using “blastn” from BLAST+ (v2.9.1) [42] with the sequence similarity search against *nt* and UniVec databases (both were downloaded in December 2020). The cleaned contigs were then uploaded to NCBI for an additional check of contamination. Assembly completeness was estimated using BUSCO (v3.1.0) [43] with the sequence similarity search against the arthropod single-copy gene set (arthropoda\_odb9: n = 1367). To estimate the mapping rate from raw reads, both short and long genomic reads as well as RNA-Seq short reads were aligned back to the genome assembly using Minimap2.

#### 4.3. Genome Annotation

Three essential genomic elements of *P. zhonghuajia* genome: repetitive elements, non-coding RNAs (ncRNAs), and protein-coding genes were annotated. To annotate repeats, RepeatModeler v2.0.1 [44] with LTR search process (-LTRStruct) was used to generate a de novo repeat library, which was then combined with Dfam 3.1 [45] and RepBase-20181026 [46] databases to form a custom repeat library. The repeat-masked genome assembly was produced using RepeatMasker v4.0.9 [47].

The MAKER v2.31.10 [48] pipeline was used to predict protein-coding genes by integrating ab initio, transcript- and homology-based evidence. The ab initio prediction was generated by a BRAKER v2.1.5 [49] pipeline to train Augustus v3.3.3 [50] and GeneMarkES/ET/EP 4.48\_3.60\_lic [51] with the utilization of RNA-Seq data and protein sequences to increase prediction accuracy. The input alignments from mapping RNA-seq data to the genome assembly were produced using HISAT2 v2.2.0 [52] and the arthropod protein sequences were obtained from OrthoDB10 v1 database [53]. The genome-guided assembler StringTie v2.1.2 [54] was used to assemble transcripts as transcriptome evidence to integrate in MAKER. The protein sequences that were also utilized by MAKER for final prediction were downloaded from NCBI including the sequences from *Drosophila melanogaster* (Insecta), *Daphnia magna* (Crustacea), *Ixodes scapularis*, *Varroa destructor*, *Tetranychus urticae*, and *Dermatophagoides pteronyssinus* (Acari). Gene functional annotation was performed using Diamond v0.9.24 [55] from searching against the UniProtKB database with the sensitive mode “-more-sensitive -e 1e<sup>-5</sup>”. InterProScan 5.41–78.0 [56] was used to search against the databases including Pfam [57], SMART [58], Gene3D [59], Superfamily [60] and CDD [61]. The protein domains, gene ontology (GO), and gene pathways (KEGG, Reactome) were annotated using ggnog-mapper v2.0.1 [62] from searching against ggnog v5.0 [63].

NcRNAs including rRNA, snRNA, and miRNA were identified using infernal v1.1.3 [64] by searching sequence similarity against Rfam database. tRNAs were predicted using tRNAscan-SE v2.0.6 [65] and the high-confident sequences were maintained as the final tRNA set by the tRNAscan-SE script “EukHighConfidenceFilter”.

#### 4.4. Gene Ontology Analysis and Species Evolution

The protein sequences from 11 species across three orders (Araneae: Stegodyphus dumicola; Scorpiones: Centruroides sculpturatus; Arachnoidea: Dermatophagoides pteronyssinus, Dinothrombium tinctorium, Galendromus occidentalis, Ixodes scapularis, Leptotrombidium deliense, Sarcoptes scabiei, T. urticae, Tropilaelaps mercedesae, and Varroa destructor) were downloaded from NCBI in December 2020 together with the annotated protein set from *P. zhonghuajia* for gene family orthology inference using OrthoFinder v2.3.8 [66] after aligning sequences using Diamond. The resulting single-copy orthologs from OrthoFinder were used for phylogenetic analysis. The sequence alignment as input for phylogenetic construction was generated using the following steps: (1) aligning orthologous protein sequences using MAFFT v7.394 [67] with “L-INS-I”; (2) filtering ambiguous aligned regions using BMGE v1.12 [68] with the parameters “-m BLOSUM90 -h 0.4”; (3) Concatenating all the protein alignments generated above using FASconCAT-G v1.04 [69] as the input

for phylogenetic construction. The species phylogeny was constructed using IQ-TREE v2.0-rc2 [70] with the parameters “-m MFP -mset LG -msub nuclear -rclusterf 10 -B 1000 -alrt 1000 -symtest-remove-bad -symtestpval 0.10”. The estimated time of species divergence was calculated using MCMCTree from PAML v4.9j [71] with parameters “clock = 2, BDparas = 1 1 0.1, kappa\_gamma = 6 2, alpha\_gamma = 1 1, rgene\_gamma = 2 20 1, sigma2\_gamma = 1 10 1”. There were four fossil evidences downloaded from the PBDB database (<https://www.paleobiodb.org/navigator/>, accessed on 14 August 2022) as calibrations used in this estimation above: *Allopalaeophonius caledonicus* (4.305–4.438) from the order Scorpiones as the root, *Pseudoprotacarus scoticus* (4.076–4.192) from Arachnida, *Carbolohmannia maimaiphilus* (3.114–3.232) from Acariformes, and *Deinocroton draculi* (0.935–1.455) from Mesostigmata.

#### 4.5. Identification of Gene Family Expansion and Contraction

The gene family expansion and contraction in *P. zhonghuajia* genome compared with the 11 species used for phylogenetic construction were estimated using CAFÉ v4.2.1 [72] with the model of single birth–death parameter lambda and a significance level of 0.01 ( $p = 0.01$ ). The identified significantly expanded gene families were then assigned with GO and KEGG categories using R package clusterProfiler v3.10.1 [73] with the default parameters ( $p = 0.01$  and  $q = 0.05$ ).

#### 4.6. Phylogeny Construction and Gene Expression of Toxin-Related Gene Families

The *P. zhonghuajia* neurotoxin, dermonecrotic toxin, and agatoxin protein sequences were predicted from the genome assembly using BITACORA v1.2 [74] based on homology searches using the sequences from Chelicerata and Myriapoda downloaded from NCBI RefSeq (Supplementary Table S11) and confirmed by searching against protein database using the online blastp ([https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE\\_TYPE=BlastSearch&LINK\\_LOC=blasthome](https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome), accessed on 14 August 2022). The HMM profiles generated from HMMER v3.2.1 [75] using “hmmbuild” were used in BITACORA. Multiple alignments were performed using Geneious prime V 2021.1.1 (created by Biomatters. Available from <https://www.geneious.com>, accessed on 5 November 2020) with the method of Clustal Omega. FastTree [76] was used to construct neurotoxin, dermonecrotic toxin, and agatoxin protein phylogenies based on the maximum likelihood method. We used RNA sequencing data to detect the expression of toxin-related genes in seven-day-pregnant mites and one-day-old mites. The estimation of gene expression levels was performed using Salmon [77] with the Gaussian axial fluctuation (GAF) model to generate normalized read counts. The gene expression heatmap was generated using the heatmap function (heatmap) from the R package NMF [78] (Supplementary Figure S2).

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/xxx/s1>, Table S1: Genome survey based on k-mer distribution for *pyemotes zhonghuajia* (excel file), Table S2: Repeat annotation in the *Pyemotes zhonghuajia* genome (excel file), Table S3: Annotations of non-coding RNAs in the *Pyemotes zhonghuajia* genome (excel file), Table S4: Significantly expanded gene families (excel file), Table S5: Expanded and contracted gene families in each species. Table S6: The identity of neurotoxins of *P. zhonghuajia* with *P. tritici* (excel file), Table S7: The identity of agatoxins of *P. zhonghuajia* with other species (excel file), Table S8: The identity of dermonecrotic of *P. zhonghuajia* with other species (excel file), Table S9: Information of the 24 toxin-related genes, Table S10: Transcript per million (TPM) values (excel file), Table S11: Gene IDs of the reference species, Table S12: The accession numbers of all the sequences used in the phylogenies (excel file). Figure S1: GO (A) and KEGG (B) function enrichment of significantly expanded gene families. Figure S2: Heatmap of the identified toxin-related genes (pdf file).

**Author Contributions:** Conceptualization, Y.-F.S. and J.-F.L.; methodology, Y.-F.S., M.-F.Y., J.-F.L. and S.Y.; software, Y.-F.S., C.W. and B.Y.; resources, Y.-F.S., S.Y., L.-T.L., M.-F.Y., L.-C.Y. and J.-F.L.; writing—original draft preparation, Y.-F.S.; writing—review and editing, J.-F.L., C.W. and Y.-F.S.; supervision, J.-F.L. and C.W.; project administration, J.-F.L.; funding acquisition, J.-F.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China (32060637 and 31770694), Guizhou Province Science and Technology Innovation Talent Team Project (Qian Ke He Pingtai Rencai-CXTD [2021]004), High-level Talent Innovation and Entrepreneurship Funding Project in Guizhou Province ([2021]01), the Growth Project of Youth Talent in Ordinary Universities in Guizhou Province ([2021]079), Natural Science Special Project in Guizhou University (Special post, [2020]-02) and Hebei Provincial Key R&D Program (20326517D).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Genome assembly and raw sequencing data have been deposited at the NCBI under the accessions JACCHO000000000 and SRR12261228—SRR12261230, respectively. Genome annotations are available at the Figshare under the link: [https://figshare.com/articles/dataset/A\\_Long-Read\\_Genome\\_Assembly\\_of\\_a\\_Native\\_Mite\\_in\\_China\\_Pyemotes\\_Zhonghuaajia\\_Yu\\_Zhang\\_He\\_Prostigmata\\_Pyemotidae\\_Item/20518008](https://figshare.com/articles/dataset/A_Long-Read_Genome_Assembly_of_a_Native_Mite_in_China_Pyemotes_Zhonghuaajia_Yu_Zhang_He_Prostigmata_Pyemotidae_Item/20518008) (accessed on 18 August 2022).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Yu, L.; Zhang, Z.Q.; He, L. Two New Species of *Pyemotes* Closely Related to *P. Tritici* (Acari: *Pyemotidae*). *Zootaxa* **2010**, *2723*, 1–40. [[CrossRef](#)]
2. He, L.M.; Jiao, R.; Xu, C.X.; Hao, B.F.; Han, J.C.; Yu, L.C. Application of mtDNA COI Gene Sequence in Identification of *Pyemotes*. *J. Hebei Agric. Sci.* **2010**, *14*, 46–50.
3. Liu, J.F.; Tian, T.A.; Li, X.L.; Chen, Y.C.; Yu, X.F.; Tan, X.F.; Zhu, Y.; Yang, M.F. Is *Pyemotes Zhonghuaajia* (Acari: *Pyemotidae*) a Suitable Biological Control Agent against the Fall Armyworm *Spodoptera Frugiperda* (Lepidoptera: *Noctuidae*)? *Syst. Appl. Acarol.* **2020**, *25*, 649–657. [[CrossRef](#)]
4. Li, L.; He, L.; Yu, L.; He, X.Z.; Xu, C.; Jiao, R.; Zhang, L.; Liu, J. Preliminary Study on the Potential of *Pyemotes Zhonghuaajia* (Acari: *Pyemotidae*) in Biological Control of *Aphis Citricola* (Hemiptera: *Aphididae*). *Syst. Appl. Acarol.* **2019**, *24*, 1116–1120. [[CrossRef](#)]
5. De Sousa, A.H.; Mendonça, G.R.Q.; Lopes, L.M.; Faroni, L.R.D. Widespread Infestation of *Pyemotes Tritici* (Acari: *Pyemotidae*) in Colonies of Seven Species of Stored-Product Insects. *Genet. Mol. Res.* **2020**, *19*, 1–5. [[CrossRef](#)]
6. Tian, T.A.; Yu, L.; Sun, G.J.; Yu, X.F.; Li, L.; Wu, C.X.; Chen, Y.C.; Yang, M.F.; Liu, J.F. Biological Control Efficiency of an Ectoparasitic Mite *Pyemotes Zhonghuaajia* on Oriental Armyworm *Mythimna Separata*. *Syst. Appl. Acarol.* **2020**, *25*, 1683–1692.
7. Chen, Y.C.; Tian, T.A.; Chen, Y.H.; Yu, L.C.; Hu, J.F.; Yu, X.F.; Liu, J.F.; Yang, M.F. The Biocontrol Agent *Pyemotes zhonghuaajia* Has the Highest Lethal Weight Ratio Compared with Its Prey and the Most Dramatic Body Weight Change during Pregnancy. *Insects* **2021**, *12*, 490. [[CrossRef](#)]
8. Tomalski, M.D.; Bruce, W.A.; Travis, J.; Blum, M.S. Preliminary Characterization of Toxins from the Straw Itch Mite, *Pyemotes Tritici*, Which Induce Paralysis in the Larvae of a Moth. *Toxicon* **1988**, *26*, 127–132. [[CrossRef](#)]
9. Han, J.C.; He, L.M.; Jiao, R.; Hao, B.F.; Xu, Z.X.; Yu, L.C. Analysis of the toxin gene analogs cloned from *Pyemotes phloeosinus* sp. nov. *J. Hebei Agric. Sci.* **2008**, *12*, 72–74.
10. Lu, H.L.; Li, L.T.; Yu, L.C.; He, L.M.; Ouyang, G.C.; Liang, G.W.; Lu, Y.Y. Ectoparasitic mite, *Pyemotes zhonghuaajia* (Prostigmata: *Pyemotidae*), for biological control of Asian citrus psyllid, *Diaphorina citri* (Hemiptera: *Liviidae*). *Syst. Appl. Acarol.* **2019**, *24*, 520–524. [[CrossRef](#)]
11. Guo, X.; Xu, Z.; Xiong, D.P. Study of utilizing *Pyemotes zhonghuaajia* to control *Semanotus bifasciatus* beetles. *Chin. Bull. Entomol.* **2010**, *47*, 529–532.
12. Li, Y.C.; Huang, C.Y.; Xia, Z.R.; Zhou, J.S.; Li, J.S.; Sun, Y.Z.; Xu, Y.; Wu, S.Q.; Zhang, F.P. Preliminary Study on Biocontrol Potential of *Pyemotes Zhonghuaajia* on *Monochamus Alternatus* Hope. *Genom. Appl. Biol.* **2019**, *38*, 2516–2521.
13. Zhang, Z.S.; Xiong, D.P.; Cheng, W. Control of Stem Borers by a Parasitoid, *Pyemotes tritici* Lagreze-Fossot & Montane. *Chin. J. Biol. Control.* **2008**, *01*, 1–6.
14. Deng, S.Q.; Chen, J.T.; Li, W.W.; Chen, M.; Peng, H.J. Application of the scorpion neurotoxin AaIT against insect pests. *Int. J. Mol. Sci.* **2019**, *20*, 3467. [[CrossRef](#)]
15. Harvey-Samuel, T.; Xu, X.; Lovett, E.; Dafa'alla, T.; Walker, A.; Norman, V.C.; Carter, R.; Teal, J.; Akilan, L.; Leftwich, P.T.; et al. Engineered expression of the invertebrate-specific scorpion toxin AaHIT reduces adult longevity and female fecundity in the diamondback moth *Plutella xylostella*. *Pest Manag. Sci.* **2021**, *77*, 3154–3164. [[CrossRef](#)]
16. Li, H.; Xia, Y. Improving the secretory expression of active recombinant AaIT in *Pichia pastoris* by changing the expression strain and plasmid. *World J. Microbiol. Biotechnol.* **2018**, *34*, 104. [[CrossRef](#)]
17. Tianpei, X.; Li, D.; Qiu, P.; Luo, J.; Zhu, Y.; Li, S. Scorpion peptide LqhIT2 activates phenylpropanoid pathways via jasmonate to increase rice resistance to rice leafrollers. *Plant Sci.* **2015**, *230*, 1–11. [[CrossRef](#)]
18. Zlotkin, E.; Eitan, M.; Bindokas, V.P.; Adams, M.E.; Moyer, M.; Burkhart, W.; Fowler, E. Functional duality and structural uniqueness of the depressant insect-selective neurotoxins. *Biochemistry* **1991**, *30*, 4814–4821. [[CrossRef](#)]

19. Kroemer, J.A.; Bonning, B.C.; Harrison, R.L. Expression, delivery and function of insecticidal proteins expressed by recombinant baculoviruses. *Viruses* **2015**, *7*, 422–455. [[CrossRef](#)]
20. Windley, M.J.; Herzig, V.; Dziemborowicz, S.A.; Hardy, M.C.; King, G.F.; Nicholson, G.M. Spider-venom peptides as bioinsecticides. *Toxins* **2012**, *4*, 191–227. [[CrossRef](#)]
21. Chaves-Moreira, D.; Senff-Ribeiro, A.; Wille, A.C.M.; Gremski, L.H.; Chaim, O.M.; Veiga, S.S. Highlights in the knowledge of brown spider toxins. *J. Venom. Anim. Toxins Incl. Trop. Dis.* **2017**, *23*, 6. [[CrossRef](#)] [[PubMed](#)]
22. Adams, M.E. Agatoxins: Ion channel specific toxins from the American funnel web spider, *Agelenopsis aperta*. *Toxicon* **2004**, *43*, 509–525. [[CrossRef](#)] [[PubMed](#)]
23. Nentwig, W.; Kuhn-Nentwig, L. Main components of spider venoms. In *Spider Ecophysiology*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 191–202.
24. Kuhn-Nentwig, L.; Stöcklin, R.; Nentwig, W. Venom composition and strategies in spiders: Is everything possible? *Adv. Insect Phys.* **2011**, *40*, 1–86.
25. Tomalski, M.D.; Hutchinson, K.; Todd, J.; Miller, L.K. Identification and Characterization of Tox21A: A Mite CDNA Encoding a Paralytic Neurotoxin Related to TxP-I. *Toxicon* **1993**, *31*, 319–326. [[CrossRef](#)]
26. Burden, J.P.; Hails, R.S.; Windass, J.D.; Suner, M.M.; Cory, J.S. Infectivity, Speed of Kill, and Productivity of a Baculovirus Expressing the Itch Mite Toxin Txp-1 in Second and Fourth Instar Larvae of *Trichoplusia Ni*. *J. Invertebr. Pathol.* **2000**, *75*, 226–236. [[CrossRef](#)] [[PubMed](#)]
27. Yan, Y.; Zhang, N.; Liu, C.; Wu, X.; Liu, K.; Yin, Z.; Zhou, X.; Xie, L. A Highly Contiguous Genome Assembly of a Polyphagous Predatory Mite *Stratiolaelaps scimitus* (Womersley) (Acari: Laelapidae). *Genome Biol. Evol.* **2021**, *13*, evab011. [[CrossRef](#)] [[PubMed](#)]
28. Hollenstein, K.; Dawson, R.J.P.; Locher, K.P. Structure and mechanism of ABC transporter proteins. *Curr. Opin. Struct. Biol.* **2007**, *17*, 412–418. [[CrossRef](#)] [[PubMed](#)]
29. Jimenez-Acosta, F.; Planas, L.; Penneys, N. Demodex mites contain immunoreactive lipase. *Arch. Dermatol.* **1989**, *125*, 1436–1437. [[CrossRef](#)]
30. Ando, T.; Homma, R.; Ino, Y.; Ito, G.; Miyahara, A.; Yanagihara, T.; Kimura, H.; Ikeda, S.; Yamakawa, H.; Iwaki, M.; et al. Trypsin-like protease of mites: Purification and characterization of trypsin-like protease from mite faecal extract *Dermatophagoides farinae*. Relationship between trypsin-like protease and Der f III. *Clin. Exp. Allergy* **1993**, *23*, 777–784. [[CrossRef](#)]
31. Chaim, O.M.; Sade, Y.B.; Da Silveira, R.B.; Toma, L.; Kalapothakis, E.; Chávez-Olórtegui, C.; Mangili, O.C.; Gremski, W.; Von Dietrich, C.P.; Nader, H.B.; et al. Brown Spider Dermonecrotic Toxin Directly Induces Nephrotoxicity. *Toxicol. Appl. Pharmacol.* **2006**, *211*, 64–77. [[CrossRef](#)]
32. Krantz, G.W. Dolichocybe Keiferi, A New Genus and New Species of Pyemotid Mite, with a Description of a New Species of *Siteroptes* (Acarina: Pyemotidae). *Ann. Entomol. Soc. Am.* **1957**, *50*, 259–264. [[CrossRef](#)]
33. Shen, Z.; Jacobs-Lorena, M. Evolution of Chitin-Binding Proteins in Invertebrates. *J. Mol. Evol.* **1999**, *48*, 341–347. [[CrossRef](#)] [[PubMed](#)]
34. Peigneur, S.; Tytgat, J. Toxins in Drug Discovery and Pharmacology. *Toxins* **2018**, *10*, 126. [[CrossRef](#)] [[PubMed](#)]
35. Kuhn-Nentwig, L.; Langenegger, N.; Heller, M.; Koua, D.; Nentwig, W. The Dual Prey-Inactivation Strategy of Spiders-In-Depth Venomic Analysis of *Cupiennius salei*. *Toxins* **2019**, *11*, 167. [[CrossRef](#)]
36. Bushnell, B. BBtools. Available online: <https://sourceforge.net/projects/bbmap/> (accessed on 4 April 2020).
37. Vurture, G.W.; Sedlazeck, F.J.; Nattestad, M.; Underwood, C.J.; Fang, H.; Gurtowski, J.; Schatz, M.C. GenomeScope: Fast reference-free genome profiling from short reads. *Bioinformatics* **2017**, *33*, 2202–2204. [[CrossRef](#)]
38. Kolmogorov, M.; Yuan, J.; Lin, Y.; Pevzner, P.A. Assembly of long errorprone reads using repeat graphs. *Nat. Biotechnol.* **2019**, *37*, 540–546. [[CrossRef](#)]
39. Guan, D.; McCarthy, S.A.; Wood, J.; Howe, K.; Wang, Y.; Durbin, R. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **2020**, *36*, 2896–2898. [[CrossRef](#)]
40. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **2018**, *34*, 3094–3100. [[CrossRef](#)]
41. Hu, J.; Fan, J.; Sun, Z.; Liu, S. NextPolish: A fast and efficient genome polishing tool for long read assembly. *Bioinformatics* **2020**, *36*, 2253–2255. [[CrossRef](#)]
42. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: Architecture and applications. *BMC Bioinform.* **2009**, *10*, 1. [[CrossRef](#)]
43. Waterhouse, R.M.; Seppely, M.; Simao, F.A.; Manni, M.; Ioannidis, P.; Klioutchnikov, G.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Mol. Biol. Evol.* **2018**, *35*, 543–548. [[CrossRef](#)] [[PubMed](#)]
44. Flynn, J.M.; Hubley, R.; Goubert, C.; Rosen, J.; Clark, A.G.; Feschotte, C.; Smit, A.F. RepeatModeler2 for Automated Genomic Discovery of Transposable Element Families. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 9451–9457. [[CrossRef](#)] [[PubMed](#)]
45. Hubley, R.; Finn, R.D.; Clements, J.; Eddy, S.R.; Jones, T.A.; Bao, W.; Smit, A.F.A.; Wheeler, T.J. The Dfam Database of Repetitive DNA Families. *Nucleic Acids Res.* **2016**, *44*, D81–D89. [[CrossRef](#)]
46. Bao, W.D.; Kojima, K.K.; Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **2015**, *6*, 11. [[CrossRef](#)]
47. Repeat Masker Open-4.0. 2013–2015. Available online: <http://www.repeatmasker.org> (accessed on 8 January 2020).

48. Holt, C.; Yandell, M. MAKER2: An annotation pipeline and genome- database management tool for second-generation genome projects. *BMC Bioinform.* **2011**, *12*, 491. [[CrossRef](#)]
49. Hoff, K.J.; 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. [[CrossRef](#)] [[PubMed](#)]
50. Tanke, M.; Steinkamp, R.; Waack, S.; Morgenstern, B. AUGUSTUS: A web server for gene finding in eukaryotes. *Nucleic Acids Res.* **2004**, *32*, W309–W312.
51. Lomsadze, A.; Ter-Hovhannisyan, V.; Chernoff, Y.O.; Borodovsky, M. Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Res.* **2005**, *33*, 6494–6506. [[CrossRef](#)] [[PubMed](#)]
52. Kim, D.; Paggi, J.M.; Park, C.; Bennett, C.; Salzberg, S.L. Graph-Based Genome Alignment and Genotyping with HISAT2 and HISAT-Genotype. *Nat. Biotechnol.* **2019**, *37*, 907–915. [[CrossRef](#)]
53. Kriventseva, E.V.; Kuznetsov, D.; Tegenfeldt, F.; Manni, M.; Dias, R.; Simão, F.A.; Zdobnov, E.M. 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*, D807–D811. [[CrossRef](#)]
54. Kovaka, S.; Zimin, A.V.; Pertea, G.M.; Razaghi, R.; Salzberg, S.L.; Pertea, M. Transcriptome Assembly from Long-Read RNA-Seq Alignments with StringTie2. *Genome Biol.* **2019**, *20*, 278. [[CrossRef](#)] [[PubMed](#)]
55. Buchfink, B.; Xie, C.; Huson, D.H. Fast and Sensitive Protein Alignment Using DIAMOND. *Nat. Methods.* **2014**, *12*, 59–60. [[CrossRef](#)] [[PubMed](#)]
56. Finn, R.D.; Finn, R.D.; Attwood, T.K.; Babbitt, P.C.; Bateman, A.; Bork, P.; Bridge, A.J.; Chang, H.Y.; Dosztányi, Z.; El-Gebali, S.; et al. InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res.* **2017**, *45*, D190–D199. [[CrossRef](#)]
57. El-Gebali, S.; Mistry, J.; Bateman, A.; Eddy, S.R.; Luciani, A.; Potter, S.C.; Qureshi, M.; Richardson, L.J.; Salazar, G.A.; Smart, A.; et al. The Pfam Protein Families Database in 2019. *Nucleic Acids Res.* **2019**, *47*, D427–D432. [[CrossRef](#)] [[PubMed](#)]
58. Letunic, L.; Bork, P. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res.* **2018**, *46*, D493–D496. [[CrossRef](#)] [[PubMed](#)]
59. Lewis, T.E.; Sillitoe, I.; Dawson, N.; Lam, S.D.; Clarke, T.; Lee, D.; Orengo, C.; Lees, J. Gene3D: Extensive Prediction of Globular Domains in Proteins. *Nucleic Acids Res.* **2018**, *46*, D435–D439. [[CrossRef](#)] [[PubMed](#)]
60. Wilson, D.; Pethica, R.; Zhou, Y.; Talbot, C.; Vogel, C.; Madera, M.; Chothia, C.; Gough, J. SUPERFAMILY—Sophisticated Comparative Genomics, Data Mining, Visualization and Phylogeny. *Nucleic Acids Res.* **2009**, *37*, 380–386. [[CrossRef](#)]
61. Marchler-Bauer, A.; Bo, Y.; Han, L.; He, J.; Lanczycki, C.J.; Lu, S.; Chitsaz, F.; Derbyshire, M.K.; Geer, R.C.; Gonzales, N.R.; et al. CDD/SPARCLE: Functional Classification of Proteins via Subfamily Domain Architectures. *Nucleic Acids Res.* **2017**, *45*, D200–D203. [[CrossRef](#)]
62. Huerta-Cepas, J.; Forslund, K.; Coelho, L.P.; Szklarczyk, D.; Jensen, L.J.; von Mering, C.; Bork, P. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* **2017**, *34*, 2115–2122. [[CrossRef](#)]
63. Huerta-Cepas, J.; Szklarczyk, D.; Heller, D.; Hernández-Plaza, A.; Forslund, S.K.; Cook, H.; Mende, D.R.; Letunic, I.; Rattei, T.; Jensen, L.J.; 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*, D309–D314. [[CrossRef](#)]
64. Nawrocki, E.P.; Eddy, S.R. Infernal 1.1: 100-Fold Faster RNA Homology Searches. *Bioinformatics* **2013**, *29*, 2933–2935. [[CrossRef](#)] [[PubMed](#)]
65. Chan, P.P.; Lowe, T.M. tRNAscan-SE: Searching for tRNA genes in genomic sequences. In *Methods in Molecular Biology*; Humana Press: Totowa, NJ, USA, 2019; Volume 1962, pp. 1–14.
66. Emms, D.M.; Kelly, S. OrthoFinder: Phylogenetic Orthology Inference for Comparative Genomics. *Genome Biol.* **2018**, *20*, 238. [[CrossRef](#)] [[PubMed](#)]
67. Katoh, K.; Standley, D.M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [[CrossRef](#)] [[PubMed](#)]
68. Criscuolo, A.; Gribaldo, S. BMGE (Block Mapping and Gathering with Entropy): A new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.* **2010**, *10*, 210. [[CrossRef](#)] [[PubMed](#)]
69. Kück, P.; Longo, G.C. FASconCAT-G: Extensive Functions for Multiple Sequence Alignment Preparations Concerning Phylogenetic Studies. *Front. Zool.* **2014**, *11*, 81. [[CrossRef](#)]
70. Minh, B.Q.; Schmidt, H.A.; Chernomor, O.; Schrempf, D.; Michael, D.; Haeseler, A.; Von Lanfear, R. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **2020**, *37*, 1530–1534. [[CrossRef](#)]
71. Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* **2007**, *24*, 1586–1591. [[CrossRef](#)]
72. Han, M.V.; Thomas, G.W.C.; Lugo-Martinez, J.; Hahn, M.W. Estimating Gene Gain and Loss Rates in the Presence of Error in Genome Assembly and Annotation Using CAFE 3. *Mol. Biol. Evol.* **2013**, *30*, 1987–1997. [[CrossRef](#)]
73. Yu, G.; Wang, L.G.; Han, Y.; He, Q.Y. ClusterProfiler: An R Package for Comparing Biological Themes among Gene Clusters. *OMICS A J. Integr. Biol.* **2012**, *16*, 284–287. [[CrossRef](#)]
74. Vizueta, J.; Sánchez-Gracia, A.; Rozas, J. BITACORA: A comprehensive tool for the identification and annotation of gene families in genome assemblies. *Mol. Ecol. Resour.* **2020**, *20*, 1445–1452. [[CrossRef](#)]
75. Potter, S.C.; Eddy, S.R.; Park, Y.; Lopez, R.; Finn, R.D.; Hmmer, T. HMMER Web Server: 2018 Update. *Nucleic Acids Res.* **2018**, *46*, 200–204. [[CrossRef](#)] [[PubMed](#)]

- 
76. Price, M.N.; Dehal, P.S.; Arkin, A.P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* **2010**, *5*, e9490. [[CrossRef](#)] [[PubMed](#)]
  77. Patro, R.; Duggal, G.; Love, M.I.; Irizarry, R.A.; Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **2017**, *14*, 417–419. [[CrossRef](#)]
  78. Gaujoux, R.; Seoighe, C. A Flexible R Package for Nonnegative Matrix Factorization. *BMC Bioinform.* **2010**, *11*, 367. [[CrossRef](#)]