

Article

Genomic Analysis Reveals the Role of New Genes in Venom Regulatory Network of Parasitoid Wasps

Bo Zhang ^{1,2}, Yifan Bu ^{1,2}, Jiqiang Song ^{1,2}, Bo Yuan ^{1,2}, Shan Xiao ^{1,2} , Fang Wang ^{1,2} , Qi Fang ^{1,2}, Gongyin Ye ^{1,2} , Yi Yang ^{1,2,*} and Xinhai Ye ^{3,4,*} 

¹ State Key Laboratory of Rice Biology and Breeding, Zhejiang University, Hangzhou 310058, China; zhangbo2021@zju.edu.cn (B.Z.); buyifan@zju.edu.cn (Y.B.)

² Ministry of Agricultural and Rural Affairs Key Laboratory of Molecular Biology of Crop Pathogens and Insects, Zhejiang University, Hangzhou 310058, China

³ College of Advanced Agriculture Science, Zhejiang A&F University, Hangzhou 311300, China

⁴ Zhejiang Key Laboratory of Biology and Ecological Regulation of Crop Pathogens and Insects, Zhejiang A&F University, Hangzhou 311300, China

* Correspondence: yylqy@zju.edu.cn (Y.Y.); yexinhai@zafu.edu.cn (X.Y.)

Simple Summary: Parasitoid wasps are insects that lay eggs on or inside other arthropods, evolving extraordinary abilities to avoid their hosts' defenses and use their resources. While the new genes are thought to drive adaptive evolution, their exact roles remain unknown. This study identified a set of new genes that emerged during the evolution of Pteromalidae wasps. Most of these new genes formed by gene duplication, while others emerged de novo birth. These genes are shorter, simpler in structure, and primarily enriched in reproductive organs and venom glands. A key finding revealed that one new gene acts as a central hub, coordinating with older genes to regulate venom-related gene network instead of making venom proteins directly. This shows how new genes collaborate with existing ones to create evolutionary innovations in parasitoid wasps.

Abstract: New genes play a critical role in phenotypic diversity and evolutionary innovation. Parasitoid wasps, a highly abundant and diverse group of insects, parasitize other arthropods and exhibit remarkable evolutionary adaptations, such as evading host immune responses and exploiting host resources. However, the specific contributions of new genes to their unique traits remain poorly understood. Here, we identified 480 new genes that emerged after the *Nasonia-Pteromalus* divergence. Among these, 272 (56.7%) originated through DNA-mediated duplication, representing the largest proportion, followed by 77 (16.0%) derived from RNA-mediated duplication and 131 (27.3%) that arose de novo. Comparative analysis revealed that these new genes generally have shorter coding sequences and fewer exons compared to single-copy older genes conserved in the seven parasitoid wasps. These new genes are predominantly expressed in the reproductive glands and exhibit venom gland-biased expression. Notably, gene co-expression network analysis further identified that a new gene may act as a hub by interacting with older genes to regulate venom-related networks rather than directly encoding venom proteins. Together, our findings provide novel insights into the role of new genes in driving venom innovation in parasitoid wasps.



Received: 5 March 2025

Revised: 22 April 2025

Accepted: 3 May 2025

Published: 7 May 2025

Citation: Zhang, B.; Bu, Y.; Song, J.; Yuan, B.; Xiao, S.; Wang, F.; Fang, Q.; Ye, G.; Yang, Y.; Ye, X. Genomic Analysis Reveals the Role of New Genes in Venom Regulatory Network of Parasitoid Wasps. *Insects* **2025**, *16*, 502. <https://doi.org/10.3390/insects16050502>

Copyright: © 2025 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/](https://creativecommons.org/licenses/by/4.0/)

<https://creativecommons.org/licenses/by/4.0/>).

Keywords: parasitoid wasp; new gene; venom; evolution; gene regulatory network

1. Introduction

The emergence of new genes has been proposed as a key driver of phenotypic diversity and innovation across the tree of life [1,2]. New genes arise in specific genomic loci within a species at a particular point in evolutionary time where they did not previously exist [3]. The birth of new genes occurs through diverse molecular mechanisms, encompassing gene duplication, transposable element protein domestication, lateral gene transfer, gene fusion, and de novo origination [4]. These diverse mechanisms have facilitated widespread new gene origination across taxa, from plants to animals [5–7]. In primates, the brain and testis serve as evolutionary hotspots for the recruitment of new genes, likely contributing to the development of unique traits, biological functions, and behaviors [8,9]. The birth of organ-specific new genes significantly enhances our understanding of the frequency with which these new genes contribute to phenotypic evolution.

The vast species diversity and evolutionary complexity of insects provide a critical framework for investigating genetic novelty and the role of new genes in adaptation. A well-known new gene, *Jingwei*, which was the first reported new gene in *Drosophila*, arising through retrotransposition and neofunctionalization. *Jingwei* acquired a novel expression pattern specifically in the testis [10]. Since this discovery, numerous studies have demonstrated that new genes can influence male courtship behaviors or fertility-related functions over various evolutionary timescales in *Drosophila*. Examples include *sphinx*, *nsr*, *Umbrea*, *saturn*, and *atlas* [11–14]. In addition, *Lushu*, as a new gene identified in *Plutella xylostella*, also exhibited the male-biased expression pattern, which potentially results in enhancing sperm competition [15]. While most new genes show prominent roles in male reproductive organs, some exceptions have been observed. For instance, a new duplicate gene of parasitoid wasp *Nasonia vitripennis* (*VenomY*) affected detoxification and immunity genes in envenomated fly hosts [16]. These findings suggest that new genes can act as reservoirs of genetic innovation, contributing to diverse biological functions and driving organ-specific adaptations.

Parasitoid wasps represent a highly diverse and abundant group of insects that obligately parasitize other arthropods [17,18]. Among the diverse biological characteristics of parasitoid wasps, venom is particularly hypothesized to represent key evolutionary innovations that have facilitated the ecological success of parasitoid wasps. Recent studies have revealed venom genes in parasitoid wasp have exhibited a high turnover rate, primarily driven by the co-option of other pre-existing genes [19,20]. However, systematic research on the role of new genes in the origination and evolution of venom in parasitoid wasps remains limited.

In this study, we performed whole-genome syntenic alignments and identified 480 new genes in the parasitoid wasp *Pteromalus puparum*, which diverged after *Nasonia-Pteromalus* lineage split approximately 8 million years ago (MYA). Our findings reveal that the evolutionary patterns of new genes in parasitoid wasps, including origination mechanisms and tissue-specific expression, are comparable to those observed in *Drosophila* and mammals. Notably, these new genes exhibit tissue-specific expression, being predominantly expressed in reproductive glands and displaying venom-biased expression profiles. Although the new genes did not directly exhibit venom-related functions, they acted as hub genes within venom-related gene networks by interacting with older genes. These findings suggest that new genes may play a pivotal role in driving the innovative evolution of venom in parasitoid wasps.

2. Materials and Methods

2.1. Data Collection

We collected seven high-quality genomes of parasitoid wasps in the family Pteromalidae, including *P. puparum*, *P. venustus*, *P. qinghaiensi*, *N. vitripennis*, *A. calandrae*, *P. vindemmiae*, and *T. elegans* (Supplementary Table S1). The completeness of the genome assemblies was assessed using the Benchmarking Universal Single-Copy Orthologs (BUSCO v5.7.1) with the insecta_odb10 database [21]. The protein-coding sequences were downloaded from the corresponding data sources, and the longest transcript was considered representative of the gene with annotations for multiple isoforms.

2.2. Phylogenetic Analysis

To construct the phylogenetic relationships, we utilized protein sequences from seven species and clustered them into orthologous groups using OrthoFinder v2.2.7 [22]. The 3798 one-to-one orthogroups (also referred to as single-copy genes) shared among the seven parasitoid wasp species were selected for constructing the phylogenetic tree. Protein sequences were aligned and filtered using MAFFT v7 and trimAl v1.2 with default parameters [23,24]. The alignments for each orthogroup were then concatenated to generate a supergene, which was utilized for subsequent tree construction. The phylogenetic tree was constructed using maximum likelihood (ML) with IQ-TREE v2.1.2, employing the best-fit model (JTT + F + I + R9) estimated by ModelFinder [25,26]. Statistical support for the phylogenetic tree was evaluated using Ultrafast bootstrap analysis with 1000 replicates. Divergence times were estimated using r8s v1.81 [27], with calibration time points based on previous research [28].

2.3. Dating the Protein-Coding Genes Within Parasitoid Wasps

We utilized a whole-genome synteny-based pipeline (SBP) to identify new genes [6], selecting *P. puparum* as the reference species and generating alignments with the other six species using LASTZ v1.04.03 (<https://lastz.github.io/lastz/> accessed on 4 March 2025). The genes were assigned branch numbers ranging from 0 (indicating genes shared by all selected parasitoid wasps) to 6 (indicating genes specific to *P. puparum*) to denote the age of each gene, with higher branch numbers corresponding to younger genes. New genes were defined as those that originated after the *Nasonia-Pteromalus* split (~8 MYA) and lacked a corresponding syntenic locus in *N. vitripennis*, *A. calandrae*, *P. vindemmiae*, and *T. elegans*, following the stringent criterion outlined by previous study [6]. In detail, we excluded genes where more than 70% exonic regions overlapped with repetitive elements and removed the genes with patchy phylogenetic distribution. Finally, to validate the reliability of the new genes, we performed BLASTP alignments for the 480 new genes against the NCBI RefSeq Hymenoptera protein sequence database, using the following criteria: sequence identity > 20% and E-value < 10^{-5} . The identified new genes were classified into three categories of origin mechanisms [6]: DNA-mediated duplication, RNA-mediated duplication (retrogenes), and de novo birth. Retrogenes were identified as intronless genes derived from parental genes with at least one intron; otherwise, they were classified as gene duplications. To enhance identification accuracy of de novo genes originating from ancestrally non-coding regions, we implemented an additional pipeline [5] to detect the stepwise origination process of ORFs with the following sequential criteria: (1) ORFs must lack homologs in *T. elegans* and possess no more than one homolog in each species; (2) ORFs must show no more than one homologous DNA sequence per species, with at least one ortholog demonstrating more than 20% sequence coverage. Subsequently, we validated the structural annotation and orthology relationship of candidate de novo genes using TOGA [29] with parameters (–cb 3,5 –cjin 500).

2.4. Selection Analysis

After performing all-against-all BLASTP v2.12, we made the inference of old duplicate paralogs assigned to branch 0–3 and parent–child duplicate relationship. To examine the selection pressure exerted on the new duplicate genes, we compared them with their closest parental paralogous genes to calculate the ratio of non-synonymous substitution rate (Ka) to synonymous substitution rate (Ks). According to gene age and alignment, we made the inference of old duplicate paralogs. Initially, we generated protein-coding sequence alignments between parental and child genes using MAFFT v7 based on the dating inference [23]. Subsequently, we used the PAL2NAL tool to convert the protein alignments into codon-level alignments [30]. The paralogous Ka/Ks tests were conducted utilizing the PAMLv4.9 package [31], and the likelihood ratio test (LRT) was employed to calculate the *p*-value, assuming Ka/Ks = 0.5. Genes with Ka > 0.5, Ks > 5, or Ks values exceeding 1.5 times the interquartile range of the Ks distribution were excluded. Duplicate genes with Ka/Ks < 0.5 and a *p*-value < 0.05 were considered to be under negative selection, indicating evolutionary constraints on both copies.

2.5. Expression Pattern of New Genes

We downloaded the RNA-Seq data of testis, ovary, venom gland, and salivary gland at 3.5 days and 5.5 days and gut data at 3.5 days and 5.5 days (Supplementary Table S1). We employed Fastp v0.23.4 to eliminate adapters and low-quality bases [32]. The filtered reads were mapped to the *P. puparum* genome using Bowtie2 v2.2.9, and the output was processed with RSEM v1.3.3 to generate transcript per million (TPM) values [33,34]. To analyze the expression patterns of the new genes, we compared them with one-to-one single-copy genes shared by the seven species (older genes) across various tissues. Genes with TPM > 1 in at least one tissue were classified as expressed. The tissue-specific index (τ) was calculated as follows:

$$\tau = \frac{\sum_{i=1}^N (1 - Xi)}{N - 1}; Xi = \frac{Xi}{\max_{1 \leq i \leq n} (Xi)}$$

where Xi is the gene expression level in tissue i , and N is the tissue numbers. Genes with $\tau > 0.85$ were classified as tissue-specific expressed genes. To further quantify expression abundance and breadth across three gene-age groups (potential branch 0–1, branch 2–3, and branch 4–6), we calculated the maximum expression observed in the seven tissues and counted the number of tissues in which genes were expressed.

2.6. Co-Expression Network Analysis

We used the R package WGCNA v 1.72 to construct the weighted gene co-expression networks [35]. A total of 14,622 genes were used to build the co-expression network with the parameters as follows: network type = unsigned, soft power = 9, module identification method = dynamic tree cut, minimum module size = 30, and the threshold to merge modules with a high similarity = 0.5. Based on the expression patterns, we clustered all selected genes into 17 modules. Principal component analysis was performed on the genes within each module, and the value of the first principal component, termed module eigengene (ME), was used to represent the overall level of gene expression in the module. We treated tissues as traits and calculated the correlation between MEs and traits, along with the corresponding *p*-values, to identify key modules associated with specific traits. For each module, gene significance (GS) was defined as the absolute value of the correlation between the gene and the trait. The selected network regulated by the new genes was visualized using Cytoscape v3.9.1 [36]. Gene Ontology (GO) enrichment analysis was conducted using the R package clusterProfiler [37].

2.7. Total RNA Isolation, cDNA Synthesis, and Real-Time Quantitative PCR

Total RNA was isolated using RNAiso Plus (Takara Bio, Otsu, Japan). Then, the first-strand cDNA was prepared from 1.0 µg total RNA by reverse transcription using the TransScript One-Step gDNA Removal and cDNA Synthesis SuperMix Kit (TransGen Biotech, Beijing, China). The resulting PCR products were then sequenced. To quantify the expression of the new gene *Ppup071090.1* across various tissues of *P. puparum*, RT-qPCR was performed in testes, ovaries, venom glands, larval salivary glands, and carcasses (without ovaries and venom glands), and the primers for RT-qPCR are detailed in Supplementary Table S7. This was conducted using the Bio-Rad CFX 96 Real-Time Detection System (Bio-Rad, Hercules, CA, USA) with ChamQ SYBR Color qPCR Master Mix (Vazyme, Nanjing, China), following the manufacturer's instructions. The relative expression levels were normalized to the reference gene 18s using the $2^{-\Delta\Delta CT}$ method [38]. All quantitative data were expressed as the mean \pm standard error of the mean (SEM) of three independent biological replicates. Statistical significance was performed using a one-way analysis of variance followed by Tukey's honestly significant difference test for multiple comparisons. This statistical analysis was conducted on R software v4.4.2.

3. Results

3.1. Identification and Origin of New Genes in the Parasitoid Wasps

To identify and investigate the origin of new genes in parasitoid wasps, we screened the genomes of seven species across the Pteromalidae family, including *Pteromalus puparum*, *P. venustus*, *P. qinghaiensis*, *Nasonia vitripennis*, *Anisopteromalus calandrae*, *Pachycrepoideus vindemmiae*, and *Theocolax elegans*. *T. elegans* and *P. vindemmiae* were strategically chosen as outgroups to establish evolutionary context, with *N. vitripennis* and *A. calandrae* serving as closely related ingroup taxa. These seven genomes are of high quality, with average percentage of complete single-copy BUSCO genes of 97.3% (Supplementary Table S2). Due to the deep evolutionary timescales involved, tracing the processes of new gene formation is inherently challenging. However, focusing on closely related species can offer insights over shorter evolutionary timescales [39]. To this end, we first reconstructed a phylogenetic tree using 3798 single-copy genes identified across the seven species. Our analysis indicated that these parasitoid wasps shared close evolutionary relationships, with an estimated divergence time of approximately 80 million years (MY), and the most closely related species diverged around 1.49 MY ago. Thus, this group of selected Pteromalidae wasps represents an excellent model for systematically investigating the formation of new genes in parasitoid wasps.

To minimize any confusion arising from the frequent genomic rearrangements in Hymenoptera evolution and potential inconsistencies in gene annotation quality, we employed a whole-genome alignment-based approach to identify new genes in parasitoid wasps [6]. Using the LASTZ-MULTIZ pipeline [6,40], we aligned the genomes of the seven parasitoid wasp species. *P. puparum* was selected as the reference genome for its high-quality assembly, comprehensive gene annotations supported by diverse RNA sequencing datasets (e.g., IsoSeq, CAGE-Seq, and PAS-Seq), and its status as an emerging model organism with increasing functional studies [41–43]. For each of the 17,656 protein-coding genes in *P. puparum*, we assigned an origin age to corresponding phylogenetic tree branches based on ortholog presence or absence across the other species. Genes that originated after recent *Nasonia*-*Pteromalus* split (~8 MYA) were classified as “new genes”. In total, we identified 480 new genes that emerged during the evolution of Pteromalidae wasps, accounting for 2.71% of the entire gene set in *P. puparum* (Figure 1A). We observed a high rate of new gene birth in the Pteromalidae lineage, with 5.34 new genes emerging per MY (480/89.83 MY). Specifically, 102 genes emerged after the divergence of *P. puparum* and

P. venustus around 1.49 MYA, corresponding to a birth rate of 68.46 new genes per MY. This rate is notably higher than those observed in *Drosophila* and other lineages [44], indicating that the rapid formation of new genes may play a crucial role in the adaptive evolution of parasitoid wasps.

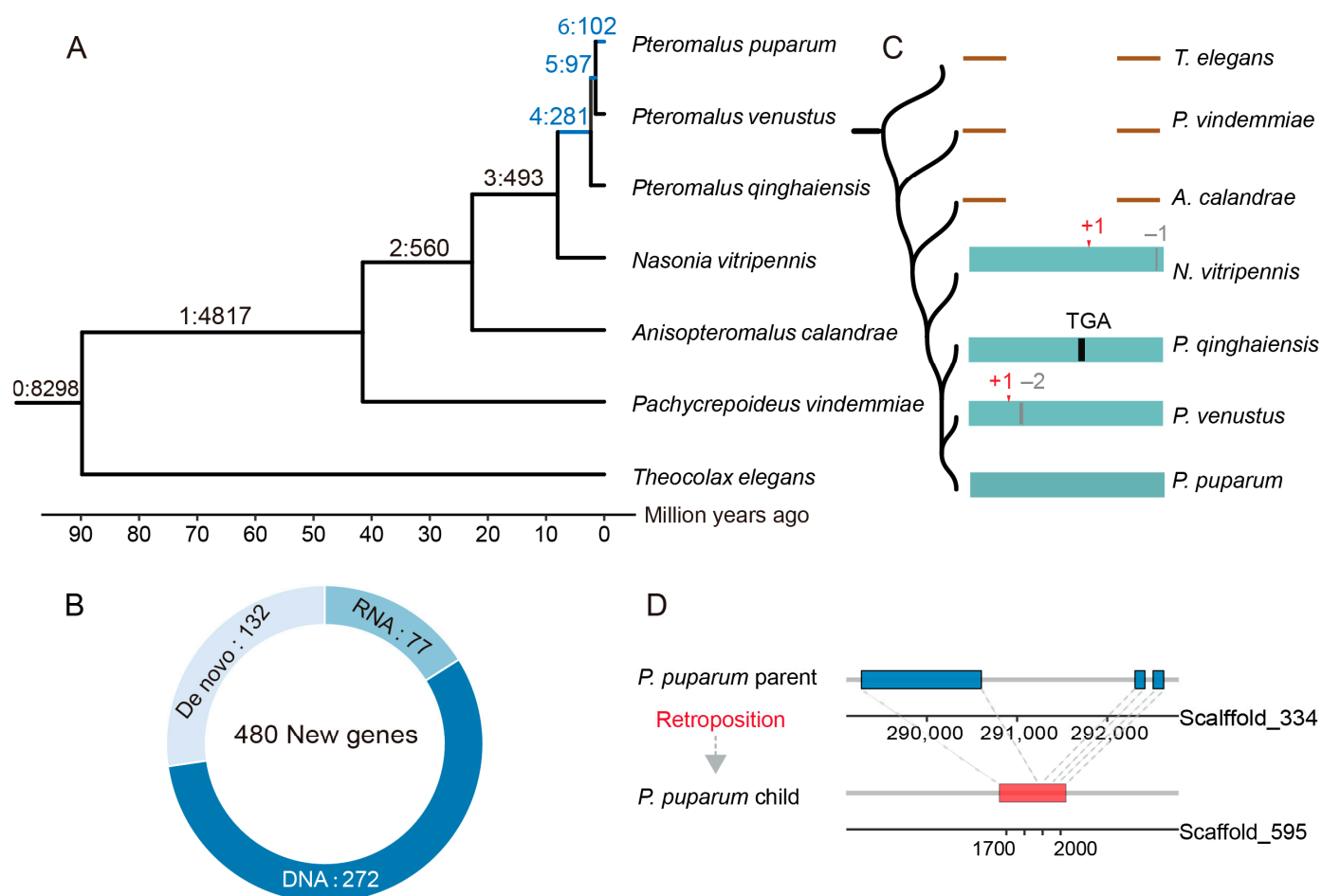


Figure 1. New genes identified in parasitoid wasps. **(A)** Distribution of new genes in the phylogenetic tree. Branches 0–6 were used as the distinct gene-age groups based on the synteny-based pipeline. Genes that emerged after *Nasonia*–*Pteromalus* split (branch 4–6) were identified as the new genes. Branches 4–6 (blue): Number of new genes. **(B)** Numbers and proportions of the new genes originated from DNA-mediated duplication, RNA-mediated duplication (retroposition), and de novo birth. **(C)** Stepwise origination process for an example of a de novo gene. The black bar represents a premature stop codon, the red arrow represents the frameshift insertion, and the grey bar indicates the frameshift deletion. Inserted bases are marked in red with their count shown. Deleted bases are marked in gray with their count shown. **(D)** The origination process for an example of retrotransposon. The blue or red box indicates exon.

We next classified the new genes into three categories based on their mechanisms of origin, including DNA-mediated duplication, RNA-mediated duplication, and de novo birth [6]. Over half of the new genes (56.67%, 272/480) originated through DNA-mediated duplication, followed by RNA-mediated duplication (16.0%, 77/480) and de novo birth (27.29%, 131/480) (Figure 1B, Supplementary Table S3). This pattern was consistent with findings in other lineages, where most new genes arose through DNA duplication [7,44,45]. As another major mechanism of gene duplication, RNA-mediated duplication (retroposition) is a process where mRNAs are reverse-transcribed into DNAs and then insert back into a new position on the genome. A hallmark of RNA-mediated duplication is the transition from a multi-exon ancestral gene to a single-exon new gene. These duplicates retain

sequences from their parent genes and contribute to phenotypic evolution through various mechanisms, including neofunctionalization, hypofunctionalization, subfunctionalization, and gene dosage regulation [46]. For example, a retrogene in *P. puparum* (~1.49 MYA), which is lineage-specific, is a single-exon gene derived from a three-exon parental gene (Figure 1D).

Recent studies have also emphasized the role of de novo genes in driving rapid protein diversity across various taxa, including *Drosophila* and mammals [8,47]. In our study, we identified several de novo genes during the evolution of parasitoid wasps. Unlike orphan genes (lineage-specific genes lacking homologous sequences in distantly related species), our de novo genes were traced to ancestral non-coding DNA sequences. This was facilitated by our dataset of closely related species, which allowed us to reconstruct the stepwise origination processes of de novo genes. By comparing novel open reading frames (ORFs) with their closest outgroup non-coding sequences, we identified key mutations, such as indels (insertions and deletions) and substitutions, that transformed non-coding DNA into coding sequences. For instance, *Ppup035000.1* (putative uncharacterized protein) represents a stepwise de novo gene formation process involving multiple frameshifts and substitutions. Orthologous non-coding sequences for *Ppup035000.1* were identified in both *N. vitripennis* and *P. venustus*, which exhibited frameshift mutations. Additionally, a premature stop codon was detected in closely related species, namely *P. qinghaiensis*, further illustrating this stepwise transformation (Figure 1C).

Overall, our genome-wide alignment-based approach yielded a high-confidence dataset of new genes involved in evolution of parasitoid wasps. This resource provides a foundation for investigating the functional and evolutionary significance of new genes in driving phenotypic innovation within parasitoid wasps.

3.2. The Structural and Evolutionary Analysis of the New Genes

Studies across diverse lineages, including *Drosophila*, plants, and mammals, have shown that new genes are typically short in sequence and predominantly consist of single exon [5,16]. We conducted a comparative analysis of coding sequence (CDS) length by categorizing genes into three age groups: branch 0–1 (oldest genes), branch 2–3 (middle-aged genes), and branch 4–6 (new genes). The oldest genes in branch 0–1 showed a significantly longer CDS length comparing to the middle-aged genes and the new genes (one-sided, unpaired Wilcoxon test, $p < 0.05$). Similarly, the middle-aged genes had significantly longer CDS lengths than the new genes (one-sided, unpaired Wilcoxon test, $p < 0.05$) (Figure 2A).

We also compared the number of exons among the three age groups. The oldest genes had significantly more exons than both the middle-aged and new genes (one-sided, unpaired Wilcoxon test, $p < 0.05$). And the middle-aged genes also had more exons than the new genes, with this difference being statistically significant (one-sided, unpaired Wilcoxon test, $p < 0.05$) (Figure 2B). Among the 480 new genes, 280 were single-exon genes, 155 (55.36%) of which originated from RNA-mediated duplication or de novo processes. This aligns with expectations for RNA-mediated duplicates, which are commonly single-exonic. However, a number of new genes (45.9%) originating from DNA duplication are also single-exonic, which warrants further investigation.

After gene duplication, newly duplicated genes might be an immediate source of functional novelty under the selective pressure to survive genetic erosion [48]. Our dating results provide a framework for identifying new duplicate genes (child) and their closest old paralogs (parent), enabling us to explore whether new duplicates evolve under distinct evolutionary pressures compared to their parental copies. Since protein functional divergence follows from the ratio of Ka/Ks non-synonymous substitutions, we used PAML package to perform paralogous Ka/Ks test applied in 94 parent–child gene pairs [49].

After filtering genes with outlier values (see details in methods), 19 out of 94 pairs of tested parent–child copies (20.21%) were significantly lower than 0.5, indicating functional constraint on both parent–child genes ($Ka/Ks < 0.5$, $p < 0.05$) (Supplementary Table S4). In contrast, 249 out of 878 pairs (29.1%) of old duplicate genes in branch 0–3 were under negative selection ($Ka/Ks < 0.5$, $p < 0.05$), which reflects those old duplicate genes generally undergo stronger selective constraints (Supplementary Table S5). The proportion of old genes under negative selection was significantly higher than the new duplicate genes. The previous studies in primates and flatfishes also showed the same results, whereas the proportion of new duplicate genes under negative selection in our study is positioned between that of primates and flatfishes [6,7]. This indicates new duplicates experienced a unique selection pressure within parasitoid wasps.

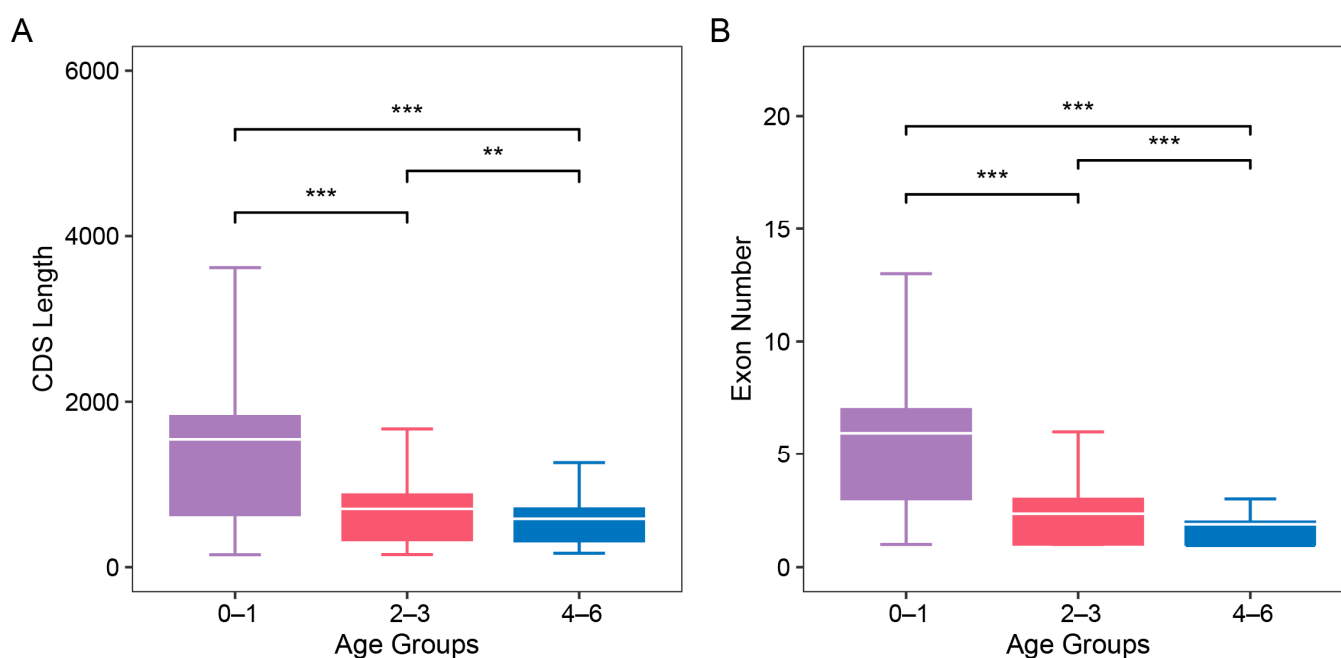


Figure 2. The CDS length (A) and exon numbers (B) of three age groups (branch 0–1, branch 2–3, and branch 4–6). The white bar indicates the average of each age group. Wilcoxon’s test was used to calculate significance between age groups. p -value between age groups was calculated by Wilcoxon’s test, with significance levels indicated as follows: ** $p < 0.01$; *** $p < 0.001$.

3.3. Tissue-Specific Expression Pattern of New Genes

Comparative multi-transcriptome analyses of *Drosophila* and mammalian adult tissues have suggested that new genes tend to be testis-specific, whereas older genes are more commonly ubiquitously expressed or exhibit specificity in somatic tissues [50–52]. This observation led to the proposal of the “out of the testis” hypothesis, which posits that new genes initially gain functionality in the testis, possibly due to its permissive transcriptional regulation. This implied that new genes acquire new functions by going through unique expression alterations. To explore the expression pattern of the new genes in *P. puparum*, we calculated the expression level (TPM) of all annotated genes across seven RNA-Seq datasets, including samples from yellow-pupae testis, adult ovary, adult venom gland, adult gut, and larval salivary gland. We found that new genes had significantly lower expression levels in all the tissues compared to single-copy older genes, which are defined as one-to-one orthologs across seven parasitoid species (Figure 3A). To ensure reliable tissue-specific analysis, we filtered out genes with TPM < 1 across all tissues, leaving 72 new genes for which tissue-specific tau (τ) scores were calculated. Tau scores range from 0 (ubiquitous expression) to 1 (strong tissue specificity) (Supplementary Table S6). The tissue specificity

of 72 new genes was significantly higher than that of the 3544 single-copy older genes conserved in the seven species (one-sided, unpaired Wilcoxon test, $p < 0.05$) (Figure 3B). Among 56 of 72 new genes exhibited tissue-bias expression, 12 of the 56 new genes (21.4%) exhibited testis-specific expression ($\tau > 0.85$), with 6 originating from DNA-mediated duplication and more than half having TPM values below 10. In contrast, five of six de novo genes demonstrated higher testis-bias expression, with median TPM values exceeding 10 (Supplementary Table S6). Interestingly, a larger proportion of new genes exhibited tissue-specific expression biased toward the ovary (23 out of 56 new genes, 41.07%) or venom gland (15 out of 56 new genes, 26.79%). Among the 15 new genes with venom bias, 13 new genes (branch 4 to branch 6) originating from DNA-mediated and RNA-mediated duplication showed low expression levels. Therefore, we investigated the changes in expression patterns of new duplicates and their parental genes. Among 13 parent–child gene pairs with a median TPM greater than one in multiple tissues, six child genes exhibited a shift in tissue specificity compared to their respective parent genes. For instance, the child gene *SPOPL* in branch 4 (Speckle-type POZ protein-like) demonstrated a maximum expression in the venom gland compared to other tissues despite a tau score of less than 0.85 (TPM = 7.98). In contrast, the parent gene *CON* (Connectin) displayed a higher expression level (TPM > 100) in ovary and testis.

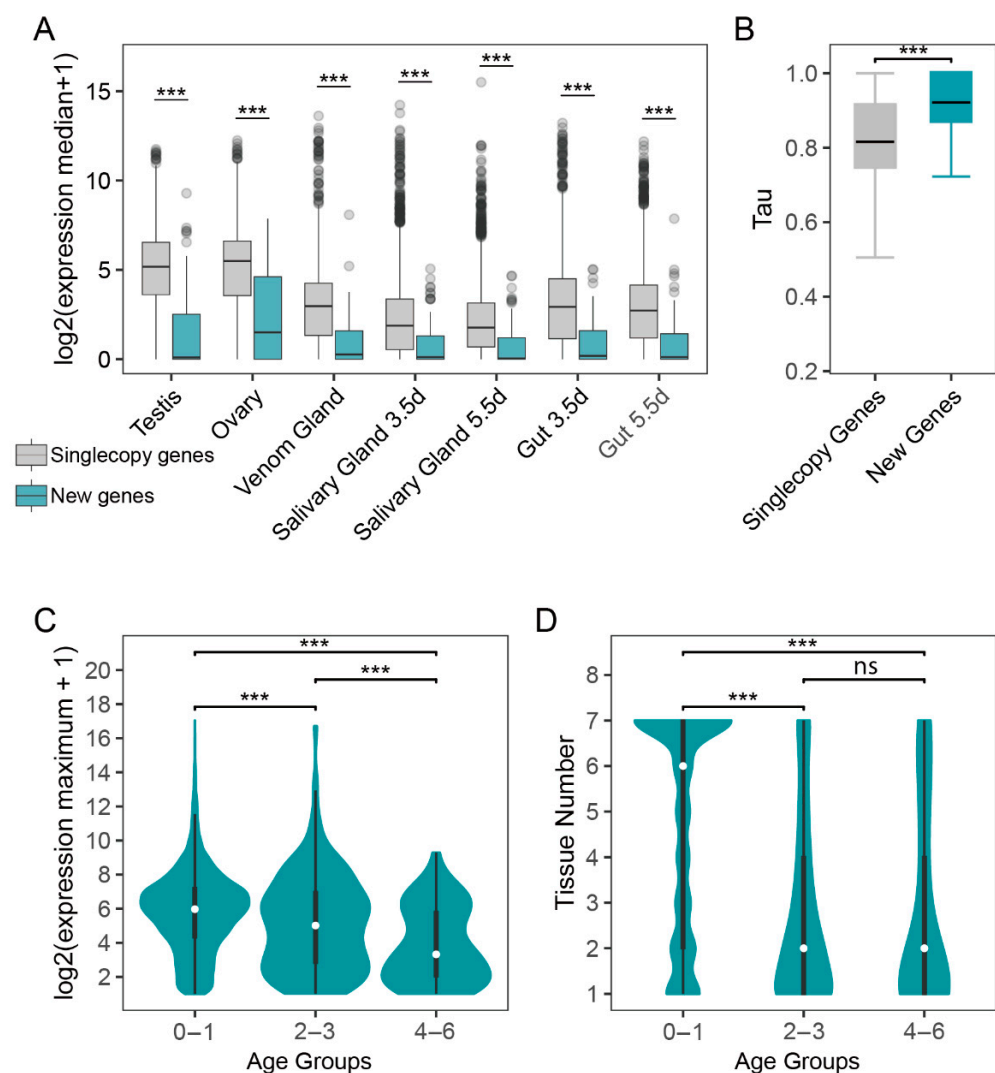


Figure 3. Expression patterns of the new genes. (A) The log-scale expression of new gene and single copy genes for different tissues. The relative gene expression was calculated using log2 (median

TPM + 1) (B) The tissue-specific score between new genes and old singleton genes. The black bar indicates the average. (C,D) Log2-based maximum expression levels across seven tissues and the number of tissues where genes were expressed (TPM > 1). Interquartile ranges are represented by black bars, while the violin curve illustrates the probability density of the data, with the median value indicated by a white dot. Genes were categorized into three groups based on their age identified in Figure 1A. *p*-value between age groups was calculated by Wilcoxon's test, with significance levels indicated as follows: ns: not significant. *** $p < 0.001$.

Furthermore, both venom-bias de novo genes exhibited a low-to-high expression trend: a de novo gene in branch 4 with TPM > 100 and the other de novo gene in branch 6 with TPM < 10. Thus, we discovered that the gene expression patterns varied across different gene age groups by analyzing transcription profiles in seven tissues of *P. puparum*. Specifically, new genes were expressed in only one tissue, with a median TPM value of less than one, while the oldest gene groups (branch 0–1) exhibited expression across six tissues, with a median TPM value exceeding 5 (Figure 3C,D). When comparing expression abundance and breadth, the oldest genes showed significantly higher expression levels than both middle-aged and new genes (one-sided, unpaired Wilcoxon test, $p < 0.05$) (Figure 3C,D). The middle-aged genes also show significant differences from new genes in terms of expression levels and the number of expressed tissues (one-sided, unpaired Wilcoxon test, $p > 0.05$). This suggests the presence of an age-dependent expression trend in parasitoid wasps, a phenomenon also observed in previous studies on new genes in mammals, fish, and plants [5,7,45].

3.4. A New Gene Serves as a Hub Gene of Venom Gene Regulatory Network

Upon carefully examining the new genes we identified, alongside the 179 venom genes previously reported [41], we found no overlap between the two groups. This suggests that the new genes may not yet directly contribute to the innovation of the venom composition in parasitoid wasps. Therefore, we hypothesize that the new genes expressed in the venom gland may play a regulatory role associated with venom gene expression. To test this hypothesis, we utilized seven RNA-Seq datasets from the studied tissues (testis, ovary, venom gland, salivary gland, and gut) to construct a weighted gene co-expression network. This analysis clustered 14,622 genes into 17 modules based on expression similarity, with module sizes ranging from 53 to 9542 genes (Figure 4A). The turquoise module (MEturquoise), comprising 9542 genes, showed a strong association with the ovary and testis ($0.6 \leq R^2 \leq 0.8$, $p < 0.05$). Remarkably, 59.72% (43/72) of 72 expressed new genes were identified within this module. Among these, 27.90% (12 genes) exhibited testis-biased expression, while 51.16% (22 genes) displayed ovary-biased expression ($\tau > 0.85$) (Figure 4C). Of the 17 modules analyzed, only one module showed a strikingly strong correlation with the venom gland ($R^2 = 0.94$, $p < 0.05$), identified as the blue module, which contains 1749 genes and is closely associated with the venom gene regulatory network (Figure 4A). Within this module, 334 of the 1749 genes exhibited venom-biased expression, indicating that older genes play a significant role in the venom co-expression network. The percentage of the new genes in the blue module was higher than the genome-wide percentage of the new genes (Figure 4B). Specifically, 19 new genes were presented in this venom-related gene network, with 15 exhibiting venom-specific expression, suggesting their potential roles in venom-related functions (Figure 4C). To evaluate the contribution of new genes to the venom-related network, we calculated module membership (kME) values, ranging from -1 to 1 , to quantify each gene's connectivity with other genes in the module. Genes with high kME values, known as hub genes, occupy central positions within the network. Based on their high connectivity, we identified a new gene (*Ppup071090.1*) as a hub gene, which originated from a three-exon parental gene through RNA-mediated duplication (Figure 4D,E). To

validate the reliability of this new hub gene, we first conducted RT-PCR analysis, which confirmed the specific presence of *Ppup071090.1* in ovarian and venom gland tissues (Supplementary Figure S2). Subsequent RT-qPCR quantification revealed the expression levels of *Ppup071090.1* were significantly higher in the ovary and venom gland compared to testis, larval salivary gland, and carcasses (without ovary and venom gland) ($p < 0.05$) (Supplementary Figure S3).

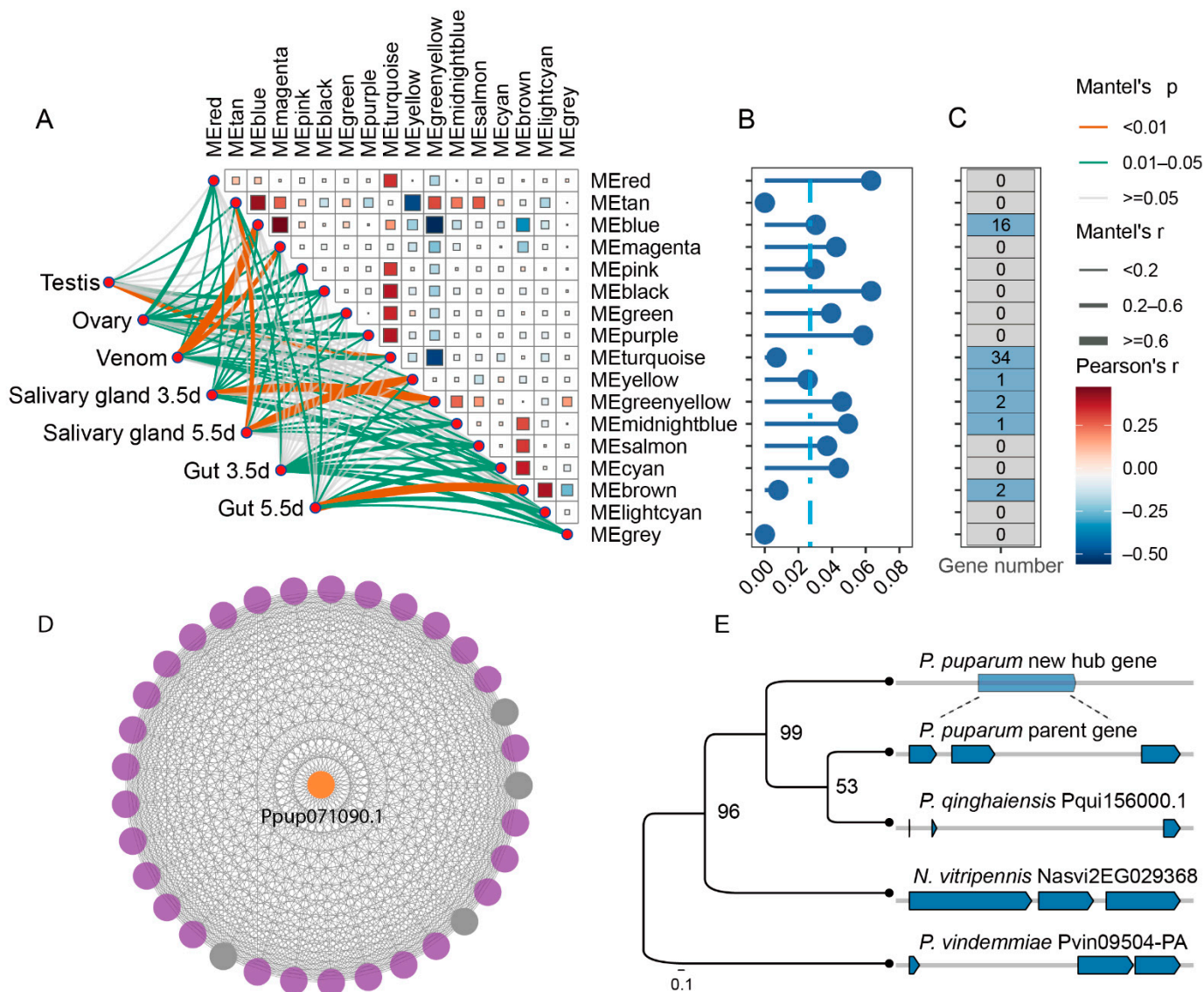


Figure 4. Expression network of new and old genes across the different tissues. (A) Gene weight co-expression network. (B) The percentage of the new genes in the different modules. The vertical blue line represents the genome-wide percentage of the new genes (2.71%), and the proportion of new genes are indicated by blue points. (C) Heatmap of the tissue-specific index of the new genes across different modules. (D) The venom-related network was visualized, highlighting a hub gene with the highest degree of connections and its associated interactions. In the plot, orange node represents the core new gene, purple nodes represent venom-biased genes, and grey nodes represent non-biased genes. (E) Exon-intron structure of parental genes (blue) and new hub (orange) gene.

Further GO enrichment analysis of the genes in the hub-related gene network revealed their involvement in pathways such as protein N-linked glycosylation, nucleosome assembly, and nucleosome organization (Supplementary Figure S1). Notably, 41 venom genes (41/179, 23% of all venom genes) were identified in this network [41]. These findings

indicate that the new gene has not been directly utilized as a venom protein-coding gene; instead, it may function as a central hub involved in the regulation of venom genes.

4. Discussion

New gene origination plays a major driving force for understanding evolutionary mechanisms underlying new traits and adaptive functions. An excellent example of this is the venom system of parasitoid wasps. As significant evolutionary innovations, venom offers an exceptional opportunity to investigate how new genes contribute to the phenotypic adaptations of these wasps to their diverse natural hosts. In this study, we analyzed a representative group of parasitoid wasps to explore the role of new genes, uncovering insights into their origination mechanisms and evolutionary roles. Our results show that gene duplication is the dominant origin of the new genes in parasitoid wasps, consistent with its role as a critical evolutionary process for generating genetic novelty [53]. Similar mechanisms have been observed in other taxa, such as *Drosophila*, mammals, and flatfishes [6,7,44]. Notably, the evolutionary origins of intron-free new genes derived from ancestrally intron-free parental genes remain a persistent challenge in the identification of retrogenes. Furthermore, more than half of the child duplicates retain a similar expression pattern to their parent paralogous genes. This supports the existence of “responsive backup circuits” across various species, where a redundant gene copy is upregulated when its paralog is subjected to an inactivating perturbation [46].

Combining gene age analysis with expression patterns, we identified the testis as a hotspot for new gene origination, consistent with its high transcriptional activity and permissive chromatin state [54,55]. However, our results reveal that new genes in parasitoid wasps are not solely restricted to the testis. Many new genes also expressed in venom gland, emphasizing the role of venom systems as evolutionary innovations that recruit new genes over short evolutionary time scales. A plausible explanation for this rapid turnover in venom-related genes lies in the diverse parasitism strategies of parasitoid wasps, which require continuous updates to their venom repertoire to adapt to a variety of hosts [20,41]. Although no new genes were found to have become venom proteins in *P. puparum*, several acquired the venom expression and were integrated as hub genes within existing venom-related gene networks. These hub genes participated in metabolic pathways and interacted extensively with older genes. Interestingly, older genes predominated within the venom-related co-expression networks, suggesting that they play a more pervasive role in maintaining the stable regulation of venom functions. A similar regulatory mechanism has been observed in human-specific gene networks associated with brain development [56,57]. For new genes to acquire functional relevance, they often become incorporated into pre-existing genetic interaction networks. This integration enables them to gain associated biological activities. New genes, with their shorter coding sequences, often feature intrinsically disordered regions and low-complexity sequences, characteristics that increase their binding flexibility and adaptability [58,59]. These advantageous properties facilitate their rapid incorporation into gene networks as hubs, where they interact with a wide range of partners.

In summary, we identified new genes across a group of representative parasitoid wasps, shedding light on their roles in the rapid evolution of venom systems. Our findings emphasize the importance of new genes in driving venom evolution and highlight the molecular mechanisms behind their emergence and integration. These results provide valuable insights into the evolutionary dynamics of new gene functions and represent a significant contribution to our understanding of venom evolution in parasitoid wasps.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/insects16050502/s1>. Supplementary Table S1: List of species and data resources used in this study. Supplementary Table S2: BUSCO scores of the seven species. Supplementary Table S3: List of new genes. Supplementary Table S4: Duplicated new genes under natural selection. Supplementary Table S5: Duplicated older genes under natural selection. Supplementary Table S6: Seventy-two new genes expressed in more than one tissue with median TPM > 1. Supplementary Table S7: The primers for qRT-PCR. Supplementary Figure S1: GO enrichment analysis of the genes in the hub-related gene network. Supplementary Figure S2: Validation of new hub gene *Ppup071090.1* expression pattern through RT-PCR analysis in various tissues. Supplementary Figure S3: Relative expression levels of new hub gene *Ppup071090.1* across various tissues.

Author Contributions: Conceptualization, X.Y., Y.Y. and G.Y.; methodology, B.Z. and Y.Y.; software, B.Z. and Y.B.; validation, J.S.; formal analysis, B.Z., F.W., Q.F. and Y.Y.; investigation, B.Z., Y.B., B.Y., S.X. and Y.Y.; resources, B.Z., Y.B., B.Y., S.X., F.W., Q.F. and Y.Y.; data curation, B.Z. and B.Y.; writing—original draft preparation, B.Z.; writing—review and editing, B.Z.; visualization, B.Z.; supervision, X.Y., Y.Y. and G.Y.; project administration, X.Y., Y.Y. and G.Y.; funding acquisition, X.Y., Y.Y. and G.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Key Program of National Natural Science Foundation of China (NSFC) (grant no. 32330085 to G.Y.) and the Program of NSFC (grant no. 32202376 and grant no. 32472631 to X.Y.; grant no. 32302428 to Y.Y.).

Data Availability Statement: Raw data are provided in spreadsheets and can be downloaded at Supplementary Materials.

Acknowledgments: We are grateful to Yong Zhang, working at the Institute of Zoology, Chinese Academy of Sciences, and Qingzhu Zhang for their assistance in the methodology of discovering new genes.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Long, M.; Betrán, E.; Thornton, K.; Wang, W. The origin of new genes: Glimpses from the young and old. *Nat. Rev. Genet.* **2003**, *4*, 865–875. [\[CrossRef\]](#)
- Kaessmann, H. Origins, evolution, and phenotypic impact of new genes. *Genome Res.* **2010**, *20*, 1313–1326. [\[CrossRef\]](#) [\[PubMed\]](#)
- Rödelsperger, C.; Prabh, N.; Sommer, R.J. New gene origin and deep taxon phylogenomics: Opportunities and challenges. *Trends Genet.* **2019**, *35*, 914–922. [\[CrossRef\]](#) [\[PubMed\]](#)
- Chen, S.D.; Krinsky, B.H.; Long, M.Y. New genes as drivers of phenotypic evolution. *Nat. Rev. Genet.* **2013**, *14*, 645–660, Erratum in *Nat. Rev. Genet.* **2013**, *14*, 744. [\[CrossRef\]](#)
- Zhang, L.; Ren, Y.; Yang, T.; Li, G.W.; Chen, J.H.; Gschwend, A.R.; Yu, Y.; Hou, G.X.; Zi, J.; Zhou, R.; et al. Rapid evolution of protein diversity by *de novo* origination in *Oryza*. *Nat. Ecol. Evol.* **2019**, *3*, 679–690. [\[CrossRef\]](#)
- Shao, Y.; Chen, C.Y.; Shen, H.; He, B.Z.; Yu, D.Q.; Jiang, S.; Zhao, S.L.; Gao, Z.Q.; Zhu, Z.L.; Chen, X.; et al. GenTree, an integrated resource for analyzing the evolution and function of primate-specific coding genes. *Genome Res.* **2019**, *29*, 682–696. [\[CrossRef\]](#)
- Li, H.R.; Chen, C.Y.; Wang, Z.K.; Wang, K.; Li, Y.X. Wang W. Pattern of new gene origination in a special fish lineage, the flatfishes. *Genes* **2021**, *12*, 1819. [\[CrossRef\]](#)
- An, N.A.; Zhang, J.; Mo, F.; Luan, X.K.; Tian, L.; Shen, Q.S.; Li, X.S.; Li, C.Q.; Zhou, F.Q.; Zhang, B.Y.; et al. *De novo* genes with an lncRNA origin encode unique human brain developmental functionality. *Nat. Ecol. Evol.* **2023**, *7*, 264–276. [\[CrossRef\]](#)
- Zhang, Y.E.; Long, M.Y. New genes contribute to genetic and phenotypic novelties in human evolution. *Curr. Opin. Genet. Dev.* **2014**, *29*, 90–96. [\[CrossRef\]](#)
- Long, M.Y.; Langley, C.H. Natural selection and the origin of *Jingwei*, a chimeric processed functional gene in *Drosophila*. *Science* **1993**, *260*, 91–95. [\[CrossRef\]](#)
- Rivard, E.L.; Ludwig, A.G.; Patel, P.H.; Grandchamp, A.; Arnold, S.E.; Berger, A.; Scott, E.M.; Kelly, B.J.; Mascha, G.C.; Bornberg-Bauer, E.; et al. A putative evolved gene required for spermatid chromatin condensation in *Drosophila melanogaster*. *PLoS Genet.* **2021**, *17*, e1009787. [\[CrossRef\]](#) [\[PubMed\]](#)
- Ding, Y.; Zhao, L.; Yang, S.A.; Jiang, Y.; Chen, Y.A.; Zhao, R.P.; Zhang, Y.; Zhang, G.J.; Dong, Y.; Yu, H.J.; et al. A young duplicate gene plays essential roles in spermatogenesis by regulating several Y-linked male fertility genes. *PLoS Genet.* **2010**, *6*, e1001255. [\[CrossRef\]](#)

13. Dai, H.Z.; Chen, Y.; Chen, S.D.; Mao, Q.Y.; Kennedy, D.; Landback, P.; Eyre-Walker, A.; Du, W.; Long, M.Y. The evolution of courtship behaviors through the origination of a new gene in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 7478–7483. [\[CrossRef\]](#)
14. Ross, B.D.; Rosin, L.; Thomae, A.W.; Hiatt, M.A.; Vermaak, D.; de la Cruz, A.F.A.; Imhof, A.; Mellone, B.G.; Malik, H.S. Stepwise evolution of essential centromere function in a neogene. *Science* **2013**, *340*, 1211–1214. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Zhao, Q.; Zheng, Y.H.; Li, Y.Y.; Shi, L.P.; Zhang, J.; Ma, D.N.; You, M.S. An orphan gene enhances male reproductive success in *Plutella xylostella*. *Mol. Biol. Evol.* **2024**, *41*, msae142. [\[CrossRef\]](#)
16. Martinson, E.O.; Siebert, A.L.; He, M.; Kelkar, Y.D.; Doucette, L.A.; Werren, J.H. Evaluating the evolution and function of the dynamic Venom Y protein in ectoparasitoid wasps. *Insect Mol. Biol.* **2019**, *8*, 499–508. [\[CrossRef\]](#)
17. Burke, G.R.; Sharanowski, B.J. Parasitoid wasps. *Curr. Biol.* **2024**, *34*, R483–R488. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Ye, X.H.; Yang, Y.; Zhao, X.X.; Fang, Q.; Ye, G.Y. The state of parasitoid wasp genomics. *Trends Parasitol.* **2024**, *40*, 914–929. [\[CrossRef\]](#)
19. Ye, X.H.; Yang, Y.; Zhao, C.; Xiao, S.; Sun, Y.; He, C.; Xiong, S.J.; Zhao, X.X.; Zhang, B.; Lin, H.W.; et al. Genomic signatures associated with maintenance of genome stability and venom turnover in two parasitoid wasps. *Nat. Commun.* **2022**, *13*, 6417. [\[CrossRef\]](#)
20. Martinson, E.O.; Mrinalini; Kelkar, Y.D.; Chang, C.H.; Werren, J.H. The evolution of venom by co-option of single-copy genes. *Curr. Biol.* **2017**, *27*, 2007–2013. [\[CrossRef\]](#)
21. Manni, M.; Berkeley, M.R.; Seppey, M.; Simao, F.A.; Zdobnov, E.M. BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* **2021**, *38*, 4647–4654. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Emms, D.M.; Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **2019**, *20*, 238. [\[CrossRef\]](#)
23. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Capella-Gutierrez, S.; Silla-Martinez, J.M.; Gabaldon, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **2009**, *25*, 1972–1973. [\[CrossRef\]](#)
25. Minh, B.Q.; Schmidt, H.A.; Chernomor, O.; Schrempf, D.; Woodhams, M.D.; Von Haeseler, A.; Lanfear, R. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic Era. *Mol. Biol. Evol.* **2020**, *37*, 1530–1534. [\[CrossRef\]](#)
26. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.F.; von Haeseler, A.; Jermini, L.S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **2017**, *14*, 587–589. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Sanderson, M.J. r8s: Inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **2003**, *19*, 301–302. [\[CrossRef\]](#)
28. Peters, R.S.; Krogmann, L.; Mayer, C.; Donath, A.; Gunkel, S.; Meusemann, K.; Kozlov, A.; Podsiadlowski, L.; Petersen, M.; Lanfear, R.; et al. Evolutionary history of the Hymenoptera. *Curr. Biol.* **2017**, *27*, 1013–1018. [\[CrossRef\]](#)
29. Kirilenko, B.M.; Munegowda, C.; Osipova, E.; Jebb, D.; Sharma, V.; Blumer, M.; Morales, A.E.; Ahmed, A.W.; Kontopoulou, D.G.; Hilgers, L.; et al. Integrating gene annotation with orthology inference at scale. *Science* **2023**, *380*, 368–386. [\[CrossRef\]](#)
30. Suyama, M.; Torrents, D.; Bork, P. PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **2006**, *34*, W609–W612. [\[CrossRef\]](#)
31. Yang, Z.H. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **2007**, *24*, 1586–1591. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Chen, S.F.; Zhou, Y.Q.; Chen, Y.R.; Gu, J. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **2018**, *34*, 884–890. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Langdon, W.B. Performance of genetic programming optimised Bowtie2 on genome comparison and analytic testing (GCAT) benchmarks. *Biodata Min.* **2015**, *8*, 1. [\[CrossRef\]](#)
34. Li, B.; Dewey, C.N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **2011**, *12*, 323. [\[CrossRef\]](#)
35. Langfelder, P.; Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **2008**, *9*, 559. [\[CrossRef\]](#)
36. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [\[CrossRef\]](#)
37. Wu, T.Z.; Hu, E.Q.; Xu, S.B.; Chen, M.J.; Guo, P.F.; Dai, Z.H.; Feng, T.Z.; Zhou, L.; Tang, W.L.; Zhan, L.; et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* **2021**, *2*, 100141. [\[CrossRef\]](#)
38. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. *Methods* **2001**, *25*, 402–408. [\[CrossRef\]](#)
39. Wright, C.J.; Smith, C.W.J.; Jiggins, C.D. Alternative splicing as a source of phenotypic diversity. *Nat. Rev. Genet.* **2022**, *23*, 697–710. [\[CrossRef\]](#)

40. Zhang, Y.E.; Vibranovski, M.D.; Landback, P.; Marais, G.A.; Long, M. Chromosomal redistribution of male-biased genes in mammalian evolution with two bursts of gene gain on the X chromosome. *PLoS Biol.* **2010**, *8*, e1000494. [\[CrossRef\]](#)
41. Ye, X.; He, C.; Yang, Y.; Sun, Y.H.; Xiong, S.; Chan, K.C.; Si, Y.; Xiao, S.; Zhao, X.; Lin, H.; et al. Comprehensive isoform-level analysis reveals the contribution of alternative isoforms to venom evolution and repertoire diversity. *Genome Res.* **2023**, *33*, 1554–1567. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Ye, X.; Yang, Y.; Fang, Q.; Ye, G. Genomics of insect natural enemies in agroecosystems. *Curr. Opin. Insect Sci.* **2024**, *68*, 101298. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Ye, X.; Yan, Z.; Yang, Y.; Xiao, S.; Chen, L.; Wang, J.; Wang, F.; Xiong, S.; Mei, Y.; Wang, F.; et al. A chromosome-level genome assembly of the parasitoid wasp *Pteromalus puparum*. *Mol. Ecol. Resour.* **2020**, *20*, 1384–1402. [\[CrossRef\]](#)
44. Zhou, Q.; Zhang, G.J.; Zhang, Y.; Xu, S.Y.; Zhao, R.P.; Zhan, Z.B.; Li, X.; Ding, Y.; Yang, S.A.; Wang, W. On the origin of new genes in *Drosophila*. *Genome Res.* **2008**, *18*, 1446–1455. [\[CrossRef\]](#)
45. Chen, C.Y.; Yin, Y.; Li, H.R.; Zhou, B.T.; Zhou, J.; Zhou, X.F.; Li, Z.P.; Liu, G.C.; Pan, X.Y.; Zhang, R.; et al. Ruminant-specific genes identified using high-quality genome data and their roles in rumen evolution. *Sci. Bull.* **2022**, *67*, 825–835. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Kuzmin, E.; Taylor, J.S.; Boone, C. Retention of duplicated genes in evolution. *Trends Genet.* **2022**, *38*, 59–72. [\[CrossRef\]](#)
47. Zheng, E.B.; Zhao, L. Protein evidence of unannotated ORFs in *Drosophila* reveals diversity in the evolution and properties of young proteins. *eLife* **2022**, *11*, e78772. [\[CrossRef\]](#)
48. Vakirlis, N.; Carvunis, A.R.; McLysaght, A. Synteny-based analyses indicate that sequence divergence is not the main source of orphan genes. *eLife* **2020**, *9*, e53500. [\[CrossRef\]](#)
49. Betrán, E.; Thornton, K.; Long, M. Retroposed new genes out of the X in *Drosophila*. *Genome Res.* **2002**, *12*, 1854–1859. [\[CrossRef\]](#)
50. Assis, R.; Bachtrog, D. Neofunctionalization of young duplicate genes in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 17409–17414. [\[CrossRef\]](#)
51. Guschanski, K.; Warnefors, M.; Kaessmann, H. The evolution of duplicate gene expression in mammalian organs. *Genome Res.* **2017**, *27*, 1461–1474. [\[CrossRef\]](#) [\[PubMed\]](#)
52. Peng, J.H.; Zhao, L. The origin and structural evolution of *de novo* genes in *Drosophila*. *Nat. Commun.* **2024**, *15*, 810. [\[CrossRef\]](#) [\[PubMed\]](#)
53. Magadum, S.; Banerjee, U.; Murugan, P.; Gangapur, D.; Ravikesavan, R. Gene duplication as a major force in evolution. *J. Genet.* **2013**, *92*, 155–161. [\[CrossRef\]](#)
54. Neme, R.; Tautz, D. Fast turnover of genome transcription across evolutionary time exposes entire non-coding DNA to gene emergence. *eLife* **2016**, *5*, e09977. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Witt, E.; Benjamin, S.; Svetec, N.; Zhao, L. Testis single-cell RNA-seq reveals the dynamics of *de novo* gene transcription and germline mutational bias in *Drosophila*. *eLife* **2019**, *8*, e47138. [\[CrossRef\]](#)
56. Zhang, W.Y.; Landback, P.; Gschwend, A.R.; Shen, B.R.; Long, M.Y. New genes drive the evolution of gene interaction networks in the human and mouse genomes. *Genome Biol.* **2015**, *16*, 202. [\[CrossRef\]](#)
57. Zu, J.; Gu, Y.X.; Li, Y.; Li, C.T.; Zhang, W.Y.; Zhang, Y.E.; Lee, U.; Zhang, L.; Long, M.Y. Topological evolution of coexpression networks by new gene integration maintains the hierarchical and modular structures in human ancestors. *Sci. China Life Sci.* **2019**, *62*, 594–608. [\[CrossRef\]](#)
58. Dunker, A.K.; Cortese, M.S.; Romero, P.; Iakoucheva, L.M.; Uversky, V.N. The roles of intrinsic disorder in protein interaction networks. *FEBS J.* **2005**, *272*, 5129–5148. [\[CrossRef\]](#)
59. Dosztányi, Z.; Chen, J.; Dunker, A.K.; Simon, I.; Tompa, P. Disorder and sequence repeats in hub proteins and their implications for network evolution. *J. Proteome Res.* **2006**, *5*, 2985–2995. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.