



OPEN

## A pipeline contributes to efficient identification of salivary proteins in short-headed planthopper, *Epeurysa nawaii*

Xiao-Jing Wang<sup>1,3</sup>, Qiao Li<sup>2,3</sup>, Zhuang-Xin Ye<sup>1</sup> & Hai-Jian Huang<sup>1</sup>✉

Saliva, an oral secretion primarily originating from salivary glands (SGs), exert critical roles in the ongoing evolutionary interaction between insects and plants. However, identifying insect salivary components poses challenges due to the tiny size of insects, low secretion amounts, and the propensity for degradation after secretion. In this study, we developed a transcriptome-based approach to comprehensively analyze the salivary proteins of the short-headed planthopper, *Epeurysa nawaii*, a species with unique feeding habits on bamboo. A total of 165 salivary proteins were identified, with 114 secretory genes highly and specifically expressed in SGs. Consistent with most phloem-feeding insects, digestive enzymes, calcium-binding proteins, oxidoreductases, and a few previously reported salivary effectors were ubiquitously distributed in *E. nawaii* saliva. However, we also identified a substantial portion of salivary proteins exhibiting taxonomy specificity, including 60 *E. nawaii*-specific and 62 Delphacidae-specific proteins. These taxonomy-restricted proteins potentially play a role in insect adaptation to specific host plants. Our study provides an efficient pipeline for salivary protein identification and serves as a valuable resource for the functional characterization of effectors.

Saliva, an oral secretion primarily originating from salivary glands (SGs), plays crucial roles in the intricate interactions between insects and their host plants. It is mainly comprised of bioactive proteins with diverse functions, ranging from overcoming plant defenses to facilitating nutrient digestion, enabling insects to feed and thrive successfully<sup>1,2</sup>. For example, salivary sheath protein salivary-4 from *Nilaparvata lugens* enabled insect stylet anchored in a fixed point and sealed wounded plant tissues<sup>3</sup>; salivary protein Bt56 from *Bemisia tabaci* suppress plant defense by leveraging SA-JA crosstalk<sup>4</sup>; salivary proteases from *Acyrtosiphon pisum* are capable of degrading sieve-tube proteins<sup>5</sup>. After secretion, the insect saliva directly contacts with plant tissues, and intimates associated with plant physiology. A few salivary proteins have been reported to experience a high evolutionary rate<sup>6,7</sup>. Noteworthy, some of proteins even exhibit taxonomy/species-specificity, which can be only found in specific taxa<sup>7</sup>. Understanding the insect salivary proteins will provide a wealth of information about the complex co-evolving arms race between insects and plants, and might be useful in pest management.

In recent decades, with the development of sequencing technology, the salivary proteins of several herbivorous insects have been unveiled. Generally, insect saliva can be identified using two methods: direct analysis of secreted saliva by a proteomic approach and indirect analysis of SG extracts by a transcriptomic approach<sup>8–11</sup>. Although the former method is expected to provide an accurate representation of the proteins that are secreted into artificial diets, some important salivary proteins might be omitted<sup>12</sup>. For example, the LsSP1 protein, which is validated to be secreted and exerts essential roles in insect feeding, fails to be detected in secreted saliva using proteomic analysis<sup>13</sup>. Similarly, the salivary protein Bt56 can be predicted by transcriptomic analysis but not by proteomic analysis<sup>4,14</sup>. In proteomic analysis, trypsin is usually used for protein digestion, exclusively cleaving peptide chains at the carboxyl side of the amino acid lysine or arginine<sup>15</sup>. Therefore, proteins with fewer lysine or arginine residues might have been omitted using this method. Additionally, insect salivary proteins are usually of short amino acid length, which might have a lower chance of being detected by the proteomic approach<sup>16</sup>. When analyzing the salivary proteins that were validated to be secreted, a few conserved properties were discovered,

<sup>1</sup>State Key Laboratory for Managing Biotic and Chemical Threats to the Quality and Safety of Agro-products, Institute of Plant Virology, Ningbo University, Ningbo, China. <sup>2</sup>Animal and Plant Quarantine Service, Technology Center of Wuhan Customs District, Wuhan, China. <sup>3</sup>These authors contributed equally: Xiao-Jing Wang and Qiao Li. ✉email: huanghajian@nbu.edu.cn

including the secretory property represented by a signal peptide at the N-terminus and the tissue-specific property represented by high and specific expression in the SG<sup>17,18</sup>. These properties contribute to the efficient and accurate identification of salivary proteins.

The Delphacidae, commonly referred to as planthoppers, encompass over 2100 described species worldwide (<https://sites.udel.edu/planthoppers>, accessed on December 2023). Most species within this family feed on gramineous plants, with some posing significant threats as agricultural pests. Notably, the brown planthopper (*N. lugens*, BPH), small brown planthopper (*Laodelphax striatellus*, SBPH), and white-backed planthopper (*Sogatella furcifera*, WBPH) are well-known Delphacidae species, causing substantial economic damage to rice production in Asia<sup>19</sup>. Presently, our understanding of Delphacidae is primarily limited to these three rice planthoppers, while other planthopper species, despite their considerable size, remain poorly investigated. For instance, with the exception of the three rice planthoppers, scant transcriptomic data is available for other planthopper species in NCBI SRA database.

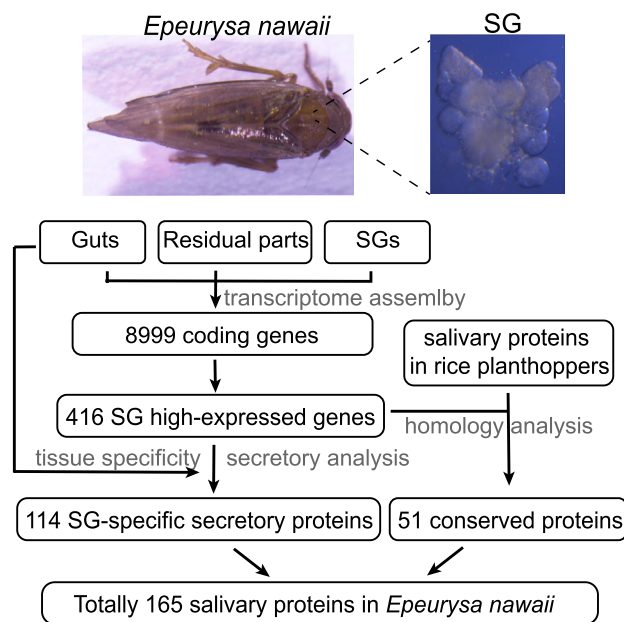
The short-headed planthopper, *Epeurysa nawaii* Matsumura, is widely distributed in Asia. In contrast to the three rice planthoppers, which predominantly feed on grass plants, *E. nawaii* primarily feeds on and damage to bamboo<sup>20</sup>. It is interesting to compare salivary proteins between insects with distinct feeding habits. In this study, we conducted a comprehensive analysis of the salivary proteins of *E. nawaii* (Fig. 1). Furthermore, we compared the salivary proteins of *E. nawaii* with those of the other three planthoppers and other herbivorous insects, revealing both shared and unique saliva compounds.

## Material and methods

### Sample preparation and transcriptomic sequencing

*E. nawaii* used in this study were gathered from a bamboo grove at Ningbo University (Ningbo, China) in May 2023. Before dissection, the insects underwent a 5 s anesthetization in CO<sub>2</sub> and were subsequently transferred onto ice. Insect SGs and guts were dissected in a phosphate-buffered saline (PBS) solution (137 mM NaCl, 2.68 mM KCl, 8.1 mM Na<sub>2</sub>HPO<sub>4</sub> and 1.47 mM KH<sub>2</sub>PO<sub>4</sub> at pH7.4) using a pair of forceps (Ideal-Tek, Switzerland) under a SZ2-ILST Stereomicroscope (Olympus, Shinjuku, Japan). Twenty insects, including male adults and female adults, were used for tissue dissection.

The isolated SGs, guts, and residual parts were homogenized in RNAiso plus (TaKaRa, Dalian, China) using FastPrep-24 5G (MP Biomedicals, Santa Ana, California, USA). The residual part sample was collected by carefully removing SGs and guts in the PBS solution under a SZ2-ILST Stereomicroscope. Twenty SGs, twenty guts, and ten residual parts were used, respectively. Total RNA extraction followed the manufacturer's protocols, and the RNA samples were forwarded to Novogene Institute (Novogene, Beijing, China) for transcriptomic sequencing. Briefly, mRNA isolation from total RNA employed poly-T oligo-attached magnetic beads. mRNA fragmentation occurred under elevated temperature using divalent cations in NEBNext First Strand Synthesis Reaction Buffer (5X). First strand cDNA synthesis utilized a random hexamer primer and M-MuLV Reverse Transcriptase. Subsequently, the second strand cDNA synthesis was carried out using DNA Polymerase I and RNase H. Blunt



**Figure 1.** Pipeline used in analyzing salivary proteins in *Epeurysa nawaii*. Transcriptomic sequencing was performed in salivary glands (SGs), gut, and the residual parts excluding SGs and guts. The output reads were combined for Trinity assembly. Genes with an FPKM value > 100 in the SGs were defined as SG high expressed genes. Two gene categories were classified as salivary proteins: (1) secretory genes that were highly and specifically expressed in SGs; (2) SG high-expressed genes that have homology in saliva of related insect species. Totally 165 salivary proteins were identified following this pipeline.

ends were generated by converting the remaining overhangs using exonuclease/polymerase activities. Following adenylation of DNA fragment 3' ends, NEBNext Adaptors with hairpin loop structures were ligated to prepare fragments for hybridization. To select cDNA fragments within the preferred length of 250–300 bp, library fragments were purified using the AMPure XP system (Beckman Coulter, Beverly, USA). Subsequently, 3 µl of USER Enzyme (NEB, USA) treated the size-selected, adaptor-ligated cDNA, incubating at 37 °C for 15 min, followed by 5 min at 95 °C, before PCR amplification. PCR employed Phusion High-Fidelity DNA polymerase, Universal PCR primers, and Index Primer. Finally, PCR products were purified using the AMPure XP system, and library quality was assessed using the Agilent Bioanalyzer 2100 system.

The index-coded samples were clustered using the acBot Cluster Generation System, with the TruSeq PE Cluster Kit v3-cBot-HS (Illumina), following the manufacturer's guidelines. After cluster generation, library preparations underwent sequencing on an Illumina Novaseq6000 platform, producing 150 bp paired-end reads. One library was constructed for each sample, and no biological replicate was performed. The output data were submitted to the National Genomics Data Center under accession number: PRJCA022401.

### Transcriptome assembly

Clean reads were obtained using the internal software Fastp version 0.23.1<sup>21</sup>, which involved the removal of low-quality reads containing adapters, empty reads, or sequences with unknown bases "N" from the raw data. The quality of the reads was assessed using the FastQC program v0.11.3<sup>22</sup>.

For transcriptome assembly, the left files (read1 files) from all samples were consolidated into a single large left.fq file, and the right files (read2 files) were consolidated into a single large right.fq file. Transcriptome assembly was then conducted based on the left.fq and right.fq files using Trinity<sup>23</sup>. The min\_kmer\_cov parameter was set to 2 by default, and all other parameters were set to their default values.

### Tissue-specific expressions analysis

Clean reads obtained from transcriptomic sequencing were aligned to the assembled unigenes using RNA-Seq by Expectation Maximization (RSEM)<sup>24</sup>. The quantification of relative transcript levels was performed using the expected number of Fragments Per Kilobase of transcript sequence per Million base pairs sequenced (FPKM), providing an accurate representation of transcript abundance. As this study lacked biological replicates, the read counts underwent additional normalization using the Trimmed Mean of M values method<sup>25</sup>. Differentially expressed unigenes in the SGs\_vs\_guts and SGs\_vs\_residue comparisons were identified based on predefined criteria: a fold change ratio greater than 10 and a q-value lower than 0.005<sup>14</sup>. The SG-high expressed genes were subjected to Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis using TBtools software v1.069770<sup>26</sup>.

### Gene annotation and protein prediction

The assembled unigenes underwent analysis through the BLASTX algorithm by querying the GenBank databases. Gene Orthology (GO) and KEGG Orthology (KO) annotations for these unigenes were established using Blast2GO (<http://www.blast2go.org/>) and InterProScan software (<http://www.ebi.ac.uk/Tools/pfa/ipscan/>), respectively. Additionally, the unigenes were cross-referenced with the Clusters of Orthologous Groups (COG) database (<http://www.ncbi.nlm.nih.gov/COG/>).

The coding sequence of each gene was analyzed by searching against the NCBI NR, SwissProt, COG, and KEGG databases. Subsequently, the remaining unigenes underwent further prediction using TransDecoder<sup>27</sup>.

### Analysis of secretory proteins

The prediction of signal peptides was conducted using the SignalP 6.0 Server (<https://services.healthtech.dtu.dk/services/SignalP-6.0/>). For transmembrane domain prediction, each amino acid sequence containing a signal peptide was submitted to the TMHMM Server v. 2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>). Proteins lacking transmembrane domains or having only one transmembrane domain, which partially overlapped with the predicted signal peptide, were classified as secreted proteins.

### Identification of taxonomy-specific proteins

To identify potential Delphacidae-specific proteins, the salivary proteins were subject to BLAST search against the predicted proteins in *Drosophila melanogaster*<sup>28</sup>, *Acyrtosiphon pisum*<sup>29</sup>, *Bemisia tabaci*<sup>30</sup>, and *Riptortus pedestris*<sup>31</sup>, employing a cutoff E-value of 10<sup>-10</sup>, respectively. Only genes with no homologous gene in above species were defined as the Delphacidae-specific genes. To identify potential *E. nawaii*-specific proteins, the Delphacidae-specific proteins were subject to BLAST search against the predicted proteins in *N. lugens*<sup>32</sup>, *L. striatellus*<sup>33</sup>, and *S. furcifera*<sup>34</sup> with a cutoff E-value of 10<sup>-10</sup>, respectively. Only genes with no homologous gene in three rice planthopper species were defined as the *E. nawaii*-specific genes.

### Comparative analysis of insect saliva proteins

The salivary proteins from aphids (*Acyrtosiphon pisum*, *Macrosiphum euphorbiae*, *Myzus persicae*, *Diuraphis noxia*, *Schizaphis graminum*, and *Megoura viciae*)<sup>35–40</sup>, planthoppers (*Nilaparvata lugens*, *Laodelphax striatellus*, *Sogatella furcifera*)<sup>18,41,42</sup>, whiteflies (*B. tabaci*)<sup>14</sup>, and psyllid (*D. citri*)<sup>11</sup> were extracted following published sequence information. The identification of both shared and species-specific salivary proteins among different arthropod species was accomplished by BLASTing *E. nawaii* salivary proteins against above sequences with a cutoff E-value of 10<sup>-10</sup>.

## Results

### Overview of *E. nawai* transcriptome

To depict the gene expression patterns across various tissues in *E. nawai*, cDNA libraries were prepared for SGs, guts, and residual parts excluding SGs and guts. High-throughput sequencing of these samples generated approximately 45 million paired-end (PE) reads per library. Raw reads underwent preprocessing steps, including sequence duplication removal and GC content analysis using FastQC software (Table 1). The summaries of the total clean/raw reads, the total number of bases in the clean/raw data, and the error rate of average base sequencing, Q20, and Q30 are presented in Table 1.

Due to the absence of an available *E. nawai* genome, the output reads were combined for Trinity assembly, resulting in a total of 18,395 unigenes with a mean length of 1678 bp and an N50 of 2200 bp (Table S1). Subsequently, initial coding sequence predictions were performed by searching against the NCBI NR, SwissProt, COG, and KEGG databases, yielding significant BLASTX hits for 11,478 unigenes. Further predictions were made using TransDecoder, identifying an additional 4304 putative proteins from the remaining unigenes. In total, our analysis predicted 15,782 coding genes in the transcriptome assembly (Table S2), providing a foundational basis for subsequent analyses.

### Functional analysis of high-expressed genes in SGs

We conducted an in-depth analysis of highly expressed genes in *E. nawai* SGs to elucidate their primary functions. Genes with an FPKM value exceeding 100 in SGs were considered highly expressed. A total of 416 genes met this criterion (Table S3). Notably, a significant proportion of these genes were ribosomal proteins, which mainly associated with protein synthesis. Additionally, SGs exhibited a notable presence of functionally unknown or hypothetical proteins, hinting at potential associations with the rapid evolution of salivary genes. Furthermore, genes previously validated to be secreted in rice planthoppers, such as mucin-like protein, vitellogenin, annexin-like protein, and salivap-3, were highly expressed in *E. nawai* SGs. Remarkably, over 16% of highly expressed SG genes featured a signal peptide (Table S3), suggesting their potential delivery into saliva through the eukaryotic endoplasmic reticulum–Golgi pathway.

To gain insights into the functions of these highly expressed SG genes, KEGG pathway enrichment analysis was performed. Five pathways showed significant enrichment, including energy metabolism (31 proteins), ribosome (72 proteins), translation (76 proteins), oxidative phosphorylation (29 proteins), and exosome (23 proteins) (Fig. 2). These results suggest that SGs are primarily involved in active protein synthesis and secretion.

### Identification of salivary proteins in *E. nawai*

Salivary proteins typically exhibit conserved characteristics, including high and specific expression in SG, the presence of a signal peptide at the N-terminus, and sequence conservation in related insect species<sup>17,18</sup>. In our study, we investigated the expression patterns of 416 SG-high expressed genes in SGs, guts, and residual parts. The results revealed that 180 genes displayed SG-specificity, with 43 among the Top50 high-expressed SG genes being specifically expressed in SGs (Table S3). Regarding secretory prediction, 35.1% (146 out of 416) of genes harbored a signal peptide at the N-terminus (Table S3). Notably, this proportion significantly increased in the Top50 genes, with 62% predicted to be secretable. Consequently, based on the above criteria, we identified a total of 114 secretory genes that were highly and specifically expressed in SGs (Table S3).

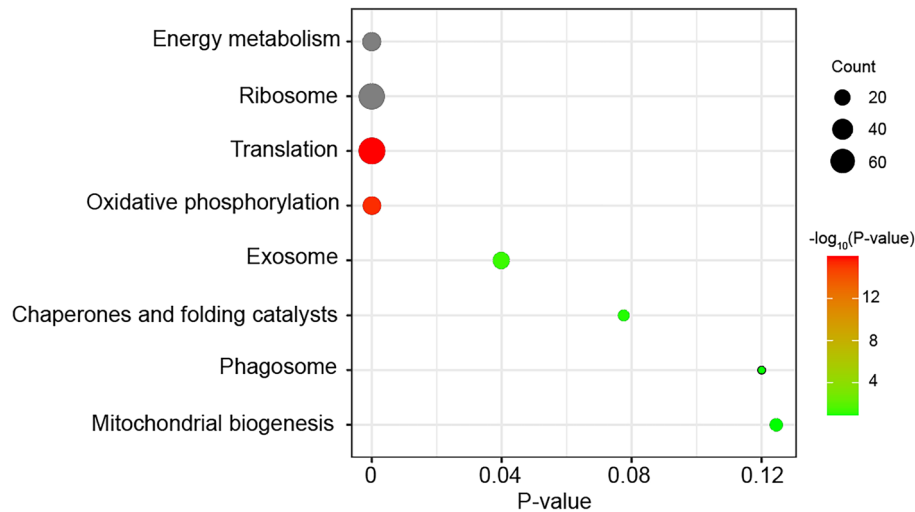
For the remaining 302 SG-high expressed genes, we investigated whether their homologues were present in the secreted saliva of three previously reported planthopper species. This analysis led to the identification of an additional 51 proteins, with 17 specifically expressed in SGs and 7 containing a signal peptide. Notably, the salivary effectors vitellogenin and annexin, validated for their role in impairing plant defenses<sup>43,44</sup>, were identified using this method, despite not being specifically expressed in SGs. Similarly, the mucin-like protein and salivap-3 protein, which did not predictably have a signal peptide, were also considered potential proteins of *E. nawai* saliva. Collectively, we classified 165 proteins as salivary proteins of *E. nawai*, forming the basis for subsequent analyses (Table S4).

### Comprehensive analysis of salivary proteins in different insects

According to gene annotation, 123 salivary proteins were annotated as hypothetical proteins, uncharacterized proteins, or unannotated (Table S4). To assess their species-specificity, the 165 salivary proteins were subjected to a BLAST search against protein repertoire predicated by genome in various insects, including *N. lugens*, *L. striatellus*, *S. furcifera*, *A. pisum*, *B. tabaci*, *R. pedestris*, *H. vitripennis*, and *D. melanogaster*, with a cutoff E-value of  $10^{-10}$ . The results showed that 60 salivary proteins were specific to *E. nawai*, lacking homologues in other

Sample	Raw reads	Clean reads	Raw base (G) <sup>a</sup>	Clean base (G) <sup>b</sup>	Error (%) <sup>c</sup>	Q20 (%) <sup>d</sup>	Q30 (%) <sup>e</sup>	GC content (%)
SG	50,569,634	49,798,772	7.59	7.47	0.03	97.70	93.60	45.89
Gut	46,676,596	45,777,776	7.00	6.87	0.03	97.84	93.72	42.27
Residue	43,216,632	42,265,910	6.48	6.34	0.03	97.47	93.20	41.40

**Table 1.** Summary of statistics from Illumina sequencing. <sup>a</sup>Raw Base, The total number of bases in the raw data. <sup>b</sup>Clean Base, The total number of bases in the clean data. <sup>c</sup>Error, Error rate of average base sequencing. <sup>d</sup>Q20, Percentage of bases with Phred quality scores > 20. <sup>e</sup>Q30, Percentage of bases with Phred quality scores > 30.



**Figure 2.** Enrichment analysis of salivary gland (SG)-high expressed genes. The SG -high expressed genes were underwent Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis using TBools software<sup>26</sup>. Enriched P values were calculated according to one-sided hypergeometric test.

insects, even closely related Delphacidae species (Table S4). Additionally, 62 salivary proteins exhibited Delphacidae-specificity, with no homologues found outside the Delphacidae. In contrast, only 43 salivary proteins were prevalently existed across various insects (Table S4).

To date, saliva collection on artificial diets has been performed in several Hemiptera insects, including rice planthoppers (*N. lugens*, *S. furcifera*, and *L. striatellus*), aphids (*A. pisum*, *M. euphorbiae*, *M. persicae*, *D. noxia*, and *S. graminum*), psyllid (*D. citri*), and whitefly (*B. tabaci*). The salivary proteins of *E. nawaiii* were compared with those of these 10 species. Seventy-nine *E. nawaiii* proteins had homologues in the saliva of other insects, categorizable into 53 groups based on their gene annotation (Table S5). Seven proteins were widely distributed in insect saliva, including aminopeptidase, carbonic anhydrase, glyceraldehyde-3-phosphate dehydrogenase, elongation factor, apolipoporphins, heat shock protein, and ribosomal protein (Fig. 3). Thirty proteins were exclusively found in the saliva of Delphacidae species (Table S5). For instance, the salivary sheath protein mucin and calcium-binding protein annexin were identified across all planthopper species. Salivap-5 and PI-PLC X

	Nil	Las	Sof	Mae	Scg	Myp	Acp	Mev	Din	Bet	Dic
Aminopeptidase	✓		✓			✓	✓	✓	✓	✓	✓
Carbonic anhydrase	✓	✓		✓	✓	✓	✓	✓			✓
Glyceraldehyde-3-phosphate dehydrogenase	✓		✓	✓		✓	✓			✓	✓
Elongation factor	✓	✓	✓			✓	✓	✓		✓	
Apolipoporphins			✓	✓		✓	✓	✓	✓	✓	
Heat shock cognate protein	✓		✓	✓						✓	✓
Ribosomal protein	✓	✓		✓						✓	✓
Actin	✓		✓							✓	✓
Arginine kinase	✓			✓						✓	✓
ATP synthase	✓			✓						✓	✓
Peroxiredoxin	✓							✓		✓	✓
Tubulin	✓			✓						✓	✓
Polyubiquitin protein	✓			✓						✓	✓

**Figure 3.** Comparative analysis of salivary proteins in *Epeurysa nawaiii* with other Hemiptera species. Nil, *Nilaparvata lugens*; Lao, *Laodelphax striatellus*; Sof, *Sogatella furcifera*; Mae, *Macrosiphum euphorbiae*; Scg, *Schizaphis graminum*; Myp, *Myzus persicae*; Acp, *Acyrtosiphon pisum*; Mev, *Megoura viciae*; Din, *Diuraphis noxia*; Bet, *Bemisia tabaci*; Dic, *Diaphorina citri*. *E. nawaiii* salivary proteins with their homologous in saliva of more than four species were displayed.

domain-containing protein were identified in *E. nawai*, *N. lugens*, and *S. furcifera*, but not in *L. striatellus*. Salivap-3 and salivap-4 were exclusively identified in *E. nawai* and *N. lugens* (Table S5).

## Discussion

In this study, we conducted a comprehensive analysis of the salivary proteins of *E. nawai* using a transcriptome-based pipeline. Our focus was on secretory proteins that were highly and specifically expressed in SG, as well as SG-high expressed proteins that had homologues in the saliva of other planthopper species. A total of 165 proteins were identified as salivary proteins, with the majority exhibiting Delphacidae-specificity, and some being specific to *E. nawai*. Our findings present an efficient pipeline for the identification of salivary proteins, establishing a robust foundation for understanding the mechanisms of host manipulation by salivary effectors.

Saliva plays critical roles in digestion, detoxification, and manipulation of plant defenses, attracting increasing attention in recent years. However, identifying insect salivary proteins poses challenges due to the tiny size of insects, low secretion amounts, and the propensity for degradation after secretion<sup>12</sup>. In most cases, insect saliva is identified by directly analyzing secreted saliva in artificial diets through a proteomic approach or indirectly analyzing SG extracts via a transcriptomic approach<sup>8–11</sup>. In proteomic analyses, while salivary proteins have been revealed in several insect species, the number of identified salivary proteins varies greatly, ranging from less than ten in *Metopolophium dirhodum* to more than two hundred in *D. citri*<sup>11,45</sup>. This variability may be attributed to differences in saliva collection procedures and the sensitivity of mass spectrometry analysis. In our study, utilizing transcriptomic sequencing and bioinformatic analysis, we successfully identified the primary functions of SG and the potential salivary proteins it secretes (Fig. 1). Core salivary proteins, including digestive enzymes (aminopeptidase, cathepsin, and venom protease), calcium-binding proteins (annexin and EF-hand domain-containing protein), oxidoreductases (peroxiredoxin and glyceraldehyde-3-phosphate dehydrogenase), and reported salivary effectors (vitellogenin and carbonic anhydrase), were identified<sup>44,46–48</sup>. These results affirm the efficiency and reliability of our pipeline in identification of salivary proteins.

According to previous research, the majority of salivary proteins are typically secretable and specifically expressed in SG<sup>18</sup>. Our analysis identified 114 proteins that met these criteria. However, a few proteins were highly and specifically expressed in SG without a detectable signal peptide. This discrepancy may be attributed to partial sequence assembly or inaccurate protein prediction. Additionally, unconventional secretion pathways and limitations in signal peptide-based predictions could contribute to the absence of signal peptides in some salivary proteins<sup>49</sup>. Although most salivary proteins displayed SG-specificity, exceptions were noted in some salivary effectors. For instance, salivary maltase and trypsin-like protease, involved in extra-oral digestion, were specifically expressed in guts<sup>18,50</sup>. Vitellogenin, a salivary effector suppressing plant defenses, was mainly expressed in ovaries and fat bodies<sup>44,48</sup>. While our study employed homology analysis to include conserved proteins found in saliva of closely related species, further investigation is warranted to capture potentially omitted salivary proteins.

The intriguing feature of *E. nawai* saliva lies in the abundance of taxonomy-specific proteins, known as orphan proteins. Saliva plays a crucial role in the evolutionary battle between plants and herbivorous insects<sup>1,2</sup>. Orphan salivary genes have been reported as critical for insect survival, such as Bsp9 in whiteflies for suppressing plant defenses<sup>51</sup>, salivap-4 in planthoppers for salivary sheath formation<sup>3</sup>, and C002 in aphids foraging and feeding<sup>52</sup>. In *E. nawai*, 60 salivary proteins exhibited *E. nawai*-specificity, with no homologues found even in closely related Delphacidae species (Table S4). Given the distinct feeding habits of *E. nawai* compared to *N. lugens*, *S. furcifera*, and *L. striatellus*, which primarily feed on rice plants, our findings suggest that feeding habits may influence the composition of salivary proteins. The origin of species-specific proteins is a crucial aspect of insect evolution<sup>53</sup>. While previous research demonstrated horizontal gene transfer as a mechanism for recruiting orphan salivary proteins<sup>6,54</sup>, no homologues of *E. nawai*-specific proteins were found in other organisms outside the Insecta, warranting further research to uncover their origin and function.

In conclusion, our study provided an efficient and reliable pipeline for identifying salivary proteins. The widespread identification of digestive enzymes, calcium-binding proteins, and oxidoreductases indicate the conserved role of saliva in mediating plant–insect interactions. The abundant presence of orphan salivary proteins underscores the importance of these proteins for insects to adapt to specific host plants.

## Data availability

The output transcriptomic data were submitted to the National Genomics Data Center under accession number: PRJCA022401.

Received: 22 January 2024; Accepted: 12 March 2024

Published online: 14 March 2024

## References

- Elzinga, D. A. & Jander, G. The role of protein effectors in plant-aphid interactions. *Curr. Opin. Plant Biol.* **16**, 451–456 (2013).
- Miles, P. W. Aphid saliva. *Biol. Rev. Camb. Philos. Soc.* **74**, 41–85 (1999).
- Huang, H. J. *et al.* A salivary sheath protein essential for the interaction of the brown planthopper with rice plants. *Insect Biochem. Mol. Biol.* **66**, 77–87 (2015).
- Xu, H. X. *et al.* A salivary effector enables whitefly to feed on host plants by eliciting salicylic acid-signaling pathway. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 490–495 (2019).
- Furch, A. C., Van Bel, A. J. & Will, T. Aphid salivary proteases are capable of degrading sieve-tube proteins. *J. Exp. Bot.* **66**, 533–539 (2015).
- Boulain, H. *et al.* Fast evolution and lineage-specific gene family expansions of aphid salivary effectors driven by interactions with host-plants. *Genome Biol. Evol.* **10**, 1554–1572 (2018).
- Rao, W. *et al.* Secretome analysis and in planta expression of salivary proteins identify candidate effectors from the brown planthopper *Nilaparvata lugens*. *Mol. Plant Microbe Interact.* **32**, 227–239 (2019).

8. Fu, W. *et al.* Screening candidate effectors of the bean bug *Riptortus pedestris* by proteomic and transcriptomic analyses. *Front. Ecol. Evol.* **9**, 760368 (2021).
9. Hattori, M., Komatsu, S., Noda, H. & Matsumoto, Y. Proteome analysis of watery saliva secreted by green rice leafhopper, *Nephotettix cincticeps*. *PLoS ONE* **10**, e0123671 (2015).
10. Wim Jonckheere, W. D. *et al.* The salivary protein repertoire of the polyphagous spider mite *Tetranychus urticae*: a quest for effectors. *Mol. Cell. Proteomics* **15**, 3594–3613 (2016).
11. Wu, Z. Z., Qu, M. Q., Chen, M. S. & Lin, J. T. Proteomic and transcriptomic analyses of saliva and salivary glands from the Asian citrus psyllid, *Diaphorina citri*. *J. Proteomics* **238**, 104136 (2021).
12. Van Bel, A. J. & Will, T. Functional evaluation of proteins in watery and gel saliva of aphids. *Front. Plant Sci.* **7**, 1840 (2016).
13. Huang, H. J. *et al.* Planthopper salivary sheath protein LsSP1 contributes to manipulation of rice plant defenses. *Nat. Commun.* **14**, 737 (2023).
14. Huang, H. J. *et al.* Identification of salivary proteins in the whitefly *Bemisia tabaci* by transcriptomic and LC-MS/MS analyses. *Insect Sci.* **28**, 1369–1381 (2021).
15. Olsen, J. V., Ong, S. E. & Mann, M. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol. Cell. Proteomics* **3**, 608–614 (2004).
16. Dong, Y. *et al.* The mirid bug *Apolygus lucorum* deploys a glutathione peroxidase as a candidate effector to enhance plant susceptibility. *J Exp Bot* **71**, 2701–2712 (2020).
17. Huang, H. J. *et al.* Identification of *Riptortus pedestris* salivary proteins and their roles in inducing plant defenses. *Biology* **10**, 753 (2021).
18. Huang, H. J. *et al.* Screening and functional analyses of *Nilaparvata lugens* salivary proteome. *J. Proteome Res.* **15**, 1883 (2016).
19. Norowi, M., Sriratanasak, W., Xia, J. & Lu, Z. G. *Situation of planthoppers in Asia* (International Rice research Institute, 2010).
20. Yang, L., Chen, Y. & Chen, H. Notes on planthoppers infesting bamboo in Guizhou. *J. Mt. Agric. Biol.* **18**, 154–161 (1999).
21. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
22. Andrews, S. *FastQC a Quality Control Tool for High Throughput Sequence Data* (Babraham Institute, 2014).
23. Grabherr, M. G. *et al.* Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.* **29**, 644 (2011).
24. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **12**, 323 (2011).
25. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).
26. Chen, C. *et al.* TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* **13**, 1194–1202 (2020).
27. Haas, B., Papanicolaou, A. & Yassour, M. (2017) TransDecoder. *TransDecoder*.
28. Hoskins, R. A. *et al.* Sequence finishing and mapping of *Drosophila melanogaster* heterochromatin. *Science* **316**, 1625–1628 (2007).
29. Li, Y., Park, H., Smith, T. E. & Moran, N. A. Gene family evolution in the pea aphid based on chromosome-level genome assembly. *Mol. Biol. Evol.* **36**, 2143–2156 (2019).
30. Xie, W. *et al.* Genome sequencing of the sweetpotato whitefly *Bemisia tabaci* MED/Q. *GigaScience* **6**, gix018 (2017).
31. Huang, H. J. *et al.* Chromosome-level genome assembly of the bean bug *Riptortus pedestris*. *Mol Ecol Resour.* **21**, 2423–2436 (2021).
32. Ye, Y. X. *et al.* Chromosome-level assembly of the brown planthopper genome with a characterized Y chromosome. *Mol. Ecol. Resour.* **21**, 1287–1298 (2021).
33. Hu, Q. L. *et al.* Chromosome-level assembly, dosage compensation and sex-biased gene expression in the small brown planthopper, *Laodelphax striatellus*. *Genome Biol. Evol.* **14**, evac160 (2022).
34. Ye, Y.-X., Li, D.-T., Zhang, S.-Y., Shen, Z.-C. & Zhang, C.-X. Chromosome-level genome assembly and sex-specific differential transcriptome of the white-backed planthopper, *Sogatella furcifera*. *Curr. Genomics* **23**, 400–411 (2022).
35. Carolan, J. C., Fitzroy, C. I. J., Ashton, P. D., Douglas, A. E. & Wilkinson, T. L. The secreted salivary proteome of the pea aphid *Acyrtosiphon pisum* characterised by mass spectrometry. *Proteomics* **9**, 2457–2467 (2010).
36. Chaudhary, R., Atamian, H. S., Shen, Z., Briggs, S. P. & Kaloshian, I. Potato aphid salivary proteome: Enhanced salivation using resorcinol and identification of aphid phosphoproteins. *J. Proteome Res.* **14**, 1762–1778 (2015).
37. Harmel, N. *et al.* Identification of aphid salivary proteins: A proteomic investigation of *Myzus persicae*. *Insect Mol. Biol.* **17**, 165–174 (2010).
38. Nicholson, S. J., Hartson, S. D. & Puterka, G. J. Proteomic analysis of secreted saliva from Russian Wheat Aphid (*Diuraphis noxia* Kurd.) biotypes that differ in virulence to wheat. *J. Proteomics* **75**, 2252–2268 (2012).
39. Nicholson, S. J. & Puterka, G. J. Variation in the salivary proteomes of differentially virulent greenbug (*Schizaphis graminum* Rondani) biotypes. *J. Proteomics* **105**, 186–203 (2014).
40. Vandermoten, S. *et al.* Comparative analyses of salivary proteins from three aphid species. *Insect Mol. Biol.* **23**, 67–77 (2014).
41. Huang, H. J., Lu, J. B., Li, Q., Bao, Y. Y. & Zhang, C. X. Combined transcriptomic/proteomic analysis of salivary gland and secreted saliva in three planthopper species. *J. Proteomics* **172**, 25–35 (2018).
42. Liu, X., Zhou, H., Zhao, J., Hua, H. & He, Y. Identification of the secreted watery saliva proteins of the rice brown planthopper, *Nilaparvata lugens* (Stål) by transcriptome and shotgun LC-MS/MS approach. *J. Insect Physiol.* **89**, 60–69 (2016).
43. Chen, C. *et al.* An ANNEXIN-Like protein from the cereal cyst nematode heterodera avenae suppresses plant defense. *PLoS ONE* **10**, e0122256 (2015).
44. Ji, R. *et al.* Vitellogenin from planthopper oral secretion acts as a novel effector to impair plant defenses. *New Phytol.* **232**, 802–817 (2021).
45. Rao, S. A., Carolan, J. C. & Wilkinson, T. L. Proteomic profiling of cereal aphid saliva reveals both ubiquitous and adaptive secreted proteins. *PLoS ONE* **8**, e57413 (2013).
46. Guo, H. *et al.* Salivary carbonic anhydrase II in winged aphid morph facilitates plant infection by viruses. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2222040120 (2023).
47. Huang, H. J., Zhang, C. X. & Hong, X. Y. How does saliva function in planthopper-host interactions?. *Arch. Insect Biochem. Physiol.* **100**, e21537 (2019).
48. Zeng, J. *et al.* The N-terminal subunit of vitellogenin in planthopper eggs and saliva acts as a reliable elicitor that induces defenses in rice. *New Phytol.* **238**, 1230–1244 (2023).
49. Nickel, W. & Rabouille, C. Mechanisms of regulated unconventional protein secretion. *Nature* **10**, 148–155 (2009).
50. Cohen, A. C. Extra-oral digestion in predaceous terrestrial Arthropoda. *Annu. Rev. Entomol.* **40**, 85–103 (1995).
51. Wang, N., Zhao, P. & Ma, Y. A whitefly effector Bsp9 targets host immunity regulator WRKY33 to promote performance. *Philos. Trans. R. Soc. B* **8**, 20180313 (2019).
52. Mutti, N. S. *et al.* A protein from the salivary glands of the pea aphid, *Acyrtosiphon pisum*, is essential in feeding on a host plant. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 9965–9969 (2008).
53. Tautz, D. & Domazet-Lošo, T. The evolutionary origin of orphan genes. *Nat. Rev. Genet.* **12**, 692–702 (2011).
54. Wang, Y. Z. *et al.* Horizontally transferred salivary protein promotes insect feeding by suppressing ferredoxin-mediated plant defenses. *Mol. Biol. Evol.* **40**, msad221 (2023).

## Acknowledgements

This work was supported by the Natural Science Foundation of Zhejiang Province (LDQ24C140001: H.J.H) and the National Key Research and Development Plan in the 14th five year plan (2021YFD1401100: H.J.H).

## Author contributions

H.-J.H. planned and designed the research. X.-J.W., Q.L., and Z.-X.Y. performed experiments and analyzed data. H.-J.H. drafted the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-56896-4>.

**Correspondence** and requests for materials should be addressed to H.-J.H.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024, corrected publication 2024