

# Chromosome-level genome assembly of the bean bug *Riptortus pedestris*

Hai-Jian Huang<sup>1</sup>, Yu-Xuan Ye<sup>2</sup>, Zhuang-Xin Ye<sup>1</sup>, Xiao-Tian Yan<sup>1</sup>, Xin Wang<sup>1</sup>, Zhong-Yan Wei<sup>1</sup>, Jian-Ping Chen<sup>1</sup>, Jun-Min Li<sup>1</sup>, Zong-Tao Sun<sup>1</sup>, and Chuan-Xi Zhang<sup>2</sup>

<sup>1</sup>Affiliation not available

<sup>2</sup>Zhejiang University

December 13, 2020

## Abstract

The bean bug (*Riptortus pedestris*) causes great economic losses of soybeans by piercing and sucking pods and seeds. Although *R. pedestris* has become the focus of numerous studies associated with insect–microbe interactions, plant–insect interactions, and pesticide resistance, a lack of genomic resources has limited deeper insights. In this study, we report the first *R. pedestris* genome at the chromosomal level using PacBio, Illumina, and Hi-C technologies. The assembled genome was 1.193 Gb in size with a contig N50 of 13.97 Mb. More than 95.7% of the total genome bases were successfully anchored to 6 unique chromosomes, with the scaffold N50 reaching 181.34 Mb. Genome resequencing of male and female individuals and chromosomal staining demonstrated that the sex chromosome system of *R. pedestris* is XO, and the shortest chromosome is the X chromosome. In total, 21,562 protein-coding genes were predicted, 21,320 of which were validated as being expressed in different tissues or different developmental stages. Evolutionary analysis demonstrated that *R. pedestris* and *Oncopeltus fasciatus* formed a sister group and split 35 million years ago. Additionally, a 5.04 Mb complete genome of symbiotic *Serratia marcescens* Rip1 was assembled, and the virulence factors that account for successful colonization in the host midgut were identified. The high-quality *R. pedestris* genome provides a valuable resource for further research, as well as for the pest management of bug pests.

## 2 Introduction

The bean bug, *Riptortus pedestris* (Fabricius) (Hemiptera: Heteroptera) is a notorious pest of soybean crops in Asia (Figure 1A). They obtain nutrients and water by penetrating their sucking mouthpart into leaves, stems, pods, and seeds, leading to significant reductions in seed quality (Bae et al. 2014; Rahman and Lim 2017). Although losses caused by *R. pedestris* have been mainly reported for leguminous crops, the insects also damage rice, sesame, fruit trees, flowers, and medicinal plants (Bharathimeena and Sudharma 2009; Kang 2003). In recent years, increased attention has been paid to *R. pedestris* due to its potential roles in causing soybean *Zhengqing*, which is a staygreen syndrome that has resulted in yield losses from 0 to nearly 100% in the Huang-Huai-Hai River basin in China (Li et al. 2019). In field cage experiments, as well as under laboratory conditions, *R. pedestris* feeding was found to cause staygreen-like syndromes, with increased numbers of abnormal seeds, more empty pods, a lack of leaf senescence, and decreased yields in infested soybeans (Figure 1B) (Li et al. 2019; Li 2020). Insecticides, such as fenitrothion, bifenthrin, etofenprox, and acetamiprid have been widely used to control *R. pedestris* (Alim and Lim 2010; Jung et al. 2018). However, the chemical control of *R. pedestris* is often incompatible with natural enemies and causes pest resurgence, leading to nonnegligible environmental problems.

In addition to causing economic losses, *R. pedestris* is also an ideal model for investigating symbiotic microbe–insect associations (Takeshita and Kikuchi 2017). Its midgut crypts are filled with numerous environmentally acquired symbionts (Jung and Lee 2019). Among them, the symbiotic *Burkholderia*, which can be easily

cultured and genetically manipulated outside the host, has been best characterized (Ohbayashi et al. 2019; Takeshita and Kikuchi 2017). Different from most insect–microbe symbiotic systems, *Burkholderia* cannot be vertically transmitted. Instead, they reside within the soil, and *R. pedestris* orally acquires the symbiont from surrounding environments (Kikuchi et al. 2011). Although *Burkholderia* symbionts are not essential for insect survival and *R. pedestris* can be reared in aseptic conditions, aposymbiotic insects exhibit retarded growth, smaller body sizes and weights, and lower fecundity (Kikuchi and Fukatsu 2014; Kikuchi et al. 2007). Strikingly, recent findings demonstrated that some insecticide-degrading *Burkholderia* mediate insecticide resistance in *R. pedestris*, providing the insects with a potential advantage in insecticide-sprayed crop fields (Itoh et al. 2018; Kikuchi et al. 2012). Currently, the complete genome sequences of several *Burkholderia* strains have been reported (Takeshita et al. 2014; Takeshita et al. 2018). It is necessary to unveil the genomic information of their insect hosts as the construction of a high-quality genome can provide new insights into the mechanisms of insect–microbe interactions.

In this study, we report the first genome of a bean bug in Coreoidea with a determined sex chromosome. The genome was *de novo* assembled using PacBio and Illumina sequencing technology as well as Hi-C technology to assist in chromosome-level assembly. The genomic resource obtained for *R. pedestris* will benefit further research associated with the insect–microbe interactions, insect–plant interactions, and pesticide resistance of insects, and it provide invaluable information for controlling bug pests.

### 3 Material and methods

#### Insect strain

The *R. pedestris* bean bugs sequenced in this study were originally collected in a soybean *Zhengqing* field in Suzhou, China in 2019. The insects were reared on soybean plants (strain: Wandou 27) at  $26 \pm 0.5$  °C with humidity of  $50 \pm 5\%$  under a 16/8 h (light/dark) photoperiod. Prior to DNA/RNA extraction, the colony was inbred following a sib-sib mating protocol for three generations as previously described (Xue et al. 2015).

#### Genome sequencing and assembly

Genomic DNA was isolated from three virgin female adults using the DNeasy Blood & Tissue Extraction Kit (Qiagen Inc., Valencia, CA, USA) based on the manufacturer’s protocol. After the determination of the DNA quality and quantity, a paired-end sequencing library (150 bp in length) was constructed and sequenced using the Illumina NovaSeq6000. In addition, a 20 kb library was constructed and sequenced on 17 cells using the Sequel™ Sequencing Plate 1.2 on the Pacific Biosciences Sequel platform at Berry Genomic Corporation in Beijing, China.

*De novo* genome assembly was performed using Canu-SMARTdenovo methods (Schmidt et al. 2017). Correction was performed by Canu with the default parameters (Koren et al. 2017). The resulting reads were assembled using SMARTdenovo (-k 21 -J 3000 -t 20) (Hu et al. 2020). Nextpolish was used for genome polishing (Hu et al. 2020).

#### Hi-C sequencing and chromosome-length genome assembly

Hi-C libraries were constructed from a pool of fifty 2<sup>nd</sup> instar nymphs according to the method previously described (Lieberman-Aiden et al. 2009). Briefly, the samples were fixed with 1% formaldehyde for 10 min at room temperature. Then, a 2.5 M-glycine solution was added to obtain a final concentration of 0.2 M. The sample was incubated for 5 min at room temperature and then on ice for 15 min to stop cross-linking completely. After centrifugation, the cross-linked sample was sent to Annoroad Gene Technology Co. Ltd, Beijing, China for sequencing.

The genomic contigs were scaffolded using the 3D *de novo* assembly (3D-DNA) pipeline (Dudchenko et al. 2017). Briefly, the Hi-C reads were aligned to the draft genome assembly using Juicer; a 3D-DNA analysis was run to generate a candidate assembly; the candidate assembly was reviewed using Juicebox Assembly Tools (JBAT); and a high-quality chromosome-length genome assembly was generated after JBAT review.

## Chromosomal staining

The salivary glands were dissected from male and female *R. pedestris* 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 pH 7.4) under a stereomicroscope (COIC, Chongqing, China). The samples were fixed in 4% paraformaldehyde (Sangon Biotech, Shanghai, China) for 1 h. After washing in PBS 3 times, the samples were triturated mechanically, and 20  $\mu$ l DAPI solution (Abcam, Cambridge, USA) was added to stain the chromosomes. Fluorescence images were examined using a Leica confocal laser-scanning microscope (Leica, Heidelberg, Germany).

## Transcriptomic sequencing

Third generation transcriptomic sequencing was performed by Oxford Nanopore technology (ONT). Four samples, including eggs, nymphs, male, and female *R. pedestris* were collected. Oxford PromethION 2D amplicon libraries were generated according to the Nanopore community protocol using library preparation kit SQK-LSK109 (Oxford Nanopore, Oxford, UK) and sequenced on R9 flow cells to generate fast5 files. All generated fast5 reads were then basecalled in Guppy v3.2.10 with the default options to yield fastq files (Bolognini et al. 2019). The fastq reads for each sample were filtered using Nanofilt v2.5.0 with options -l 100 -q 7 (De Coster et al. 2018). The full length reads were detected using Pypochopper with arguments -b primer.fa -i raw.fq -t 200 -s 98 (<https://github.com/nanoporetech/pypochopper>). Pinfish was run with default parameters, and finally yielded polished consensus reads. Second generation Illumina sequencing was performed as we previously described (Huang et al. 2020). Insect samples from 6 different tissues and 37 different development stages were analyzed.

## Genome annotation

Genome annotation was performed using the LoReAn annotation pipeline (Cook et al., 2019). Briefly, the transcriptomic data generated above were aligned to the genome using the Program to Assemble Spliced Alignments (PASA) (Haas et al. 2008) and the Genomic Mapping and Alignment Program (GMAP) (Wu and Watanabe 2005). For the protein sequence map, GeneWise (Birney et al. 2004) was employed. For ab initio gene prediction, SNAP (Korf 2004), Augustus (Stanke et al. 2006), MAKER2 (Holt and Yandell 2011), GeneMark-ET (Lomsadze et al. 2014), and BRAKER1 (Hoff et al. 2016) were employed separately. The Evidence Modeler (EVM) (Haas et al. 2008) was used to combine the outputs of previous tools to generate the combined annotation model. PASA was used to update the gene models by identifying untranslated regions (UTRs) to generate a final annotation.

## Sex chromosome identification

To identify the potential sex chromosome, paired-end sequencing libraries were constructed for male and female adult *R. pedestris*. The resulting DNA reads were separately mapped to the genomic scaffolds using Bowtie2 (Langmead and Salzberg 2012). Then, Sequence Alignment/Map tools (SAMtools) was used to remove the low-quality mapped reads, and the mapped reads per million (MRPM) of each chromosome in males and females were calculated (Li et al. 2009). The sex chromosomes were determined according to the female/male ratio of MRPM, as previously described (Ma et al. 2020).

## Weighted gene co-expression network analysis (WGCNA)

The WGCNA was performed based on the WGCNA package in R (3.2.2.) (Langfelder and Horvath 2008). The parameters were set as follows: min module size = 50 and ME miss thread = 0.15. Only the enrichment index >0.85 was considered to be significantly clustered.

## Phylogenetic analysis and divergence time estimation

Protein data from 12 hemipteran insects (*Bemisia tabaci*, *Acyrtosiphon pisum*, *Diaphorina citri*, *Homalodisca vitripennis*, *Laodelphax striatellus*, *Nilaparvata lugens*, *Gerris buenoi*, *Halyomorpha halys*, *Onopeltus fasciatus*, *Cimex lectularius*, *Rhodnius prolixus*, and *Triatoma rubrofasciata*) and *Drosophila melanogaster* were retrieved from the NCBI database, GigaDB database, VectorBase (Table S1). After filtering redundant alternative splicing events, the protein dataset containing non-redundant transcripts was

used to find the homologous pairs of sequences by the all-vs-all BLASTp algorithm with an E-value of  $<1e-5$ . The BLASTp result was then converted into a normalized similarity matrix and processed using OrthoMCL v2.0.9 (Li et al., 2003). Protein families were identified by Markov chain clustering (MCL-14-137).

A Phylogenetic tree was constructed using single copy orthologs in each species (1:1:1 genes identified by OrthoMCL analysis), and *D. melanogaster* was used to root the tree. Sequence alignment was performed by MAFFT (v7.310) (Katoh and Standley 2013), and conserved amino-acid sites were identified by TrimAl (v1.2rev59) (Capella-Gutiérrez et al. 2009). We used Model finder in IQ-TREE (v1.6.10) to determine the best model, and the phylogenetic tree was constructed using the maximum likelihood method with 1000 bootstraps (Nguyen et al. 2015). The divergence time was estimated based on fossil collection records from the Paleobiology Database ([www.paleobiology.org](http://www.paleobiology.org)) using the MCMCtree program in PAML v4.9e (Yang 1997).

### 3 Results and Discussion

#### Genome assembly

We adopted second-generation and third-generation sequencing for the comprehensive analysis of the *R. pedestris* genome. A total of 118.9 Gb clean data was generated using the Illumina platform. The k-mer (k=19) analysis indicated that the *R. pedestris* genome had high heterozygosity and the estimated size was 1.04 Gb (Figure S1). Using the PacBio RS II platform, a total of 132 Gb clean data was generated with an N50 read length of 18.1 kb (Table S2). Based on the sequencing data above, we obtained a contig-level assembly that contained 934 contigs and 1,211,875,501 bases. The contig N50 reached 13,968,850 bp, which is significantly longer than the genome assemblies of other bugs (Table 1; Table S3).

Then, we used Hi-C long-range scaffolding to improve our assembly. A total of 105 Gb raw data was generated, and 82-Gb filtered clean reads were used for the following analysis (Table S4). As a result, more than 95.7% of the total genome bases were successfully anchored to 6 unique chromosomes, with the chromosomal lengths ranging from 162.5 Mb to 285.2 Mb (Figure 2A; Figure 3). Chromosomal staining was performed to validate the Hi-C result, and we identified 12 chromosomes (n=5A+XX) in female *R. pedestris* and 11 chromosomes (n=5A+X) in male *R. pedestris* (Figure 2B). A previous work demonstrated that the sex chromosome system of most insects in the family Coreidae is XO (Muramoto 1978). The concordant results between Hi-C assembly and chromosomal staining confirmed the accuracy of our genome assembly and indicated that *R. pedestris* is XO-determinant.

There were 745 scaffolds that failed to anchor. These unlocalized scaffolds were further blasted against the bacteria database to identify the potential bacteria that resided within *R. pedestris*. A total of 168 scaffolds showed high homology to known bacteria, and these bacterial-origin scaffolds were used for further analysis (Table S5). Taking these results together, we assembled the *R. pedestris* genome (containing 6 chromosomes and 577 unlocalized scaffolds) with a final size of 1.193 Gb (GC content: 33.8%) and a scaffold N50 of 181.3 Mb (Table 1).

To assess the completeness of the genome assembly, the assembled genome was subjected to BUSCO analyses against insect, arthropod, metazoan, and eukaryote datasets. The results showed that 94.7% (against the insect dataset) to 96.5% (against the arthropod dataset) of BUSCO genes could be identified in the *R. pedestris* genome, suggesting the completeness and high quality of our genome assembly (Figure S2). There were 2,774,084 repetitive sequences identified, accounting for 56.0% of the *R. pedestris* genome (Table S6). Additionally, we identified 1,266,644 SSRs in the genome assembly (Table S7).

#### Identification of the *Serratia marcescens* genome

*R. pedestris* is well known for its intimate association with *Burkholderia*, which provides beneficial effects on insect development (Takeshita and Kikuchi 2017). Our genome assembly did not detect any *Burkholderia*-origin sequence, indicating the absence of *Burkholderia* in this insect strain. There were 168 bacterial-origin scaffolds from *Serratia*, *Enterobacter*, *Kluyvera*, and *Wolbachia* in our genome assembly (Table S5). *Serratia* was reported to confer host resistance to invading parasitoids (Bai et al. 2019), while *Wolbachia*

supplies essential nutrition for host reproduction (Ju et al. 2020). Although the insect used for genome sequencing was free of *Burkholderia*, we did not observe any abnormal phenotype previously described in aposymbiotic insects (Kikuchi et al. 2007). We presumed that additional bacteria, other than *Burkholderia*, provide specific nutritional or other contributions to *R. pedestris*, which is an aspect that deserves further investigation.

Among the bacterial-origin scaffolds, the longest scaffold (sc00000) could be successfully assembled into a circle, and we mainly focused on this complete bacteria genome in the present study. Phylogenetic analysis showed that the sc00000 was closely related to *Serratia marcescens* Db11 (Figure 4). *S. marcescens* is a gram-negative bacillus that belongs to *Enterobacteriaceae*. It has a wide host range, including plants, insects and animals (Grimont and Grimont 1978). In *R. pedestris*, *S. marcescens* was reported to colonize the insect midgut, but its function remained unknown (Lee et al. 2017). Our study, for the first time, assembled the complete genome of *S. marcescens* from *R. pedestris*, and we tentatively named this bacterial strain as *S. marcescens* Rip1 (hereafter: Rip1). The complete genome of Rip1 comprises a single circular chromosome of 5.04 Mbp, with an overall G+C content of 59.4% (Figure 4). Some *Serratia* strains, such as *S. marcescens* CAV1761 and *S. marcescens* AS-1, contain additional plasmids (Sakuraoka et al. 2019). However, no sequence homologous to these plasmids was found in Rip01. Overall, the genomes of Rip1 and Db11 are highly conserved and essentially collinear (Figure S3).

Previous work demonstrated that *S. marcescens* secretes virulence factors to counter antimicrobial proteins produced by *R. pedestris*, thus escaping host immune defenses (Chen et al. 2017; Lee et al. 2017). We identified at least 15 antibiotic-resistance genes and 32 putative virulence factors in Rip1 (Figure 4, Table S8, Table S9). These genes confer antibiotic resistance to a wide variety of antibiotics, including fosfomycin, mupirocin, polymyxin, aminocoumar, beta-lactam, isoniazid, and fluoroquinolone, and might be key in the successful colonization of *S. marcescens* in the host midgut.

#### Genome annotation

Oxford Nanopore technology (ONT) was used to aid the genome annotation. A total of 22 Gb clean data in 25,876,282 ONT subreads was obtained, with an N50 read length of 1,027 bp (Table S10). In total, we identified 21,562 non-redundant protein-coding genes using the LoReAn annotation pipeline. The number of predicted genes in *R. pedestris* was similar to those in *G. buenoi* and *O. fasciatus* but much higher than that in *C. lectularius*. Quality control showed that the predicted gene set was able to recover 92.0% (against the insect dataset) to 96.7% (against the eukaryote dataset) of BUSCO genes (Figure S4). The majority (21,139 genes, 98.0%) of the predicted genes were distributed within the 6 chromosomes. Among the 21,562 predicted genes, 17,909 genes were successfully annotated by at least one database in the NCBI Nr, Swiss-Prot, KOG, KEGG, GO, and InterPro domain databases (Figure S5-S8; Table S11). Additionally, we identified 10 ncRNAs, 51 rRNA, 4,103 tRNA, 221 small nuclear RNA, 80 microRNA, and 3,202 long ncRNAs in the *R. pedestris* genome (Table S12).

#### Expression patterns of predicted genes

RNA-seq data from 6 different tissues and 37 different development stages were analyzed (Table S13), and a total of 21,320 genes were determined to be transcribed (Supplementary files 1). For tissue-specific analysis, there were 498, 21, 401, 254, 2389, and 3181 genes found to be mainly expressed in the midgut, fat body, carcass, muscle, testis, and ovary, respectively (Figure S9A). Enrichment analysis demonstrated that most midgut-specific genes were involved in metabolism, exosomes, and proteasomes (Figure S10A), while most carcass-specific genes were involved in structural constituents of the cuticle, environmental adaption, amino acid metabolism, odorant binding and sensory perception (Figure S10B). The muscle-specific genes were mainly associated with the cytoskeleton, ion channels, lipid metabolism, and protein kinases (Figure S10C). There were abundant genes specifically expressed in gonads. For testis-specific analysis, the enriched pathways were mainly annotated as being associated with signaling and cellular processes, transporters, carotenoid biosynthesis, chromosome proteins, and cytochrome P450s (Figure S10D). For ovary-specific analysis, enriched pathways were associated with chromosome proteins, DNA metabolism, glycosyltransferase,

and glycan metabolism (Figure S10E). Additionally, we used LC-MS/MS to identify proteins authentically expressed in these 6 tissues (Figure S9B). A total of 2,622 proteins was successfully identified, with a unipep[?]2 (Supplementary files 2).

For the developmental analysis, a weighted gene co-expression network analysis (WGCNA) successfully assigned 9,887 genes into 32 modules (Figure S11). Among them, 6,684 genes (belonging to three modules) were more highly expressed in the egg period than during other stages. There were 1,467 and 227 genes that showed the highest expression in the male and female adult period, respectively. Additionally, we identified 352 genes (belonging to three modules) associated with insect molting. For example, insulin-like growth factor-binding protein 7, replication factor C, and endocuticle structural glycoprotein were more highly expressed before insect molting, while N-acetylneuraminase lyase, partitioning defective 3, and cytochrome P450 6a14 were more highly expressed during insect molting (Supplementary files 1). These data provide a good resource for future functional studies of *R. pedestris*.

#### Identification of the sex chromosomes and autosomes

To identify the potential sex chromosome, Illumina reads from male and female *R. pedestris* were mapped to the reference genome, respectively. In the XO system, males and females have only an X chromosome, and the MRPM of the X chromosome in females should be nearly two-fold higher than that in males (Ma et al. 2020). Based on this presumption, the shortest chromosome was determined to be the X chromosome, with nearly double the MRPM in females compared to that in males (1,414,872 versus 859,724). For the other 5 chromosomes, the female/male ratio of the MRPM ranged from 0.87 to 0.97, which is suggestive of autosomes (Figure 3, Table S14).

#### Gene family identification and phylogenetic tree construction

To investigate the phylogenetic position of *R. pedestris*, protein data from 12 hemipteran insects were retrieved from public database, and *D. melanogaster* was used to root the tree. A total of 23,850 gene families were clustered by OrthoMCL, and we selected 963 single copy genes with 356,827 reliable sites to construct the phylogenetic tree. The results showed that *R. pedestris* is a sister taxon to the milkweed bug *O. fasciatus*, and the 2 species diverged from their common ancestor at approximately 35 million years ago (Figure 5). Three hematophagous bugs (*R. prolixus*, *C. lectularius*, and *T. rubrofasciata*) form a sister lineage, and they clustered with three herbivorous bugs in Heteroptera. There were 27, 216, and 45 gene families identified as Sternorrhyncha-specific, Auchenorrhyncha-specific, and Heteroptera-specific, respectively. Gene-family expansion/contraction analysis demonstrated a higher frequency of gene contraction in Heteroptera. Compared with *O. fasciatus*, 817 gene families in *R. pedestris* underwent expansion, while 1130 gene families in *R. pedestris* underwent contraction (Figure 5).

#### Chromosome synteny

*T. rubrofasciata* (680 Mb in genome size), which is the only chromosomal-level bug genome available in the Heteroptera suborder, was reported to have 13 chromosomes, while *R. pedestris* has only 6 chromosomes. Synteny analysis demonstrated some fusion or fission events between the two bug species (Figure 6). For example, *R. pedestris* Chr1 contains segments from *T. rubrofasciata* Chr3, Chr5, Chr6, Chr7, and Chr11, and *R. pedestris* Chr2 contains segments from *T. rubrofasciata* Chr1 and Chr4. The sex chromosome of *R. pedestris* showed high synteny with *T. rubrofasciata* Chr12 and Chr13. It is likely that Chr12 or Chr13 might be the sex chromosome of *T. rubrofasciata*, which deserves further investigation.

## 4 Conclusion

In conclusion, we present the first chromosomal-level genome for an herbivorous species from the Heteroptera suborder. *R. pedestris* XO-determinant, with the shortest chromosome being the sex chromosome. The assembled genome sequence and the gene expression patterns in different tissues or different development stages provides a valuable opportunity to understand the molecular mechanisms underlying insect-microbe interactions, plant-insect interactions, and pesticide resistance.

## DATA AVAILABILITY STATEMENT

Raw reads used for genome assembly have been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) with the accession numbers: SRR12977074. The assembled *Riptortus pedestris* genome has been deposited in the NCBI genome (accession numbers: JADPXZ000000000) under the BioProject PRJNA671796. The assembled *Serratia marcescens* Rip1 genome has been deposited in the NCBI genome (accession numbers: SUB8465613) under the BioProject PRJNA674496. Genome resequencing data have been deposited in SRA with the accession numbers of SRR12897696 (female individual) and SRR12977074 (male individual). Transcriptomic data have been deposited in SRA with the accession numbers of SRR12903501 (egg), SRR12903500 (nymph), SRR12903499 (male), and SRR12903498 (female).

## ACKNOWLEDGEMENTS

This work was financially supported by the China National Novel Transgenic Organisms Breeding Project (2019ZX08004-004), Ningbo Science and Technology Innovation 2025 Major Project (2019B10004), and K.C. Wong Magna Fund in Ningbo University.

## CONFLICT OF INTEREST

The authors have no competing interests to declare.

## AUTHOR CONTRIBUTIONS

C.X.Z., J.M.L., Z.T.S. and J.P.C. conceived and designed the whole project. Y.X.Y., X.T.Y., X.W., and H.J.H analyzed the assembled and analyzed the genomic data. H.J.H. wrote the manuscript. Z.Y.W. help collect the sample and maintained the insect strain. Z.T.S. provide the valuable suggestion for this study.

**Table 1 Assembly features for genomes of *Riptortus pedestris* and other Heteroptera species**

Feature	Riptortus pedestris	Oncopeltus fasciatus	Halyomorpha halys	Rhodnius prolixus	Gerris buenoi	Cimex lectularius	Triatoma rubrofasciata
Total size (Mb)	1.193	1,099	1,150	702	1,150	713	680
Level	Chr.	Scaf.	Scaf.	Scaf.	Scaf.	Scaf.	Chr.
Scaffold	181,335	340	802	1,088	344	947	97,329
N50 (Kb)							
No.	6	15,991	16,639	16,537	20,268	13,151	13
Scaf/Chr.							
Contig	13,968	4	18	36	4	13	2,710
N50 (Kb)							
BUSCO	96.5%	98.9%	96.7%	96.6%	97.100%	99.1%	98.2%
Genes	21,562	19,616	24,450	15,076	20,949	24,255	12,691
GC content	33.96%	32.7%	30.90%	27.1%	32.46%	29.94%	—

## References

Alim, MA and Lim, UT (2010). Ecotoxicological effect of insecticides on *Ooencyrtus nezarae* (Hymenoptera: Encyrtidae) an encyrtid parasitoid of *Riptortus pedestris* (Hemiptera: Alydidae). In: *Book Ecotoxicological effect of insecticides on Ooencyrtus nezarae (Hymenoptera: Encyrtidae) an encyrtid parasitoid of Riptortus pedestris (Hemiptera: Alydidae)*. doi: 10.1080/09583157.2013.851170

Bae, SD, Kim, HJ and Mainali, BP (2014). Infestation of *Riptortus pedestris* (Fabricius) decreases the nutritional quality and germination potential of soybean seeds. *J Asia-Pac Entomol* , **17** : 477-481. doi: 10.1016/j.aspen.2014.04.006

- Bai, L, Wang, L, Vega-Rodríguez, J, Wang, G and Wang, S (2019). A gut symbiotic bacterium *Serratia marcescens* renders mosquito resistance to plasmodium infection through activation of mosquito immune responses. *Front Microbiol*, **10** . doi: 10.3389/fmicb.2019.01580
- Bharathimeena, T and Sudharma, K (2009). Alternate host range and management of *Riptortus pedestris* F. infesting cowpea. *Annal. Plant Protect Sci* ,**17** : 26-31.
- Birney, E, Clamp, M and Durbin, R (2004). GeneWise and genomewise. *Genome Res* , **14** : 988-995. doi: 10.1101/gr.1865504
- Bolognini, D, Bartalucci, N, Mingrino, A, Vannucchi, AM and Magi, A (2019). NanoR: A user-friendly R package to analyze and compare nanopore sequencing data. *PLoS One*,**14** : e0216471. doi: 10.1371/journal.pone.0216471
- Capella-Gutiérrez, S, Silla-Martínez, JM and Gabaldón, T (2009). TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*,**25** : 1972-1973. doi: 10.1093/bioinformatics/btp348
- Chen, S, Blom, J and Walker, ED (2017). Genomic, physiologic, and symbiotic characterization of *Serratia marcescens* strains isolated from the mosquito *Anopheles stephensi* . *Front Microbiol* , **8** : 1483. doi: 10.3389/fmicb.2017.01483
- De Coster, W, D’Hert, S, Schultz, DT, Cruts, M and Van Broeckhoven, C (2018). NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics*,**34** : 2666-2669. doi: 10.1093/bioinformatics/bty149
- Dudchenko, O, Batra, SS, Omer, AD, Nyquist, SK, Hoeger, M, Durand, NC, Shamim, MS, Machol, I, Lander, ES and Aiden, AP (2017). *De novo* assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* , **356** : 92-95. doi: 10.1126/science.aal3327
- Grimont, PA and Grimont, F (1978). The genus *Serratia*. *Annu. Rev. Microbiol* , **32** : 221-248. doi: 10.1007/0-387-30746-X\_11
- Haas, BJ, Salzberg, SL, Zhu, W, Pertea, M, Allen, JE, Orvis, J, White, O, Buell, CR and Wortman, JR (2008). Automated eukaryotic gene structure annotation using EVIDENCEModeler and the program to assemble spliced alignments. *Genome Biol* , **9** : R7. doi: 10.1186/gb-2008-9-1-r7
- Hoff, KJ, Lange, S, Lomsadze, A, Borodovsky, M and Stanke, M (2016). BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics*,**32** : 767-769. doi: 10.1093/bioinformatics/btv661
- Holt, C and Yandell, M (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC bioinformatics*,**12** : 491. doi: 10.1186/1471-2105-12-491
- Hu, J, Fan, J, Sun, Z and Liu, S (2020). NextPolish: a fast and efficient genome polishing tool for long read assembly. *Bioinformatics* . doi: 10.1101/2020.04.17.035287
- Huang, HJ, Ye, ZX, Lu, G, Zhang, CX, Chen, JP and Li, JM (2020). Identification of salivary proteins in the whitefly *Bemisia tabaci* by transcriptomic and LC-MS/MS analyses. *Insect Sci* . doi: 10.1111/1744-7917.12856
- Itoh, H, Hori, T, Sato, Y, Nagayama, A, Tago, K, Hayatsu, M and Kikuchi, Y (2018). Infection dynamics of insecticide-degrading symbionts from soil to insects in response to insecticide spraying. *ISME J* , **12** : 909-920. doi: 10.1038/s41396-017-0021-9
- Ju, JF, Bing, XL, Zhao, DS, Guo, Y, Xi, Z, Hoffmann, AA, Zhang, KJ, Huang, HJ, Gong, JT, Zhang, X and Hong, XY (2020). *Wolbachia* supplement biotin and riboflavin to enhance reproduction in planthoppers. *ISME J* , **14** : 676-687. doi: 10.1038/s41396-019-0559-9
- Jung, M, Kim, S, Kim, HG and Lee, DH (2018). Lethal and sublethal effects of synthetic insecticides on the locomotory and feeding behavior of *Riptortus pedestris* (Hemiptera: Alydidae) under laboratory conditions.

*J Asia-Pac Entomol*, **21** : 179-185. doi: 10.1016/j.aspen.2017.11.019

Jung, M and Lee, D-H (2019). Abundance and diversity of gut-symbiotic bacteria, the genus *Burkholderia* in overwintering *Riptortus pedestris* (Hemiptera: Alydidae) populations and soil in South Korea. *PLoS One*, **14** : e0218240. doi: 10.1371/journal.pone.0218240

Kang, C (2003). Review on true bugs infesting tree fruits, upland crops, and weeds in Korea. *J. Appl. Entomol*, **4** : 269-277.

Katoh, K and Standley, DM (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*, **30** : 772-780. doi: 10.1093/molbev/mst010

Kikuchi, Y and Fukatsu, T (2014). Live imaging of symbiosis: spatiotemporal infection dynamics of a GFP-labelled *Burkholderia* symbiont in. *Mol Ecol*, **23** : 1445-1456. doi: 10.1111/mec.12479

Kikuchi, Y, Hayatsu, M, Hosokawa, T, Nagayama, A, Tago, K and Fukatsu, T (2012). Symbiont-mediated insecticide resistance. *PNAS*, **109** : 8618-8622. doi: 10.1073/pnas.1200231109

Kikuchi, Y, Hosokawa, T and Fukatsu, T (2007). Insect-microbe mutualism without vertical transmission: a stinkbug acquires a beneficial gut symbiont from the environment every generation. *Appl Environ Microb*, **73** : 4308-4316. doi: 10.1128/AEM.00067-07

Kikuchi, Y, Hosokawa, T and Fukatsu, T (2011). Specific developmental window for establishment of an insect-microbe gut symbiosis. *Appl Environ Microb*, **77** : 4075-4081. doi: 10.1128/AEM.00358-11

Koren, S, Walenz, BP, Berlin, K, Miller, JR, Bergman, NH and Phillippy, AM (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res*, **27** : 722-736. doi: 10.1101/gr.215087.116

Korf, I (2004). Gene finding in novel genomes. *BMC bioinformatics*, **5** : 59. doi: 10.1186/1471-2105-5-59

Langfelder, P and Horvath, S (2008). WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics*, **9** : 559. doi: 10.1186/1471-2105-9-559

Lee, DJ, Lee, JB, Jang, HA, Ferrandon, D and Lee, BL (2017). An antimicrobial protein of the *Riptortus pedestris* salivary gland was cleaved by a virulence factor of *Serratia marcescens*. *Dev Comp Immunol*, **67** : 427-433. doi: 10.1016/j.dci.2016.08.009

Li, H, Handsaker, B, Wysoker, A, Fennell, T, Ruan, J, Homer, N, Marth, G, Abecasis, G, Durbin, R and Subgroup, G (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, **25** : 2078-2079. doi: 10.1093/bioinformatics/btp352

Li, K, Zhang, X, Guo, J, Penn, H, Wu, T, Li, L, Jiang, H, Chang, L, Wu, C and Han, T (2019). Feeding of *Riptortus pedestris* on soybean plants, the primary cause of soybean staygreen syndrome in the Huang-Huai-Hai river basin. *The Crop Journal*, **7** : 360-367. doi: CNKI:SUN:CROP.0.2019-03-008

Li, W, Gao, Y, Hu, Y, Li, S, Bi, R, Zhang, J, Shi, S (2020). Effects of the *Riptortus pedestris* on staygreen syndrome and yield of soybean. *Soybean Science*, **39** : 116-122. doi: 10.11861/j. issn.1000-9841.2020.01.0116

Lieberman-Aiden, E, Van Berkum, NL, Williams, L, Imakaev, M, Ragoczy, T, Telling, A, Amit, I, Lajoie, BR, Sabo, PJ and Dorschner, MO (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, **326** : 289-293. doi: 10.1126/science.1181369

Lomsadze, A, Burns, PD and Borodovsky, M (2014). Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Res*, **42** : e119-e119. doi: 10.1093/nar/gku557

Longmead, B and Salzberg, S (2012). Fast gapped-read alignment with Bowtie2. *Nat Methods*, **9**:357-359. doi: 10.1038/nmeth.1923

- Ma, W, Le, X, Hua, H, Chen, M, Guo, M, He, K, Zhao, J and Li, F (2020). Chromosomal-level genomes of three rice planthoppers provide new insights into sex chromosome evolution. *Mol Ecol Resour* , doi: 10.1111/1755-0998.13242
- Muramoto, N (1978). A chromosome study of thirty Japanese heteropterans (Heteroptera). *Genetica*, **49** : 37-44. doi: 10.1007/BF00187812
- Nguyen, L-T, Schmidt, HA, Von Haeseler, A and Minh, BQ (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*, **32** : 268-274. doi: 10.1093/molbev/msu300
- Ohbayashi, T, Futahashi, R, Terashima, M, Barriere, Q, Lamouche, F, Takeshita, K, Meng, XY, Mitani, Y, Sone, T, Shigenobu, S, Fukatsu, T, Mergaert, P and Kikuchi, Y (2019). Comparative cytology, physiology and transcriptomics of *Burkholderia insecticola* in symbiosis with the bean bug *Riptortus pedestris* and in culture. *ISME J*, **13** : 1469-1483. doi: 10.1038/s41396-019-0361-8
- Rahman, MM and Lim, UT (2017). Evaluation of mature soybean pods as a food source for two pod-sucking bugs, *Riptortus pedestris* (Hemiptera: Alydidae) and *Halyomorpha halys* (Hemiptera: Pentatomidae). *PLoS One*, **12** : e0176187. doi: 10.1371/journal.pone.0176187
- Sakuraoka, R, Suzuki, T and Morohoshi, T (2019). Distribution and genetic diversity of genes involved in quorum sensing and prodigiosin biosynthesis in the complete genome sequences of *Serratia marcescens* . *Genome Biol Evol*, **11** : 931-936. doi: 10.1093/gbe/evz046
- Schmidt, MH-W, Vogel, A, Denton, AK, Istace, B, Wormit, A, van de Geest, H, Bolger, ME, Alseekh, S, Mass, J and Pfaff, C (2017). *De novo* assembly of a new *Solanum pennellii* accession using Nanopore sequencing. *Plant Cell*, **29** : 2336-2348. doi: 10.1105/tpc.17.00521
- Stanke, M, Keller, O, Gunduz, I, Hayes, A, Waack, S and Morgenstern, B (2006). AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res*, **34** : W435-W439. doi: 10.1093/nar/gkl200
- Takeshita, K and Kikuchi, Y (2017). *Riptortus pedestris* and *Burkholderia* symbiont: an ideal model system for insect-microbe symbiotic associations. *Res Microbiol*, **168** : 175-187. doi: 10.1016/j.resmic.2016.11.005
- Takeshita, K, Shibata, TF, Nikoh, N, Nishiyama, T, Hasebe, M, Fukatsu, T, Shigenobu, S and Kikuchi, Y (2014). Whole-genome sequence of *Burkholderia* sp. strain RPE67, a bacterial gut symbiont of the bean bug *Riptortus pedestris* . *Genome A*, **2** . doi: 10.1128/genomeA.00556-14
- Takeshita, K, Tamaki, H, Ohbayashi, T, Meng, X-Y, Sone, T, Mitani, Y, Peeters, C, Kikuchi, Y and Vandamme, P (2018). *Burkholderia insecticola* sp. nov., a gut symbiotic bacterium of the bean bug *Riptortus pedestris* . *Int J Syst Evol Micr*, **68** : 2370-2374. doi: 10.1099/ijsem.0.002848
- Wu, TD and Watanabe, CK (2005). GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics*, **21** : 1859-1875. doi: 10.1093/bioinformatics/bti310
- Xue, J, Zhou, X, Zhang, C-X, Yu, L-L, Fan, H-W, Wang, Z, Xu, H-J, Xi, Y, Zhu, Z-R, Zhou, W-W, Pan, P-L, Li, B-L, Colbourne, JK, Noda, H, Suetsugu, Y, Kobayashi, T, Zheng, Y, Liu, S, Zhang, R, Liu, Y, Luo, Y-D, Fang, D-M, Chen, Y, Zhan, D-L, Lv, X-D, Cai, Y, Wang, Z-B, Huang, H-J, Cheng, R-L, Zhang, X-C, Lou, Y-H, Yu, B, Zhuo, J-C, Ye, Y-X, Zhang, W-Q, Shen, Z-C, Yang, H-M, Wang, J, Wang, J, Bao, Y-Y and Cheng, J-A (2015). Genomes of the rice pest brown planthopper and its endosymbionts reveal complex complementary. *Genome Biol*, **15** : 521. doi: 10.1186/s13059-014-0521-0
- Yang, Z (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci*, **13** : 555-556. doi: 10.1093/bioinformatics/13.5.555

## Figure legends

**Figure 1.** The bean bug *Riptortus pedestris* poses a major threat to soybean production. (A) *R. pedestris* life cycle. (B) Staygreen syndromes, including increased numbers of abnormal seeds, more empty pods, a

lack of leaf senescence, and decreased yields, caused by

*R. pedestris*.

**Figure 2. Identification of chromosomes in *Riptortus pedestris*.** (A) DNA interaction heatmap generated in Hi-C interaction analysis. The shading gradient represents the chromosome interactions. (B) Chromosomic staining of female and male salivary gland cells in *R. pedestris*. The blue signal indicates the condensed chromosomes.

**Figure 3. Landscape of the assembled *Riptortus pedestris* genome.** From the outside in, the first circle shows the 6 chromosomes. The second circle showed the GC contents. The third and fourth circle showed the female and male MRPM value, respectively. The fifth circle showed the ratio of MRPM between female and male. The sixth circle showed the transcriptional reads from 37 different development stages. All data was calculated on a 1-Mb scale.

**Figure 4. Analysis of *Serratia marcescens* Rip1.** (A) Landscape of the assembled *Serratia marcescens* Rip1 genome. From the outside in, the first circle showed the nucleotide sequence positions (in Mb). The second circle showed the virulence factors and antibiotic-resistance genes in the genome. The third and fourth circle showed the predicted genes that transcribed in clockwise and anticlockwise, respectively. The fifth circle showed the GC contents on a 10-Kb scale. rfaF, ADP-heptose-LPS heptosyltransferase II; gpt, glycerol-3-phosphate transporter; PmrB, two-component system sensor histidine kinase BasS; alaS, aminocoumarin resistant gene; kdpE, two-component system response regulator KdpE; fbpC, Iron (III) ABC transporter; galE, UDP-glucose 4-epimerase; msbA, Lipid transporter ATP-binding/permease; mfd, transcription-repair coupling factor; shuU, ABC transport system; SRT-2, class C beta-lactamase; cysB, aminocoumarin resistant transcriptional regulator CysB; flg, flagella synthesis genes; katG, catalase/peroxidase HPI; omp36, porin OmpC; fepG, Iron-enterobactin ABC transporter permease; lpxD, UDP-3-O-(3-hydroxymyristoyl) glucosamine N-acyltransferase; AAC(6′)-Ic, aminoglycoside N-acetyltransferase AAC(6′)-Ic. (B) Phylogenetic relationship between *Serratias* species.

**Figure 5. Phylogenetic analysis of *Riptortus pedestris* with other insect species.** The phylogenetic position of *R. pedestris* was determined based on 963 single copy genes with 356,827 reliable sites. *D. melanogaster* was used to root the tree. 1:1:1 indicates single copy genes, and N:N:N indicates multicopy genes across 14 insect species. The Sternorrhyncha, Auchenorrhyncha, and Heteroptera indicate suborder-specific genes, respectively. The estimated species divergence time is illustrated at the bottom of the phylogenetic tree. Gene family expansion (red) and contraction (green) are illustrated at the branches and nodes of the tree.

**Figure 6. Genome synteny between *Triatoma rubrofasciata* and *Riptortus pedestris*.** Synteny analysis demonstrated fusion or fission events between the two bug species.

### Supplemental material

**File1. Gene expression pattern of *Riptortus pedestris* at different development stages and in different tissues**

**File 2. Identified proteins in different tissues by LC-MS/MS**

**Figure S1. The 19-mer depth distribution for genome size estimation analysis of *Riptortus pedestris*.**

**Figure S2. BUSCO analyses of assembled genome.** The assembled genome was subjected to BUSCO analyses against insect, arthropod, metazoan, and eukaryote datasets. C, complete BUSCOs; S, complete and single-copy BUSCOs; D, complete and duplicated BUSCOs, F: fragmented BUSCOs, M: missing BUSCOs.

**Figure S3. The chromosomes synteny of *S. marcescens* Db11 and *S. marcescens* Rip1.** The location of a gene in two genomes were determined. Each dot represents a gene.

**Figure S4. BUSCO analyses of predicted genes.** The predicted genes were subjected to BUSCO analyses against insect, arthropod, metazoan, and eukaryote datasets. C, complete BUSCOs; S, complete and single-copy BUSCOs; D, complete and duplicated BUSCOs, F: fragmented BUSCOs, M: missing BUSCOs.

**Figure S5. Venn diagram of annotated genes in *Riptortus pedestris* genome.** Genes annotated by NCBI Nr, Swiss-Prot, KOG, KEGG, and GO databases were shown.

**Figure S6. KOG function classification of predicted genes in *Riptortus pedestris* genome.**

**Figure S7. GO analysis of predicted genes in *Riptortus pedestris* genome.**

**Figure S8. KEGG analysis of predicted genes in *Riptortus pedestris* genome.**

**Figure S9. Six tissues dissected from adult *Riptortus pedestris*.** (A) Observation of carcass, testis, gut, muscle, ovary, and fat body under the stereoscope. Bar: 0.5 mm. (B) SDS-PAGE analysis of dissected tissues.

**Figure S10 Enrichment analysis of tissue-specific genes.** Enriched KEGG terms (left column) and GO terms from gut-specific (A), carcass-specific (B), muscle-specific (C), testis-specific (D), and ovary-specific (E) genes were illustrated.

**Figure S11 WGCNA analysis of *Riptortus pedestris* genes from 37 different development stages.**

**Table S1. The protein datasets retrieved from public database.**

**Table S2. The statistics of PacBio reads after quality control.**

**Table S3. The statistics of contig-level assemblies.**

**Table S4. The statistics of Hi-C data.**

**Table S5. The statistics of repetitive sequences in *Riptortus pedestris***

**Table S6. The statistics of SSR in *Riptortus pedestris***

**Table S7. Bacteria-origin scaffolds in genome-assembly**

**Table S8. Comparative analysis of antibiotic resistance genes in *Serratia* sp.**

**Table S9. Comparative analysis of predicted virulence factors in *Serratia* sp.**

**Table S10. The statistics of ONT reads after quality control.**

**Table S11. Gene functional annotation of *Riptortus pedestris***

**Table S12. Summary of ncRNA in *Riptortus pedestris***

**Table S13. RNA-seq data from 6 different tissues and 37 different development stages**

**Table S14. Mapped reads per million (MRPM) analysis of each chromosome in male and female *Riptortus pedestris***









