Chromosome length genome assembly of the redbanded stink bug, *Piezodorus guildinii* (Westwood).

Omaththage P. Perera (op.perera@usda.gov)

Research note

Keywords: Redbanded stink bug, Piezodorus guildinii, genome assembly, Pentatomidae, soybean pest

Posted Date: September 20th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-892524/v1

Abstract

Objective: The redbanded stink bug, *Piezodorus guildinii* (Hemiptera: Pentatomidae), is native to Caribbean Basin and is currently considered an invasive pest in Florida, Louisiana, Mississippi, and Texas in the southern United States. Although *P. guildinii* is an economically important invasive pest in the USA, a relatively few studies have been conducted to understand genetic, population genetic structure, and genetic basis of resistance to insecticides. The objective of this work was to obtain a high-quality genome assembly to develop genetic resources to conduct genetic, population genetic, and physiological studies of the RBSB.

Results: The genome of RBSB was sequenced with Pacific Biosciences technology followed by two rounds of scaffolding using Chicago libraries and HiC proximity ligation to obtain a high-quality assembly. The genome assembly contained 800 scaffolds larger than 1 kbp and the N50 was 170.84 Mbp. The largest scaffold was 222.22 Mbp and 90% of the genome was included in the 7 scaffolds larger than 1 Mbp. The number of megabase scaffolds also matched the number of chromosomes in this insect. The genome sequence will facilitate the development of resources to conduct studies on genetics, transcriptomics, and physiology of RBSB.

Introduction

The redbanded stink bug (RBSB), *Piezodorus guildinii* (Westwood) (Hemiptera: Pentatomidae), is native to Caribbean Basin and is currently considered an invasive pest in Florida, Louisiana, Mississippi, and Texas in the southern United States [1-3]. This insect is an important pest of soybeans and several other commercially grown crops and show varying levels of insecticide resistance [4]. Uncontrolled outbreaks of RBSB can cause significant economic damage to soybeans from early seed development stages to mature seeds [5]. Although RBSB is an economically important invasive pest in the USA, a relatively few studies have been conducted to understand genetic, population genetic structure, and genetic basis of resistance to insecticides. Biology, ecology, host plants, and pest status of this insect has been previously studied [2, 6-13]. Resistance to insecticides in RBSB has been documented but the genetic basis of the insecticide resistance in this insect is not well understood. So far, only one population genetic study has been carried out on this species using 1,337 SNP markers [14] to estimate genetic structure and to identify loci under natural selection. This study identified the presence of genetic structure separating populations in USA and Brazil. In order to develop genetic resources to conduct genetic, population genetic, and physiological studies, the genome of RBSB (Fig. S1) was sequenced with Pacific Biosciences long read technology using Chicago libraries and assembled to obtain a draft input assembly that was used for scaffolding Illumina short reads obtained from HiC proximity ligation libraries to obtain a high-quality assembly (Fig. 1, Figs. S2 and S3).

Main Text

Methods: Genomic DNA libraries for PacBio sequencing and initial assemblies were prepared by Dovetail Genomics (Scottes Valley, CA, USA) with high molecular weight DNA isolated from a field collected female RBSB. DNA was quantified using Qubit 2.0 fluorometer (Life Technologies, Carlsbad, CA) and a PacBio SMRTbell library with approximately 20 kbp was constructed using SMRTbell Express Template Prep Kit 2.0 (PacBio, Menlo Park, CA) following the manufacturer's protocol. DNA sequencing was performed on PacBio Sequel II sequencer using Sequel II 8M SMRT cells generating 108 Gb of data. Initial assembly was performed using Wtdbg2 assembler [15] and Blob Tools v1.1.1 was used to identify potential contamination based on by BLAST (v2.9) results of the assembly against the NT database. The assembly was filtered to remove potential haplotypic duplications using Purge Dups [16] to obtain purged draft assembly used for scaffolding. Scaffolding was performed using HiRise pipeline (Dovetail Genomics). Proximity ligation libraries were prepared using Dovetail Omni-C library protocol by digesting formaldehyde fixed chromatin with a DNAse I, repairing chromatin ends, and biotinylated bridge adapter ligation followed by proximity ligation of adapter containing ends. Then crosslinks were reversed, DNA was purified and was treated to remove biotin that was not internal to ligated fragments. Biotin-labelled DNA fragments were isolated using streptavidin beads and enriched by PCR. The library was sequenced on an Illumina HiSeqX platform to produce approximately 30x sequence coverage. The draft de novo assembly from PacBio reads and Dovetail Omni-C proximity ligated library reads were used as input data for for HiRise pipeline (Putnam et al., 2016). Dovetail Omni-C library sequences were aligned to draft input assembly using BWA (Li & Durbin, 2009, Lieberman-Aiden et al., 2009). The separations of Dovetail OmniC read pairs mapped within draft scaffolds were analyzed by HiRise to produce a likelihood model for genomic distance between read pairs, and the model was used to identify and break putative misjoins, to score prospective joins, and make joins above a threshold.

The genome completeness evaluation was performed with BUSCO version 5.2.2 [17]. The repeat annotation was done with RepeatModeler using the Dfam TE tools docker container version 1.4 (https://github.com/Dfam-consortium/TETools). The repeat classification was done with RepeatClassifier Version 2.0.2 and RepeatMasker [18] to identify the types of repeats in the RBSB genome. The database of repeats used for classification was Dfam 3.4 [19].

PacBio sequence reads from the shotgun library were mapped to *Nezara Viridula* (Linnaeus) mtDNA genome (Accession: EF208087.1) with relaxed parameters (length fraction 0.3 and similarity fraction 0.75) to extract reads matching mtDNA sequences that were de novo assembled to obtain contigs. A 20,764 nt contig with high similarity to the stink bug *Eurydema ventralis* (Kolenati) with 84% identity (Accession:MG584837.1) was selected to use as the reference to map the PacBio reads with a higher stringency at 85% similarity fraction and 75% length fraction. A total of 2,099 out of 9,113,332 reads were mapped to the reference with >150-fold sequence coverage across the coding regions and >75-fold sequence coverage in the control region of the 18,889 bp mitochondrial genome of RBSB.

Results and discussion: Assembled size of the genome was 1.205 Gbp with an N50 of 170.835 Mbp, N90 of 118.462 Mbp, L50 of 4, and L90 of 7. The final genome assembly contained 800 scaffolds larger than 1 Kbp and the largest scaffold was 222.218 Mbp (Table 1).

The genome assembly of RBSB was found to be highly complete for single-copy markers conserved within the Arthropoda and Hemiptera clades with 96.5% and 96.2% completeness values (Table 2). The BUSCO marker set will change as new genome assemblies such as the one reported in this work are added to the OrthoDB [20] clustering but the low duplication and fragmentation coupled with the completeness underlines the integrity of this genome assembly.

Table 1. Assembly summary and contiguity metrics for the shotgun input assembly of PacBio reads and the final scaffolding with proximity ligation (HiC) of the RBSB genome.

Input Assembly		Dovetail HiRise Assembly
Total Length (bp)	1,205,371,278	1,205,416,778
N50	7,401,307	170,835,737
L50	46	4
Largest scaffold	29,620,976	222,218,125
Number of scaffolds	1,262	807
Number of scaffolds > 1kbp	1,255	800
Number of gaps	5	460
Number of N's per 100 kbp	0.01	3.78

Table 2. BUSCO completeness statistics on the Arthropoda and Hemiptera marker set for the RBSB genome assembly. These numbers were generated with BUSCO version 5.2.2

BUSCO Database	Complete BUSCOs	Complete and single-copy BUSCOs	Complete and duplicated BUSCOs	Fragmented BUSCOs	Missing BUSCOs	Total BUSCO groups searched
Arthropoda	977 (96.5%)	962 (95%)	15 (1.5%)	10 (1%)	26 (2.5%)	1013
Hemiptera	2414 (96.2%)	2375 (94.6%)	39 (1.6%)	21 (0.8%)	75 (3%)	2510

Karyotyping of RBSB identified six pairs of autosomes and a pair of sex chromosomes designated X and Y (2n=14) [21]. Sex chromosomes were the smallest of the chromosome complement and appear to be different in size. We performed two rounds of repeat annotations. The RBSB genome was analyzed for known repeat families in *insecta* present in the DFAM 2.4 [22] (Supplementary data Table S1). LINE (11.64%) were the predominant retroelements found followed by SINEs (1.21%). The Tc1-IS630-Pogo transposon was the predominant DNA transposon repeat family (4.51%). A *de novo* repeat annotation using RepeatModeler [23] identified 2338 RepeatScout/RECON families and 181 LTR repeat families. All annotations are available to the research community at the AgriVectors portal [24].

Molecular resources currently available for RBSB in public databases are limited to 107 microsatellite sequences, six partial cytochrome oxidase I (COI) subunit sequences [25] and 17 Gbp of genotype by sequencing (GBS) data in the NCBI sequence read archive [14]. Transcriptome, genome or proteome data are not available for this species. We currently have a limited number of RNASeq data from the salivary glands of RBSB. Having a well characterized high quality genome assembly will be of great value in developing genetic markers for population genetic studies, linkage mapping, and identifying genomic regions associated with host selection and insecticide resistance. A fully annotated genome will also help in identifying regulatory elements and non-transcribed genomic regions that may influence gene expression. The complete sequence of the mitochondrial genome, which is often used in population genetics and molecular identification of insects is not currently available for this species. Sequence of the complete mitochondrial genome will provide additional DNA markers to conduct population studies that require relatively highly variable mitochondrial genes such as NADH dehydrogenases (ND) and cytochrome B.

Having an official gene set derived from a high-quality genome will also aid in conducting expression profiling experiments intended to elucidate physiological response to various food sources, insecticides, and studies directed to identify gene coregulatory networks. Large scale insect genomics projects like the i5k [26] and more recently Ag100Pest [27] have also highlighted the long-term benefits of building open access and genomics resources for the community.

Often metabolic resistance to insecticides involves enzymes such as cytochrome p450s, esterases, and glutathione transferases that degrade xenobiotics [28-31]. Heritable insecticide resistance is achieved through mutations that alter ligand binding capacity of receptors [32-34] [35, 36]. The genetic basis of host plant selection has been studied in insects using comparative genomics and transcriptomics [37-39]. With a fully annotated genome from NCBI, conducting experiments to understand regulation of gene expression, metabolic and genetic basis of insecticide resistance and host plant selection, and networks of genes co-regulated in physiological response to external stimuli could be studied with confidence.

To annotate the genome and understand gene regulation and profile expression patterns of genes during each life stage in the RBSB using this chromosomal length genome assembly, we have generated transcriptomics data for all life stages from eggs to adults. We plan to create Illumina sequencing libraries from each life stage with a total of 27 libraries to obtain a minimum coverage of 25 million (2x150 paired end) reads for each library. Oxford Nanopore long reads will also be obtained by pooling mRNA from all life stages of RBSB. Long Oxford Nanopore reads will facilitate annotation of full-length gene models in the genome as well as qualitative identification of isoforms from each life stage. The NCBI structural annotation will be followed by functional annotation to identify gene ontology terms and pathways[40]. which will be made available on the AgriVectors portal [24]. In addition, we plan to identify epigenetic modifications in the genome by sequencing the RBSB genomic DNA with Nanopore reads.

Limitations: Genomic DNA library preparation and assembly were performed with proprietary methods developed by a service provider with PacBio and Illumina sequencing versions currently available. Library construction and sequencing methods, assembly software, and other data used for comparative analysis may be updated in the future.

Declarations

Ethics approval and consent to participate: Not Applicable

Availability of data and material: All raw sequencing data and assemblies have been submitted to NCBI BioProject PRJNA686660.

Funding: This research was funded by USDA Agricultural Research Service in house research project 6066-22000-091-00D of Southern Insect Management Research Unit. Partial funding was provided by the ARS/State Potato Partnership Program.

Acknowledgements: We thank Calvin Pierce (USDA ARS SIMRU) for assistance with RBSB colony maintenance and sample preparation and Dr. Nathan Little for critically reading an earlier version of this

manuscript. The use or mention of a trademark or proprietary product does not constitute an endorsement, guarantee, or warranty of the product and does not imply its approval to the exclusion of other suitable products by the U.S. Department of Agriculture, an equal opportunity employer.

Consent for publication: Not applicable.

Competing interests: Authors declare that they have no competing interests.

Author contributions: Conceptualization, OPP; methodology, KCA, OPP, SS; formal analysis, KCA, OPP, SS; investigation, KCA, OPP, SS, resources, OPP, GVPR, LAM; data curation, OPP, SS; writing-original draft preparation, OPP, SS; writing-review and editing, OPP, SS, KCA, LAM and GVPR; project administration, OPP. All authors have read and agreed to the published version of this manuscript.

References

1. Bundy, C.S., et al., *Piezodorus guildinii (Westwood)*, in *Invasive stink bugs and related species (Pentatomoidea)*, J.E. McPherson, Editor. 2018, CRC Press: Boca Raton, FL. p. 425-452.

2. Panizzi, A. and F. Slansky Jr, *Legume host impact on performance of adult Piezodorus guildinii (Westwood)(Hemiptera: Pentatomidae).* Environ. Entomol., 1985. **14**(3): p. 237-242.

3. Panizzi, A. and F. Slansky Jr, *New host plant records for the stink bug Piezodorus guildinii in Florida (Hemiptera: Pentatomidae).* Fla. Entomol., 1985. **68**(1): p. 215-216.

4. Baur, M., et al., *Susceptibility to insecticides used for control of Piezodorus guildinii (Heteroptera: Pentatomidae) in the United States and Brazil.* J. Econ. Entomol., 2010. **103**(3): p. 869-876.

5. Correaferreira, B. and F. Moscardi, *Seasonal occurrence and host spectrum of egg parasitoids associated with soybean stink bugs.* Biol. Control, 1995. **5**(2): p. 196-202.

6. Panizzi, A.R. and J.G. Smith, *Biology of Piezodorus guildinii: oviposition, development time, adult sex ratio, and longevity.* Ann. Entomol. Soc. Am., 1977. **70**(1): p. 35-39.

7. Panizzi, A., *Performance of Piezodorus guildinii on four species of Indigofera legumes.* Entomol. Exp. Appl., 1992. **63**(3): p. 221-228.

8. Link, D., J. Panichi, and L. Concatto, *Oviposition by Piezodorus guildinii (Westwood, 1837) on bean plants.* Rev. Cent. Cienc. Rur., 1980. **10**(3): p. 271-276.

9. Cividanes, F.J. and J.R. Parra, *Ecological zoning of Nezara viridula (L.), Piezodorus guildinii (West.) and Euschistus heros (Fabr.)(Heteroptera: Pentatomidae) in four soyabean-producing states of Brazil.* Anais Soc. Entomol. Brasil, 1994. **23**(2): p. 219-226.

10. Oliveira, É.D. and A.R. Panizzi, *Performance of nymphs and adults of Piezodorus guildinii (Westwood)(Hemiptera: Pentatomidae) on soybean pods at different developmental stages.* Braz. Arch. Biol. Tech., 2003. **46**: p. 187-192.

11. Bastola, A., *Ecology and Biology of the Redbanded Stink Bug, Piezodorus guildinii (Westwood) in Louisiana*, in *LSU Doctoral Dissertations. 4147*. 2017, Louisiana State University: Baton Rouge. p. 86.

12. Zerbino, M., et al., *Overwintering of Piezodorus guildinii (Heteroptera, Pentatomidae) populations.* Neotrop. Entomol., 2020. **49**(2): p. 179-190.

13. Gomez, V.A., et al., *Biological aspects of Piezodorus guildinii (Westwood)(Hemiptera: Pentatomidae) reared in laboratory with different insects diets.* Rev. Soc. Entomol. Argentina, 2017. **72**(1-2).

14. Zucchi, M.I., et al., *Patterns of genome-wide variation, population differentiation and SNP discovery of the red banded stink bug (Piezodorus guildinii).* Sci. Rep., 2019. **9**(1): p. 1-11.

15. Ruan, J. and H. Li, *Fast and accurate long-read assembly with wtdbg2.* Nat. Methods, 2020. **17**(2): p. 155-158.

16. Guan, D., et al., *Identifying and removing haplotypic duplication in primary genome assemblies.* Bioinformatics, 2020. **36**(9): p. 2896-2898.

17. Manni, M., et al., *BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes.* arXiv preprint arXiv:2106.11799, 2021.

18. Smit, A.F.A., R. Hubley, and P. Green. *RepeatMasker Open-4.0*. 2015 [cited 2021 Auaust 15]; Available from: http://www.repeatmasker.org.

19. Storer, J., et al., *The Dfam community resource of transposable element families, sequence models, and genome annotations.* Mobile DNA, 2021. **12**(1): p. 2.

20. Kriventseva, E.V., et al., *OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs.* Nucleic Acids Res., 2019. **47**(D1): p. D807-D811.

21. Rebagliati, P.J., L.M. Mola, and A.G. Papeschi, *Karyotype and meiotic behaviour of the holokinetic chromosomes of six Argentine species of Pentatomidae (Heteroptera).* Caryologia, 2001. **54**(4): p. 339-347.

22. Hubley, R., et al., *The Dfam database of repetitive DNA families.* Nucleic Acids Res., 2016. **44**(D1): p. D81-D89.

23. Flynn, J.M., et al., *RepeatModeler2 for automated genomic discovery of transposable element families.* Proc, Nat. Acad. Sci. USA, 2020. **117**(17): p. 9451-9457.

24. Saha, S., et al., *An Open Access Resource Portal for Arthropod Vectors and Agricultural Pathosystems: AgriVectors. org.* The 1st International Electronic Conference on Entomology 2021. **doi: 10.3390/IECE-10576**.

25. Greenstone, M., P. Tillman, and J. Hu, *Predation of the newly invasive pest Megacopta cribraria (Hemiptera: Plataspidae) in soybean habitats adjacent to cotton by a complex of predators.* J. Econ. Entomol., 2014. **107**(3): p. 947-954.

26. i5K-Consortium, *The i5K Initiative: advancing arthropod genomics for knowledge, human health, agriculture, and the environment.* J. Hered., 2013. **104**(5): p. 595-600.

27. Childers, A.K., et al., *The USDA-ARS Ag100Pest Initiative: High-Quality Genome Assemblies for Agricultural Pest Arthropod Research.* Insects, 2021. **12**(7): p. 626.

28. Oakeshott, J.G., et al., *Metabolic enzymes associated with xenobiotic and chemosensory responses in Nasonia vitripennis.* Insect Mol. Biol., 2010. **19**: p. 147-163.

29. Yang, J., et al., *A Drosophila systems approach to xenobiotic metabolism.* Physiol. Genomics, 2007. **30**(3): p. 223-231.

30. Li, X., M.A. Schuler, and M.R. Berenbaum, *Molecular mechanisms of metabolic resistance to synthetic and natural xenobiotics.* Annu. Rev. Entomol., 2007. **52**: p. 231-253.

31. Amezian, D., R. Nauen, and G. Le Goff, *Transcriptional regulation of xenobiotic detoxification genes in insects-An overview.* Pestic. Biochem. Phys., 2021: p. 104822.

32. Bajda, S., et al., *A mutation in the PSST homologue of complex I (NADH:ubiquinone oxidoreductase) from Tetranychus urticae is associated with resistance to METI acaricides.* Insect Biochem Mol Biol, 2017. **80**: p. 79-90.

33. Douris, V., et al., *Investigation of the contribution of RyR target-site mutations in diamide resistance by CRISPR/Cas9 genome modification in Drosophila.* Insect Biochem Mol Biol, 2017. **87**: p. 127-135.

34. Douris, V., et al., *Using CRISPR/Cas9 genome modification to understand the genetic basis of insecticide resistance: Drosophila and beyond.* Pestic. Biochem. Physiol., 2020. **167**: p. 104595.

35. Banazeer, A., et al., *Characterization of genetic basis and realized heritability of bifenthrin-resistance selected in dusky cotton bug, Oxycarenus hyalinipennis (Costa)(Hemiptera: Lygaeidae) in Pakistan.* Crop Prot., 2021. **141**: p. 105441.

36. Hawkins, N.J., et al., *The evolutionary origins of pesticide resistance.* Biol. Rev., 2019. **94**(1): p. 135-155.

37. Wybouw, N., et al., *Long-term population studies uncover the genome structure and genetic basis of xenobiotic and host plant adaptation in the herbivore Tetranychus urticae.* Genetics, 2019. **211**(4): p. 1409-1427.

38. Schirrmann, M.K., et al., *Genomewide signatures of selection in Epichloë reveal candidate genes for host specialization.* Mol. Ecol., 2018. **27**(15): p. 3070-3086.

39. Benevenuto, J., et al., *Comparative genomics of smut pathogens: insights from orphans and positively selected genes into host specialization.* Front. Mcrobiol., 2018. **9**: p. 660.

40. Saha, S., et al., *Workflows for rapid functional annotation of diverse arthropod genomes.* bioRxiv, 2021. **doi: https://doi.org/10.1101/2021.06.12.448177**.

Figures

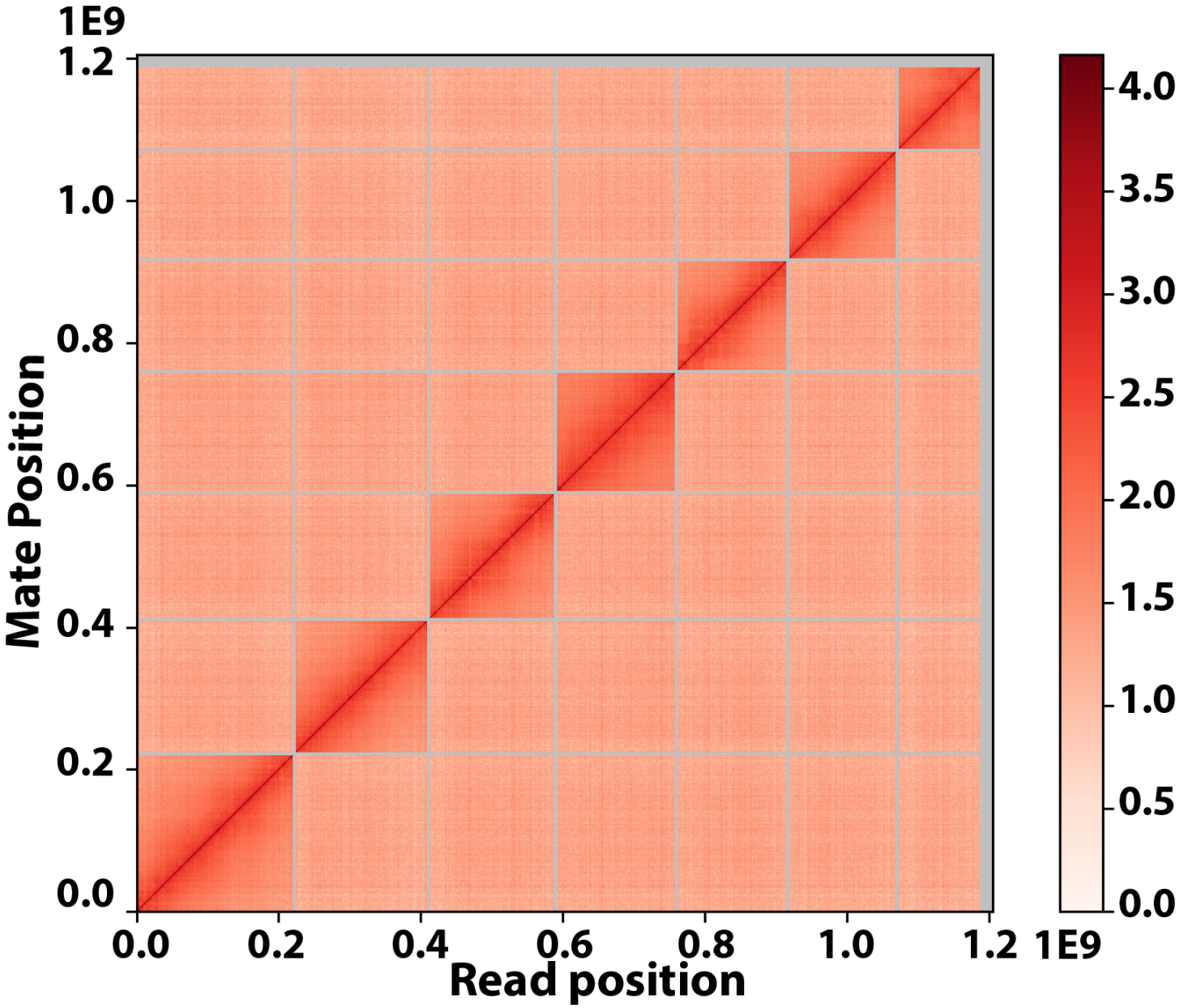

Figure 1

HiC linkage density histogram of Piezodorus guildinii genome assembly. The x and y axes in this histogram show the mapping positions of the first and second read in a read pair, respectively, grouped into bins. The intensity of color of each square represents the number of read pairs within that bin. Scaffolds less than 1 Mb were excluded from this histogram.

Supplementary Files

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

- FigS1.tif
- FigS2.tif

- FigS3.tif
- SupplementaryFigureLegends.docx
- TableS1.docx