The genome sequence of the Neotropical brown stink bug, *Euschistus heros* provides insights into population structure, demographic history and signatures of adaptation

Kumar Saurabh Singh, Erick M.G. Cordeiro, Benjamin J. Hunt, Aniruddha A. Pandit, Patricia L. Soares, Alberto S. Correa, Christoph T. Zimmer, Maria I. Zucchi, Carlos Batista, Julian A.T. Dow, Shireen-Anne Davies, Fernando Luís Cônsoli, Celso Omoto, Chris Bass

PII: S0965-1748(22)00172-2

DOI: https://doi.org/10.1016/j.ibmb.2022.103890

Reference: IB 103890

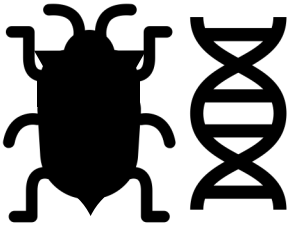
To appear in: *Insect Biochemistry and Molecular Biology*

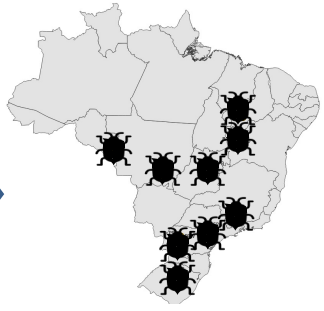
Received Date: 30 September 2022

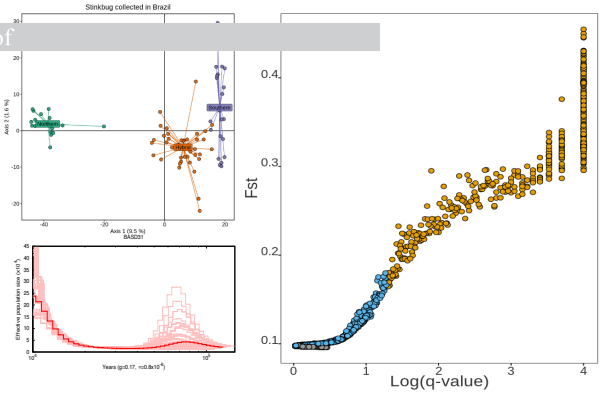
Revised Date: 3 December 2022

Accepted Date: 3 December 2022

Please cite this article as: Singh, K.S., Cordeiro, E.M.G., Hunt, B.J., Pandit, A.A., Soares, P.L., Correa, A.S., Zimmer, C.T., Zucchi, M.I., Batista, C., Dow, J.A.T., Davies, S.-A., Cônsoli, Fernando.Luí., Omoto, C., Bass, C., The genome sequence of the Neotropical brown stink bug, *Euschistus heros* provides insights into population structure, demographic history and signatures of adaptation, *Insect Biochemistry and Molecular Biology* (2023), doi: https://doi.org/10.1016/j.ibmb.2022.103890.







1. Sequencing and assembly of *E. heros* reference genome

2. Genotype by sequencing of *E. heros* populations in Brazil

3. Insights into population structure, demographic history and signatures of adaptation

The genome sequence of the Neotropical brown stink bug, *Euschistus heros* provides insights into population structure, demographic history and signatures of adaptation

Kumar Saurabh Singh[1$#], Erick M. G. Cordeiro[2#], Benjamin J. Hunt[1], Aniruddha A. Pandit[3], Patricia L. Soares[2], Alberto S. Correa[2], Christoph T. Zimmer[1,4], Maria I. Zucchi[5,6], Carlos Batista[5], Julian A. T. Dow[3], Shireen-Anne Davies[3], Fernando Luís Cônsoli[2], Celso Omoto[2], Chris Bass[1*]

[1]College of Life and Environmental Sciences, Biosciences, University of Exeter, Penryn Campus, Penryn, Cornwall, UK

[2]Departamento de Entomologia e Acarologia, Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, Brazil

[3]School of Molecular Biosciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK

[4]Syngenta Crop Protection, Werk Stein, Schaffhauserstrasse, Stein, Switzerland

[5]Institute of Biology, University of Campinas, Campinas, SP, Brazil.

[6]Secretariat of Agriculture and Food Supply of São Paulo State, APTA, UPDR-Piracicaba, São Paulo, Brazil

[$]Present address: Bioinformatics Group, Wageningen University, Droevendaalsesteeg 1, 6708 PB, Wageningen, The Netherlands

#Joint first author

*Correspondence to:

Chris Bass, email: c.bass@exeter.ac.uk, Tel. +44 1326 259084

Abstract

The Neotropical brown stink bug, *Euschistus heros*, is a major pest of soybean in South America. The importance of *E. heros* as a pest has grown significantly in recent times due to increases in its abundance and range, and the evolution of insecticide resistance. Recent work has begun to examine the genetic diversity, population structure, and genetic mechanisms of insecticide resistance in *E. heros*. However, to date, investigation of these topics has been hampered by a lack of genomic resources for this species. Here we address this need by assembling a high-quality draft genome for *E. heros*. We used a combination of short and long read sequencing to assemble an *E. heros* genome of 1.4 Gb comprising 906 contigs with a contig N50 of 3.5 MB. We leveraged this new genomic resource, in combination with genotyping by sequencing, to explore genetic diversity in populations of this species in Brazil and identify genetic loci in the genome which are under selection. Our genome-wide analyses, confirm that there are two populations of *E. heros* co-occurring in different geographical regions in Brazil, and that, in certain regions of the country these populations are hybridizing. We identify several regions of the genome as under selection, including markers associated with putative insecticide resistance genes. Taken together, the new genomic resources generated in this study will accelerate research into fundamental aspects of stinkbug biology and applied aspects relating to the sustainable control of a highly damaging crop pest.

1. Introduction

The Neotropical brown stink bug, *Euschistus heros* (F.) (Hemiptera: Pentatomidae), is an economically important pest of soybean and vegetable crops and causes severe damage via plant wilt and abortion of fruits and seeds (Sosa-Gómez et al., 2020). This species was originally considered a secondary pest with a restricted distribution (Zucchi et al., 2019). Considered rare in the Neotropics in the 1970s (Panizzi et al., 1977), *E. heros* used to be present in low numbers in central-west Brazil but expanded to southern and northeastern regions following increases in soybean cultivation in these areas (Panizzi, 2015). It is now the most abundant stink bug pest of soybean and other major crops in Brazil and has invaded Argentina and Paraguay, threatening other countries and states in South, Central and North America (Panizzi, 2015; Saluso et al., 2011; Sosa-Gómez et al., 2020). Furthermore, more recently, *E. heros* has been detected in cotton in Brazil, suggesting it has evolved the ability to utilize this plant as a host (Smaniotto and Panizzi, 2015; Zucchi et al., 2019). The control of *E. heros* relies heavily on the use of synthetic insecticides; however, in Brazil, the reliance on a restricted number of modes of action (neonicotinoids, pyrethroids and organophosphates) has created a high risk for resistance development (Boff et al., 2022; Somavilla et al., 2020; Sosa-Gómez et al., 2020; Tibola et al., 2021). Furthermore, recent work has demonstrated that field-collected strains of this species can rapidly develop moderate resistance to neonicotinoids under laboratory selection (Castellanos et al., 2019).

Should strong insecticide resistance evolve in *E. heros* in Brazil, the spread of any new resistance gene will be strongly influenced by population structure and gene flow across its range. Recent work using the genotyping by sequencing (GBS) approach identified high levels of genetic diversity in this species and strong genetic structure separating northern and southern populations with a distinct hybrid zone between them (Soares et al., 2018; Zucchi et al., 2019). The single nucleotide polymorphisms (SNPs) generated in this study were also used to identify DNA markers under strong selection. However, of the 61 loci identified as under directional selection, only 4 matched sequences in the NCBI database, precluding the

analysis of the majority of candidates. This outcome is a reflection of the paucity of genomic data available for this species. Specifically, to date, the genome of this species has not been sequenced, hampering the analysis and identification of genes of interest, such as those under selection and/or those relevant to the control of this pest.

Here we addressed this need by sequencing, assembling, and annotating the draft genome of *E. heros*. We then leveraged this resource, in combination with the genotyping by sequencing approach, to explore genetic diversity in populations of this species in Brazil and identify genetic loci in the genome which are under selection.

2. Methods

2.1 Insect strains

The EH1 strain of *E. heros* was used in this study. This is a long-term laboratory culture that is susceptible to insecticides. The culture was reared at 24°C, 55% relative humidity, with a 16/8 h (day/night) light cycle on a diet of green bean pods, *Phaseolus vulgaris*; soybean seeds, *Glycine max*; raw shelled peanuts, *Arachis hypogaea*; and sunflower seeds, *Helianthus annuus* (L.).

2.2 DNA extraction and genome sequencing

High molecular weight genomic DNA was extracted from individual adult stinkbugs, or pools of three stinkbugs, using the Genomic-tip 20/g kit (Qiagen) according to the manufacturer's instructions and eluted in tris-EDTA buffer. DNA quantity and quality was assessed by spectrophotometry using a NanoDrop (Thermo Scientific), Qubit assay (ThermoFisher) and gel electrophoresis. Sufficient DNA from a single stinkbug was obtained for the preparation of a single PCR-free paired-end library that was sequenced on a lane of an Illumina HiSeq 2500 using a 125bp paired-end read metric. DNA extracted from pools of stinkbugs were used to prepare libraries for PacBio sequencing, which were size selected for 15-20 kb+ and run on 15 SMRT cells, generating more than 130 Gb of data.

Following quality assessment of sequence data, reads were trimmed, filtered, and Illumina and PacBio data assembled with DBG2OLC (Ye et al., 2016) in a hybrid assembly approach. PacBio data was also assembled independently using the long-read assembler Flye -version 2.8.2 (Kolmogorov et al., 2019). The DBG2OLC and Flye assemblies were then merged using QuickMerge (Chakraborty et al., 2016) to produce a single more contiguous assembly. The merged assembly was then subjected to three rounds of polishing using both short-reads and long-reads. Polca -version 2.0.0 (Zimin and Salzberg, 2020) was used for short-read polishing and Racon -version 1.4.10 (Vaser et al., 2017) was used for long-read polishing. Duplicates in the polished assembly were identified and purged with Purge-haplotigs -version 1.0.4 (Roach et al., 2018). The completeness of the gene space in the assembled genome was assessed by KAT -version 2.4.2 (Mapleson et al., 2017), BUSCO (Benchmarking universal single-copy orthologs) –version 4.1.2 (Simão et al., 2015) and BLOBTools -version 1.1.1 (Laetsch and Blaxter, 2017).

2.3 Genome Annotation

Prior to gene prediction, the genome of *E. heros* was masked for repetitive elements using RepeatMasker –v4.0.7 (Tarailo-Graovac and Chen, 2009). RepeatModeler –v1.0.11 (Smit and Hubley, n.d.) was also used to identify repetitive sequences and transposable elements. The RepBase –v24.05 (Bao et al., 2015) library was then merged with the repeats predicted by RepeatModeler and used to mask the genome. Protein coding genes were predicted using GeneMark-ES –v4.68 (Borodovsky and Lomsadze, 2011) and AUGUSTUS –v3.4.0 (Stanke and Morgenstern, 2005) implemented in the BRAKER -2.1.6 (Brůna et al., 2021) pipeline using protein datasets and RNA-seq alignments as evidence. Publicly available *E. heros* RNA-seq datasets (BugAtlas.org) were individually mapped against the repeat masked genome using STAR –v2.7.1 (Dobin et al., 2013). The bam files from individual samples were then combined and fed into BRAKER with Uniprot ecdysozoa (downloaded 27/01/22) and *Halyomorpha halys* (Sparks et al., 2020) proteomes in --etp mode. The BRAKER supplementary script selectSupportedSubsets.py was used to filter the annotation file for gene models with either

full or partial hint support, followed by filtering using AGAT v0.6.0 (Dainat, 2022) to remove incomplete gene models and genes shorter than 200 bases. Functional annotation of the predicted gene models was performed based on homology searches against the NCBI nr and Interpro databases using BLAST2GO mapping and annotation features in OmicsBox -v2.0.36 (Conesa et al., 2005).

2.4 Ortholog Analysis

The proteome derived from the final annotation was compared to 6 other insect genomes. The proteomes of *E. heros, Acyrthosiphon pisum* 22Mar2018_4r6ur, *Bemisia tabaci* ASM185493v1, *H. halys* v2.0, *Myzus persicae* G006 v3, *Oncopeltus fasciatus* v1.2 and *Triboleum castaneum* v5.2 were downloaded from NCBI or USDA i5k Workspace and analysed with the *E. heros* proteome using Orthofinder -v2.5.4 (Emms and Kelly, 2015) to define orthologous groups of gene families. A rooted species tree was generated with Orthofinder using the STRIDE algorithm (Emms and Kelly, 2017) and converted to ultrametric using a supplementary Orthofinder script and a root divergence age of 420 MYA (Johnson et al., 2018). The matrix of orthogroups generated by Orthofinder was filtered using CAFE tutorial scripts prior to modelling gene family evolution using CAFE -v4.2.1 (Han et al., 2013).

2.5 Study sites for population genetic analysis

Adults of *E. heros* were sampled from 13 sites (total of 79 individuals on soybean plants) during the period between 12/2015 and 07/2016 (**Figure S1**). Upon arrival at the laboratory, the samples were separated in different tubes, and preserved in 98% ethanol at -80ºC prior to molecular analyses.

2.6 DNA extraction and Genotyping by sequencing

DNA was extracted from the head and leg tissues of each insect using a cetyltrimethylammonium bromide 2% (CTAB)-based protocol (Soares et al., 2018). The DNA integrity and concentration were first assessed by agarose gel (0.8% w/v) electrophoresis,

followed by fluorometric (i.e., Qubit system®) and spectrophotometric methods (i.e., NanoDrop®). Only DNA samples with a yield of >15 ng/µl were used for sequencing.

Genotyping by sequencing (GBS) was performed using the *PstI* and *MseI* restriction enzymes using the method described by Elshire *et al*. (Elshire et al., 2011). The selection of these restriction enzymes was based on previous work (Zucchi et al., 2019). Digested DNA was used to prepare Illumina libraries which were sequenced on an Illumina HiSeq 2500 (100 bp single-end reads) at the Animal Genome Centre at USP/ESALQ.

2.7 Demultiplexing, genotyping and SNPs filtering

The *E. heros* DNAseq data was demultiplexed and cleaned using *process-radtags* in STACKS v.2.4 (Catchen et al., 2013). Single-end reads were first checked for DNA quality, and adapter contamination using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Clean reads were, aligned to the reference genome created in this study using the Burrows-Wheeler aligner (BWA) v.0.7.15 (Li and Durbin, 2009) '*mem*' function. Aligned files were then sorted and converted to BAM format using SAMtools 1.10 (Li et al., 2009). Using a *population* module, SNPs were filtered by a minimum allele frequency (*min-maf*) of <5%, maximum observed heterozygosity of 75%, and maximum missing rate of 20%. Minimum coverage (x3), maximum number of alleles of 2, and markers under Hardy-Weinberg equilibrium were selected using VCFtools v.4.0 (Danecek et al., 2011).

2.8 Population Genomic Analyses

Using only biallelic markers under HWE genome wide (i.e., not considering collection sites), population structure and admixture were first evaluated using ADMIXTURE -version1.23 (Alexander and Lange, 2011), with unsupervised testing employing 12 values of K with *cv* = 5 and 3 replications. The most likely number of clusters was determined by the lowest cross-validation error value. We also implemented a non-model-based method, Principal Component Analysis (PCA), in R using *Adegenet* and *ade4* (Dray and Dufour, 2007; Jombart,

2008). In addition, a heatmap of the pairwise Euclidian distances between sampled locations was generated using the inbuilt functionality of R and a minimum spanning network (MSN) analysis using neutral SNPs were generated in R using the poppr package (Kamvar et al., 2014).

Finally, maximum-likelihood (ML) phylogeny was used to evaluate topology relationships of samples belonging to different populations. The ML-based phylogenetic tree was constructed by converting VCF formatted variants file to PHYLIP format using a custom python script. The final alignment had 79 sequences of 5096 bp in length. Phylogeny was estimated using IQTree -version 1.6 (Nguyen et al., 2015) using the TVMe+R5 as the best fit substitution model, according to Bayesian Information Criterian (BIC), with correction for ascertainment bias and 1000 bootstrap replicates (*-B* 1000 *-alrt* 1000 *-bnni -T* AUTO).

To infer past population dynamics and demographic changes over evolutionary time scales, the demography history of *E. heros* was inferred using the Pairwise sequentially Markovian Coalescent (PSMC) method (Li and Durbin, 2011). PSMC input files were generated using a custom python script that converts pstack output, from the STACKS pipeline, directly to psmcfa format. PSMC was run with the total number of iterations set to 30 (*-N* 30), T max (*-t*) set to 15, initial mutation/recombination ratio (*-r*) set to 5 and time bin parameter (*-p*) set to “4 + 30*2”. To scale effective population size parameter (*Ne*) we used a generation time of 0.17 and an average mutation rate of $0.8x10^{-8}$ per nucleotide per generation.

Using the whole dataset, the number of loci under natural selection was estimated by comparing the different sampled locations using BAYESCAN v.2.1 (Foll and Gaggiotti, 2008), using a prior odd of 100:1 (neutral model: selective model) and false discovery rate (FDR) of 5% (all other parameters as default). Comparison between lineages (40 samples composed of non-admixtured individuals, 21,928 variants used*)*, among all sampled sites (79 samples and 18,922 variants), and between sites within southern lineage (20 southern samples and 32,036 variants) were made to identify putative markers under selection. Finally, gene set

enrichment analysis (GSEA), implemented in a 10,000 bp window from SNPs under selection, was used to identify functional classes of genes that are overrepresented in the candidate gene set compared to the complete set using Blast2GO and Fisher's exact test (Conesa et al., 2005). Additionally, pairwise Fixation index (F_{ST}) was calculated using VCFtools (Danecek et al., 2011), using a 50,000 bp window to aid in the detection of SNP outliers and putative gene candidates in the surrounding region.

2.9 Data availability

The *E. heros* genome assembly is available at the National Center for Biotechnology Information (NCBI) under the BioProject accession PRJNA489772. All scripts associated with the alignment of GBS data and other downstream analyses are available here: https://github.com/kumarsaurabh20/Stinkbug_v2.0_popgenome.

3. Results and Discussion

3.1 The genome of E. heros

Sequencing of an inbred laboratory colony of *E. heros* using a combination of short Illumina reads (100X) and long single molecule sequencing (PacBio, 100X) generated more than 184 GB of data. This was assembled to generate an *E. heros* genome assembly of 1.3 Gb, somewhat higher than genome size estimates derived from k-mer analysis of the Illumina reads (**Fig. S2**), but consistent with the genome size of other Pentatomidae (Crowley and Barclay, 2021; Saha et al., 2022; Sparks et al., 2020). The final assembly comprised 906 contigs > 1 kb with an N50 of 3.5 MB (**Table 1**). The completeness of the gene space in the assembled genome was assessed using the BUSCO pipeline (Simão et al., 2015) with ~95% of Arthropoda and Insecta test genes identified as complete in the assembly (**Table 1**, **Fig. S2**). BRAKER2 (Brůna et al., 2021) generated 28,978 gene models of which 24,033 were BLAST annotated (excluding sequences with no BLAST hits or "uncharacterized protein" hits). Taken together these metrics are indicative of a high-quality draft genome with good gene representation and completeness. Thus, this new genomic resource for *E. heros* effectively

addresses the paucity of genomic data available for this species and complements the small but growing number of other pentatomid genome assemblies which have been recently published (Crowley and Barclay, 2021; Saha et al., 2022; Sparks et al., 2020).

To gain further insight into the gene complement of *E. heros* the proteome derived from the final annotation was compared to those of 6 other insect species, *A. pisum*, *B. tabaci*, *H. halys*, *M. persicae*, *O. fasciatus* and *T. castaneum*. Ortholog analysis generated 18,311 gene clusters (**Fig. 1**). Among these, 5,549 gene clusters were found in all species of which 1,275 consisted entirely of single-copy genes. A total of 527 genes were specific to *E. heros*, while 10,067 genes were shared between *E. heros* and *H. halys*, and 6,945, 7,196, 6,892, 8,844 and 7,149 genes are shared between *E. heros* and *A. pisum*, *B. tabaci*, *M. persicae, O. fasciatus and T. castaneum*, respectively.

Modelling of global gene gain and loss revealed a gene turnover rate of 0.0024 gains and losses per gene per million years in *E. heros*, similar to that reported for *D. melanogaster* (0.0023 duplications/gene/million years) (Lynch and Conery, 2000). Estimation of gene gain and loss in gene families across the 7 arthropod species revealed a positive average expansion (0.204) in *E. heros*, with a greater number of genes gained (5006, with 1123 gene families expanded) than lost (2003, with 1696 gene families contracted) (**Table S1**). This compares with *H. halys* which also has a positive (0.3814) average expansion resulting from a higher number of genes gained (7670, with 3605 gene families expanded) than lost (2049, with 1293 gene families contracted). Gene ontology (GO) enrichment analysis of genes specific to the stinkbug clade, identified GO categories related to sensory perception of taste, odorant receptor activity and mannan metabolic processes as significantly enriched (FDR < 0.05) (**Table S2**). These enriched terms relate to host plant localisation and utilisation and thus may be of applied relevance to the importance of *E. heros* and *H. halys* as agricultural pests. Related to this, stinkbug-specific genes represent good candidates to target in the development of future control interventions as they may facilitate the design of pest-specific controls. Finally, a total of 210 gene families were identified as rapidly evolving in *E. heros*

with genes involved in aspartic-type endopeptidase activity, serine-type carboxypeptidase activity, DNA integration and transposon activity significantly enriched (**Table S3**).

3.2 Population structure in E. heros populations in Brazil

We leveraged the genome assembly of *E. heros* to explore genetic diversity in populations of *E. heros* sampled from 13 sites across Brazil **(Fig. S1)**. The genotyping by sequencing approach was used for SNP discovery with the data obtained mapped to the reference assembly generated in this study. Combining all the GBS-based samples, a total of 396,022,893 BAM records were processed. After the processing step, 175,745,183 primary alignments (44%) were retained that represent 1505790.5 records/sample (16.0% - 60.4%). A total of 649,050 loci were compiled with effective per-sample coverage of 13.8x (stdev=11.6x) and a mean number of 90.2 sites per locus.

The first two components from the Principal Component Analysis (PCA) captured 15.53% of the total variance (PC1 = 13.18% and PC2 = 1.75) and grouped the 79 samples into three clusters corresponding to the northern, southern, and central regions of Brazil (**Fig. 2B).** A maximum likelihood (ML) phylogenetic analysis of neutrally evolving SNPs supported the patterns observed in PCA with samples clustering on the basis of geographic location (**Fig. 3**). Northern samples formed a distinct grouping in the phylogenetic tree with samples from the southern and central regions forming somewhat less distinct groupings. Admixture analysis partitioned genetic variation into 2 primary genetic clusters (i.e., optimal $K = 2$, **Fig. 2A**), and suggested that samples from the central region are hybrids of the northern and southern populations (**Fig. 2A**). Both minimum spanning network and heatmap analysis confirmed that hybrids are genetically closer to the southern than to northern lineage suggesting a higher degree of gene flow between these two groups (**Fig. 2C, D)**. Interrogation of the reference genome sequence at twenty randomly selected neutral positions that were found to consistently discriminate between the northern and southern lineages revealed that the genotype calls at all loci matched those of the southern lineage. This strongly suggests

that the reference assembly produced in this study represents the southern lineage of *E. heros*.

Taken together our data support the findings of previous analyses of genetic diversity in *E. heros* in Brazil conducted using non-reference approaches. Specifically, earlier studies of genetic diversity in this species using microsatellite markers revealed high genetic diversity and low structuration among the populations of *E. heros* studied (Husch et al., 2018). However, more recent work genotyping populations of *E. heros* using mitochondrial and nuclear gene markers or GBS has suggested there are two divergent lineages of *E. heros* in South America (Soares et al., 2018; Zucchi et al., 2019). Our reference-guided analysis corroborates these findings and provides additional evidence that this species is comprised of two well-differentiated populations in Brazil experiencing ongoing admixture. Our findings are also consistent with recent biological investigation of northern and southern lineages of *E. heros* in Brazil. Specifically, studies of reproductive behaviour and outcome have shown that while the northern and southern lineage can mate and produce viable offspring, southern lineage females and males preferred to mate with co-strain individuals whereas northern lineage insects did not show strain preference (Hickmann et al., 2021). This assortative mating may, in part, explain the genetic differentiation between the two lineages observed in our analyses. Finally, our finding that the level of genetic differentiation between hybrids and the southern lineages is significantly lower than between hybrids and the northern lineages provides new evidence that supports the trend observed in previous studies (Zucchi et al., 2019).

3.3 Population demographic reconstruction

Pairwise Sequential Markovian Coalescent Analysis (PSMC) was used to examine the temporal change in *E. heros* population size in Brazil. This revealed population growth over the last millennium (**Fig. 4**) coinciding with the beginning of the Holocene, which corresponds to the current geographical epoch, and after the last glacial period. The southern population

suffered a sharp decline in population size, experiencing a bottleneck, at around 10,000 and 100,000 years ago during the Pleistocene epoch. The bottleneck period coincides with the last glacial periods following Pleistocene terminations. Currently, the southern population is almost three times larger than the northern population. Interestingly the most marked variation in population size was observed for the hybrid population with the three representative hybrid populations varying by an order of magnitude in current effective population size. The consistent expansion observed in all populations of *E. heros* in our study likely reflects the intensification of agriculture in Brazil, especially in relation to soybean production but also cotton and bean. It has been previously hypothesized that the northern and southern populations of *E. heros* are exchanging adaptations in admixture zones in Brazil, and the marked variation in the effective population size of hybrid populations might reflect the emergence of more or less well adapted genotypes. Our results are largely consistent with the findings of a previous study which used mitochondrial and nuclear gene markers to investigate the demographic history of *E. heros* populations in Brazil and Paraguay (Soares et al., 2018; Zucchi et al., 2019). This study inferred that the two lineages diverged in the Pliocene (4.5 Myr), that the northern lineage is older and more diverse than the southern lineage and that populations are expanding in size and range but at different rates, strongly affected by environmental variables. Our results, which combine information from a much larger number of loci than employed previously, corroborate the notion that *E. heros* populations are expanding in Brazil in a demographic process that started with changes in environmental conditions and has been accelerated by anthropogenic factors related to agricultural intensification. Specifically, the hybrid zone is located in the largest soybean and cotton producing region in Brazil, this would have provided a green bridge for northern and southern lineages to connect and hybridise and may have contributed to the emergence of new adaptations. As the populations identified in this study are well structured, suggestive of limited gene flow, these adaptations will likely remain primarily in the hybrid zone and only spread slowly to other locations. In this regard, these findings have implications for the control of this species with insecticides. Specifically, the marked population structure in *E. heros* in Brazil

observed in this and previous studies suggests that resistance alleles that emerge in populations may increase in frequency rapidly within lineages/populations but spread only slowly between them.

3.4 Genome-wide selection scan and identification of SNPs under selection

To explore the genomic landscape of divergence between lineages and among the geographically separated populations of *E. heros*, and identify candidate genomic regions exhibiting signatures of selection, we performed a selection scan based on a Bayesian approach to model allele frequency differences (Foll and Gaggiotti, 2008). When considering two groups (northern vs southern pure lineages), no putative markers under selection were found considering the neutral model 100x more likely than positive selection model. In contrast, when the 13 sampled locations were compared, a total of 729 markers were flagged as putative candidate regions under positive selection (**Table S4**). Gene enrichment analysis of genes within 5 kb windows of the flagged markers showed that genes related to sodium channel activity (GO:0005272) are over-represented (**Table S5**). Of the candidate genes identified, three were potentially related to insecticide use. These included genes encoding two cytochrome P450s and the voltage-gated sodium channel, all within a 5 kb range of significant makers (**Table S4**). Furthermore, an association between the frequency of the SNPs linked to these candidate genes and geographic location of the *E. heros* populations sampled was observed, with samples representing northern and southern lineages exhibiting marked divergence in frequency of the alternate nucleotide at each position, and hybrids displaying intermediate SNP frequencies (**Fig. 5B)**. This finding may suggest local adaptation within the northern and southern lineages of *E. heros* and significant differences in allele frequencies at various locations. Related to this, in a recent area-wide survey of the likelihood of insecticide control failure in *E. heros* in the state of Goiás in central Brazil, a higher risk for beta-cyfluthrin control failure was detected in southern regions, whereas in the northern area of the state a higher risk of control failure was detected for imidacloprid (Tuelher et al., 2018).

The identification of two candidate genes, *CYP4G15* and *CYP6D5,* encoding cytochrome P450s is significant as this class of enzymes has been implicated in metabolic resistance to insecticides in a wide range of insect pests (Nauen et al., 2022). Furthermore, in the case of *E. heros* the P450 inhibitor piperonyl butoxide (PBO) has been shown to significantly increase the susceptibility of field populations to pyrethroid insecticides, implicating P450 enzymes in the detoxification of this insecticide class (Boff et al., 2022). Enhanced P450 activity has also been implicated in the resistance of two laboratory-selected strains to imidacloprid (Castellanos et al., 2019). P450s of the CYP4G subfamily catalyse the synthesis of cuticular hydrocarbons (CHC) (Feyereisen, 2020), and have been shown to be highly expressed in the oenocytes, specialized secretory cells in the epidermis (Balabanidou et al., 2016; Chung et al., 2009). In addition to their essential role in CHC biosynthesis P450s belonging to this subfamily have been implicated in resistance to insecticides in several insect species (Feyereisen, 2020). More specifically, the regulation of CHC production by CYP4G enzymes has been shown to affect insecticide penetration, and hence mediate resistance via reduced uptake of insecticide through the insect cuticle (Balabanidou et al., 2016; Kefia et al., 2019). This mechanism has been shown to confer resistance to pyrethroids (Balabanidou et al., 2016), which are one of the primary insecticides used for the control of *E. heros*. In the case of *CYP6D5*, P450 genes belonging to this subfamily have been shown to confer resistance to xenobiotics, including insecticides, in other insects. For example, the upregulation of *CYP6D1* in houseflies, *Musca domestica*, confers resistance to pyrethroids (Scott et al., 1998) and *CYP6D5* of *D. melanogaster* has been linked to resistance to caffeine (Najarro et al., 2015). Thus, both P450s associated with markers under selection in *E. heros* represent promising candidates for future functional analyses of their role in insecticide resistance and/or detoxification.

The identification of the voltage-gated sodium channel gene in association with markers under selection in our analysis is significant as this gene encodes the target protein of pyrethroid insecticides. Conserved mutations have been identified in this gene in a range of insect

species that confer resistance to this insecticide class (ffrench-Constant et al., 2016). Unfortunately, no GBS reads were identified in this study mapping to the known mutation positions. Thus, future screening for mutations at these 'resistance hotspots' in *E. heros* is warranted.

4. Conclusions

We present the first reference genome sequence assembly for *E. heros* addressing a critical resource gap in the research toolkit for this species. Measures of assembly contiguity (N50, L50 etc.) and gene representation (BUSCO and OrthoDB assessment) are indicative of a contiguous assembly with a high degree of gene content completeness.

Leveraging the new *E. heros* assembly to explore the genomic landscape of divergence among populations of this important pest across Brazil supported the presence of two populations occurring in different geographical areas in Brazil corresponding to northern and southern regions. Our analysis also reveal that the two populations are hybridizing at the Brazilian cerrado. These hybrids are sufficiently differentiated to suggest they have the potential to form a third population. Both the northern and southern populations have expanded in size since the beginning of the Holocene. The confirmation of strong population structure in *E. heros* in Brazil is of applied significance as it has the potential to influence the flow of genes between populations of this species that impact its status as a pest. For example, alleles conferring resistance to insecticides. In this regard our genome-wide selection analysis identified markers under selection in comparisons of *E. heros* populations that are in close association with genes encoding putative insecticide resistance genes. These candidates represent promising avenues for future investigation of the mechanisms of insecticide detoxification and resistance in this species.

Acknowledgments

This project has received funding from the Biotechnology and Biological Sciences Research Council, UK under the BBSRC-FAPESP Joint Pump-Priming Awards for AMR and Insect

Pest Resistance in Livestock and Agriculture (Grant Ref: BB/R022623/1 and 2017/50455-5) and BBSRC-FAPESP Newton Award for AMR and insect pest resistance in agriculture and livestock (Grant Ref: BB/S018719/1 and 2018/21155-6). Erick Cordeiro was supported by a post-doctoral fellowship from FAPESP (Grant Ref: 2018/20668-0). We thank Emma Randall for technical assistance with DNA extraction.

References

Alexander, D.H., Lange, K., 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics 12, 246.

Balabanidou, V., Kampouraki, A., MacLean, M., Blomquist, G.J., Tittiger, C., Juárez, M.P., Mijailovsky, S.J., Chalepakis, G., Anthousi, A., Lynd, A., Antoine, S., Hemingway, J., Ranson, H., Lycett, G., Vontas, J., 2016. Cytochromes P450 associated with insecticide resistance catalyse cuticular hydrocarbon production in Anopheles gambiae. Proc. Natl. Acad. Sci. U. S. A. 113, 9268–9273.

Bao, W., Kojima, K.K., Kohany, O., 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob. DNA 6, 11. https://doi.org/10.1186/s13100-015-0041-9

Boff, J.S., Reis, A.C., Patricia, da S.G., Pretto, V.E., Garlet, C.G., Melo, A.A., Bernardi, O., 2022. The Effect of Synergistic Compounds on the susceptibility of *Euschistus heros* (Hemiptera: Pentatomidae) and *Chrysodeixis includens* (Lepidoptera: Noctuidae) to pyrethroids. Environ. Entomol. 51, 421–429. https://doi.org/10.1093/ee/nvac005

Borodovsky, M., Lomsadze, A., 2011. Eukaryotic gene prediction using GeneMark.hmm-E and GeneMark-ES. Curr. Protoc. Bioinformatics 35, Unit 4.6.1-10. https://doi.org/10.1002/0471250953.bi0406s35

Brůna, T., Hoff, K.J., Lomsadze, A., Stanke, M., Borodovsky, M., 2021. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. NAR Genom. Bioinform. 3, lqaa108. https://doi.org/10.1093/nargab/lqaa108

Castellanos, N.L., Haddi, K., Carvalho, G.A., de Paulo, P.D., Hirose, E., Guedes, R.N.C., Smagghe, G., Oliveira, E.E., 2019. Imidacloprid resistance in the Neotropical brown stink bug *Euschistus heros*: selection and fitness costs. J. Pest Sci. 92, 847–860. https://doi.org/10.1007/s10340-018-1048-z

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., Cresko, W.A., 2013. Stacks: an analysis tool set for population genomics. Mol. Ecol. 22, 3124–3140. https://doi.org/10.1111/mec.12354

Chakraborty, M., Baldwin-Brown, J.G., Long, A.D., Emerson, J.J., 2016. Contiguous and accurate *de novo* assembly of metazoan genomes with modest long read coverage. Nucleic Acids Res 44, e147.

Chung, H., Sztal, T., Pasricha, S., Sridhar, M., Batterham, P., Daborn, P.J., 2009. Characterization of *Drosophila melanogaster* cytochrome P450 genes. Proc. Natl. Acad. Sci. U. S. A. 106, 5731–5736. https://doi.org/10.1073/pnas.0812141106

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., Robles, M., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21, 3674–3676.

Crowley, L., Barclay, M.V.L., 2021. The genome sequence of the bishop's mitre shieldbug, *Aelia acuminata* (Linnaeus, 1758). Wellcome Open Res. 6, 320. https://doi.org/10.12688/wellcomeopenres.17400.1

Dainat, J., 2022. AGAT: Another Gff Analysis Toolkit to handle annotations in any GTF/GFF format. Zenodo. https://www.doi.org/10.5281/zenodo.3552717

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., 2011. 1000 Genomes Project Analysis Group, The variant call format and VCFtools. Bioinformatics 27, 2156–2158.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. https://doi.org/10.1093/bioinformatics/bts635

Dray, S., Dufour, A.B., 2007. The ade4 package: implementing the duality diagram for ecologists. J. Stat. Softw. 22, 1–20.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., Mitchell, S.E., 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6, e19379. https://doi.org/10.1371/journal.pone.0019379

Emms, D.M., Kelly, S., 2017. STRIDE: Species tree root inference from gene duplication events. Mol. Biol. Evol. 34, 3267–3278. https://doi.org/10.1093/molbev/msx259

Emms, D.M., Kelly, S., 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16, 157. https://doi.org/10.1186/s13059-015-0721-2

Feyereisen, R., 2020. Origin and evolution of the CYP4G subfamily in insects, cytochrome P450 enzymes involved in cuticular hydrocarbon synthesis. Mol. Phylogenet. Evol. 143, 106695. https://doi.org/10.1016/j.ympev.2019.106695

ffrench-Constant, R.H., Williamson, M.S., Davies, T.G.E., Bass, C., 2016. Ion channels as insecticide targets. J. Neurogenet. 30, 163–177.

Foll, M., Gaggiotti, O., 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A bayesian perspective. Genetics 180, 977–993. https://doi.org/10.1534/genetics.108.092221

Han, M. v., Thomas, G.W.C., Lugo-Martinez, J., Hahn, M.W., 2013. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. Mol. Biol. Evol. 30, 1987–1997. https://doi.org/10.1093/molbev/mst100

Hickmann, F., Cordeiro, E.G., Soares, P.L., Aurélio, M.S.L., Schwertner, C.F., Corrêa, A.S., 2021. Reproductive patterns drive the gene flow and spatial dispersal of *Euschistus heros* (Hemiptera: Pentatomidae). J. Econ. Entomol. 114, 2346–2354. https://doi.org/10.1093/jee/toab190

Husch, P.E., Ferreira, D.G., Seraphim, N., Harvey, N., Silva-Brandão, K.L., Sofia, S.H., Sosa-Gómez, D.R., 2018. Structure and genetic variation among populations of *Euschistus heros* from different geographic regions in Brazil. Entomol. Exp. Appl. 166, 191–203. https://doi.org/10.1111/eea.12666

Johnson, K.P., Dietrich, C.H., Friedrich, F., Beutel, R.G., Wipfler, B., Peters, R.S., Allen, J.M., Petersen, M., Donath, A., Walden, K.K.O., Kozlov, A.M., Podsiadlowski, L., Mayer, C., Meusemann, K., Vasilikopoulos, A., Waterhouse, R.M., Cameron, S.L., Weirauch, C., Swanson, D.R., Percy, D.M., Hardy, N.B., Terry, I., Liu, S., Zhou, X., Misof, B., Robertson, H.M., Yoshizawa, K., 2018. Phylogenomics and the evolution of hemipteroid insects. Proc. Natl. Acad. Sci. U. S. A. 115, 12775–12780. https://doi.org/10.1073/pnas.1815820115

Jombart, T., 2008. Adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24, 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Kamvar, Z.N., Tabima, J.F., Grünwald, N.J., 2014. *Poppr*: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2, e281. https://doi.org/10.7717/peerj.281

Kefia, M., Balabanidou, V., Douris, V., Lycett, G., Feyereisen, R., Vontas, J., 2019. Two functionally distinct CYP4G genes of *Anopheles gambiae* contribute to cuticular hydrocarbon biosynthesis. Insect Biochem. Mol. Biol. 110, 52–59.

Kolmogorov, M., Yuan, J., Lin, Y., Pevzner, P.A., 2019. Assembly of long, error-prone reads using repeat graphs. Nat. Biotechnol. 37, 540–546.

Laetsch, D.R., Blaxter, M.L., 2017. BlobTools: Interrogation of genome assemblies. F1000Res 6, 1287.

Li, H., Durbin, R., 2011. Inference of human population history from individual whole-genome sequences. Nature 475, 493–496. https://doi.org/10.1038/nature10231

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., Subgroup, 1000 Genome Project Data Processing, 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079.

Lynch, M., Conery, J.S., 2000. The evolutionary fate and consequences of duplicate genes. Science (1979) 290, 1151–1155. https://doi.org/10.1126/science.290.5494.1151

Mapleson, D., Accinelli, G.G., Kettleborough, G., Wright, J., Clavijo, B.J., 2017. KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. Bioinformatics 33, 574–576.

Najarro, M.A., Hackett, J.L., Smith, B.R., Highfill, C.A., King, E.G., Long, A.D., Macdonald, S.J., 2015. Identifying loci contributing to natural variation in xenobiotic resistance in Drosophila. PLoS Genet. 11, e1005663. https://doi.org/10.1371/journal.pgen.1005663

Nauen, R., Bass, C., Feyereisen, R., Vontas, J., 2022. The role of cytochrome P450s in insect toxicology and resistance. Annu. Rev. Entomol. 67, 105–124.

Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274. https://doi.org/10.1093/molbev/msu300

Panizzi, A.R., 2015. Growing problems with stink bugs (Hemiptera: Heteroptera: Pentatomidae): *Species Invasive to the U.S. and Potential Neotropical Invaders*. Am. Entomol. 61, 223–233. https://doi.org/10.1093/ae/tmv068

Panizzi, A.R., Corrêa, B.S., Gazzoni, D.L., Oliveira, E.B., Newman, G.G., Turnipseed, S.G., 1977. Insetos da soja no Brasil. Embrapa, CNPSo, Londrina, PR, Boletim Técnico 1, 20.

Roach, M.J., Schmidt, S.A., Borneman, A.R., 2018. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. BMC Bioinformatics 19, 460.

Saha, S., Allen, K.C., Mueller, L.A., Reddy, G.V.P., Perera, O.P., 2022. Chromosome length genome assembly of the redbanded stink bug, *Piezodorus guildinii* (Westwood). BMC Res. Notes 15, 115. https://doi.org/10.1186/s13104-022-05924-5

Saluso, A., Xavier, L., Silva, F.A.C., Panizzi, A.R., 2011. An invasive pentatomid pest in Argentina: neotropical brown stink bug, *Euschistus heros* (F.) (Hemiptera: Pentatomidae). Neotrop. Entomol. 40, 704–5.

Scott, J.G., Liu, N., Wen, Z., 1998. Insect cytochromes P450: diversity, insecticide resistance and tolerance to plant toxins. Comp. Biochem. Physiol. C Pharmacol. Toxicol. Endocrinol. 121, 147–155. https://doi.org/10.1016/S0742-8413(98)10035-X

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E. v, Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212.

Smaniotto, L.F., Panizzi, A.R., 2015. Interactions of selected species of stink bugs (Hemiptera: Heteroptera: Pentatomidae) from leguminous crops with plants in the neotropics. Fla. Entomol. 98, 7–17. https://doi.org/10.1653/024.098.0103

Smit, A.F.A., Hubley, R., n.d. RepeatModeler Open-1.0. 2008–2015. http://www.repeatmasker.org.

Soares, P.L., Cordeiro, E.M.G., Santos, F.N.S., Omoto, C., Correa, A.S., 2018. The reunion of two lineages of the Neotropical brown stink bug on soybean lands in the heart of Brazil. Sci. Rep. 8, 2496. https://doi.org/10.1038/s41598-018-20187-6

Somavilla, J.C., da S Gubiani, P., Reis, A.C., Führ, F.M., Machado, E.P., Bernardi, O., 2020. Susceptibility of *Euschistus heros* and *Dichelops furcatus* (Hemiptera: Pentatomidae) to insecticides determined from topical bioassays and diagnostic doses for resistance monitoring of *E. heros* in Brazil. Crop Prot. 138, 105319. https://doi.org/10.1016/j.cropro.2020.105319

Sosa-Gómez, D.R., Corrêa-Ferreira, B.S., Kraemer, B., Pasini, A., Husch, P.E., Delfino Vieira, C.E., Reis Martinez, C.B., Negrão Lopes, I.O., 2020. Prevalence, damage, management and insecticide resistance of stink bug populations (Hemiptera: Pentatomidae) in commodity crops. Agric. For. Entomol. 22, 99–118. https://doi.org/10.1111/afe.12366

Sparks, M.E., Bansal, R., Benoit, J.B., Blackburn, M.B., Chao, H., Chen, M., Cheng, S., Childers, C., Dinh, H., Doddapaneni, H.V., Dugan, S., Elpidina, E.N., Farrow, D.W., Friedrich, M., Gibbs, R.A., Hall, B., Han, Y., Hardy, R.W., Holmes, C.J., Hughes, D.S.T., Ioannidis, P., Cheatle Jarvela, A.M., Johnston, J.S., Jones, J.W., Kronmiller, B.A., Kung, F., Lee, S.L., Martynov, A.G., Masterson, P., Maumus, F., Munoz-Torres, M., Murali, S.C., Murphy, T.D., Muzny, D.M., Nelson, D.R., Oppert, B., Panfilio, K.A., Paula, D.P., Pick, L., Poelchau, M.F., Qu, J., Reding, K., Rhoades, J.H., Rhodes, A., Richards, S., Richter, R., Robertson, H.M., Rosendale, A.J., Tu, Z.J., Velamuri, A.S., Waterhouse, R.M., Weirauch, M.T., Wells, J.T., Werren, J.H., Worley, K.C., Zdobnov, E.M., Gundersen-Rindal, D.E., 2020. Brown marmorated stink bug, *Halyomorpha halys* (Stål), genome: putative underpinnings of polyphagy, insecticide resistance potential and biology of a top worldwide pest. BMC Genomics 21, 227. https://doi.org/10.1186/s12864-020-6510-7

Stanke, M., Morgenstern, B., 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 33, W465–W467. https://doi.org/10.1093/nar/gki458

Tarailo-Graovac, M., Chen, N., 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr. Protoc. Bioinformatics 25. https://doi.org/10.1002/0471250953.bi0410s25

Tibola, C.M., Silva, L., Sgubin, F., Omoto, C., 2021. Monitoring resistance of *Euschistus heros* (Fabricius) (Hemiptera: Pentatomidae) to insecticides by using encapsulated artificial diet bioassay. Insects 12, 599. https://doi.org/10.3390/insects12070599

Tuelher, E.S., da Silva, É.H., Rodrigues, H.S., Hirose, E., Guedes, R.N.C., Oliveira, E.E., 2018. Area-wide spatial survey of the likelihood of insecticide control failure in the neotropical brown stink bug *Euschistus heros*. J. Pest Sci. (2004) 91, 849–859. https://doi.org/10.1007/s10340-017-0949-6

Vaser, R., Sović, I., Nagarajan, N., Šikić, M., 2017. Fast and accurate *de novo* genome assembly from long uncorrected reads. Genome Res. 27, 737–746.

Ye, C., Hill, C.M., Wu, S., Ruan, J., Ma, Z. (Sam), 2016. DBG2OLC: Efficient assembly of large genomes using long erroneous reads of the third generation sequencing technologies. Sci. Rep. 6, 31900. https://doi.org/10.1038/srep31900

Zimin, A. v., Salzberg, S.L., 2020. The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies. PLoS Comput. Biol. 16, e1007981. https://doi.org/10.1371/journal.pcbi.1007981

Zucchi, M.I., Cordeiro, E.M.G., Wu, X., Lamana, L.M., Brown, P.J., Manjunatha, S., Viana, J.P.G., Omoto, C., Pinheiro, J.B., Clough, S.J., 2019. Population genomics of the Neotropical brown stink bug, *Euschistus heros*: The most important emerging insect pest to soybean in Brazil. Front. Genet. 10. https://doi.org/10.3389/fgene.2019.01035

Figure legends

Figure 1. Phylogenomic analysis of the genome of *E. heros* and 6 other arthropod species. (A) Phylogenetic relationship and gene orthology of *E. heros* and other arthropods. SC indicates common orthologs with the same number of copies in different species, MC indicates common orthologs with different copy numbers in different species. SpS indicates species-specific paralogs, UC indicates all genes which were not assigned to a gene family, SS and AS indicate clade-specific genes. (B) Gene families shared by selected species. (C) Species dated phylogenetic tree and gene family evolution. Numbers on the branch indicate counts of gene families that are expanding (green), contracting (red) and rapidly evolving (blue).

Figure 2. *Euschistus heros* population structure and admixture. (A) Admixture analysis for three K values. (B) Principal component analysis; colors represent sampled location showed on the map (right side). (C) Minimum spanning network (MSN) analysis. Each sample sequenced is represented as a node, and the genetic distance is represented by the edges. (D) Heatmap of the pairwise distance between sampled locations. Darker color indicates further distances.

Figure 3. Maximum likelihood phylogeny of *E. heros* samples collected from across Brazil. Phylogeny was generated using SNPs from GBS data. The alignment had 79 sequences with 5096 columns, 3481 distinct patterns, 3395 parsimony-informative, 142

singleton sites and 1559 constant sites. TVMe+R5 substitution model was chosen as the Best-fit model according to BIC (Bayesian Information Criterion). Bootstrap values were generated utilizing 1000 samples from the ultrafast bootstrap function implemented in the IQTree.

Figure 4. Population demographic reconstruction of *E. heros* populations in Brazil. Inferred historical population sizes using pairwise sequential Markovian coalescent analysis. The x-axis is time in years and the y-axis the effective population size. Analysis was performed for three northern populations (BASD31, BALE13 and PIB7), three southern populations (PRQS37, SPA10 and PRC34) and three hybrid populations (ROC31, MTR39 and GOPB22).

Figure 5. Signatures of selection in comparisons of *E. heros* populations in Brazil. (A) Putative markers under neutral, balancing and positive selection when all locations were contrasted, (B) SNP frequency at each sampled location of three putative candidates within 1 kb of the flagged SNP markers.

Table 1. *E. heros* genome assembly statistics

Assembly Statistic	Value
Assembly size / Total bases	1311287533 bp
Total number of Contigs (> = 1 kb)	906
Contig N50	3469525 bp
Largest contig	24980295 bp
Contig L50	107
Mean Length	1447337 bp
Total complete BUSCOs (Arthropoda)	96%
Total gene content	28,978
Total genes annotate	24,033

Fig. 1

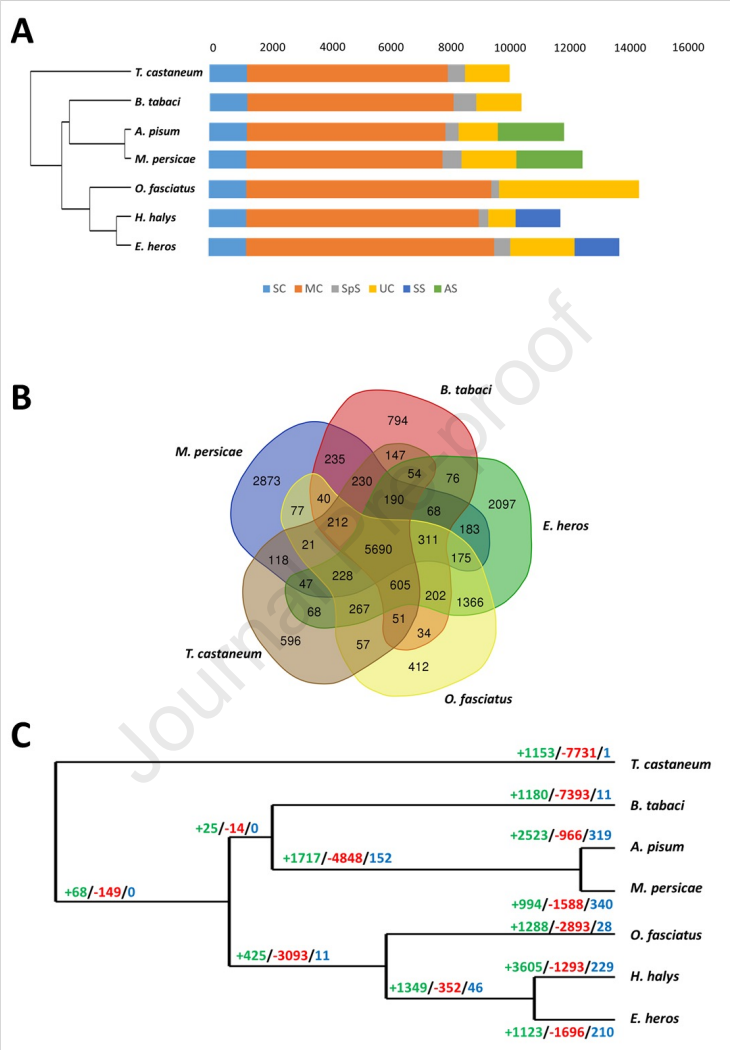


Fig. 2

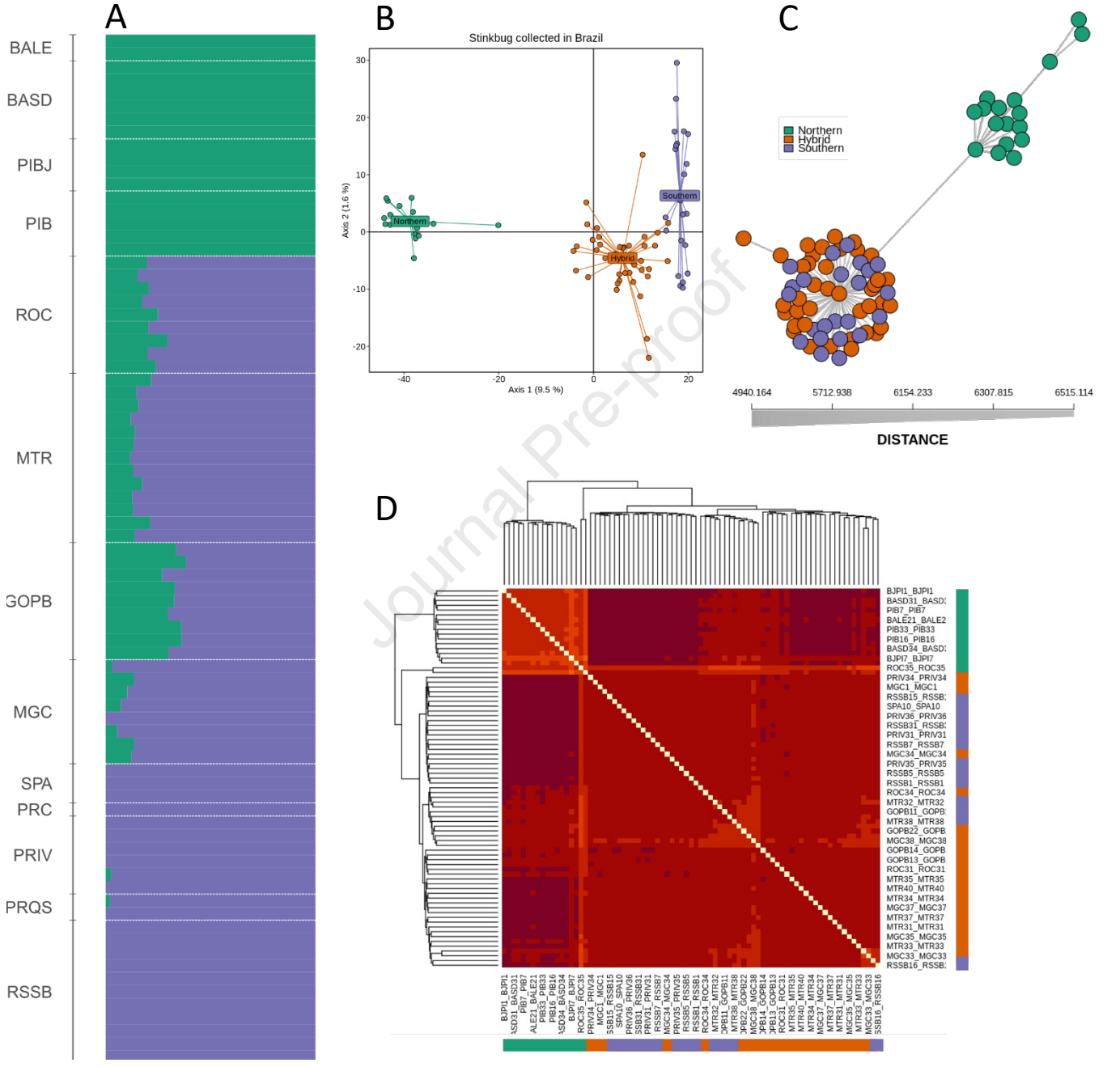


Fig. 3

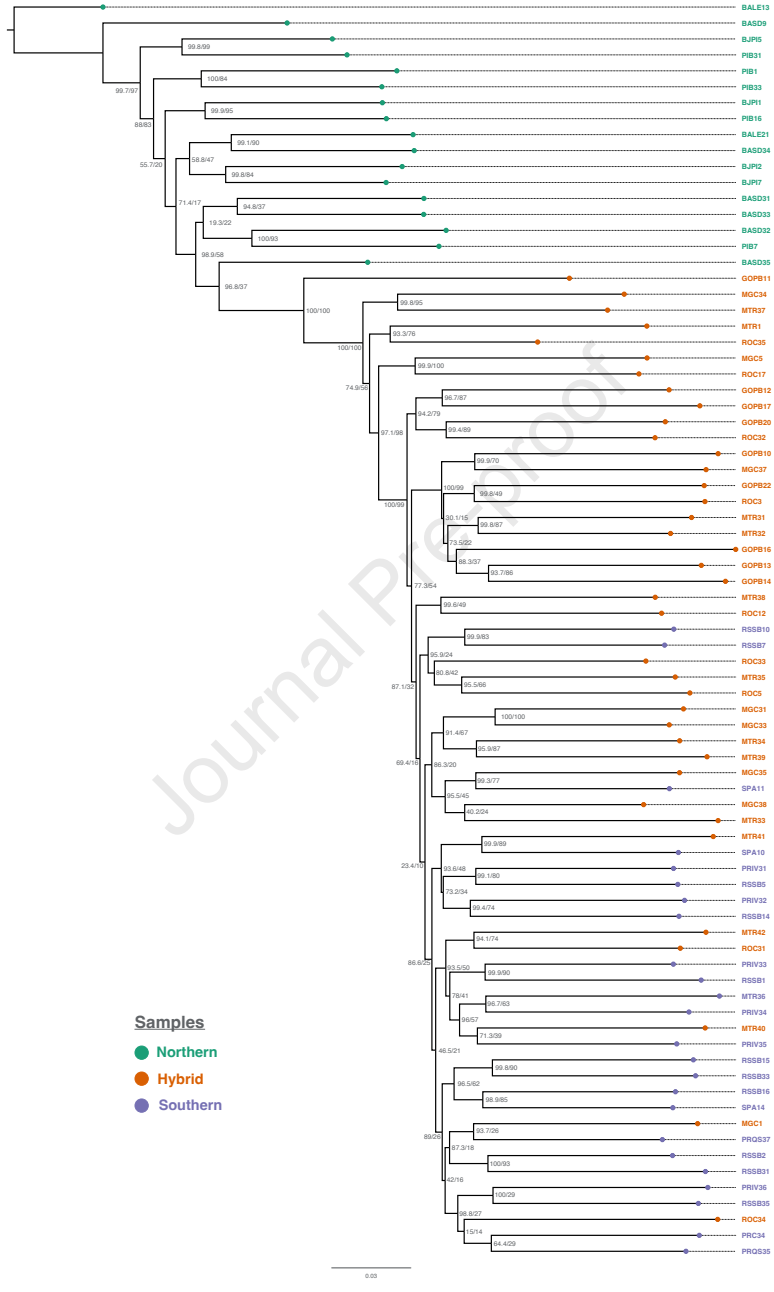


Fig. 4

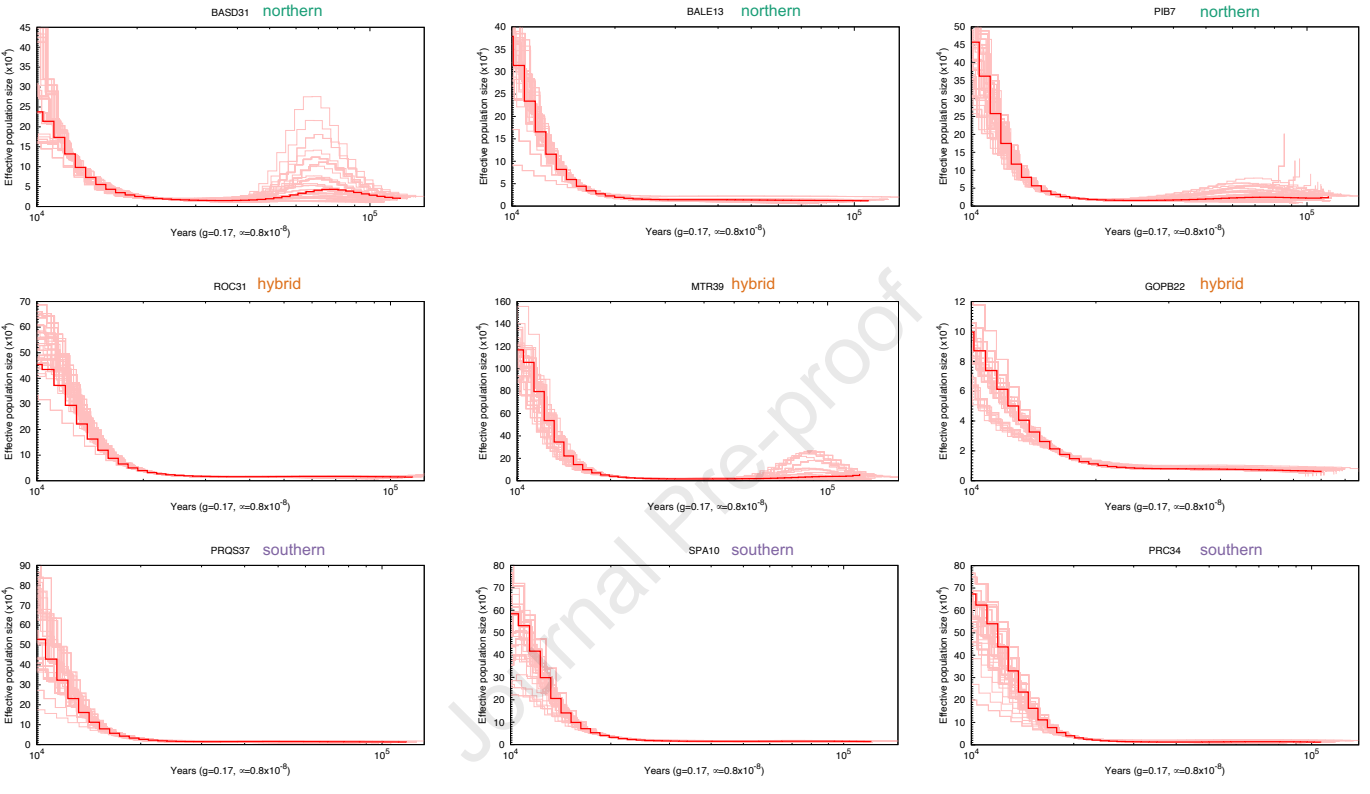
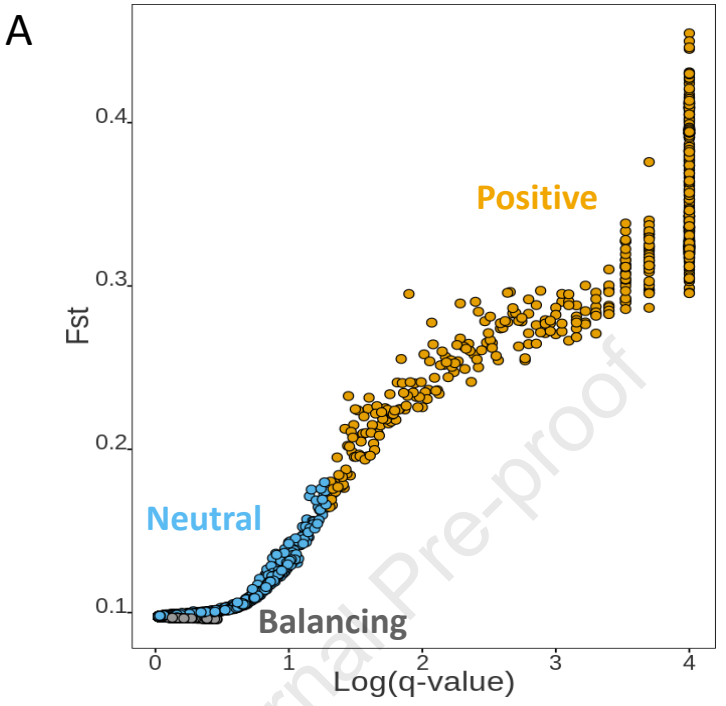
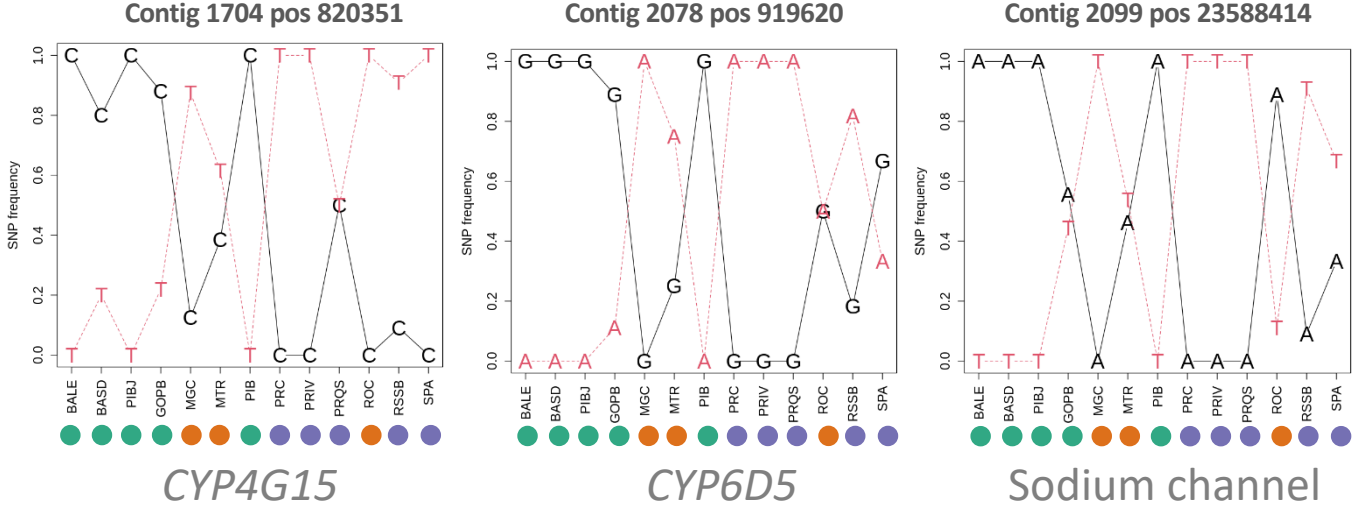


Fig. 5



B

SNP candidate frequency at each location



The genome of the Neotropical brown stink bug, *Euschistus heros* was sequenced, assembled, and annotated.

The reference genome and genotyping by sequencing was used to explore genetic diversity in populations of *E. heros* in Brazil.

Two populations of *E. heros* were identified in Brazil that are hybridizing in certain regions of the country.

Several regions of the genome were identified as under selection, including markers associated with putative insecticide resistance genes.