Functional & Integrative Genomics (2022) 22:611–624
https://doi.org/10.1007/s10142-022-00852-w

ORIGINAL ARTICLE

Whole genome sequencing of spotted stem borer, *Chilo partellus*, reveals multiple genes encoding enzymes for detoxification of insecticides

Mukesh K. Dhillon[1] · Jagdish Jaba[2] · Pallavi Mishra[3] · Mir Asif Iquebal[3] · Sarika Jaiswal[3] · Aditya K. Tanwar[1] · Nareshkumar Bharat[2] · Naveen Arora[2] · Suraj Prasad Mishra[2] · Shyam Prasad Gogineni[4] · Fazil Hasan[1] · Anil Rai[3] · Dinesh Kumar[3,5] · Hari C. Sharma[2]

Received: 24 November 2021 / Revised: 20 March 2022 / Accepted: 22 March 2022 / Published online: 15 April 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Spotted stem borer, *Chilo partellus*, is the most important constraint for increasing the production and productivity of maize and sorghum, the two major coarse cereals in Asia and Africa. The levels of resistance to this pest in the cultivated germplasm are low to moderate, and hence, farmers have to use insecticides for effective control of this pest. However, there is no information on the detoxification mechanisms in *C. partellus*, which is one of the constraints for deployment of appropriate insecticides to control this pest. The ability to detoxify insecticides varies across insect populations, and hence, we sequenced different populations of *C. partellus* to identify and understand detoxification mechanisms to devise appropriate strategies for deployment of different insecticides for controlling this pest. Larval samples were sequenced from three different cohorts of *C. partellus* using the Illumina HiSeq 2500 platform. The data were subjected to identify putative genes that are involved in detoxification on insecticides in our cohort insect species. These studies resulted in identification of 64 cytochrome P450 genes (CYP450s), and 36 glutathione S-transferases genes (GSTs) encoding metabolic detoxification enzymes, primarily responsible for xenobiotic metabolism in insects. A total of 183 circadian genes with > 80% homolog and 11 olfactory receptor genes that mediate chemical cues were found in the *C. partellus* genome. Also, target receptors related to insecticide action, 4 acetylcholinesterase (AChE), 14 γ-aminobutyric acid (GABA), and 15 nicotinic acetylcholine (nAChR) receptors were detected. This is the first report of whole genome sequencing of *C. partellus* useful for understanding mode of action of different insecticides, and mechanisms of detoxification and designing target-specific insecticides to develop appropriate strategies to control *C. partellus* for sustainable crop production.

Keywords Maize · Sorghum · Spotted stem borer · *Chilo partellus* · Genome assembly · Insecticide detoxification

✉ Dinesh Kumar
dinesh.kumar@icar.gov.in

[1] Division of Entomology, ICAR-Indian Agricultural Research Institute, New Delhi 110012, India

[2] International Crops Research Institute for the Semi-Arid Tropics, Patancheru, Hyderabad 502324, Telangana, India

[3] Centre for Agricultural Bioinformatics, ICAR-Indian Agricultural Statistics Research Institute, New Delhi 110012, India

[4] ICAR-Indian Institute of Millets Research, Rajendranagar, Hyderabad 500030, Telangana, India

[5] Department of Biotechnology, School of Interdisciplinary and Applied Sciences, Central University of Haryana, Mahendergarh 123031, Haryana, India

Introduction

Spotted stem borer, *Chilo partellus* (Crambidae: Lepidoptera) first described by Swinhoe in 1885, is one of the most widely distributed species of *Chilo* in the tropics. *Chilo partellus* is native of Asia, but in the 1950s it got established in East Africa, and since then has spread to southern and central Africa (IAPC 1985). Recently, it posed a serious risk of invasion in Americas, Australia, China, Europe, New Zealand, and West Africa due to congenial environment, host crops, and agronomic practices (Yonow et al., 2017). It is an established pest of maize, sorghum, sugarcane, and rice, and has also been recorded from small millets and wild grasses (CABI 2021), causing 18 to 25% yield losses in maize and sorghum in Asia and Africa (Dhaliwal et al. 2015). Further,

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

the presence, abundance, and intensity of infestation by *C. partellus* are influenced by both biotic and abiotic factors, wherein the morphological, behavioral, physiological, and bioecological variation makes it to survive under diverse and adverse agroecological conditions (Kfir et al., 2002; Dhillon et al. 2021a, b).

Several techniques such as sex pheromones, cultural control, intercropping and habitat management, manipulation of sowing dates, host plant resistance, and biological and chemical control have been explored for the management of *C. partellus*. However, these management techniques are highly influenced by various biotic and abiotic factors resulting in their inconsistent effectiveness in space and time. Over the past five decades, several sources of resistance have been identified in sorghum and maize against *C. partellus* (Kanta et al., 1997; Sharma et al. 2003, 2007; Rakshit et al. 2008; Dhillon and Gujar 2013); however, there has been limited success in development and deployment of resistant varieties against this pest. Genetically, diverse crop plants/species receive differential herbivory by the insects under different geographical and environmental conditions (Chown and Terblanche 2006; Sezonlin et al. 2006; Williams and Howells 2018). This results in change in resistant or susceptible reaction at different locations as a result of genotype × environment interactions, and/or existence of genetically diverse populations under different agroecological conditions (Beyeneet al., 2011; Rizvi et al. 2021; Dhillon et al. 2021b). The genome-wide identification and evolution of specific gene families have been done in woodland strawberry and cucumber (Haider et al., 2019, 2021).

Currently, a total of 1590 genome insect genome-sequencing projects are available in National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov). The i5K initiative, also known as "the Manhattan Project of Entomology," aimed at sequencing the genomes of 5,000 insects to revolutionize the understanding about insects in terms of health, food, and economic security (Robinson et al. 2011). Biological studies have been successful in understanding various aspects of *C. partellus* such as bioecology (Dhillon and Hasan 2017), phenological variation (Dhillon et al. 2021a), diapause (Dhillon et al. 2019), reproductive physiology (Dhillon and Hasan 2018), biochemical profiles (Dhillon et al. 2021b; Tanwar et al. 2021), and mating systems (Dhillon et al. 2020), including interaction with the host plants (Samal et al., 2021). However, there had been limited success in deciphering underlying mechanisms and finding novel solutions for the management of *C. partellus* due to non-availability of genomic resources. The whole genome sequencing could help in understanding the role of specific genes involved in production of heat- or cold-shock proteins to adapt to diverse and adverse environmental conditions. Such information is available only for a few insect species such as silkworm, *Bombyx mori* L. (Mita et al. 2004; Xia et al. 2004; Zhang et al. 2014); Antarctic midge, *Belgica antarctica* (Diptera: Chironomidae) (Kelley et al. 2014); and *Chilo suppressalis* (Walker) (Yin et al. 2014; Ma et al. 2020).

The biological experiments are good at understanding various dimensions of insect life; however, underlying mechanisms, their functions and the genes responsible can only be best understood by getting their genome decoded. Several lepidopteran insects have been targeted to generate their genomic information which is publicly available in their databases such as Manduca Base (http://agripestbase.org/manduca/), Heliconius homepage (http://www.heliconius.org), KAIKObase (Shimomura et al. 2009), SilkDB (Duan et al., 2010), MonarchBase (Zhan and Reppert 2013), KONAGAbase (Jouraku et al. 2013), DBM-DB (Tang et al. 2014), and ChiloDB (Yin et al. 2014). These have been used by the scientific community for various aspects of their life processes such as phylogeny, endocrine system, metabolism, diapause physiology, and novel insect pest management strategies. Though the genome study of *Chilo suppressalis* has been reported (Yin et al. 2014), *Chilo partellus* genome is yet to be sequenced. Besides this, not much information is available on olfactory genes, circadian genes, and olfactory receptor genes which are required in insect-pest management. Therefore, to generate the genomic resources, we undertook whole genome sequencing of C. partellus to get new insights into various biological processes, mechanisms of insecticide detoxification, insecticide resistance, discovery and synthesis of target site-specific insecticide molecules, and other appropriate strategies for controlling this pest for sustainable crop production.

Materials and methods

Insect source

We used three different cohorts of *C. partellus* larval samples obtained from three diverse sources, i.e., laboratory-established strain (sample 1: New Delhi: 28.6139° N, 77.2090° E; AMSL: 216 m), North India (sample 2: Hisar: 29°10′N; 75°46′E; AMSL 215.2 m), and South India (sample 3: Coimbatore: 11.0168° N, 76.9558° E; AMSL: 411 m). These three cohort larval samples had two biological replicates, thus making six test samples for whole genome sequencing.

Genomic DNA extraction, library preparation, and whole genome sequencing

The above mentioned six larval samples from sample 1, sample 2, and sample 3 were used to obtain DNA samples. Total genomic DNA was extracted with Invitogen PureLink

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Genomic DNA Mini Kit (Thermo Fisher Scientific, USA). The DNA quality was checked and quantified by NanoDrop and Qubit Florimeter, Thermo Scientific, USA. The DNA library was prepared using NEBNext® Ultra™ DNA Library Prep Kit for Illumina® (NEB #E7645) as per the standard protocols (New England Biolabs).

The whole genome sequencing of the genomic DNA samples of six *C. partellus* larval samples from sample 1, sample 2, and sample 3 was performed on Illumina HiSeq 2500 to get paired-end reads. It was done in Rapid Run mode, with mean read length of 150 bp. Further, 500 ng of genomic DNA of test samples was enzymatically fragmented by dsDNA Fragmentase. The library was prepared using the NEBNext Ultra II kit according to the manufacturer's instructions (New England Biolabs, USA).

Preprocessing and quality assessment of raw sequencing reads

The Illumina sequencing reads of *C. partellus* were quality checked using FASTQC v0.11.5 (Andrews 2010). The fastq files were pre-processed before performing assembly. Removal of contaminated reads was performed to get the error corrected reads. The low-quality bases with Phred5 quality score of less than Q30, and the adapter sequence contaminations in raw reads were removed using PRINSEQ v0.20.4 (Schmieder and Edwards 2011) and repaired the reads using BBmap v37.66 (https://sourceforge.net/projects/bbmap).

Genome assembly and quality assessment

De novo assembly were performed using ABySS (Simpson et al. 2009), MaSuRCA (Zimin et al., 2013), SOAPdenovo2 (Luo et al. 2012), and SPAdes (Bankevich et al., 2012) for sample 1, sample 2, and sample 3. The default k-mer sizes were used for MaSuRCA and SPAdes. A range of k-mers from 31 to 95 was used for SOAPdenovo2 assembly and ABySS assembly.

Further, the improved hybrid genome assembly was achieved by merging the above three best contiguity genome assemblies of sample 1, sample 2, and sample 3 by using the Metassembler v1.5. High N50 value and minimum number of contigs increased N50 values together with longer scaffolds contributed to improving the genome coverage (Wences and Schatz, 2015). The "gaps" and "Ns" caused by repeats were measured in the mis-assemblies. In the course of meta-assembly, we minimized the gaps and other sequencing errors. We employed Quast v5.0.2 (Gurevich et al. 2013) to gather extensive assembly statistics. BUSCO v5.0 (Simao et al., 2015) was employed for assessing the genome completeness by checking the presence of conserved genes in the assembled contigs.

Gene prediction and functional annotations

We predicted CDSs from the SPAdes and Metassembler assembled contigs using Augustus v3.2.3 (Stanke and Morgenstern, 2005). The predicted genes were annotated using the BLASTX search against the NCBI non-redundant protein database (NR) and Universal Protein (UniProt) (The UniProt Consortium 2017), Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al. 2021), and Gene Ontology (GO). For functional enrichment and biological interpretation of the genes, the GO and KEGG ontology terms for genes were mapped using the Blast2GO software (Conesa and Götz 2008). The significantly enriched pathways for genes in this study were determined by the Blast2GO software.

Identification of detoxification genes, circadian genes, and olfactory receptor genes and their sequence homology distribution

The genes related to insecticide action and detoxification were curated from various databases and repositories such as UniProt (https://www.uniprot.org) and NCBI database (https://www.ncbi.nlm.nih.gov), encoding detoxification metabolic enzymes namely cytochrome P450s (CYP450s) and glutathione S-transferases (GSTs), and the target receptors related to insecticide action, namely acetylcholinesterases (AChEs), nicotinic acetylcholine receptors (nAChRs), and gamma-aminobutyric acid (GABA receptors). The predicted genes were annotated using BLASTx within the non-redundant (NR) NCBI nucleotide database at the cut-off *e*-value of 0.05 by searching reference sequences. Figure 1 shows the schematic illustration of sequencing and analysis of the three different cohorts of these *C. partellus* larval samples.

Identification of SSR markers in *Chilo partellus* genome

The assembled genome was used for SSRs mining using MISA (Microsatellite Analysis) tool (http://pgrc.ipk-gatersleben.de/misa/). We identified mono- to hexa-nucleotides microsatellites using the criteria of at least 10 repeats for mono-nucleotide, 6 repeats for di-nucleotides, five repeats for tri- and tetra-nucleotide, and 4 repeats for penta- and hexa-nucleotide (Thiel et al. 2003).

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

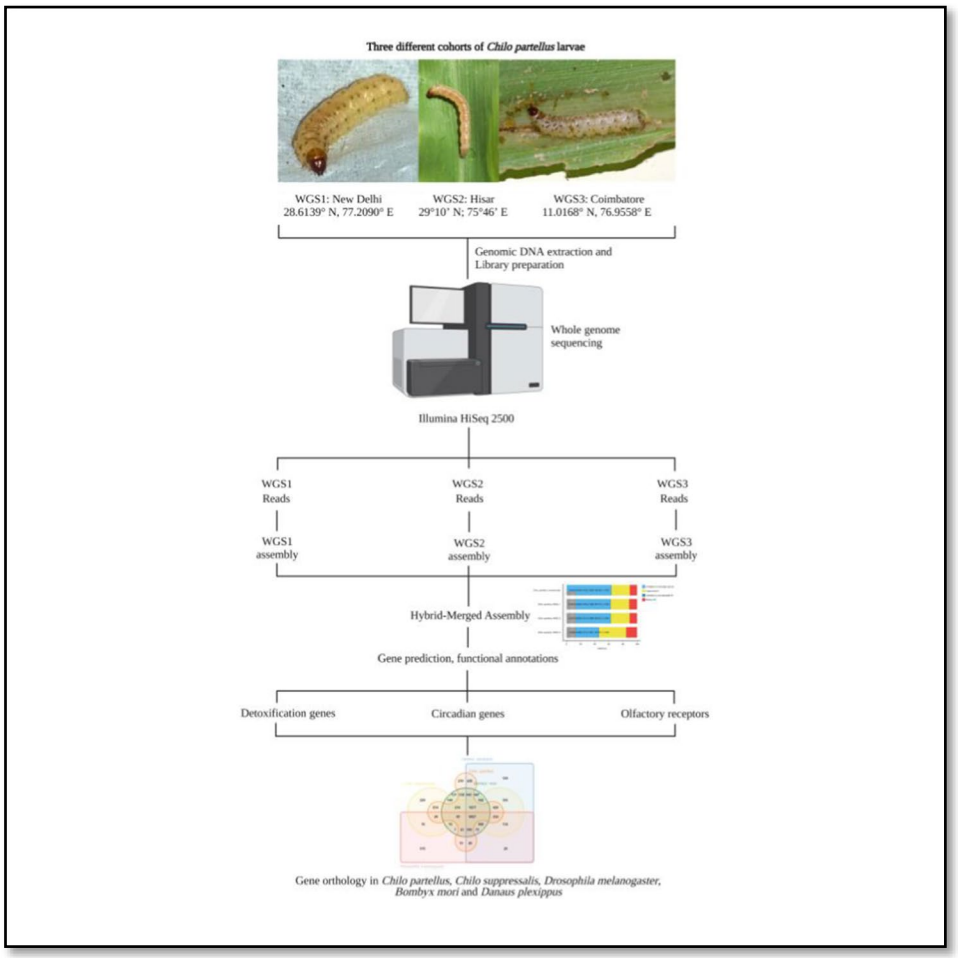

Fig. 1 Schematic illustration of the whole genome sequencing and analysis overview of the three different cohorts of *C. partellus* larval samples obtained from three diverse sources (WGS1= Sample 1; WGS2=Sample 2; WGS3=Sample 3)

Results

Preprocessing and quality assessment of raw sequencing reads

After quality check of Illumina reads of *C. partellus* using FASTQC v0.11.5, these were pre-processed, removing the low-quality bases with Phred5 quality score, $Q < 30$. After dropping 8.32%, 7.53%, and 6.87% of the reads from sample 1, sample 2, and sample 3, respectively, ~ 32 GB (91.68%), ~ 42 GB (92.47%), and 38 GB (93.13%) cleaned reads were considered for assembly. Summary of the raw sequencing and cleaned reads is provided in Table 1.

Table 1 Read statistics before and after pre-processing

Sample	Read orientation	Total number of reads	Total number of bases (Mb)	Total number of cleaned reads	Total number cleaned bases (Mb)
Sample 1	R1	35,013,914	5252.09	32,099,998	4233.83
	R2	35,013,914	5252.09	32,099,998	4233.83
Sample 2	R1	45,448,931	6817.34	42,025,151	5275.21
	R2	45,448,931	6817.34	42,025,151	5275.21
Sample 3	R1	40,940,595	6141.09	38,126,834	5,136.21
	R2	40,940,595	6141.09	38,126,834	5,136.21

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Genome assembly and quality assessment

Improved hybrid draft genome assembly through reconciliation algorithm

De novo assembly performed for sample 1, sample 2, and sample 3 using various assemblers showed the SPAdes assembly to perform better. So, for the further downstream analysis, SPAdes assembly was selected, since this had better contiguity than all other assemblies generated (Supplementary File 1).

The reconciliation assembly approach was employed in the present work to refine the incomplete draft genome assemblies of sample 1, sample 2, and sample 3. The Metassembler was employed for merging the three assemblies to get one hybrid, and a better assembly. Sample 1 consisted of 162,790 contigs (N50 3914), sample 2 146,416 contigs (N50 4225), and sample 3 323,106 contigs (N50 1839). After pooling these sequences, Metassembler was performed for hybrid assembly, resulting into 102,935 reads with the N50 value 4466 (Table 2). However, the improved draft version of assembly was generated by using an iterative merging strategy of Metassembler v1.5 by merging sample 1, sample 2, and sample 3 to obtain paired-end reads of *C. partellus*. For the parameters measured, default values suggested in manuals were used. The Metassembler implemented the reconciliation algorithm to refine and obtain the reconstructed genome.

Completeness of the improved hybrid assembly

BUSCO (Benchmarking Universal Single-Copy Orthologs) was employed for quantitative assessment of the assembly and its annotation (Simao et al. 2015). Of the total 1,013 BUSCO groups that were searched, the meta-assembly contained 645 complete (C), 639 complete and single-copy (S), 6 complete and duplicated (D), 262 fragmented (F), and 106 missing (M) BUSCOs. Comparatively sample 1, sample 2, and sample 3, assemblies were 62.58% (S: 61.99%, D: 0.59%, F: 26.25%, M: 11.15%), 62.88% (S: 62.19%, D: 0.69, F: 26.55%, M: 10.56%), and 45.90% (S: 45.21%, D: 0.69%, F: 38.59%, M: 15.49%) complete, respectively (Table 3). The gene completeness score as measured by BUSCO increased in the improved assembly, while the numbers of fragmented and missing BUSCO genes were reduced (Fig. 2). This genome comparison can be used to help such draft assemblies to obtain the finished genome assembly.

Gene prediction and functional annotations

The FGENESH module of the Molquest v.4.5 software package (Salamov and Solovyev 2000) and Augustus were employed, and 21,378 genes were predicted for the improved meta-assembly (Table 4). A number of predicted genes were found relatively lower in sample 1, sample 2, and sample 3 assemblies, but greater in the improved one. Gene identification from the Metassembler dataset against the *Rhodnius prolixus* Stäl as a model suggested 21,378

Table 2 Assembly statistics of genomes of three different populations of *C. partellus* and improved hybrid assembly

Assembly statistics	Sample 1	Sample 2	Sample 3	Improved hybrid assembly
Contigs (> = 0 bp)	162,790	146,416	323,106	102,935
Contigs (> = 1000 bp)	104,493	102,326	134,196	88,892
Contigs (> = 5000 bp)	18,285	19,975	4292	18,260
Contigs (> = 10,000 bp)	3523	3930	176	3519
Contigs (> = 25,000 bp)	77	88	0	77
Contigs (> = 50,000 bp)	1	1	0	1
Total length (> = 0 bp)	383,266,348	385,246,634	386,881,630	332,507,095
Total length (> = 1000 bp)	345,748,227	354,725,565	279,381,634	321,528,655
Total length (> = 5000 bp)	147,542,127	162,392,947	27,770,148	147,365,308
Total length (> = 10,000 bp)	47,556,389	53,480,132	2,095,132	47,499,098
Total length (> = 25,000 bp)	2,361,723	2,632,691	0	2,361,723
Total length (> = 50,000 bp)	52,319	59,923	0	52,319
Contigs	145,130	137,851	237,966	102,202
Largest contig	52,319	59,923	22,089	52,319
Total length	376,702,623	382,067,382	354,247,675	332,205,333
GC (%)	34.24	34.34	35.04	34.27
N50	3914	4225	1839	4466

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Table 3 BUSCO assembly statistics of *C. partellus* genome

BUSCO assembly statistics	Sample 1	Sample 2	Sample 3	Improved hybrid assembly
Complete BUSCOs (C)	634	637	465	645
Complete and single-copy BUSCOs (S)	628	630	458	639
Complete and duplicated BUSCOs (D)	6	7	7	6
Fragmented BUSCOs (F)	266	269	391	262
Missing BUSCOs (M)	113	107	157	106
Total BUSCO groups searched	1013	1013	1013	1013

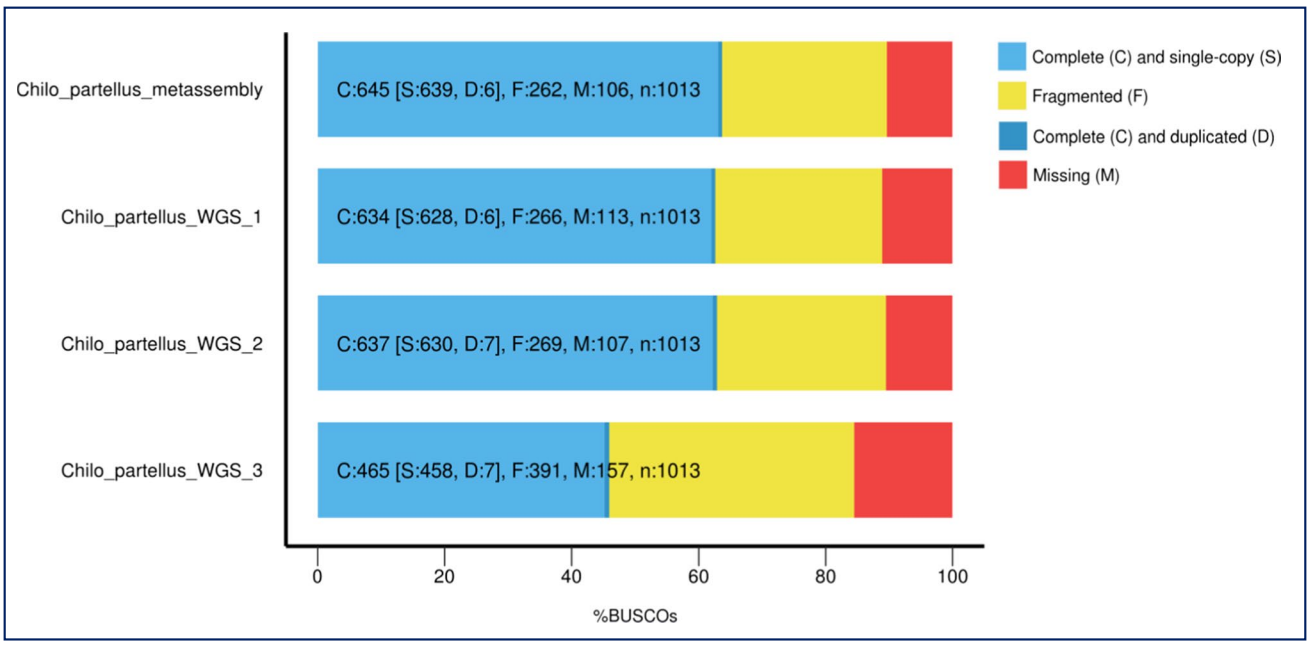

Fig. 2 Comparative gene completeness statistics using the BUSCO for all the three different *C. partellus* populations with the merged genome assembly

Table 4 Gene search from different populations and the improved draft genome of *C. partellus*

Parameter	Sample 1	Sample 2	Sample 3	Metassembler
No. of genes predicted	12,706	12,762	10,760	21,378

total genes in the improved version, but showed a lower number of genes with another dataset, which was included in merging and used to obtain the improved one (Mesquita et al. 2015). Variable number of predicted genes observed in the draft assemblies can be attributed to split genes and over estimation during gene finding. The over estimation of genes often results when fragmented single genes are present on multiple contigs or scaffolds. Improvements in gap filling and read mapping depth resulted in completeness of the number of genes in meta-assembly. The predicted total gene number was greater in meta-assembly than in other draft assemblies (Table 4).

Identification of detoxification genes, circadian genes, and olfactory receptor genes and their sequence homology distribution

Identification of detoxification genes

Insects are faced with numerous toxins (xenobiotics) as they go through the life cycle, some produced naturally by plants (allelochemicals), and some introduced by humans (insecticides). To survive the natural toxins, insects have evolved various detoxification mechanisms. These mechanisms allow the insects to withstand insecticides, and the level and the mode of action greatly differ. This results in differences

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

in susceptibility to insecticides in different populations of *C. partellus*. Knowledge of detoxification mechanisms is important to deploy different resistance mechanisms in crop plants, and to select effective insecticides for pest management under field conditions. Therefore, using blast search and filtering by searching for motifs using hmmsearch, which produce a full list of hmmsearch and HMMERHEAD command line options (Johnson et al. 2010), resulted in identification of 64 cytochrome P450 genes (CYP450s), and 36 glutathione S-transferases genes (GSTs), which are members of the major multigene enzyme families primarily responsible for xenobiotic metabolism, i.e., metabolism of insecticides and plant-toxic allelochemicals (Bao et al. 2012). In addition, insecticide targets such as 4 acetylcholinesterase (AChE), 14 γ-aminobutyric acid receptor (GABA), and 15 nicotinic acetylcholine receptor (nAChR) genes were also identified in *C. partellus* (Supplementary Table 1).

Identification of circadian genes and olfactory receptor genes

A total of 3,291 reference protein sequences of insect circadian genes were obtained from the UniProt database (https://www.uniprot.org) and CGDB (circadian genes a database) (http://cgdb.biocuckoo.org/index.php), which included ortholog gene species, *Drosophila melanogaster*, *Bombyx mori*, *Chilo suppressalis*, and *Danaus plexippus*. Manually annotated well-studied circadian genes for (1) period (PER), (2) timeless (TIM), (3) clock (CLK), (4) cycle (CYC), and (5) cryptochrome (CRY genes) were also identified by using conserved domains within proteins. We found a total of 183 (> 80% homolog) and 1541 (> 40% homolog) circadian genes from this analysis, which were manually annotated (Supplementary Table 2). Eleven olfactory receptor (OR) genes (Supplementary Table 2) were also found in the *C. partellus* genome against a 66 well-curated insect OR gene family (https://github.com/sdk15/insectOR) responsible for chemosensory processes such as locating food, shelter, mates, and oviposition sites.

Sequence homology distribution of detoxification, circadian and olfactory receptor genes

To annotate these genes, we searched reference sequences using BLASTx within the non-redundant (NR) NCBI nucleotide database using a cut-off *e*-value of 0.05. A total of 5,427 genes (25.38%) did not match any annotated sequences due to short nucleotide length. However, 15,951 genes (74.61%) displayed annotated BLASTx hits. The *e*-value distributions for the 15,951 annotated genes showed that 47.2% of the sequences had significant homology matches in the NCBI database. The similarity distribution showed that 10.1% of the sequences had greater than 80% homology, 57.42% of the sequences had above 60% homology, 21.67% of the sequences had above 40% homology, and 10.81% of the sequences had less than 40% homology. A species distribution analysis revealed that majority of the sequences closely resembled to other Lepidopterans, viz., C. *suppressalis*, followed by *Helicoverpa armigera* (Hubner), *Ostrinia furnacalis* (Guenée), and *Papilio xuthus* L. (Fig. 3; Supplementary Table 3). Approximately 66.09% of the genes had top homology matches with *C. suppressalis* (Supplementary Table 4).

Functional annotation of predicted genes

The identified genes were annotated by against NCBI non-redundant protein database, UniProt, KEGG, and GO. A total of 20,818 genes (42.53%) were successfully annotated, including 20,001 (40.87%) from the NR database, 13,980 (28.56%) from the UniProt database, 7,943 (16.23%) from the KEGG database, and 8,239 (16.83%) from the GO database. The GO distribution for the genes was classified into three categories (139 subcategories): biological processes (69 subcategories), molecular function (38 subcategories), and cellular components (32 subcategories) (Supplementary Table 5).

For the biological processes category, clusters were highly represented where categories relating to "biosynthetic process" were enriched for 3,869 genes, followed by "cellular nitrogen compound metabolic process" for 2,625 genes, "signal transduction" for 2,566 genes, "cellular protein modification process" for 2,559 genes, "DNA metabolic process" for 1,751 genes, "biological process" for 1,688 genes, "lipid metabolic process" for 1,196 genes, "response to stress" for 10,19 genes, "immune system process" for 104 genes, and "protein folding" for 97 genes. For the cellular components category, the clusters relating to "cellular component" had an enrichment of 8,565 genes, "nucleus" for 3,513 genes, and "protein-containing complex" for 2,220 genes. For the molecular functions category, clusters relating to "ion binding" were 8,678 genes, "molecular function" for 4,295 genes, "oxidoreductase activity" for 3,036 genes, "transmembrane transporter activity" for 2,570 genes, "peptidase activity" for 2,437 genes, "kinase activity" for 1,890 genes, "DNA binding" for 1,754 genes, "enzyme regulator activity" for 1,115 genes, and "enzyme binding" for 217 genes (Table 5; Fig. 4).

To further evaluate the annotation for identified detoxification gene, a total of 133 detoxification genes were enriched with 471 Gene Ontology terms. We examined the annotated genes that possessed the functional classifications. Out of total 133 total genes, the complete list of annotation with respect to protein families, PRINTS, PROSITE,

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

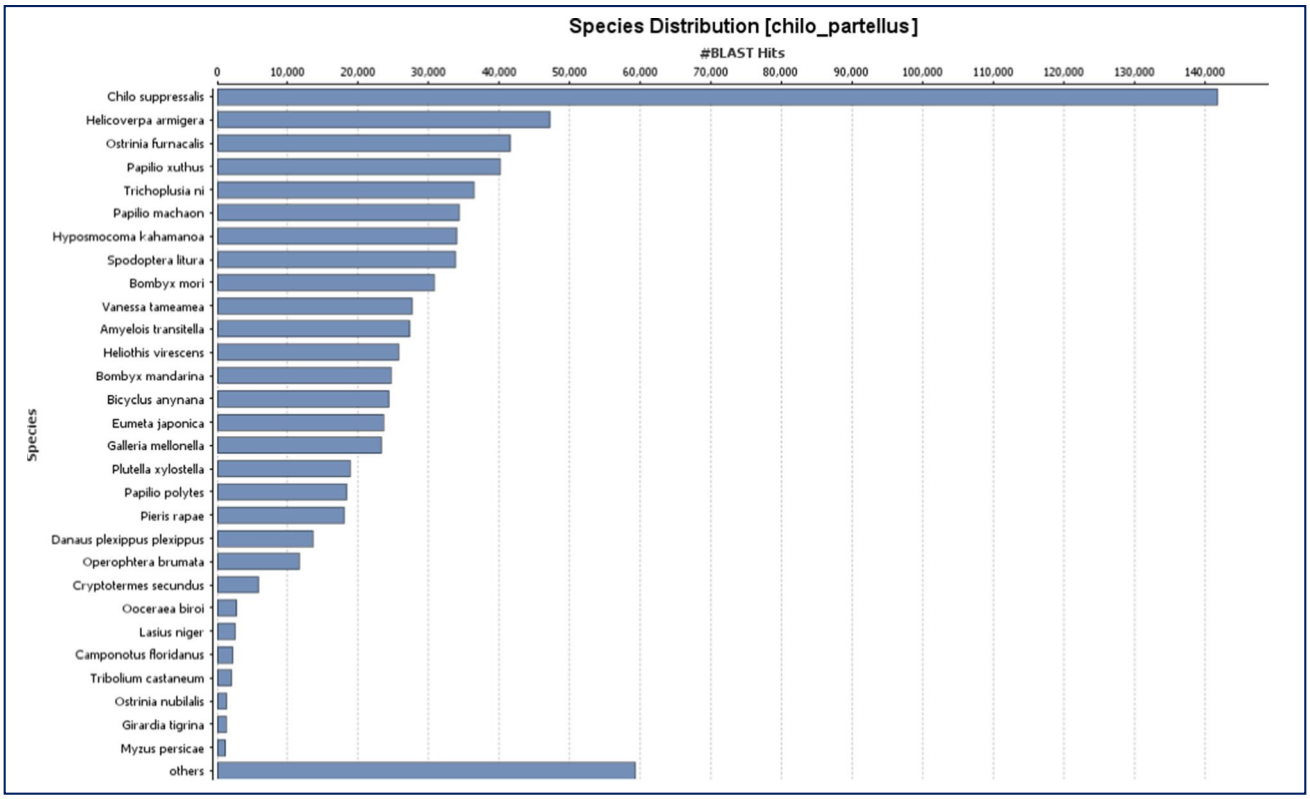

Fig. 3 Sequence homology distribution for *Chilo partellus* genes

Table 5 Gene Ontology terms identified in each category for identified detoxification genes

Category	Number of terms
Biological processes	43
Molecular functions	448
Cellular components	86

InterPro, Pfam, and PANTHER are listed in the Supplementary Table 1.

KEGG Ontology annotation-based functional enrichment and metabolic pathway analysis

According to the KEGG pathway analysis, we found 6,560 significantly enriched genes in pathways using the associated KOG (KEGG Ontology of Genes) annotation terms. Moreover, with these associated pathway IDs, 468 enzymes were found, and their effect on corresponding metabolism was also explored. The major gene-enriched pathways included "Purine metabolism, Thiamine metabolism" (28.80%), "Drug metabolism—other enzymes" (7.25%), "Glycerolipid metabolism" (3.35%), "Folate biosynthesis" (2.62%), "Porphyrin and chlorophyll metabolism" (2.39%), "Arginine and proline metabolism" (1.92%), and "Cutin, suberine, and wax biosynthesis" (1.92%). All extensive pathway maps have been provided in Supplementary Table 6.

Gene orthology–based analysis

The formations of orthologs are the key steps in finding gene evolution. We identified unique and shared gene families and proteomes in *C. partellus* and *C. suppressalis* (GenBank Accession: GCA_004000445.1). We detected 90.60% shared gene orthology between *C. partellus* and *C. suppressalis* (Fig. 5). Further comparison of a contiguous region of the species revealed that 5.76% and 3.63% of the sequences were not shared between *C. suppressalis* and *C. partellus*. The species form 9,365 clusters, 1,634 orthologous clusters, and 7,731 single-copy gene clusters. Comparison of proteomes revealed that 8,485 gene clusters were common for *C. partellus* and *C. suppressalis* (Supplementary Table 7).

Identification of SSR markers in Chilo partellus genome

A total of 63,816 SSRs were identified in the 102,935 sequences of assembled genome. Out of which, 41,036 were mononucleotides, 14,378 dinucleotides, 5,941 trinucleotides,

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

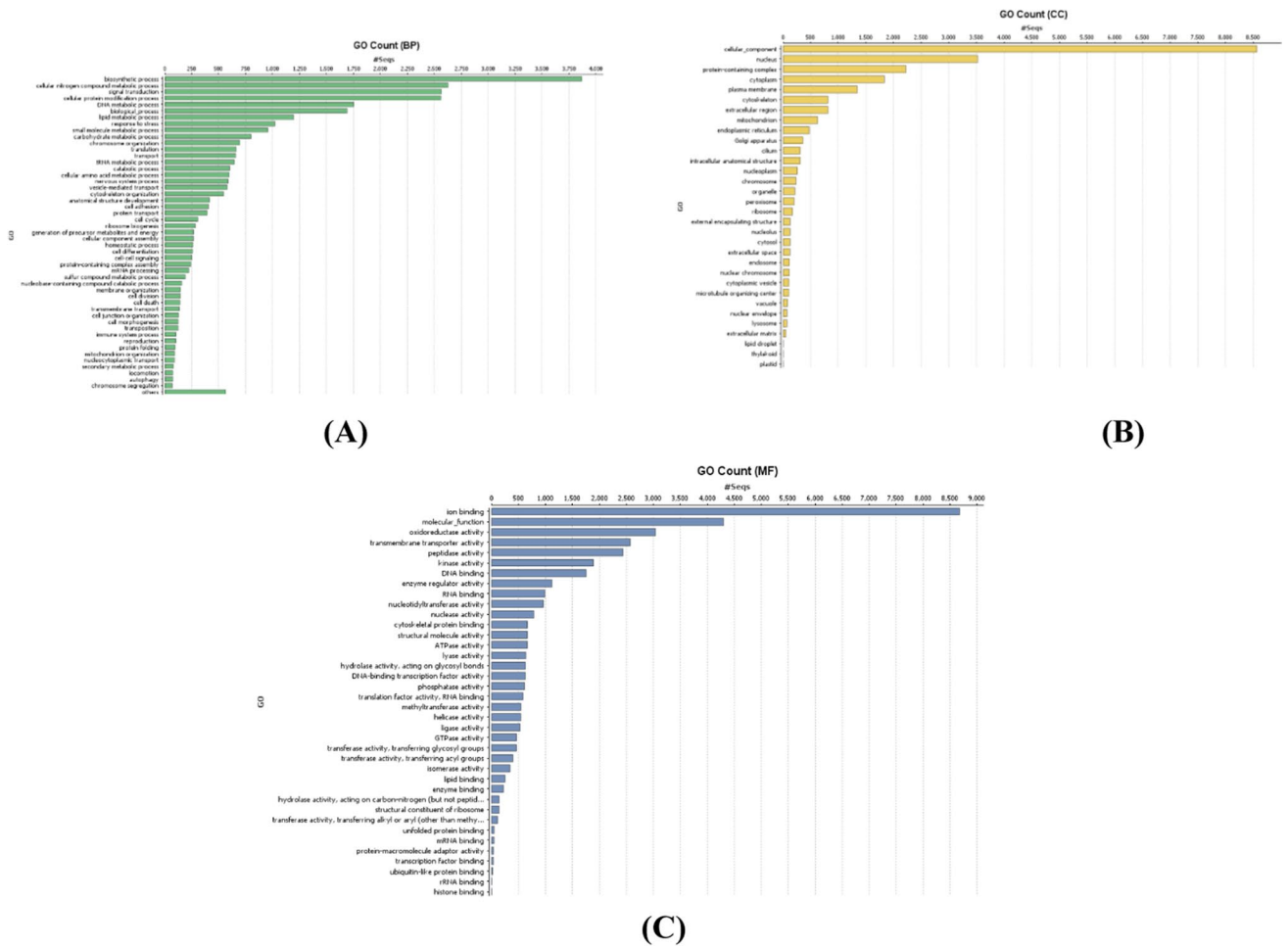

Fig. 4 GO distribution of *C. partellus* predicted genes. **A** Biological process. **B** Cellular component. **C** Molecular function

2,215 tetranucleotides, 196 pentanucleotides, and 50 hexanucleotides (Table 6, Fig. 6). Among the 63,816 microsatellites, A/T-type (97.55%) microsatellites were most common in the genome, followed by AC/GT (45.45%), and AAT/ATT (63.03%) AAAC/GTTT (36.97%), AAATG/ATTTC (21.93%), and AAGGAT/ATCCTT (28%) in mononucleotides, dinucleotides, trinucleotides, tetranucleotides, pentanucleotides, and hexanucleotides, respectively. The details about all the identified SSRs for frequency of identified SSR motifs and frequency of classified repeat types have been provided in Supplementary Table 8.

Discussion

The draft genome assembly of spotted stem borer, *C. partellus*, reveals its genome size ≈332.2 Mb, which is larger than genomes of *D. melanogaster*-180 Mb (Adams et al. 2000), *Danaus plexippus*-248 Mb (Zhan et al. 2011), *Heliconius melpomene* (L)-266 Mb (Dasmahapatra et al. 2012), and *Papilio polytes* L-227 Mb (Nishikawa et al. 2015), and smaller than other important lepidopteran insect pests of maize like *Busseola fusca* (Fuller)-492.9 Mb (Hardwick et al. 2019), *C. suppressalis*-824 Mb (Yin et al. 2014), and *Spodoptera frugiperda* (J.E. Smith)-371 Mb (Gouin et al. 2017). The differences in the genome size of different insects could be due to number of repetitive elements in the genome and other factors like physiology, range of host plants, and environmental adaptation (Xue et al., 2014). The GC content of *C. partellus* genome (34.27%) was almost similar to other important insect pests such as *Nilaparvata lugens* (Stål) (37.59%) (Xue et al. 2014), *Laodelphax striatellus* (Fallén) (34.5%) (Zhu et al. 2017), *Sogatella furcifera* (Horváth) (31.6%) (Evans and Gundersen-Rindal 2003), and *S. frugiperda* (32.97%) (Kakumani et al. 2014). Furthermore, the constitution of draft genome assembly of *C. partellus* resulted in 102,935 sequences with N50 length of 4466, had 645 complete and 639 complete and single-copy orthologs, and a total of 21,378 genes analogous to *R. prolixus*, which can be

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

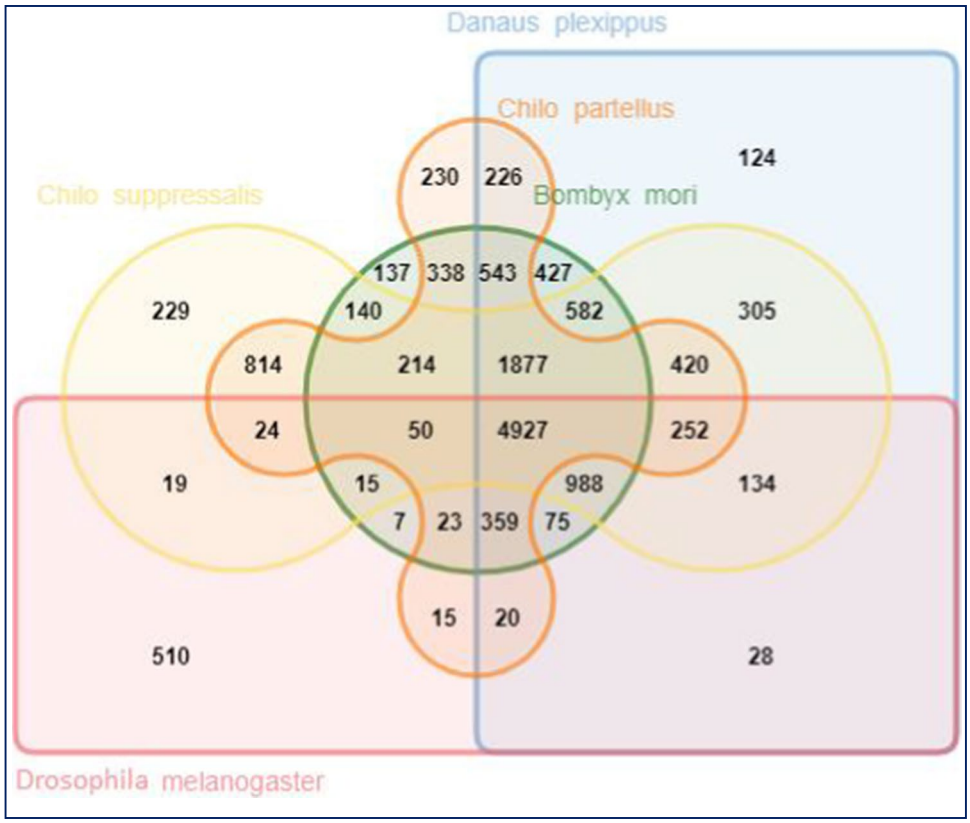

Fig. 5 Venn diagram (Edwards plot) showing the overlap of orthologous genes found within five species, viz., *Chilo partellus*, *Chilo suppressalis*, *Drosophila melanogaster*, *Bombyx mori*, and *Danaus plexippus*

helpful in obtaining the finished genome assembly. Since this is the first maiden effort to sequence the *C. partellus* genome, the genome assembly coverage is comparatively lower than these discussed insects.

The gene sequence homology found 15,951 genes in *C. partellus* with 74.61% of the sequences closely resembling to other lepidopterans, viz., *C. suppressalis*, *H. armigera*, *O. furnacalis*, and *P. xuthus*, having top homology with *C. suppressalis* genes (66.09%). All these compared pest insects are plant borers/foliage feeders and are common in many aspects like visual, chemosensory, feeding behavior, digestive machinery, and metamorphosis, while differ in their host preference. Furthermore, most of the insects including these resembling lepidopterans have evolved strategies for survival under adverse environmental conditions by entering diapause. *Chilo suppressalis*, the most resembling insect species in terms of gene sequence homology to *C. partellus*, undergo facultative diapause at the onset of autumn (Inoue and Kamano 1957), while the other maize borers like *Diatraea grandiosella* Dyar and *B. fusca* undergo facultative summer diapause under tropical and subtropical conditions (Usua 1973; Kikukawa and Chippendale 1983). Unlike these maize borers, *C. partellus* undergo both winter and summer diapause in the larval stage (Dhillon et al. 2019), and could be having other additional untapped genes ascribing to dual diapause which are not captured in the present draft genome assembly of *C. partellus*. *H. armigera* has also been reported to undergo both winter and summer diapause, but in the pupal stage (Wilson et al. 1979; Wu and Guo 1995; Liu et al. 2006).

The insect behavioral activities which regulate key biological events like locomotion, courtship, mating behavior, seasonal adaptations, egg-laying, and photoperiodism

Table 6 The microsatellites (SSRs) identified from assembled genome of *C. partellus*

Total number of sequences examined	102,935
Total size of examined sequences (bp)	332,507,095
Total number of identified SSRs	63,816
Number of SSR containing sequences	41,655
Number of sequences containing more than 1 SSR	14,487
Number of SSRs present in compound formation	3726
Distribution to different repeat type classes	
Unit size	Number of SSRs
Monomer	41,036
Dimer	14,378
Trimer	5941
Tetramer	2215
Pentamer	196
Hexamer	50

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

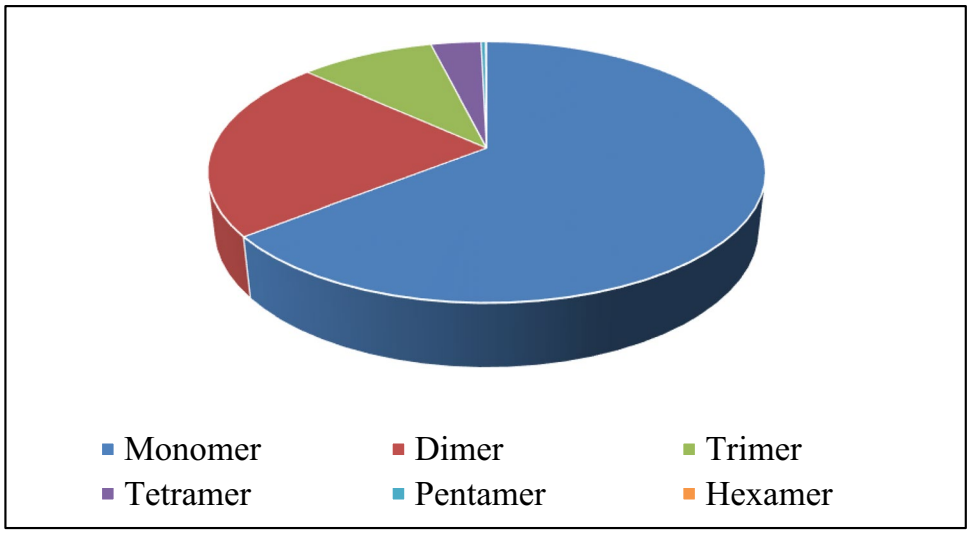

Fig. 6 Graphical distribution of SSR-motif types

mainly depend on the circadian rhythms (Saunders 2002). Further, the olfactory receptor genes help in perceiving the chemical cues, while circadian genes play an important role in staggering the biological activities to prevailing situation and adaption to adverse environmental conditions. A total of 3,291 reference protein sequences of insect circadian genes orthologous to *D. melanogaster*, *B. mori*, *C. suppressalis*, and *D. plexippus* were identified, of which 183 genes have > 80% homology and 1541 genes with > 40% homology circadian genes including CRY genes along with 11 olfactory receptor genes in the *C. partellus* genome, which could be responsible for chemosensory processes, varying reaction to diverse host plants and genotypes, and adaptation under diverse agro-ecological conditions. The presence of circadian rhythm genes like CRY genes in *C. partellus* like that in other lepidopterans including *C. suppressalis* (Kattupalli et al., 2021) indicates the correctness of our draft genome assembly. Furthermore, the phylogenies are important for addressing various biological questions such as relationships among species as well as genes, the origin and spread of the pest, and demographic changes among different *Chilo* species. Comparison of proteomes in the present study revealed 8,485 gene clusters with 90.60% shared gene orthology, and a unique and shared gene (GenBank Accession: GCA_004000445.1) between *C. partellus* and *C. suppressalis*. Further, a total of 63,816 microsatellites were also identified in the present assembled genome, from which reliable SSRs could be designed for their deployment in deciphering the population structure of *C. partellus* under diverse agro-ecologies. The SSRs mined can be of much use for population structure and diversity analysis. They can also be used for pedigree construction as well as DNA signature for sub-species level. Such information are required in insect management.

The functional gene annotation of *C. partellus* predicted genes and detoxification genes successfully annotated a total of 20,818 genes (42.53%) encompassing biological processes, molecular function, and cellular components, of which 133 were functional detoxification genes. The KEGG metabolic pathway analysis found 6,560 significantly enriched genes encompassing major gene-enriched pathways, which indicate that the combination of the detoxification genes with metabolic pathway genes could play a significant role in xenobiotic detoxification in *C. partellus*. Although the frequency of the metabolic pathway genes in the present hybrid assembly was low, more such gene families and functional genes could be identified once the complete genome assembly of *C. partellus* is available. Further, the insects have evolved specialized detoxification mechanisms, assisted by several detoxifying genes such as ABC transporters, cytochrome P450, cytochrome P450 monooxygenase, glutathione S transferase, carboxylesterase, cholinesterase, and UDP-glycosyl transferase, to deal with xenobiotic defense and to withstand insecticides and/or display varying levels of susceptibility to the insecticides. The CYP450 genes have been reported to play role in inactivation of endogenous toxic compounds in several insects (Ono et al. 2006; Després et al. 2007). Like in *C. partellus*, other insects have also been found to harbor several xenobiotic detoxifying genes like 130 CYP450 and 6 GSTs in *C. suppressalis*, nine GST genes in *S. furcifera*, and 13 GST genes have been identified in *N. lugens* (Zhou et al. 2013), which have the potential to detoxify both endogenous and xenobiotic compounds. The identified detoxification genes may also play a role in residual detoxification by bioweapon catalytic scavengers (Paidi et al. 2021). From the present draft genome assembly of *C. partellus*, we found 64 CYP450s, 36GSTs, 4 AChE, 14 GABA, and 15 nAChR genes, which could

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

be useful in understanding the detoxification mechanism and xenobiotic metabolism and designing target-specific insecticides to control the ecologically diverse *C. partellus* populations.

Conclusions

This is the first report of whole genome sequencing of *C. partellus* providing significant information on one of the devastating pests of coarse cereals, sorghum, and maize, *C. partellus*. The genes related to mode of action of insecticides and the detoxification mechanisms were identified, including 64 cytochrome P450 genes (CYP450s), 36 glutathione S-transferases (GSTs), 4 acetylcholinesterase (AChE), 14 γ-aminobutyric acid receptor (GABA), and 15 nicotinic acetylcholine receptor (nAChR) genes. This information will be useful for deployment of selective insecticides for the control of *C. partellus* for sustainable crop production. The discovered SSR markers can be used for population structure, diversity analysis, and DNA signature for sub-species level. These findings have immense potential use in insect-pest management in endeavor of better crop productivity.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10142-022-00852-w.

Funding This study was financially supported by the National Agricultural Science Fund, Indian Council of Agricultural Research, New Delhi, India (NASF/ABP-5017/2016–17), and Indian Council of Agricultural Research, Ministry of Agriculture and Farmers' Welfare, Govt. of India (CABin grant F. no. Agril. Edn.4–1/2013-A&P).

Data availability Whole genome sequencing of spotted stem borer, *Chilo partellus*, reveals multiple genes encoding enzymes for detoxification of insecticides (BioProject: PRJNA778570; Biosample: SAMN22965316, SAMN22965317, and SAMN22965318 with SRA ID: SRR16896023: laboratory-established strain of *Chilo partellus*; SRR16896022: North India strain of *Chilo partellus*; SRR16896021: South India strain of *Chilo partellus*).

Declarations

Conflict of interest The authors declare no competing interests.

References

Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD et al (2000) The genome sequence of *Drosophila melanogaster*. Science 287:2185–2195

Andrews S (2010) FASTQC. A quality control tool for high throughput sequence data

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Comput Mol Cell Biol 19(5):455–477

Bao Y-Y, Wang Y, Wu J-W, Zhao D, Xue J, Zhang B-Q, Shen Z-C, Zhang C-X (2012) De novo intestine-specific transcriptome of the brown planthopper Nilaparvata lugens revealed potential functions in digestion, detoxification and immune response. Genomics 99(4): 256–264

Beyene Y, Mugo S, Mutinda C, Tefera T, Karaya H, Ajanga S, Shuma J, Tende R, Kega V (2011) Genotype by environment interactions and yield stability of stem borer resistant maize hybrids in Kenya. Afr J Biotech 10(23):4752–4758

CABI (2021) Invasive Species Compendium. Wallingford, UK: CAB International. www.cabi.org/isc

Chown SL, Terblanche JS (2006) Physiological diversity in insects: ecological and evolutionary contexts. Advances in Insect Physiology 33:50–152

Conesa A, Götz S (2008) Blast2GO: A comprehensive suite for functional analysis in plant genomics. Int J Plant Genom 2008:619832

Dasmahapatra KK, Walters JR, Briscoe AD, Davey JW, Whibley A, Nadeau NJ, Zimin AV, Salazar C, Ferguson LC, Martin SH et al (2012) Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature 487:94

Després L, David JP, Gallet C (2007) The evolutionary ecology of insect resistance to plant chemicals. Trends Ecol Evol 22:298–307

Dhaliwal GS, Jindal V, Mohindru B (2015) Crop losses due to insect pests: Global and Indian scenario. Indian J Entomol 77(2):165–168

Dhillon MK, Hasan F (2018) Consequences of diapause on post-diapause development, reproductive physiology and population growth of Chilo partellus (Swinhoe). Physiol Entomol 43:196–206

Dhillon MK, Hasan F, Tanwar AK, Bhadauriya AS (2019) Factors responsible for aestivation in spotted stem borer, Chilo partellus (Swinhoe). J Exp Zool A 331(6):326–340

Dhillon MK, Gujar GT (2013) Maize genotypes identified with resistance to spotted stem borer, *Chilopartellus*andfavourable agronomic traits. Annals of Plant Protection Sciences 21:224–228

Dhillon MK, Hasan F (2017) Temperature-dependent development of diapausing larvae of *Chilo partellus* (Swinhoe) (Lepidoptera: Crambidae). J Therm Biol 69:213–220

Dhillon MK, Hasan F, Tanwar AK, Jaba J, Singh N, Sharma HC (2020) Genetic regulation of diapause and associated traits in *Chilo partellus* (Swinhoe). Sci Rep 10 (1), 1793 (1-11).https://doi.org/10.1038/s41598-020-58640-0

Dhillon MK, Tanwar AK, Kumar S, Hasan F, Sharma S, Jaba J, Sharma HC (2021b) Biological and biochemical diversity in different biotypes of spotted stem borer, *Chilo partellus* (Swinhoe) in India. Sci Rep 11:5735. https://doi.org/10.1038/s41598-021-85457-2

Dhillon MK, Tanwar AK, Hasan F, Bhadauriya AS (2021a) Phenotypic variation in adults of *Chilo partellus* (Swinhoe) from diverse ecological regions of. India. Acta Zoologica 00:1–11. (Published online first). https://doi.org/10.1111/azo.12377

Duan J, Li R, Cheng D et al (2010) SilkDB v2.0: a platform for silkworm (*Bombyx mori*) genome biology. Nucleic Acids Res 38:D453–D456

Evans JD, Gundersen-Rindal D (2003) Beenomes to Bombyx: future directions in applied insect genomics. Genome Biol 4:1–4

Gouin A, Bretaudeau A, Nam K, Gimenez S, Aury JM, Duvic B, Hilliou F, Durand N, Montagné N, Darboux I et al (2017) Two genomes of highly polyphagous lepidopteran pests (*Spodoptera frugiperda*, Noctuidae) with different host-plant ranges. Sci Rep 7:1–12

Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Haider MS, Khan N, Pervaiz T, Zhongjie L, Nasim M, Jogaiah S, Mushtaq N, Jiu S, Jinggui F (2019) Genome-wide identification, evolution, and molecular characterization of the PP2C gene family in woodland strawberry. Gene 702:27–35

Haider MS, Britto SD, Nagaraj G, Gurulingaiah B, Shekhar R, Ito S, Jogaiah S (2021) Genome-wide identification, diversification, and expression analysis of lectin receptor-like kinase (LecRLK) gene family in cucumber under biotic stress. Int J Mol Sci 22(12):6585

Hardwick KM, Ojwang AME, Stomeo F, Maina S, Bichang'a, G., Calatayud, P.A. et al (2019) Draft genome of *Busseola fusca*, the maize stalk borer, a major crop pest in Sub-Saharan Africa. Genome Biol Evol 11(8):2203–2207

Inter-African Phytosanitary Council (IAPC) (1985) *Chilo partellus* (Swinhoe). Distribution maps of major crop pests and diseases in Africa, Map No. 199. Inter-African Phytosanitary Council, Yaounde, Cameroon

Inoue T, Kamano S (1957) The effect of photoperiod and temperature on the induction of diapause in the rice stem borer, *Chilo suppressalis* Walker. Japanese Journal of Applied Entomology and Zoology 1:100–105

Johnson LS, Eddy SR, Portugaly E (2010) Hidden Markov model speed heuristic and iterative HMM search procedure. BMC Bioinformatics 11:431

Jouraku A, Yamamoto K, Kuwazaki S et al (2013) KONAGAbase: a genomic and transcriptomic database for the diamond back moth *Plutella Xylostella*. BMC Genomics 14:464

Kakumani PK, Malhotra P, Mukherjee SK, Bhatnagar RK (2014) A draft genome assembly of the armyworm, *Spodoptera frugiperda*. Genome 104:134–143

Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M (2021) KEGG: integrating viruses and cellular organisms. Nucleic Acids Res 49(D1):D545–D551

Kanta, U., Dhillon, B.S. and Sekhon, S.S. (1997). Evaluation and development of maize germplasm for resistance to spotted stem borer. In: Mihm JA (Ed) Insect Resistant Maize: Recent Advances and Utilization. Proceedings of an International Symposium CIMMYT, Mexico, pp 246–254

Kattupalli D, Barbadikar KM, Balija V, Ballichatla S, R, A., Padmakumari, A.P., Saxena, S., Gaikwad, K., Yerram, S., Kokku, P. et al (2021) The draft genome of yellow stem borer, an agriculturally important pest, provides molecular insights into its biology, development and specificity towards rice for infestation. Insects 12:563

Kelley JL, Peyton JT, Fiston-Lavier AS, Teets NM, Yee MC, Johnston JS,…, Denlinger DL (2014) Compact genome of the Antarctic midge is likely an adaptation to an extreme environment. Nature Communications *12*(5), 4611.https://doi.org/10.1038/ncomms5611

Kfir R, Overholt WA, Khan ZR, Polaszek A (2002) Biology andmanagement of economically important lepidopteran cereal stemborers in Africa. Annu Rev Entomol 47:701–731

Kikukawa S, Chippendale GM (1983) Seasonal adaptation of populations of the southwestern corn borer, *Diatraea grandiosella* from tropical and temperate regions. J Insect Physiol 29:561–567

Liu ZD, Gong PY, Wu KJ, Sun JH, Li DM (2006) A true summer diapause induced by high temperatures in the cotton bollworm, *Helicoverpa armigera* (Lepidoptera: Noctuidae). J Insect Physiol 52:1012–1020

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu SM, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam TW, Wang J (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience 1(1):18

Ma W, Zhao X, Yin C, Jiang F, Du X, Chen T, Zhang Q, Qiu L, Xu H, Joe Hull J, Li G, Sung W, Li F, Lin Y (2020) A chromosome-level genome assembly reveals the genetic basis of cold tolerance in a notorious rice insect pest, Chilo suppressalis. Mol Ecol Resour 20:268–282. https://doi.org/10.1111/1755-0998.13078

Mita K, Kasahara M, Sasaki S, Nagayasu Y, Yamada T, Kanamori H,…, Sasaki T (2004) The genome sequence of silkworm, *Bombyx mori*. DNA Res 11:27–35. https://doi.org/10.1093/dnares/11.1.27

Nishikawa H, Iijima T, Kajitani R, Yamaguchi J, Ando T, Suzuki Y, Sugano S, Fujiyama A, Kosugi S, Hirakawa H et al (2015) A genetic mechanism for female-limited Batesian mimicry in Papilio butterfly. Nat Genet 47:405–409

Ono H, Rewitz KF, Shinoda T, Itoyama K, Petryk A, Rybczynski R, Jarcho M, Warren JT, Marqués G, Shimell MJ et al (2006) Spook and Spookier code for stage-specific components of the ecdysone biosynthetic pathway in Diptera. Dev Biol 298:555–570

Paidi MK, Satapute P, Haider MS, Udikeri SS, Ramachandra YL, Vo DN, Govarthanan M, Jogaiah S (2021) Mitigation of organophosphorus insecticides from environment: Residual detoxification by bioweapon catalytic scavengers. Environ Res 200:111368

Rakshit S, Kaul J, Dass S, Singh R, Sekhar JC, Singh SB (2008) Catalogue of Indian maize inbred lines. *Technical Bulletin No.* 2008/3. Directorate of Maize Research, Pusa Campus, New Delhi 110 012, India. 40 pp.

Rizvi SAH, George J, Reddy GVP, Zeng X, Guerrero A (2021) Latest developments in insect sex pheromone research and its application in agricultural pest management. Insects 12(6):484. https://doi.org/10.3390/insects12060484

Robinson GE, Hackett KJ, Purcell-Miramontes M, Brown SJ, Evans JD, Goldsmith MR, Lawson D, Okamuro J, Robertson HM, Schneider DJ (2011) Creating a buzz about insect genomes. Science 331:1386

Salamov A, Solovyev V (2000) Ab initio gene finding in Drosophila genomic DNA. Genome Res 10:516–522

Samal I, Dhillon MK, Tanwar AK, Kumar S, Hasan F (2021) Biological performance and amino acid profiles of different geographical *Chilo partellus* populations on diverse maize genotypes. *Entomologia Generalis* (In press). https://doi.org/10.1127/entomologia/2021/1203

Saunders DS (2002) Insect clocks. Elsevier, Amsterdam, The Netherlands

Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. Bioinformatics 27(6):863–864

Sezonlin M, Dupas S, Le Rü B, Le Gall P, Moyal P, Calatayud PA, Giffard I, Faure N, Silvain JF (2006) Phylogeography and population genetics of the maize stalk borer *Busseola fusca* (Lepidoptera, Noctuidae) in sub-Saharan Africa. Mol Ecol 15(2):407–420

Sharma HC, Dhillon MK, Pampapathy G, Reddy BVS (2007) Inheritance of resistance to spotted stem borer, *Chilo partellus* in sorghum, *Sorghum bicolor*. Euphytica 156:117–128

Sharma HC, Taneja SL, Kameswara Rao N, Prasada Rao KE (2003) Evaluation of sorghum germplasm for resistance to insect pests. *Information Bulletin No.* 63. International Crops Research Institute for the Semi-Arid Tropics, Patancheru 502 324, Andhra Pradesh, India. 177 pp

Shimomura M, Minami H, Suetsugu Y et al (2009) KAIKObase: an integrated silkworm genome database and datamining tool. BMC Genomics 10:486

Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212

Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I (2009) ABySS: a parallel assembler for short read sequence data. Genome Res 19(6):1117–1123

Stanke M, Morgenstern B (2005) AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints.

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Nucleic acids research, 33(Web Server issue), W465–W467. https://doi.org/10.1093/nar/gki458

Tang W, Yu L, He W, Yang G, Ke F, Baxter SW, You S, Douglas CJ, You M (2014) DBM-DB: the diamond back moth genome database. Database (Oxford), bat087. https://doi.org/10.1093/database/bat087

Tanwar AK, Kirti JS, Kumar S, Dhillon MK (2021) The amino acid and lipophilic profiles of *Chilo partellus* (Swinhoe) larvae fluctuates with diapause. Journal of Experimental Zoology-a: Ecological and Integrative Physiology 335(7):595–601

The UniProt Consortium (2017) UniProt: the universal protein knowledgebase. Nucleic Acids Res 45(D1):D158–D169

Thiel T, Michalek W, Varshney R, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). Theor Appl Genet 106:411–422

Usua EJ (1973) Induction of diapause in the maize stemborer, *Busseola fusca*. Entomol Exp Appl 16:322–328

Williams RS, Howells JM (2018) Effects of intraspecific genetic variation and prior herbivory in an old-field plant on the abundance of the specialist aphid *Uroleuconnigro tuberculatum* (Hemiptera: Aphididae). Environ Entomol 47:422–431

Wilson AGL, Lewis T, Cunningham RB (1979) Overwintering and spring emergence of *Heliothis armigera* (Hübner) (Lepidoptera: Noctuidae) in the Namoi Valley, New South Wales. Bull Entomol Res 69:97–109

Wu KM, Guo YY (1995) Diapause inducing factors of cotton bollworm, *Helicoverpa armigera* (Hübner). Acta Phytotaxonomica Sinica 22:331–336

Xia QY, Zhou ZY, Lu C, Cheng DJ, Dai FY, Li B et al (2004) A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). Science 306:1937–1940

Xue J, Zhou X, Zhang CX, Yu LL, Fan HW, Wang Z, Xu HJ, Xi Y, Zhu ZR, Zhou WW et al (2014) Genomes of the rice pest brown planthopper and its endosymbionts reveal complex complementary contributions for host adaptation. Genome Biol 15:1–20

Yin C, Liu Y, Liu J, Xiao H, Huang S, Lin Y, Han Z, Li F (2014) ChiloDB: a genomic and transcriptome database for an important rice insect pest *Chilo suppressalis*. *Database*, 2014, 1–7. bau065. https://doi.org/10.1093/database/bau065

Yonow T, Kriticos DJ, Ota N, Van Den Berg J, Hutchison WD (2017) The potential global distribution of *Chilo partellus*, including consideration of irrigation and cropping patterns. J Pest Sci 90(2):459–477

Zhan S, Reppert SM (2013) MonarchBase: the monarchbutterfly genome database. Nucleic Acids Res 41:D758–D763

Zhan S, Merlin C, Boore JL, Reppert SM (2011) The monarch butterfly genome yields insights into long-distance migration. Cell 147:1171–1185

Zhang Z, Yin CL, Liu Y, Jie WC, Lei WJ, Li F (2014). iPathCons and iPathDB: an improved insect pathway construction tool and the database. Database-Oxford, *2014*, bau105. https://doi.org/10.1093/database/bau105.

Zhou WW, Liang QM, Xu Y, Gurr GM, Bao YY, Zhou XP, Zhang CX, Cheng J, Zhu ZR (2013) Genomic insights into the glutathione S-transferase gene family of two rice planthoppers, *Nilaparvata lugens* (Stål) and *Sogatella furcifera* (Horváth) (Hemiptera: Delphacidae). PLoS ONE 8:e56604

Zhu J, Jiang F, Wang X, Yang P, Bao Y, Zhao W, Wang W, Lu H, Wang Q, Cui N et al (2017) Genome sequence of the small brown planthopper *Laodelphax striatellus*. Gigascience 6:gix109

Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA (2013) The MaSuRCA genome assembler. Bioinformatics 29(21):2669–2677

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

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).
Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.
These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.
We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.
While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.
These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.
To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.
Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.
If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com