The oriental fruit moth genome provides insights into evolutionary adaptation of tortricid pests

Wei Song^{1,2, #}, Li-Jun Cao^{1, #}, Lei-Yue¹, Shao-Kun Guo¹, Jin-Cui Chen¹, Ya-Jun Gong¹, Xu-Lei Fan¹, Shu-Jun Wei^{1, *}

1. Institute of Plant and Environmental Protection, Beijing Academy of Agriculture and Forestry Sciences, 9 Shuguanghuayuan Middle Road, Haidian District, Beijing 100097, China
2. Beijing Key Laboratory for Forest Pests Control, Beijing Forestry University, Beijing, 100083, China

songw0513@163.com

gmatjhpl@163.com

nongda-yuelel@163.com

mscgsk@163.com

chenjincui1314@126.com

gongyajun200303@163.com

xulei201210@163.com

shujun268@163.com

* Corresponding authors:

Shu-Jun Wei, Institute of Plant and Environmental Protection, Beijing Academy of Agriculture and Forestry Sciences, 9 Shuguanghuayuan Middle Road, Haidian District, Beijing 100097, China; Tel: +86 10 51503439; E-mail: shujun268@163.com

Short title: *Grapholita molesta* genome

Abstract

Moths of the family Tortricidae (Insecta: Lepidoptera) usually distribute in temperate and tropical high upland regions. The oriental fruit moth (OFM) *Grapholita molesta* is a globally important pest of stone and pome fruit from this family. In this study, we assembled a chromosome-level genome for the OFM and conducted a comparative genomic analysis with other lepidopterans. This genome was assembled to 28 chromosomes with a size of 517.71 Mb, an N50 of 19.5 Mb, and a BUSCO completeness of 97.1%. In total, 19968 protein-coding genes were predicted, among which 15269 were functionally annotated. We manually annotated ten gene families for 13 representative genomes from the Lepidopteran. In general, two tortricid moths have a moderate number of detoxification and receptor genes and the lowest number of HSP genes, congruent with their polyphagous dietary and pattern of distribution. However, the OFM has the highest number of P450, IRs and ORs among all species. Compared to the other tortricid species of the codling moth *Cydia pomonella*, the OFM has more genes in all gene families. Our results indicate that the high number of some detoxification and receptor genes may be related to the strong adaptation of OFM as a global pest. The high-quality genome of OFM provides an invaluable resource for understanding the ecology, genetics, and evolution of tortricid moths.

Keywords: Insect; genome; detoxification; receptor; fruit borer; adaptation

Introduction

The Tortricidae is one of the largest families of Lepidoptera with over 5000 described species, including many economically important pests in agriculture and forestry (van der Geest & Evenhuis, 1991). Tortricid moths cause damages to their hosts in larval stages by feeding on the leaves, fruits, buds, shoots, twigs, and stems. Moths of the family Tortricidae tend to distribute in temperate and tropical high upland regions (van der Geest & Evenhuis, 1991). However, some species expanded their distribution range as global invasive species (Kirk et al., 2013b; Wan et al., 2019).

The oriental fruit moth (OFM) *Grapholita molesta* is a cosmopolitan pest of stone and pome fruits from the Tortricidae (Neven et al., 2018; Quaintance & Wood, 1916; Torriani et al., 2010). Larvae of the OFM cause damage by boring into twigs as well as fruits (Figure 1). Population genetic analysis shows that this moth is native to East Asia and has spread into other stone-fruit growing continents in the past century (Kirk et al., 2013a; Song et al., 2018; Wei et al., 2015). Although many methods have been applied to counter outbreaks and spread of OFM, this remains a challenging work for both of the farmers and researchers (Audemard et al., 1989; Il'ichev et al., 2007; Rica et al., 2006; Rothschild, 1979; Sarker & Lim, 2018). Understanding evolutionary adaption underlying the worldwide distribution may provide insights into the management of this pest. Current studies identified pheromone gland gene, receptor genes, thioredoxin genes, detoxification genes, microRNAs from the OFM from transcriptome data to explore the possible mechanisms related to pest control (Chen et al., 2019a; Guo et al., 2017; Shen et al., 2018; Vatanparast & Kim, 2019; Wang et al., 2017). However, due to the lack of genomic resources, the molecular and genetic study on the OFM was largely delayed. Song et al. (2017) assembled a draft genome for the OFM based on sequences generated by the Illumina Miseq platform. However, because of the high fragmentation of this draft assembly, it was only applied for the development of microsatellite markers (Song et al., 2018; Song et al., 2017). A high-quality genome is essential to understand the evolutionary adaptation and develop new control approaches of OFM.

In the study, we report a chromosome-level genome of OFM. This genome was de novo assembled based on sequences obtained from the PacBio and Illumina platforms and assembled at the chromosome level assisted by the Hi-C technique. We compared genome features between OFM and another cosmopolitan pest from Tortricidae, the codling moth *Cydia pomonella* (Lepidoptera: Tortricidae) (Wan et al., 2019), to explore the evolutionary adaptation of tortricid moths. The genomic resource developed here provides an invaluable resource for understanding the genetics, ecology, and evolution of moths.

Materials and methods

Sample preparation, DNA libraries construction and sequencing

Laboratory reared strains of the OFM were used for de novo genome sequencing. The OFM strain derived from three pairs of males and females has been maintained for about 10 generations on apples in the Beijing Academy of Agriculture and Forestry Sciences. Genomic

DNA was extracted using the DNeasy tissue kit (Qiagen, Hilden, Germany) for Illumina libraries and using the MagAttract HMW DNA kit (Qiagen, Hilden, Germany) for NanoPore library construction. The Illumina libraries were sequenced using the Illumina HiSeq X Ten platform (Illumina, CA, USA) and the NanoPore library was sequenced using the Oxford NanoPore platform (Pacific Biosciences, CA, USA) (Table S1).

The Hi-C proximity ligation library was constructed following Berkum et al. (2010). The genome was digested by restriction enzyme Mbol. Fragments were sheared into 200-600 bp. The fragment ends were added A-tailing by Klenow (exo-) and then added Illumina paired-end sequencing adapter. The libraries were amplified by 12 cycles PCR and sequenced in the Illumina HiSeq X10 platform.

Transcriptome sequence

In total, 4 RNA-seq libraries were constructed for eggs, larvae, pupae and adults of the OFM (Table S1). All libraries were prepared using VAHTSTM mRNA-seq V2 Library Prep Kit for Illumina according to the manufacturer's instructions (Vazyme, Nanjing, China) and sequenced using the Illumina NovaSeq platform. The RNA-seq reads were de novo assembled and mapped to the OFM genome using TopHat (Trapnell et al., 2009) with default parameters. The assembly results of RNA-seq were used as evidence for genome structure annotation. After removing the low quality reads with Trimmomatic v0.38 (Bolger et al., 2014), the reads were mapped to the chromosome-level genome using Hisat v2.2.0 (Kim et al., 2019) and assembled with StringTie v2.1.2 (Pertea et al., 2015). FPKM (Fragments Per Kilobase per Million) values of each annotated gene in each RNA-seq were estimated with cufflinks v2.2.1 (Kim et al., 2013).

Genome size estimation, assembly, and quality assessment

The raw reads generated from the Illumina platform were filtered by Trimmomatic v0.38 (Bolger et al., 2014). To estimate the genome size of the OFM, we used paired-end sequencing reads from Illumina sequencing library to determine the distribution of K-mers. Genome size, heterozygosity and rate of duplication were estimated using the online GenomeScope v1.0 software (Vurture et al., 2017).

Long reads generated from the NanoPore platform were corrected and assembled using CANU version v1.8 (Koren et al., 2017) with default parameters. The assembled contigs were polished based on Illumina short reads using Pilon v1.22 (Walker et al., 2014). To remove the possible secondary alleles, the assembled contigs were filtered by using the pipeline Purge Haplotigs (Roach et al., 2018), resulting in a contig-level genome. The Illumina short reads sequenced from the Hi-C library were used to assemble these contigs into a chromosome-level genome using the Juicer v1.5 (Neva C. et al., 2016) and 3D de novo assembly (3D-DNA) pipeline (Dudchenko et al., 2017) (Table 1).

Completeness of each assembled version of the genome was assessed using a Benchmarking Universal Single-Copy Orthologs (BUSCO) v3.0.2 (Simao et al., 2015) analysis, based on the insecta_odb9 database (1,658 genes). We conducted a synteny analysis between OFM and another tortricid moth, the codling moth *Cydia pomonella* (Assembly accession: GCA_003425675.2) (Wan et al., 2019) and *Spodoptera litura* (Assembly accession:

GCF_002706865.1) (Cheng et al., 2017) using TBtools v0.58 (Chen et al., 2020) to check the quality of the chromosome-level genome. Short contigs in *S. litura* and *C. pomonella* genomes were removed before the blast alignment.

Repeat elements and non-coding RNAs annotation

Repeats and transposable element families in the OFM genome were first detected by RepeatMasker pipeline v4.0.7 (Tarailo-Graovac & Chen, 2009) against the Insecta repeats within RepBase Update (<http://www.girinst.org>) and Dfam database (20170127), with RMBlast v2.10.0 as a search engine, which employs complementary computational methods to build and classify consensus models of putative repeats. tRNAs were annotated by tRNAscan-SE (Lowe & Eddy, 1997) with default parameters. rRNAs were annotated by RNAmmer prediction (Lagesen et al., 2007).

Protein-coding gene annotation

The protein-coding gene in the OFM genome was structurally annotated using *ab initio*, RNA-seq-based, and homolog-based methods. First, the RNA-seq reads were aligned to the reference genome using PASA (Brian J et al., 2003) to extract the complete genes for SNAP model training using SNAP v2013-02-16 (Korf, 2004). Second, a combination of de novo prediction using SNAP and Augustus v3.2.3 (Stanke & Waack, 2003), homolog-based prediction using protein-coding genes of *Drosophila melanogaster* and *Bombyx mori*, and RNA-seq-based prediction using transcripts of OFM was conducted using Maker v3.01.03 genome annotation pipeline (Cantarel et al., 2008). For Augustus analysis, the gene model from BUSCO analysis of chromosome-level assembly was used. Third, the gene set annotated by the Maker pipeline was filtered by keeping those with an FPKM (fragments per kilobase of exon model per million reads mapped) value > 0 in any RNA-seq library. Finally, PASA was used to update the annotation results with transcripts of OFM; all predictions were further filtered using GffRead v0.11.7 implemented in Cufflinks v2.2.1 (Kim et al., 2013) to remove without or with multiple stop codons.

Functions of the protein-coding genes were annotated using eggNOG-Mapper v1.0.3 (Huerta-Cepas et al., 2017) against the database EggNOG v5.0 (Huerta-Cepas et al., 2019). Genes that can be functionally annotated by EGGNOG analysis were retained in gene structure annotation and used for further analysis.

Orthology identification and gene family expansion/contraction analysis

Protein-coding genes from available genomes of two Coleoptera, two Diptera and other 11 Lepidoptera were retrieved from the NCBI genomes database and published references for comparative analysis. Orthologs were identified using OrthoFinder v2.2.7 (Emms & Kelly, 2015) under default parameters. MAFFT v7.450 (Katoh & Standley, 2013) was employed to align amino acid sequences of 1:1:1 orthologous gene with the G-INS-I algorithm. The phylogenetic tree was inferred using an approximately-maximum-likelihood method implemented in FastTree v2.1.10 (Price et al., 2009), which is a default phylogenetic inference method for genomic data in OrthoFinder. We used r8s (Sanderson, 2003) to estimate the divergent times among species with divergence times of two nodes, i.e. *Tribolium castaneum* and *Anoplophora glabripennis* (Wang et al., 2019), *Trichoplusia ni* and *S.*

litura (Wan et al., 2019) as calibrations. The Computational Analysis of gene Family Evolution (CAFE) version 3.1 (Bie et al., 2006) was used to analyze gene family expansion and contraction. The orthologue groups that present or absent in only one organism were defined as species-specific or lost orthologues. The gene sets groups with multiple-copy displayed in one organism were defined as specific expansions.

Gene family annotation

Adaption to heat stresses usually relies on Heat shock proteins (HSP) (Bedulina et al., 2013; Colinet et al., 2013), while host plant adaptation usually employs detoxification genes to detoxify plant secondary metabolites in insects. There are five major detoxification gene families in animals, i.e., cytochrome P450 monooxygenases (P450s), glutathione-S-transferases (GSTs), ATP-binding cassette transporters (ABCs), UDP-glycosyltransferases (UGTs) and carboxyl/cholinesterases (CCEs). Chemical senses to olfaction and taste have been found to play important roles in insects to find mates, food, and oviposition sites, and to avoid harmful situations and non-host habitats (Hansson & Stensmyr, 2011). Four gene families are often involved in chemical senses, including olfactory receptor (ORs), gustatory receptor (GRs), Ionotropic receptors (IRs) and Odorant-binding proteins (OBPs). Here, we manually annotated and compared those gene families among 13 representative genomes of Lepidoptera, to explore the possible genomic signatures related to the environmental adaption of the OFM as well as the other tortricid moth, *C. pomonella*.

We used both model-based and similarity-based methods to annotate the gene families. For model-based identification, the Hidden Markov models (HMMs) were downloaded from Pfam 32.0 database (September 2018; (El-Gebali et al., 2018)) and run with HMMER v3.3 (Finn et al., 2011). The corresponding HMM model not found in Pfam database was manually trained using HMMER under the default parameters. For similarity-based identification, orthologs from *D. melanogaster*, *B. mori*, *Aedes aegypti*, *Anopheles gambiae* and *C. pomonella* search against target genomes using BLAST v2.2.31 (Altschul et al., 1990) with an e-value cutoff of $1e^{-5}$. An automatic BITACORA v1.0 (Vizueta et al., 2020) pipeline (full mode) was used to conduct the HMMER and BLAST analyses. The annotated genes were filtered manually based on gene length and the presence of conserved domains by removing genes shorter than 80 amino acids and those lack of conserved domains.

Results and discussion

Genome sequencing and assembly

By integrating both Illumina HiSeq short-read sequencing and NanoPore long-read sequencing, we generated 102.43 Gb of clean read data, representing 197.85-fold coverage of the assembled 517 Mb genome (Table S1).

Based on Illumina short sequencing reads, the estimated size of OFM genome ranges from 393 Mb to 439 Mb; heterozygosity ranges from 0.721% to 0.865%; duplication rate ranges from 1.55% to 1.73% when K value was set to 17, 21, 27 and 31 (Figure S1). The heterozygosity of OFM is lower than the *Thrips palmi* (1.01% to 1.32%) (Guo et al., 2020) and the caddisfly *Stenopsyche tienmushanensis* (1.05%-1.10%) (Luo et al., 2018), higher than the

beet armyworm, *Spodoptera exigua* (0.59%) (Zhang et al., 2019), and similar to the invading fall webworm (0.75% and 0.83%) (Wu et al., 2018).

The total length of the CANU assembly was 949.05 Mb, with an N50 equal to 850.95 Kb. After removing the possible secondary alleles, the size of the assembly was 521.6 Mb with an N50 length of 1.79Mb. Then, the contigs were polished into contig-level sequences using the short-reads again and generated a 524.93 Mb-length genome. These contigs were assembled into 28 chromosomes using Hi-C assembly, with a total length of 525.22 Mb and an N50 length of 19.5 Mb (Figure 2a). The chromosomal-level genome was finally polished by Pilon to obtain a total length of 517.71 Mb and an N50 about 19.5 Mb (Table 1).

To validate the quality of the assembled genome, we analyze the BUSCO completeness of each version of the OFM genome assembly. The completeness rises from 76.6 % to 97.1 %, with 96.1 %, 1.0 %, 1.0 %, and 1.9 % of genes were single-copy, duplicated, fragmented, and missing for the final genome assembly, respectively (Table 1). The quality of the assembly was much improved in each step, both in N50 size and BUSCO completeness. Our study showed that subsequent analyses of secondary allele removing and polish should be incorporated after assembly by traditional assembly methods.

This is the second chromosomal-level genome from the family Tortricidae following *C. pomonella* (Wan et al., 2019). Compared to the contig N50 of currently assembled genomes from the Lepidoptera (0.013 to 0.86 Mb) (Wan et al., 2019), the quality of contig level assembly is the highest in the OFM (1.79 Mb). The BUSCO completeness of chromosome-level assemblies for the lepidopteran species ranges from 91.5% to 98.5% (Wan et al., 2019), the completeness of OFM assembly is among the high-quality ones (97.1%). These results suggest that we have developed a high-efficiency genome assemble pipeline and the OFM genome assembly was of high quality in this study.

Synteny and chromosome fusion in the Tortricidae

Powerful syntenic blocks were found between the OFM and *C. pomonella* genome as well as *S. litura* (Figure 2b). Every chromosome was one-to-one synteny between OFM and *C. pomonella* genome. In *C. pomonella* genome, there are 28 autosomes and one sex chromosome W. However, we have not identified the sex chromosome W (chrom 29 in *C. pomonella*) in this OFM genome. The number of chromosomes in the OFM is the same to *C. pomonella*. In Tortricidae, the number of chromosomes ranges from 27 to 31 (SuomMainen, 1969). The subfamily Tortricinae has 30 chromosomes, while the Olethreutinae, which includes the OFM and *Cydia pomonella*, has 28 chromosomes (van der Geest & Evenhuis, 1991). The results indicate we have successfully assembled all chromosomes except for the sex chromosome W in the OFM.

Fusion event between the sex chromosome Z and an autosome corresponding to chromosome 15 in the *Bombyx mori* reference genome was reported in the *C. pomonella* (Fuková et al., 2005; Wan et al., 2019). This phenomenon is also found in the OFM genome, confirmed the common fusion event in tortricid moths (Nguyen et al., 2013). Two additional fusion events were found in the OFM, as reported in the *C. pomonella* (Wan et al., 2019). The first one is the fusion of chromosome 7 and chromosome 8 of *S. litura* into

Chromosome 2 of OFM; the second is the fusion of chromosome 5 and chromosome 22 of *S. litura* into chromosome 4 of OFM (Figure 2b).

Protein-coding, non-coding RNA genes and repetitive elements

We annotated 39274 protein-coding genes using the MAKER pipeline, of which 19478 were supported by transcriptome data. We saved 19968 genes when the final PASA pipeline was finished, with the completeness of 92.6% estimated by BUSCO. Fifteen thousand two hundred sixty-nine gene sets were functionally identified in the genome of OFM using EGGNOG (Table 1).

We predicted 61 8s_rRNAs, 63 5s_rRNA and 29786 tRNAs (including 8853 tRNAs decoding standard 20 amino acid, 14 selenocysteine tRNAs (TCA), 13 possible suppressor tRNAs (CTA, TTA), 163 tRNAs with undetermined/unknown isotypes and 20743 predicted pseudogenes, in which 203 tRNAs with intron) in the OFM genome (Table S2).

A total of 15.37 Mb repetitive elements occupying 11.01% of the OFM genome were annotated using a combination of structural information and homology prediction (Table S3). Retrotransposable elements, known to be the dominant form of repeats, constituted a large part of the genome and included the most abundant subtypes, such as long terminal repeat elements (LTRs), long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs). We predicted 41362559 bp retroelements (including 4947739 bp SINEs, 33924345 bp LINEs and 2490475 bp LTRs) in OFM genome based on Insecta RepBase libraries, which grouped in 231642 elements and occupied 7.99% of the assembly. We annotated 8670 DNA transposons (1730875 bp, 0.33 %), 19652 small RNA (4014531 bp, 0.78 %) and 135317 Tandem Repeats (TRs) (2 satellites and 135315 simple repeats). Nineteen thousand two hundred eighteen low complexity repeat elements occupying 0.18 % of the whole genome were identified (Table S3).

Comparative genome analysis and phylogenetic relationships

A total of 320445 genes among 339228 genes from the 17 species were assigned to 14425 orthologue groups. We found 1785 genes in 427 orthologs were species-specific and there were 4948 orthologs present in all species. The G50 (The number of genes in the orthogroup such that 50% of genes are in orthogroups of that size or larger) and O50 (The smallest number of orthogroups such that 50% of genes are in orthogroups of that size or larger) of all genes was 28 and 3286 respectively (Table S4). For OFM, there were 19011 genes identified and 18258 genes occupying 96% were assigned in orthologs. A total of 117 gene families were species-specific and clustered into 13 orthogroups.

Phylogenetic analysis supports the monophyly of all involved superfamilies of Lepidopteran. Molecular dating support that OFM and *C. pomonella* diverged at 85.48 million years ago (Mya) (Figure 3). There were 883 gene family expansion/contractions in OFM genome, similar to *C. pomonella* (928) as well as most other moths and butterflies (Table S5).

Detoxification, receptor and HSP genes

In general, two tortricid moths, i.e. OFM and *C. pomonella* have moderate number of detoxification and receptor genes and the lowest number of HSP genes (Table 2), congruent

with their polyphagous dietary and pattern of distribution (van der Geest & Evenhuis, 1991). The OFM has the highest number of P450, IRs and ORs among all species. Compared to the other tortricid species of the codling moth *Cydia pomonella*, the OFM has more genes in all gene families. Our results indicate that the high number of some detoxification and receptor genes may be related to the strong adaptation of OFM as a global pest.

Conclusions

We assembled a genome for the OFM, which allowed us to compare genomic changes in the evolution of Lepidoptera insects. Overall, the OFM genome provides a useful resource for understanding the genetic basis of traits that underlie the ecology of budworms and the evolutionary divergence and convergence of these insects more generally. This genomic resource may also ultimately be useful for the management of OFM, such as through the identification of novel targets for chemical control and resistance monitoring.

Data Availability Statement

The whole genome assembly has been deposited in the Genome repository (accession numbers: CP053120-CP053147) under NCBI BioProject PRJNA627114.

Author contributions

Shu-Jun Wei conceived and designed the study; Li-Jun Cao, Wei Song, Lei-Yue, Shao-Kun Guo, Jin-Cui Chen, Ya-Jun Gong conducted the experiments; Wei Song, Li-Jun Cao and Shu-Jun Wei analyzed the data; Shu-Jun Wei, Wei Song and Li-Jun Cao discussed the results and wrote the manuscript.

Acknowledgments

We thank Qiang Gao for his help in the assembly of Hi-C data. This research was supported by the National Natural Science Foundation of China (31901884), Joint Laboratory of Pest Control Research Between China and Australia (Beijing Municipal Science & Technology Commission), and Beijing Key Laboratory of Environmentally Friendly Pest Management on Northern Fruits (BZ0432).

Reference

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403-410. doi:10.1016/S0022-2836(05)80360-2
- Audemard, H., Leblon, C., Neumann, U., & Marboutie, G. (1989). Evaluation of a 7 year control experimentation on the oriental fruit moth *Cydia molesta* Busck (Lep, Tortricidae) by the mating disruption technique. *Journal of Applied Entomology-Zeitschrift Fur Angewandte Entomologie*, 108(2), 191-207.
- Bedulina, D. S., Evgen'ev, M. B., Timofeyev, M. A., Protopopova, M. V., Garbuz, D. G., Pavlichenko, V. V., . . . Zatssepina, O. G. (2013). Expression patterns and organization of the hsp70 genes correlate with thermotolerance in two congener endemic amphipod species (*Eulimnogammarus cyaneus* and *E. verrucosus*) from Lake Baikal. *Molecular Ecology*, 22(5), 1416-1430. doi:10.1111/mec.12136
- Berkum, N. L. V., Lieberman-Aiden, E., Williams, L., Imakaev, M., & Lander, E. S. (2010). Hi-C: A method to study the three-dimensional architecture of genomes. *Journal of Visualized Experiments Jove*, 39(39), 1-7.

- Bie, T. D., Cristianini, N., Demuth, J. P., & Hahn, M. W. (2006). CAFE: a computational tool for the study of gene family evolution. *Bioinformatics*, 22(10), 1269-1271. doi:10.1093/bioinformatics/btl097
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114-2120. doi:10.1093/bioinformatics/btu170
- Brian J, H., Arthur L, D., Stephen M, M., Jennifer R, W., Roger K, S., Linda I, H., . . . Owen, W. (2003). Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Research*, 31(19), 5654-5666.
- Cantarel, B. L., Korf, I., Robb, S. M., Parra, G., Ross, E., Moore, B., . . . Yandell, M. (2008). MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research*, 18(1), 188-196. doi:10.1101/gr.6743907
- Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., & Xia, R. (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Molecular Plant*, doi.org/10.1016/j.molp.2020.1006.1009. doi:10.1101/289660 %J bioRxiv
- Chen, L. H., Tian, K., Wang, G. R., Xu, X. L., He, K. H., Liu, W., & Wu, J. X. (2019a). The general odorant receptor GmolOR9 from *Grapholita molesta* (Lepidoptera: Tortricidae) is mainly tuned to eight host-plant volatiles. *Insect Science*. doi:10.1111/1744-7917.12725
- Chen, W., Yang, X., Tetreau, G., Song, X., Coutu, C., Hegedus, D., . . . Wang, P. (2019b). A high-quality chromosome-level genome assembly of a generalist herbivore, *Trichoplusia ni*. *Molecular Ecology Resources*, 19(2), 485-496. doi:10.1111/1755-0998.12966
- Cheng, T., Wu, J., Wu, Y., Chilukuri, R. V., Huang, L., Yamamoto, K., . . . Mita, K. (2017). Genomic adaptation to polyphagy and insecticides in a major East Asian noctuid pest. *Nature Ecology & Evolution*, 1(11), 1747-1756. doi:10.1038/s41559-017-0314-4
- Colinet, H., Siaussat, D., Bozzolan, F., & Bowler, K. (2013). Rapid decline of cold tolerance at young age is associated with expression of stress genes in *Drosophila melanogaster*. *The Journal of Experimental Biology*, 216(Pt 2), 253-259. doi:10.1242/jeb.076216
- Dudchenko, Olga, Batra, Sanjit, S., Omer, . . . Sarah. (2017). De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science*, 356, 92-95.
- El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., . . . Smart, A. (2018). The Pfam protein families database in 2019. *Nucleic Acids Research*, 47(D1), D427-D432.
- Emms, D. M., & Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, 16(157), 1-14. doi:10.1186/s13059-015-0721-2
- Finn, R. D., Clements, J., & Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research*, 39(Web Server issue), W29-W37. doi:10.1093/nar/gkr367
- Fuková, I., Nguyen, P., & Marec, F. e. (2005). Codling moth cytogenetics: karyotype, chromosomal location of rDNA, and molecular differentiation of sex chromosomes. *Genome*, 48(6), 1083-1092. doi:10.1139/g05-063
- Guo, S. K., Cao, L. J., Song, W., Shi, P., Gao, Y. F., Gong, Y. J., . . . Wei, S. J. (2020). Chromosome-level assembly of the melon thrips genome yields insights into evolution of a sap-sucking lifestyle and pesticide resistance. *Molecular Ecology Resources*, 10.1111/1755-0998.13189. doi:10.1111/1755-0998.13189

- Guo, Y., Chai, Y., Zhang, L., Zhao, Z., Gao, L. L., & Ma, R. (2017). Transcriptome analysis and identification of major detoxification gene families and insecticide targets in *Grapholita molesta* (Busck) (Lepidoptera: Tortricidae). *Journal of Insect Science*, 17(2), 1-14. doi:10.1093/jisesa/iex014
- Hansson, B. S., & Stensmyr, M. C. (2011). Evolution of insect olfaction. *Neuron*, 72(5), 698-711.
- Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., von Mering, C., & Bork, P. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. *Molecular Biology and Evolution*, 34(8), 2115-2122. doi:10.1093/molbev/msx148 %J Molecular Biology and Evolution
- Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernandez-Plaza, A., Forslund, S. K., Cook, H., . . . Bork, P. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, 47(D1), D309-D314. doi:10.1093/nar/gky1085
- Il'ichev, A. L., Williams, D. G., & Gut, L. J. (2007). Dual pheromone dispenser for combined control of codling moth *Cydia pomonella* L. and oriental fruit moth *Grapholita molesta* (Busck) (Lep., Tortricidae) in pears. *Journal of Applied Entomology*, 131(5), 368-376.
- Kanost, M. R., Arrese, E. L., Cao, X., Chen, Y. R., Chellapilla, S., Goldsmith, M. R., . . . Blissard, G. W. (2016). Multifaceted biological insights from a draft genome sequence of the tobacco hornworm moth, *Manduca sexta*. *Insect Biochemistry and Molecular Biology*, 76, 118-147. doi:10.1016/j.ibmb.2016.07.005
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), 772-780. doi:10.1093/molbev/mst010
- Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, 37(8), 907-915. doi:10.1038/s41587-019-0201-4
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*, 14(4), R36. doi:10.1186/gb-2013-14-4-r36
- Kirk, H., Dorn, S., & Mazzi, D. (2013a). Worldwide population genetic structure of the oriental fruit moth (*Grapholita molesta*), a globally invasive pest. *BMC Ecology*, 13. doi:10.1186/1472-6785-13-12
- Kirk, H., Dorn, S., & Mazzi, D. (2013b). Worldwide population genetic structure of the oriental fruit moth (*Grapholita molesta*), a globally invasive pest. *BMC Ecology*, 13(1), 12.
- Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptivek-mer weighting and repeat separation. *Genome Research*, 27(5), 722-736. doi:10.1101/gr.215087.116
- Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics*, 5, 59. doi:10.1186/1471-2105-5-59
- Lagesen, K., Hallin, P., Rodland, E. A., Staerfeldt, H. H., Rognes, T., & Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research*, 35(9), 3100-3108. doi:10.1093/nar/gkm160
- Lange, A., Beier, S., Huson, D. H., Parusel, R., Iglaue, F., & Frick, J. S. (2018). Genome Sequence of *Galleria mellonella* (Greater Wax Moth). *Genome Announcement*, 6(2),

- e01220-01217. doi:10.1128/genomeA.01220-17
- Lowe, T. M., & Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, 25(5), 955-964. doi:10.1093/nar/25.5.955
- Luo, S., Tang, M., Frandsen, P. B., Stewart, R. J., & Zhou, X. (2018). The genome of an underwater architect, the caddisfly *Stenopsyche tienmushanensis* Hwang (Insecta: Trichoptera). *GigaScience*, 7(12), 1-12. doi:10.1093/gigascience/giy143
- Ma, W., Zhao, X., Yin, C., Jiang, F., Du, X., Chen, T., . . . Lin, Y. (2020). A chromosome-level genome assembly reveals the genetic basis of cold tolerance in a notorious rice insect pest, *Chilo suppressalis*. *Molecular Ecology Resources*, 20(1), 268-282. doi:10.1111/1755-0998.13078
- Neva C., Durand, Muhammad S., Shamim, Ido, Machol, . . . Huntley. (2016). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Systems*, 3, 95-98. doi:10.1016/j.cels.2016.07.002
- Neven, L. G., Kumar, S., Yee, W. L., & Wakie, T. (2018). Current and future potential risk of establishment of *Grapholita molesta* (Lepidoptera: Tortricidae) in Washington State. *Environmental Entomology*, 47(2), 448-456.
- Nguyen, P., Sykorova, M., Sichova, J., Kuta, V., Dalikova, M., Capkova Frydrychova, R., . . . Marec, F. (2013). Neo-sex chromosomes and adaptive potential in tortricid pests. *Proceedings of the National Academy of Sciences of the United States of America*, 110(17), 6931-6936. doi:10.1073/pnas.1220372110
- Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, 33(3), 290-295. doi:10.1038/nbt.3122
- Price, M. N., Dehal, P. S., & Arkin, A. P. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular Biology and Evolution*, 26(7), 1641-1650. doi:10.1093/molbev/msp077
- Quaintance, A. L., & Wood, W. B. (1916). *Laspeyresia molesta*, an important new insect enemy of the peach. *Journal of Agricultural Research*, 7(8), 373-378.
- Rica, E., Lacey, L. A., Guerra, N., & Headrick, H. L. (2006). Control of the oriental fruit moth, *Grapholita molesta*, using entomopathogenic nematodes in laboratory and fruit bin assays. *Journal of Nematology*, 38(1), 168-171.
- Roach, M. J., Schmidt, S. A., & Borneman, A. R. (2018). Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics*, 19(1), 460. doi:10.1186/s12859-018-2485-7
- Rothschild, G. H. L. (1979). Comparison of methods of dispensing synthetic sx-pheromone for the control of oriental fruit moth, *Cydia molesta* (Busck) (Lepidoptera, Tortricidae), in Australia. *Bulletin of Entomological Research*, 69(1), 115-127.
- Sanderson, M. J. (2003). r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics*, 19(2), 301-302. doi:10.1093/bioinformatics/19.2.301
- Sarker, S., & Lim, U. T. (2018). Extract of *Nicotiana tabacum* as a potential control agent of *Grapholita molesta* (Lepidoptera: Tortricidae). *PLoS ONE*, 13(8), e0198302. doi:10.1371/journal.pone.0198302
- Shen, Z. J., Liu, Y. J., Gao, X. H., Liu, X. M., Zhang, S. D., Li, Z., . . . Liu, X. X. (2018). Molecular identification of two thioredoxin genes from *Grapholita molesta* and their function in resistance to emamectin benzoate. *Frontiers in Physiology*, 9, 1421.

- doi:10.3389/fphys.2018.01421
- Simao, 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(19), 3210-3212. doi:10.1093/bioinformatics/btv351
- Song, S. V., Downes, S., Parker, T., Oakeshott, J. G., & Robin, C. (2015). High nucleotide diversity and limited linkage disequilibrium in *Helicoverpa armigera* facilitates the detection of a selective sweep. *Heredity (Edinb)*, 115(5), 460-470. doi:10.1038/hdy.2015.53
- Song, W., Cao, L. J., Li, B. Y., Gong, Y. J., Hoffmann, A., & Wei, S. J. (2018). Multiple refugia from penultimate glaciations in East Asia demonstrated by phylogeography and ecological modelling of an insect pest. *BMC Evolutionary Biology*, 18, 152. doi:10.1186/s12862-018-1269-z
- Song, W., Cao, L. J., Wang, Y. Z., Li, B. Y., & Wei, S. J. (2017). Novel microsatellite markers for the oriental fruit moth *Grapholita molesta* (Lepidoptera: Tortricidae) and effects of null alleles on population genetics analyses. *Bulletin of Entomological Research*, 107(3), 349-358. doi:10.1017/S0007485316000936
- Stanke, M., & Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics*, 19 Suppl 2, ii215-225. doi:10.1093/bioinformatics/btg1080
- SuomMainen, E. (1969). Chromosome evolution in the Lepidoptera. In C. D. Darlington & K. R. Lewis (Eds.), *Chromosome Today* (Vol. 2, pp. 122-138). New York: Plenum Press.
- Tarailo-Graovac, M., & Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics*, 25(1), unit 4.10.
- Torriani, M. V., Mazzi, D., Hein, S., & Dorn, S. (2010). Structured populations of the oriental fruit moth in an agricultural ecosystem. *Molecular Ecology*, 19(13), 2651-2660. doi:10.1111/j.1365-294X.2010.04711.x
- Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, 25(9), 1105-1111. doi:10.1093/bioinformatics/btp120
- van der Geest, L. P. S., & Evenhuis, H. H. (1991). *World Crop Pests 5: Tortricid pests their biology, natural enemies and control* (L. P. S. van der Geest & H. H. Evenhuis Eds. Vol. 5). Amsterdam, The Netherlands: Elsevier.
- Vatanparast, M., & Kim, Y. (2019). Yeast engineering to express sex pheromone gland genes of the oriental fruit moth, *Grapholita molesta*. *Journal of Asia-Pacific Entomology*, 22(3), 645-654. doi:10.1016/j.aspen.2019.04.009
- Vizueta, J., Sanchez-Gracia, A., & Rozas, J. (2020). BITACORA: A comprehensive tool for the identification and annotation of gene families in genome assemblies. *Molecular Ecology Resources*. doi:10.1111/1755-0998.13202
- Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., & Schatz, M. C. (2017). GenomeScope: Fast reference-free genome profiling from short reads. *Bioinformatics*, 33(14). doi:10.1093/bioinformatics/btx153
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., . . . Earl, A. M. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE*, 9(11), e112963. doi:10.1371/journal.pone.0112963
- Wan, F. H., Yin, C. L., Tang, R., Chen, M. H., Wu, Q., Huang, C., . . . Li, F. (2019). A chromosome-level genome assembly of *Cydia pomonella* provides insights into chemical ecology and insecticide resistance. *Nature Communications*, 10,

- doi.org/10.1038/s41467-41019-12175-41469. doi:ARTN 4237
10.1038/s41467-019-12175-9
- Wang, K., Li, P., Gao, Y., Liu, C., Wang, Q., Yin, J., . . . Shu, C. (2019). De novo genome assembly of the white-spotted flower chafer (*Protaetia brevitarsis*). *GigaScience*, 8(4), 1-9. doi:10.1093/gigascience/giz019
- Wang, X., Li, Y., Zhang, J., Zhang, Q., Liu, X., & Li, Z. (2017). De novo characterization of microRNAs in oriental fruit moth *Grapholita molesta* and selection of reference genes for normalization of microRNA expression. *PLoS ONE*, 12(2), e0171120. doi:10.1371/journal.pone.0171120
- Wei, S. J., Cao, L. J., Gong, Y. J., Shi, B. C., Wang, S., Zhang, F., . . . Chen, X. X. (2015). Population genetic structure and approximate Bayesian computation analyses reveal the southern origin and northward dispersal of the oriental fruit moth *Grapholita molesta* (Lepidoptera: Tortricidae) in its native range. *Molecular Ecology*, 24(16), 4094-4111. doi:10.1111/mec.13300
- Wu, N., Zhang, S., Li, X., Cao, Y., Liu, X., Wang, Q., . . . Zhan, S. (2018). Fall webworm genomes yield insights into rapid adaptation of invasive species. *Nature Ecology and Evolution*, 3(1), 105-115. doi:10.1038/s41559-018-0746-5
- Xia, Q., Zhou, Z., Lu, C., Cheng, D., Dai, F., Li, B., . . . Biology Analysis, G. (2004). A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). *Science*, 306(5703), 1937-1940. doi:10.1126/science.1102210
- Xiang, H., Liu, X., Li, M., Zhu, Y. n., Wang, L., Cui, Y., . . . Zhan, S. (2018). The evolutionary road from wild moth to domestic silkworm. *Nature Ecology & Evolution*, 2(8), 1268-1279. doi:10.1038/s41559-018-0593-4
- You, M. S., Yue, Z., He, W. Y., Yang, X. H., Yang, G., Xie, M., . . . Wang, J. (2013). A heterozygous moth genome provides insights into herbivory and detoxification. *Nature Genetics*, 45(2), 220-225.
- Zhan, S., Merlin, C., Boore, J. L., & Reppert, S. M. (2011). The monarch butterfly genome yields insights into long-distance migration. *Cell*, 147(5), 1171-1185. doi:10.1016/j.cell.2011.09.052
- Zhang, F., Zhang, J., Yang, Y., & Wu, Y. (2019). A chromosome-level genome assembly for the beet armyworm *Spodoptera exigua* using PacBio and Hi-C sequencing. *bioRxiv*, 2019.2012.2026.889121. doi:10.1101/2019.12.26.889121

Tables

Table 1 Completeness of major versions of assembly and annotation of *Grapholita molesta* genome estimated by BUSCO

Step	Version	Software	Complete %	Single copy %	Duplication %	Fragment %	Missing %	Size (Mb) / No. gene	N50 (Mb)
Assembly	Contig assembly	CANU	83.3	57.7	25.6	7.1	9.6	949.05	0.85
	Synteny reduction	purge_haplogs	76.6	74.7	1.9	10.5	12.9	521.60	1.79
	Contig polish	PILON	96.7	93.1	3.6	1.1	2.2	524.93	1.79
	Linkage group assembly	Juicer + 3D-DNA	93.3	92.2	1.1	1.0	5.7	525.22	19.5
	Linkage group polish	PILON	97.1	96.1	1.0	1.0	1.9	517.71	19.5
Annotation	Annotation	MAKER	90.1	86.1	4.0	6.6	3.3	39,274	NA
	Filtering based on expression	In-home script	89.8	85.9	3.9	6.1	4.1	19,478	NA
	Gene structure improvement	PASA	92.6	76.6	16.0	4.4	3.0	19,968	NA
	Functional annotation	EGGNOG	92.6	76.6	16.0	4.3	3.1	15,269	NA

Note: X/Y in Genome size (Mb) / Gene number items represent total size/N50 length. NA, not available.

Table 2 Number of genes in five detoxification families across 12 genomes from the Lepidoptera

Species	P450	GST	CCE	ABC	UGT	GR	IR	OBP	OR	HSP	Reference
Pxyl	163	55	85	219	38	68	61	91	225	56	You et al. (2013)
Gmol	186	56	119	128	47	95	81	64	295	46	This study
Cpom	136*	30*	73*	47*	30*	65*	39*	50*	85*	43	Wan et al. (2019)
Dple	107	35	73	76	47	74	62	68	253	63	Zhan et al. (2011)
Tni	143	51	122	71	68	85	74	91	274	55	Chen et al. (2019b)
Slit	182	66	153	223	64	107	62	87	261	58	Cheng et al. (2017)
Harm	122	57	105	76	54	120	71	79	257	53	Song et al. (2015)
Bman	94	37	94	64	48	69	56	74	218	59	Xiang et al. (2018)
Bmor	156	51	149	108	50	85	57	93	243	56	Xia et al. (2004)
Msex	164	66	137	103	51	89	76	79	261	67	Kanost et al. (2016)
Gmel	137	44	75	72	58	95	81	64	295	55	Lange et al. (2018)
Ofur	126	48	115	112	46	93	67	75	270	59	Ma et al. (2020)

*, data from Wan et al. (2019); the other data were manually identified in our study. Pxyl, *Plutella xylostella*; Gmol, *Grapholita molesta*; Cpom, *Cydia pomonella*; Dple, *Danaus plexippus*; Tni, *Trichoplusia ni*; Slit, *Spodoptera litura*; Harm, *Helicoverpa armigera*; Bman, *Bombyx mandarina*; Bmor, *Bombyx mori*; Msex, *Manduca sexta*; Gmel, *Galleria mellonella*; Ofur, *Ostrinia furnacalis*.

Figure legends

Fig. 1 Image of four stages of *Grapholita molesta*. Egg (a), larva (b), pupa (c), adult (d), the damage symptoms on the fruit of pear (e) and the twigs of peach (f) were shown. Photographs were taken by Shu-Jun Wei.

Fig. 2 Genome-wide all-by-all Hi-C matrix of OFM (a) and Synteny blocks among *Grapholita molesta*, *Spodoptera litura* and *Cydia pomonella* genomes (b). The genome sequences were assembled using the Juicer grouped using 3D-DNA; in total 16 linkage groups were identified based on Hi-C contact, indicated by blue boxes. Only sequences anchored on chromosomes were shown in the plot. The orange, green and blue oval blocks represent chromosomes assembled in *S. litura*, *G. molesta* and *C. pomonella*; the gray lines represent the synteny blocks between the genomes. Powerful synteny blocks were identified between the three moths' genomes.

Fig. 3 Phylogenetic relationships among *Grapholita molesta* and lepidopteran insects. Values on each node indicates the estimated divergence time. Superfamilies of 13 species of Lepidoptera and orders of four outgroups were indicated in the tree.