Chromosome level draft genomes of the fall armyworm, *Spodoptera frugiperda* (Lepidoptera: Noctuidae), an alien invasive pest in China

Huan Liu[1,10]*, Tianming Lan[1,11]*, Dongming Fang[1,11]*, Furong Gui[2,9]* , Hongli Wang[1,10], Wei Guo[3], Xiaofang Cheng[4], Yue Chang[5], Shuqi He[2] , Lihua Lyu[6], Sunil Kumar Sahu[1,10], Le Chen[7], Haimeng Li[1], Ping Liu[4], Guangyi Fan[5], Tongxian Liu[8], Ruoshi Hao[9], Haorong Lu[1,10], Bin Cheng[9], Shusheng Zhu[2], Zhihui Lu[9], Fangneng Huang[12], Wei Dong [1,10], Yang Dong[2], Le Kang[3], Huanming Yang[1,10], Jun Sheng[2,9#], Youyong Zhu[2#], Xin Liu[1,10#].

1. BGI-Shenzhen, Beishan Industrial Zone, Yantian District, Shenzhen 518083, China

2. Yunnan Agricultural University, Kunming 650201, China

3. State Key Laboratory of Integrated Management of Pest Insects and Rodents, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China

4. MGI, BGI-Shenzhen, Shenzhen 518083, China

5. BGI-Qingdao, BGI-Shenzhen, Qingdao 266555, China

6. Plant protection research institute，Guangdong Academy of Agricultural Sciences, Guangzhou 510640, China

7. BGI-Yunnan, No. 389 Haiyuan Road, High-tech Development Zone, Kunming, Yunnan 650106, China

8. Plant Medical College, Qingdao Agricultural University. Qingdao 266109, China

9. Yunnan Plateau Characteristic Agriculture Industry Research Institute，Kunming 650201, China

10. State Key Laboratory of Agricultural Genomics, BGI-Shenzhen, Shenzhen 518083, China,

11. China National GeneBank, Jinsha Road, Dapeng New District, Shenzhen 518120, China,

12. Department of Entomology, Louisiana State University AgCenter

*These authors contributed equally to this work.

#Correspondence: Xin Liu(liuxin@genomics.cn), Youyong Zhu (yyzhu@ynau.edu.cn), or Jun Sheng(shengjunpuer@163.com)

Abstract:

The fall armyworm (FAW), *Spodoptera frugiperda* (J.E. Smith) is a severely destructive pest native to the Americas, but has now become an alien invasive pest in China, and causes significant economic loss. Therefore, in order to make effective management strategies, it is highly essential to understand genomic architecture and its genetic background. In this study, we assembled two chromosome scale genomes of the fall armyworm, representing one male and one female individual procured from Yunnan province of China. The genome sizes were identified as 542.42 Mb with N50 of 14.16 Mb, and 530.77 Mb with N50 of 14.89 Mb for the male and female FAW, respectively. We predicted about 22,201 genes in the male genome. We found the expansion of cytochrome P450 and glutathione s-transferase gene families, which are functionally related to the intensified detoxification and pesticides tolerance. Further population analyses of corn strain (C strain) and rice strain (R strain) revealed that the Chinese fall armyworm was most likely invaded from Africa. These strain information, genome features and possible invasion source described in this study will be extremely important for making effective strategies to manage the fall armyworms.

Key Words: Fall armyworm, *Spodoptera frugiperda*, Chromosome-level genome, Insect, Pest

1 Introduction

It has been more than 100 years since the Fall armyworm (FAW), *Spodoptera frugiperda* (J.E. Smith) was reported to damage maize and other crops in the USA[1]. It is a severely destructive agricultural pest native to Americas which survives the whole year in the tropical and subtropical area from far south

Argentina, Chile and La Pampa to far north Florida, Texas, Mexico and the Caribbean[2-5]. It cannot survive severe winters because of the lack of diapause. However, FAW has a remarkable capacity of long-distance migration, with which the FAW can fly over 100 km per night[6]. Each spring, it can migrate over 2000 km from the overwintering areas to reinvade more northern regions, even up to Canada[4,7,8]. Recently, FAW spread out from its native region and invaded into Africa in 2016 with the report in São Tomé, Bénin, Togo and Nigeria[9]. This invasion rapidly became widespread in the whole sub-Saharan Africa till October 2017[10,11]. Following this trend, FAW soon invaded into many Asian countries, including India, Yemen, Thailand, Myanmar and Sri Lanka in 2018[12-14].

Early R *et al* (2018)[5] forecasted that China is one of the most vulnerable countries of being invaded by the FAW according to the information on frequent commercial trade and passenger transportation between Africa and China. Half a year later after this forecast, in January 2019, International Plant Protection Convention (IPPC) Contact Point for China spotted FAW for the first time in Puer and Dehong city, Yunnan Province, China (https://www.ippc.int/en/news/first-detection-of-fall-armyworm-in-china/). This invasive pest has rapidly invaded many provinces in China by June 2019. Now, the FAW has been detected in large parts of the world (Figure 1).

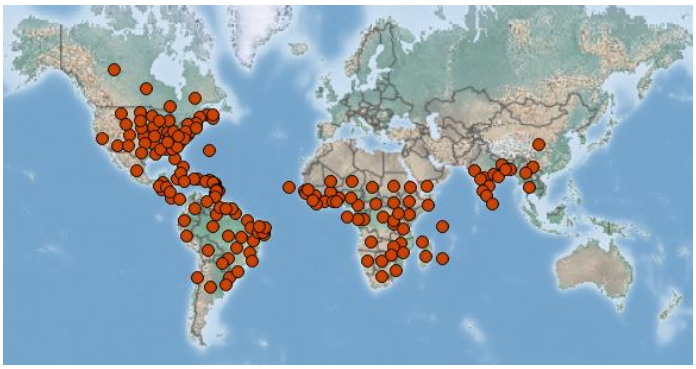

Figure 1 The distribution of the FAW all over the world[15].

There are several hosts of FAW, which mainly includes 186 plant species from 42 families[16]. Although FAW is a highly polyphagous pest, graminaceous plants are their preferred hosts, such as maize, rice, sorghum and other major agricultural species. Especially, maize is most likely to be attacked than other plant species[9]. Maize is very important food security of many countries in Asia, Africa, and Latin America (https://maize.org/projects-cimmyt-and-iita-2/), and now are facing severe threats from the infestation of the FAW. The production of maize infested by the FAW can suffer yield loss of 40% to 72%, and in some plots 100% total has been reported[17-19]. The yield loss of maize can reach 8.3 to 20.6m tons per year in just 12 African countries without any control methods for FAW,

according to the study conducted by Centre for Agriculture and Biosciences International (CABI)[20]. For Brazil alone, cost to control the FAW on maize is more than 600 million dollars per year[21]. The economic and the yield losses by FAW is a major concern worldwide.

The FAW is not a new species to science; it has been an herbivorous pest for many years. However, the mostly common used approach to mitigate the damage to crops is still the broad-spectrum insecticides[22]. The use of insecticides highly relies on the knowledge of farmers, but many farmers even do not know the name of the pest[23]. Lackof scientific guidance leads to the inappropriate use of pesticides. Besides, the effective management of FAW may require five sprays of pesticides per maize cycle[21], and many smallholder farmers cannot afford the expensive costs for it, resulting in the use of low-quality products or low dose sprays[24]. Moreover, the FAW always hides inside the stem of maize, this makes the insecticide much less effective. Although many pesticides are less harmful to the environment and humans, all these factors can lead the sublethal effects, which possibly help the FAW to evolve resistance against the pesticides[25-27]. An effective compensating management for insecticides is the use of *Bacillus thuringiensis* (Bt) toxins produced by the by the bacterium.Bt plants have been proved fatal to many insect pests, including the FAW[28-30]. The Bt toxin provides much longer protection than insecticides and less harmful to the environment and humans. Although some research reported the resistance of FAW to Bt maize[29,30], multiple genes or new gene with more or new Bt toxin expressed are still thought to have a good performance for resisting the FAW[29,30]. Biological control, including the, introduction of natural enemies and using companion cropping system[19,21,25,31,32], is also an effective way to resist the FAW.

The FAW at least consist of two morphological identical but genetically distinct subpopulations, the corn strain (C strain) and rice strain (R strain)[33-35]. The two strains have their own preferable host plants. The C strain is preferentially associated with maize and other large grasses, but the C strain prefers rice and large grasses[36,37]. The C strain is subdivided into two subgroups, the FL-type and the TX-type[38,39]. The TX-type is distributed in most of the Americas, but the FL-type is only limited to Florida and the Caribbean[3,35,40,41]. Each strain has its strain specific physiological traits, leading to some strain-specific response to biological and chemical agents[36,37,42]. The C strain larvae are more tolerant than the R strain to the methyl parathion, cypermethrin, cypermethrin and ð-endotoxin from transgenic Bt plant[43]. Therefore, the origin of FAW needs to be considered to make effective strategies to manage the FAW. The *Cytochrome c oxidase subunit I* (*COI*) and the *Triosephosphate isomerase* (*Tpi*) gene are also selected for identifying

the subtype of the FAW, but these markers cannot always give the right identifications[44].

The management of the FAW needs more detailed genetic information to help people know more about the FAW, to find more new genes to develop more effective Bt plants, and to more accurately identify the different strains for precision spraying of pesticides. Although several genomes of the FAW has been sequenced and assembled[45-47], the assemblies were fragmented. Moreover, two genomes were assembled using the Sf21[45] and Sf9[47] cell line, respectively, the resource was unique that cannot well provide a comprehensive reference. In this study, we assembled two FAW genomes to the chromosome level using two samples (one male, one female) collected from Yunnan Province, China. We analyzed the subtype of the FAW invaded Yunnan province, and also discussed the possible resource of the invaded FAW in China. We are also screening expanded gene families to seek some key genes with the function of polyphagia and tolerance to insecticides. China is the second largest corn producer after the USA; therefore, it is urgent to select a series of methods to control the FAW. This study provided key information to help make strategies to manage the FAW in China.

2 Materials and Methods

2.1 Samples and treatments

We collected seven FAW samples, including four adults, two fifth-instar larvae and two sixth-instar larvae (Table 1). The four adult FAWs were collected from Yunnan Province, China, and the larvae were collected from Guangdong Province, China. One male and one female adult individual were used for genome sequencing and assembly (Figure 2). Two other adult FAWs were used to capture the conformation of chromosomes to perform the chromosome level genome assembly. One fifth-instar larva and one sixth-instar larva were subjected for the transcriptomic studies. One sixth-instar larva was used for whole genome sequencing with 5K and 300bp insert size. All samples were intestinal and ovarian, and thoroughly cleaned before performing DNA or RNA isolation.

Table 1 Samples used in this study

Sample Identifier	Species	Location	Sex	Instars
SFynMstLFR	*Spodoptera frugiperda*	Kunming, Yunnan	Male	Adult
SFynFMstLFR	*Spodoptera frugiperda*	Kunming, Yunnan	Female	Adult
SFgdRNA 1	*Spodoptera frugiperda*	Guangzhou, Guangdong	Unknown	Fifth-instar
SFgdRNA 2	*Spodoptera frugiperda*	Guangzhou, Guangdong	Unknown	Sixth-instar
SFgdWGS	*Spodoptera frugiperda*	Guangzhou, Guangdong	Unknown	Sixth-instar

SfYnHiC	*Spodoptera frugiperda*	Kunming, Yunnan	Unknown	Adult

Figure 2 The two adult fall armyworm (FAWs) used for genome assembly. The left FAW is female, the right FAW is the male individual.

2.2 DNA isolation, library preparation, sequencing and genome assembly.

The high molecular weight DNA was extracted using the separated muscle tissue by following the protocol recommended by Cheng et al (2018)[48]. We then used the single tube long fragment read (stLFR) technology[49] to preparing the co-barcoding DNA libraries with the MGIEasy stLFR Library Prep Kit (lot number: 1000005622), and the libraries were then loaded on the sequencer for sequencing according to the protocol of MGISEQ-2000[50]. To testify the accuracy of genome assembly, we constructed a DNA library with a 5kb insert size and sequenced by BGISEQ-500 sequencer. To finally ligate the scaffolds to chromosomes, Hi-C technology[51] was used to capture the conformation of chromosomes using another two adult individuals. The primary genome was assembled using the supernova (v2.1.1)[52] software with the parameters *--maxreads=700M*. We filled gaps by use of GapCloser[53] and GapCloser stLFR (unpublished method) with the default parametes. Finally, we performed the chromosome concatenation using the Hi-C generated data by 3d-DNA pipeline[54].

2.3 Comparative genomics analysis

Identification of orthology and paralogy groups of *Spodoptera frugiperda* genes and other considered genomes were done using OrthoMCL[55] methods on the all-versus-all BLASTP alignment (e-value, <1e-5). We constructed gene families for nine species including *Bombyx mori, Danaus plexippus, Drosophila melanogaster, Heliconius melpomene, Helicoverpa armigera, Manduca sexta, Plutella xylostella, Spodoptera litura and Spodoptera frugiperda*. The phylogenetic tree, including these nine species, was constructed using the combined set of all the single copy genes.

To search for homology, we compared protein-coding genes of *Spodoptera frugiperda* to that of other species using BLASTP with an E-value threshold of 1e-5. Base on the Whole-genome BLASTP and the genome annotation results, we detected the syntenic blocks using MCscan[56]. A region with at least five syntenic genes and no more than 15 gapped genes was defined as a syntenic block.

2.4 RNA isolation, transcriptome libraries preparation and sequencing

The RNA extraction kit (RNeasy Mini Kit, Qiagen) was used for the total RNA isolation. We performed the RNase-free agarose gel to check the contamination, and then the RNA integrity and purity were measured by Agilent 2100 Bioanalyzer system (Agilent, United States) and NanoDrop Spectrophotometer (THERMO, United States), respectively. The extracted RNA was fragmented into 200-400 bp and reverse transcribed to cDNA for library preparation. The libraries were prepared to follow the manufacturer's instructions for the BGISEQ-500 sequencing platform. Pair-end 100 sequencing was performed on the BGISEQ-500 sequencer using the processed libraries.

2.4 Bioinformatics analysis for transcriptome data

Raw data were firstly processed using the Trimmomatic to filter the reads with adaptors and reads with the proportion of Ns and low-quality bases larger than 10% by SOAPfilter. Bridger software (v20141201) was used to *de novo* assemble the transcriptome, the reluctances were then removed by TGICL. The contigs were concatenated into scaffolds and further assembled to unigenes by clustering and removing redundancy.

FPKM was calculated to estimate the expression level of unigenes. In the study, the reads were mapped against the unigene library using Bowtie, and then unique mapped reads were selected for estimating the expression level by combining eXpress. Finally, DEG unigenes were selected with differential expression level with the parameter of FDR ≤ 0.01 and Fold Change ≥4.

2.5 Identifying the strains and the possible source of FAW invasion in China

We used the *Tpi* gene as a DNA maker to identify the strain of the FAW invaded into China. We identified the *Tpi* gene fragments from four FAWs, including two from Yunnan province (sequences were retrieved from the whole genome sequencing data) and two from Guangdong province (sequences were retrieved from the RNA-seq data). Eight sites (TpiE4-129, TpiE4-144, TpiE4-165, TpiE4-168, TpiE4-180, TpiE4-183, TpiE4-192, TpiE4-198) in the fourth exon of the *Tpi* gene were used for determining the strains and a possible source of the invaded FAW[57]. The fourth intron of the *Tpi* gene[57] was used for constructing the phylogenetic tree using the PhyML (v3.0)[58] software with the Maximum-Likelihood methods to assist in identifying the strains. Only two Tpi fourth introns were used for phylogenetic analysis because we cannot retrieve the introns from RNAseq data. 112 sequences of the fourth intron of the *Tpi* gene with strain information were downloaded from NCBI (Table S1). The sequence alignments were performed using the Clustal W[59].

3 Results and discussion

3.1 Chromosome level genome assembly for two FAWs

Only one individual was used for genome sequencing for both the male and female genome assembly, which was less than the number used for genome assembly by Gouin A *et al* (2017)[46]. This maximally decreased the heterozygosity level. A total of 2μg and 1.6μg total genomic DNA with the average band size larger than 20 kb were isolated for stLFR libraries preparation from the male and female FAW, respectively. Two stLFR libraries with more than 30 million barcodes were constructed for running on the BGISEQ-500 sequencer. 110.93 Gb and 91.41 Gb high-quality reads were generated for the male and female individual, respectively (Table S2). The primary assembly sizes were 542.42 Mb and 530.77 Mb for the male and female individual. The scaffold N50 and N90 for the male individual were 507.12 Kb and 6.43 Kb, and that for the female were 528.27 Kb and 5.11 Kb, respectively (Table 2). After getting the confirmation from Hi-C sequencing, we finally concatenated the scaffolds to 31 chromosomes with the scaffold N50 of 14.16 Mb and 14.89 Mb for the male and female individual, respectively. Genomes sizes assembled in this study are in the range of the lepidoptera genome sizes at 246M to 809M[60], but are larger than previously assembled FAW genomes[45-47], probably due to the use of the new stLFR technology with the super high coverage and reads quality. This is the first time to assemble the genome of the FAW to the chromosome level, which will be in no doubt to accelerate the biological studies and making the effective strategy of pest managements.

Table2 Summary of the two Spodoptera frugiperda genome assemblies

Methods	Statistics	SFynMstLFR				SFynFMstLFR			
		Scaffold Length (bp)	Number	Contig Length (bp)	Number	Scaffold Length (bp)	Number	Contig Length (bp)	Number
stLFR	N50	507,121	226	91,970	1,220	528,269	231	83,007	1,299
	Total length (bp)	542,424,128		505,703,627		530,766,122		496,217,290	
	Ratio of Ns (%)	6.77				6.51			
	GC content (%)	36.52				36.57			
HiC	N50	14,162,803	16	91,970	1,220	14,883,732	13	124,992	813
	Total length (bp)	543,659,128		505,703,627		531,931,622		496,217,290	
	Ratio of Ns (%)	6.98				6.71			
	GC content (%)	36.52				36.58			
Chromesome Level		461,198,141	31			435,876,255	31		
Chromosome Level (%)		84.83				81.94			

3.2 Evaluation of the assembly

The GC content for the genomes of the male and female individual was found to be 35.52% and 36.57% (Table 2), respectively, which showed a similar level with most closely related Lepidoptera species ranging from 31.6% to 37.7%[61]. We used the Benchmarking Universal Single-Copy Orthologs (BUSCO: version 2.0)[62] and Core Eukaryotic Genes (CEGMA)[63] to evaluated the completeness of the two assemblies. In BUSCO analysis, genomes of the male and female samples covered 95% and 94.5% complete BUSCO genes (Table S3). In the CEGMA analysis, 83.47% and 85.48% complete core eukaryotic genes were found for the two genomes (Table S4). This is better than all FAW genomes that has been published (PRJNA380964; PRJNA257248; PRJEB13110; and PRJNA344686). Besides, we also mapped the sequencing data generated from libraries of Hi-C, MatePair5K, WGS, and the RNA-seq to the assembled male genome. The mapping rates were all higher than 90% (Table 3), and the insert size were also consistent with the libraries, except for the MatePair5K, probably because the large insert size cannot ensure that the one pair reads they mapped to the same scaffold. EST sequences of the FAW were downloaded from NCBI and the transcripts were assembled without reference. We mapped these EST sequences and transcripts to the male reference genome we assembled, and the results showed that more than 90% EST sequences and more than 80% transcripts we assembled could be found on the assembled male genome (Table S5). However, it is noteworthy that the transcript from SFgdRNA 2 has a lower mapping rate than that of SFgdRNA 1. We inferred that this resulted due to the genetic differences between the C strain and the R strain, because the SFgdRNA 2

sample was identified belong to the R strain. Overall, all the above results well testified the completeness of the two genomes.

Table 3 Mapping reads against the assembled male genome using raw reads generated by different libraries

Type	HIC	MatePair	stLFR	RNA-seq	WGS
Total Mapped Reads	93.58%	93.68%	95.60%	98.98%	90.71%
Perfect Match	38.81%	38.54%	43.67%	52.14%	27.76%
Unique Match	76.39%	77.15%	83.23%	76.99%	74.77%
Total Unmapped Reads	6.42%	6.32%	4.40%	1.02%	9.29%
Total FullMapped Reads	32.48%	30.06%	71.45%	52.46%	49.40%

3.3 Annotation

We firstly used Repeat Modeler (v1.0.11), LTR finder (v1.0.5) and repeatscount (v1.0.5) methods to identify *de novo* repeat motifs by modeling *ab initio*, and these repeat motifs were added into the RepBase[64] library as known repeat elements. We then performed the RepeatMasker[65] to mask the assembly, using the combined RepBase library. Usually, Repeat elements take a substantial part of the genome and contribute as important events to genome evolution[61,66,67]. In this study, by the combination of *de novo* and homology-based searching, 153 Mbp repeat elements were finally identified for the male FAW, and accounting for 28.24% the FAW genomes.

Gene prediction was carried out by both the homology-based and *de novo* methods using repeat masked genomes. For the *de novo* prediction, we used Augustus, glimmerHMM and SNAP (Table S6). For the homology-based approaches, *Bombyx mori*, *Danaus plexippus*, *Drosophila melanogaster* and *Spodoptera litura* genomes were used for homology alignments using the TblastN. Moreover, transcripts that were predicted with RNA-seq, Gene sets were then merged to form a non-redundant gene set with GLEAN; then all annotated genes were checked and filtered manually. A total of 22201 genes were finally obtained for the male samples (Table S6).

In the final gene set we identified, we found 94.2% compete for BUSCO genes and 95.16% CEGMA genes, which were all better than the published FAW genome (Table S7, Table S8). Of these identified genes, 93.48% was confirmed that have functions (Table S9), which was facilitated the further exploration of the functions. Besides, we also found 60 miRNAs, 840 tRNAs and 197 rRNAs by using the homology prediction method.

3.4 The transcriptome analysis of the larvae

After filtering, we finally obtained 58Gb clean data with 341,526,489 cleaned reads. These reads were assembled into 72,604 contigs with the N50 of 2077bp. These contigs were further assembled into scaffolds, and the scaffolds were further assembled to 51,495 unigenes by clustering and removing redundancy. The contig number in our study are significantly higher than that in the study of Kakumani *et al*[68]. This maybe result from that we only used a single method for assembly. We also calculated the expression abundance for unigenes between the fifth-instar and sixth-instar larvae. The result showed 2,648 differentially expressed genes (DEGs). We further performed the clustering analysis to cluster genes with identical or similar expressed behaviors. Remarkable expression difference was found between the fifth-instar larvae and the sixth-instar larvae (Figure S1). This difference was in consistent with the different strains of the two larvae (we described in 3.6). However, if the different instar contributes to the differential expression, it was further confirmed by more detailed analysis.

3.5 Comparison to other published lepidopteran genomes

To further explore the detailed relationship between the FAW and its other lepidopteran relatives. We constructed a phylogenetic tree of nine genomes using 2,001 single copy genes downloaded from NCBI and insectbase (Table S10). The result showed that the *S. frugiperda* which we sequenced actually clustered with its most related species *S. litura*, which is in accordance with the study by Cheng *et al* [69](Figure 3).

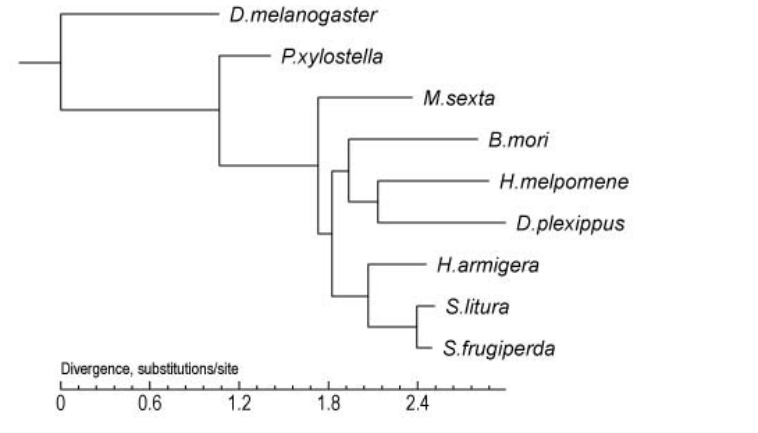

Figure 3 The phylogenetic relationships among nine lepidopteran genomes.

Through the gene family analysis, a total of 12,516 gene families were found in the *S. frugiperda* genome, including 20,012 genes. Of these gene families, 324 are specific to the *S. frugiperda* compared to the other seven species (Table S11). Then, we analyzed the 34 functional gene families of insects, finding some expanded gene families, including cytochrome p450, glutathione s-transferase, and hydrolase (Table 4). The cytochrome p450 gene family is closely related with

intensified detoxification[69], the genes in this family of the *S. frugiperda* is 200, more than that of *S. litura*, which indicated that the *S. frugiperda* was more polyphagous than *S. litura*. This is also consistent with the habits of *S. frugiperda.* The expanded glutathione s-transferase gene family was proved that could enhance the insecticides tolerance of the *S. litura*[69]. In this study, we found more genes for *S. frugiperda,* which indicated that the *S. frugiperda* was probably easier to gain resistance to pesticides. These gene families are a valuable genetic source to develop more effective pesticides or other methods to manage the FAW.

Table 4 Identified genes in gene families of 9 insects

	DME	PXY	MSE	BMO	HME	DPL	HAR	SLI	SFR
ABC transporter	55	76	40	40	35	38	36	86	66
acetylcholine receptor	18	27	20	27	24	28	22	22	27
Acetylcholinesterase	27	39	62	51	32	27	82	87	46
alkaline phosphatase	13	10	11	10	8	8	8	11	12
aminopeptidase	45	53	49	37	35	39	15	19	24
carboxylesterase	30	55	90	70	50	54	100	110	84
Chitinase	45	27	28	24	26	32	23	26	46
chloride channel	16	25	20	23	18	19	19	16	24
CTL	2	4	3	2	2	2	3	4	4
cytochrome p450	104	108	144	96	130	99	121	132	200
DNA methyltransferase	1	1	1	1	1	1	1	1	1
ecdysone receptor	1	1	3	3	3	2	3	1	3
GABA	15	16	32	9	12	15	16	14	10
glutamate-gated chloride channel	10	14	12	13	10	10	12	12	14
Glutathione s-transferase	50	28	42	25	23	25	46	47	60
glycosyltransferase	65	25	23	19	24	22	19	21	31
G protein	23	42	36	31	26	29	29	29	30
gustatory receptor	54	69	45	76	73	68	197	238	220
heat shock protein	45	38	54	35	29	46	40	48	45
hydrolase	213	250	243	209	181	204	267	310	396
immunoglobulin	65	51	70	76	56	69	71	67	80
odorant-binding protein	62	65	70	51	87	63	77	61	70
odorant receptor	67	82	79	43	73	64	87	75	75
Painless	1	1	1	1	1	1	1	1	1
pheromone	43	22	29	15	19	14	16	14	19
protease inhibitor	75	33	60	38	26	36	46	39	51
Ryanodine receptor	6	12	3	7	4	8	6	3	6
sensory neuron membrane protein	3	5	4	5	4	7	7	3	10
serpin	35	23	30	18	19	24	21	18	26
sirtuin	6	4	3	5	5	5	7	7	7

sodium channel	38	30	33	24	15	27	30	25	36
sugar transporter	17	51	42	36	40	37	41	40	59
superoxide dismutase	17	18	12	12	9	11	12	13	36
Vitellogenin receptor	8	14	15	11	9	12	16	14	15

Note: gene families come from http://www.insect-genome.com/genefamily/gene-family.php

3.6 FAW in China includes both the C and R strain, possibly invaded from Africa

There are three strain-specific sites (E4165, E4168 and E4183) in the fourth exon of the *Tpi* gene that can identify the C strain from the R strain for Western Hemisphere populations[57]. Especially, the E4183 is an effective diagnostic marker for *Tpi* gene to define the C or R strain. In this study, two Yunnan samples and one Guangdong sample were identified as the C strain and the other one Guangdong sample was identified as the R strain (Figure 4). The phylogenetic tree showed that the two Yunnan samples were clustered in the clade that consisted of all C strain individuals, which strengthened the results inferred using the strain-specific sites in the fourth exon (Figure 5). This result at least showed that the FAW invaded into China included both the strains. However, the detailed population genetic structure and the frequencies of the two strains in the Chinese population need more information from the population-level studies.

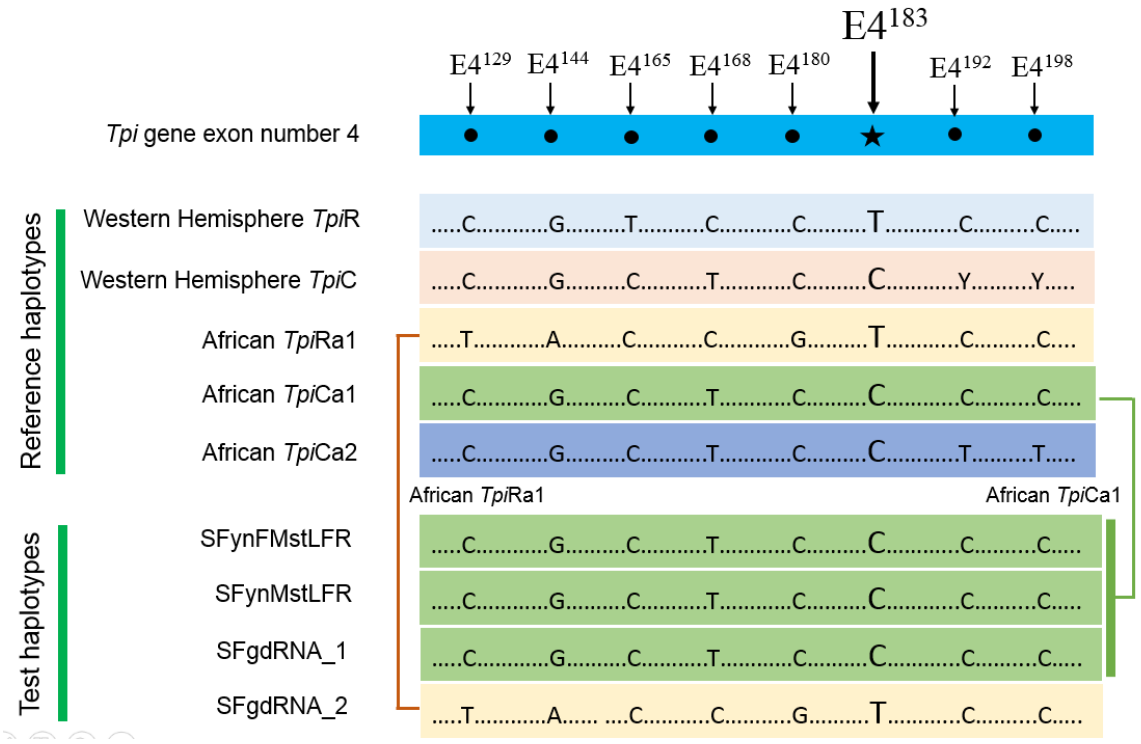

Figure 4 The identification of C and R strains.

To further confirm the possible source of the Chinese FAW, we compared the haplotype consisted of all eight polymorphisms in the fourth exon of the *Tpi* gene as shown in the Figure 4. We found that all the four individuals hold identical haplotype with the African population, including three *Tpi*Ca1 and one *Tpi*Ra1. Although the haplotype of C strain was shared by the African and the Western Hemisphere populations[57], the *Tpi*Ra1 has not been detected in any Western Hemisphere populations[57] showing the uniqueness to African populations. The finding of the *Tpi*Ra1 haplotype in the Guangdong population indicated that there were at least parts of the FAW populations in China that was invaded from Africa, probably through the frequent commercial trade and passenger transportation between Africa and China[5]. However, we cannot confirm other sources because of the small sample size we used here. The strain information and possible invasion source found in this study will be extremely important for making effective strategies to manage the FAWs in China.

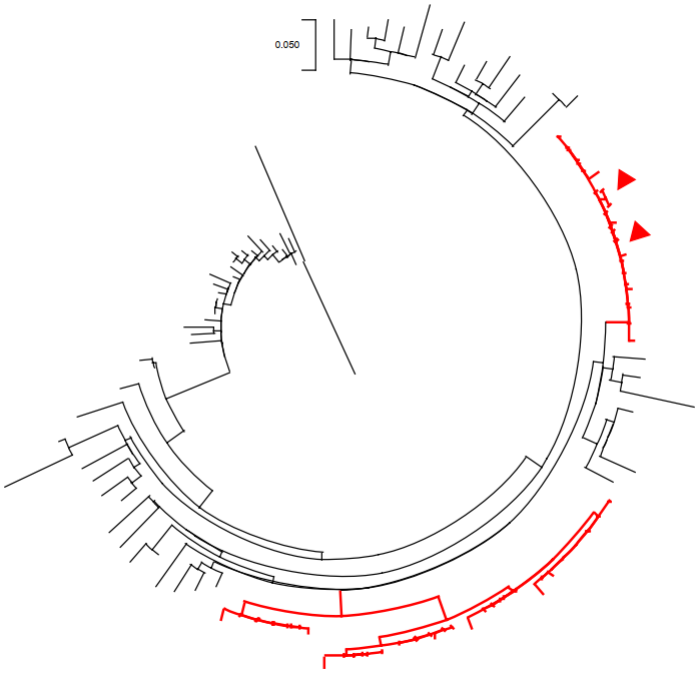

Figure 5 The phylogenetic tree to identify the FAW strains of collected from Yunnan, China. The red branched presented the C strain, the red triangles present the two Yunnan samples.

4 Conclusions

In summary, we assembled two chromosome scale genomes of the fall armyworm, representing one male and one female individual procured from Yunnan province of China. The genome sizes were identified as 542.42 Mb with N50 of 14.16 Mb, and 530.77 Mb with N50 of 14.89 Mb for the male and female FAW, respectively. . The completeness of the two genomes are better than all

previously published FAW genomes which is evident by the BUSCO and CEGMA analysis. A total of 22,201 genes were predicted in the male genome, and 12,516 gene families were found in the *S. frugiperda* genome, including 20,012 genes. Of these gene families, we found expansion of cytochrome p450 and glutathione s-transferase gene families, which were closely related to the function of intensified detoxification and pesticides tolerance. We finally identified both the R strain and C strain individuals in the Chinese population, showed that the Chinese FAW was most likely invaded from Africa. The strain information and possible invasion source found in this study will be extremely important for making the effective strategies to manage the FAWs in China.

Acknowledgements

This study was supported by the Guangdong Provincial Key Laboratory of core colletion of corp genetic resources research and application (NO.2011A091000047), Guangdong Provincial Key Laboratory of Genome Read and Write (grant no. 2017B030301011) and Key Laboratory of Genomics ,Ministry of Agriculture.

Data availability

The Raw sequencing data and the two chromosome level genome assemblies have been deposited to the CNSA (CNGB Nucleotide Sequence Archive) with accession CNP0000513 (https://db.cngb.org/cnsa/).

Author Contributions

Huanming Yang, Le Kang, Jun Sheng, Youyong Zhu, Yang Dong, Xin Liu and Huan Liu designed the research. Dongming Fang, Tianming Lan and Yue Chang, Hongli Wang, Fangneng Huang, Wei Dong and Guangyi Fan performed the data analysis. Xiaofang Chen, Haorong Lu, Ping Liu, Tongxian Liu, Rushi Hao, Bin Chen, Shusheng Zhu, Zhihui Lu and Haimeng Li performed the DNA and RNA extraction and the library preparation. Huan Liu, Tianming Lan, Yang Dong, Wei Guo, Shuqi He, Le Chen and Lihua Lyu collected the samples. Tianming Lan, Dongming Fang, Hongli Wang, Sunil Kumar Sahu and Furong Gui wrote and revised the manuscript. All the authors read and revised the final version of the manuscript.

Competing interests

The authors declare no competing interests.

1 Hinds, W. E. & Dew, J. A. The grass worm or fall army worm. *Alabama Agricultural and Experiment Station Bulletin* **186**, 25 (1915).

2 Pair, S. D., Raulston, J. R., Westbrook, J. K., Wolf, W. W. & Adams, S. D. Fall Armyworm (Lepidoptera Noctuidae) Outbreak Originating in the Lower Rio Grande Valley, 1989. *Florida Entomologist* **74**, 14 (1991).

3 Nagoshi, R. N., Meagher, R. L. & Hay-Roe, M. Inferring the annual migration patterns of fall armyworm (Lepidoptera: Noctuidae) in the United States from mitochondrial haplotypes. *Ecology and evolution* **2**, 1458-1467, doi:10.1002/ece3.268 (2012).

4 Mitchell, E. R. *et al.* Seasonal Periodicity of Fall Armyworm, (Lepidoptera Noctuidae) in the Caribbean Basin and Northward to Canada. *Journal of entomological science* **26**, 12 (1991).

5 Early, R., González-Moreno, P., Murphy, S. T. & Day, R. Forecasting the global extent of invasion of the cereal pest Spodoptera frugiperda, the fall armyworm. *NeoBiota* **40**, 25-50, doi:10.3897/neobiota.40.28165 (2018).

6 Johnson, S. J. Migration and the life history strategy of the fall armyworm, Spodoptera frugiperda in the Western Hemisphere. *Insect Sci.* **Appli 8**, 7 (1987).

7 du Plessis, H., van den Berg, J., Ota, N. & Kriticos, D. J. Spodoptera frugiperda (Fall Armyworm). *CSIRO-InSTePP Pest Geography* (2018).

8 Westbrook, J. K., Nagoshi, R. N., Meagher, R. L., Fleischer, S. J. & Jairam, S. Modeling seasonal migration of fall armyworm moths. *International journal of biometeorology* **60**, 255-267, doi:10.1007/s00484-015-1022-x (2016).

9 Goergen, G., Kumar, P. L., Sankung, S. B., Togola, A. & Tamo, M. First Report of Outbreaks of the Fall Armyworm Spodoptera frugiperda (J E Smith) (Lepidoptera, Noctuidae), a New Alien Invasive Pest in West and Central Africa. *PloS one* **11**, e0165632, doi:10.1371/journal.pone.0165632 (2016).

10 FAO. Briefing Note on FAO Actions on Fall Armyworm in Africa *http://www.fao.org/3/a-bt415e.pdf* (2017).

11 Cock, M. J. W., Beseh, P. K., Buddie, A. G., Cafa, G. & Crozier, J. Molecular methods to detect Spodoptera frugiperda in Ghana, and implications for monitoring the spread of invasive species in developing countries. *Scientific reports* **7**, 4103, doi:10.1038/s41598-017-04238-y (2017).

12 Sharanabasappa *et al.* First report of fall armyworm, Spodoptera frugiperda (J E smith) (Lepidoptera: Noctuidae), an alien invasive pest on maize in India. *Pest Management in Horticultrual Ecosystems* **24**, 7 (2018).

13 Ganiger, P. C. *et al.* Occurrence of the new invasive pest, fall armyworm, Spodoptera frugiperda (J.E. Smith) (Lepidoptera: Noctuidae), in the maize fields of Karnataka, India. *Current Science* **115**, 3 (2018).

14 Portal, C. F. A. Available online: https://www.cabi.org/isc/fallarmyworm (accessed on 6 June 2019).

15 CABI. Invasive Species Compendium. https://www.cabi.org/isc/fallarmyworm.

16 Casmuz, A. *et al.* Revisión de los hospederos del gusano cogollero del maíz, Spodoptera frugiperda (Lepidoptera: Noctuidae). *Revista de la Sociedad Entomológica Argentna* **69**, 23 (2010).

17 Murúa, G., Molina-Ochoa, J. & Coviella, C. Population dynamics of the fall armyworm, Spodoptera frugiperda (Lepidoptera: Noctuidae) and its parasitoids in northwestern Argentina. *Florida Entomologist* **89**, 8 (2006).

18 Silva-Aguayo, G. *et al.* Bioactivity of Boldo (Peumus boldusMolina) (Laurales: Monimiaceae) onSpodoptera frugiperda(J.E. Smith) andHelicoverpa zea(Boddie) (Lepidoptera: Noctuidae). *Southwestern Entomologist* **35**, 215-231, doi:10.3958/059.035.0301 (2010).

19 Wyckhuys, K. A. G. & O'Neil, R. J. Population dynamics of Spodoptera frugiperda Smith (Lepidoptera: Noctuidae) and associated arthropod natural enemies in Honduran subsistence maize. *Crop Protection* **25**, 1180-1190, doi:10.1016/j.cropro.2006.03.003 (2006).

20 Day, R. *et al.* Fall Armyworm: Impacts and Implications for Africa. *Outlooks on Pest Management* **28**, 196-201, doi:10.1564/v28_oct_02 (2017).

21 Sisay, B. *et al.* First report of the fall armyworm, Spodoptera frugiperda (Lepidoptera: Noctuidae), natural enemies from Africa. *Journal of Applied Entomology* **142**, 800-804, doi:10.1111/jen.12534 (2018).

22 Gordy, J. W., Leonard, B. R., Blouin, D., Davis, J. A. & Stout, M. J. Comparative Effectiveness of Potential Elicitors of Plant Resistance against Spodoptera frugiperda (J. E. Smith) (Lepidoptera: Noctuidae) in Four Crop Plants. *PloS one* **10**, e0136689, doi:10.1371/journal.pone.0136689 (2015).

23 Midega, C. A. O. *et al.* Farmers' perceptions of cotton pests and their management in western Kenya. *Crop Protection* **42**, 193-201, doi:10.1016/j.cropro.2012.07.010 (2012).

24 Dal Pogetto, M. H. F. A. *et al.* Corn Yield with Reduction of Insecticidal Sprayings Against Fall Armyworm Spodoptera frugiperda (Lepidoptera Noctuidae). *Journal of Agronomy* **11**, 5, doi:10.3923/ja.2012.17.21 (2012).

25 Fotso Kuate, A. *et al.* Spodoptera frugiperda Smith (Lepidoptera: Noctuidae) in Cameroon: Case study on its distribution, damage, pesticide use, genetic differentiation and host plants. *PloS one* **14**, e0215749, doi:10.1371/journal.pone.0215749 (2019).

26 Yu, S. J., Nguyen, S. N. & Abo-Elghar, G. E. Biochemical characteristics of insecticide resistance in the fall armyworm, Spodoptera frugiperda (J.E. Smith). *Pesticide Biochemistry and Physiology* **77**, 1-11, doi:10.1016/s0048-3575(03)00079-8 (2003).

27 de Bon, H. *et al.* Pesticide risks from fruit and vegetable pest management by small farmers in sub-Saharan Africa. A review. *Agronomy for Sustainable Development* **34**, 723-736, doi:10.1007/s13593-014-0216-7 (2014).

28 Sousa, F. F. *et al.* Life-History Traits of Spodoptera frugiperda Populations Exposed to Low-Dose Bt Maize. *PloS one* **11**, e0156608, doi:10.1371/journal.pone.0156608 (2016).

29 Bernardi, D. *et al.* Cross-Resistance between Cry1 Proteins in Fall Armyworm (Spodoptera frugiperda) May Affect the Durability of Current Pyramided Bt

Maize Hybrids in Brazil. *PloS one* **10**, e0140130, doi:10.1371/journal.pone.0140130 (2015).

30 Aguirre, L. A. *et al.* Evaluation of Foliar Damage bySpodoptera frugiperda(Lepidoptera: Noctuidae) to Genetically Modified Corn (Poales: Poaceae) in Mexico. *Florida Entomologist* **99**, 276-280, doi:10.1653/024.099.0218 (2016).

31 Midega, C. A. O., Pittchar, J. O., Pickett, J. A., Hailu, G. W. & Khan, Z. R. A climate-adapted push-pull system effectively controls fall armyworm, Spodoptera frugiperda (J E Smith), in maize in East Africa. *Crop Protection* **105**, 10-15, doi:10.1016/j.cropro.2017.11.003 (2018).

32 Kenis, M. *et al.* Telenomus Remus, a Candidate Parasitoid for the Biological Control of Spodoptera Frugiperda in Africa, is already Present on the Continent. *Insects* **10**, doi:10.3390/insects10040092 (2019).

33 Pashley, D. P. Host-associated Genetic Differentiation in Fall Armyworm (Lepidoptera: Noctuidae): A Sibling Species Complex? *Annals of the Entomological Society of America* **79**, 7 (1986).

34 Juárez, M. L. *et al.* Population structure ofSpodoptera frugiperdamaize and rice host forms in South America: are they host strains? *Entomologia Experimentalis et Applicata* **152**, 182-199, doi:10.1111/eea.12215 (2014).

35 Nagoshi, R. N. *et al.* Analysis of strain distribution, migratory potential, and invasion history of fall armyworm populations in northern Sub-Saharan Africa. *Scientific reports* **8**, 3710, doi:10.1038/s41598-018-21954-1 (2018).

36 Nagoshi, R. N. & Meagher, R. L. Behavior and distribution of the two fall armyworm host strains in Florida. *Florida Entomologist* **87**, 10, doi:http://dx.doi.org/10.1653/0015-4040(2004)087[0440:BADOTT]2.0.CO;2 (2004).

37 Meagher, R. L. & Nagoshi, R. N. Differential Feeding of Fall Armyworm (Lepidoptera: Noctuidae) Host Strains on Meridic and Natural Diets. *Annals of the Entomological Society of America* **105**, 462-470, doi:10.1603/an11158 (2012).

38 Nagoshi, R. N., Silvie, P. & Meagher, R. L. Comparison of Haplotype Frequencies Differentiate Fall Armyworm (Lepidoptera Noctuidae) Corn-Strain Populations from Florida and Brazil. *Molecular Entomology* **100**, 8, doi:10.1603/0022-0493(2007)100[954:cohfdf]2.0.co;2 (2007).

39 Nagoshi, R. N. *et al.* Haplotype Profile Comparisons Between Spodoptera frugiperda (Lepidoptera: Noctuidae) Populations From Mexico With Those From Puerto Rico, South America, and the United States and Their Implications to Migratory Behavior. *Journal of economic entomology* **108**, 135-144, doi:10.1093/jee/tou044 (2015).

40 Nagoshi, R. N. *et al.* Using Haplotypes to Monitor the Migration of Fall Armyworm (Lepidoptera Noctuidae) Corn-Strain Populations from Texas and Florida. *Journal of economic entomology* **101**, 8, doi:10.1603/0022-0493(2008)101[742:uhtmtm]2.0.co;2 (2008).

41 Nagoshi, R. N. *et al.* Fall armyworm migration across the Lesser Antilles and the potential for genetic exchanges between North and South American populations. *PloS one* **12**, e0171743, doi:10.1371/journal.pone.0171743 (2017).

42 Unbehend, M. *et al.* Geographic variation in sexual attraction of Spodoptera frugiperda corn- and rice-strain males to pheromone lures. *PloS one* **9**, e89255, doi:10.1371/journal.pone.0089255 (2014).

43 Adamczyk Jr, J. J., Holloway, J. W., Leonard, B. R. & Graves, J. B. Susceptibility of Fall Armyworm Collected from Different Plant Hosts to Selected Insecticides and Transgenic Bt Cotton. *Journal of Cotton Science* **1**, 8 (1997).

44 Nagoshi, R. N. Evidence that a major subpopulation of fall armyworm found in the Western Hemisphere is rare or absent in Africa, which may limit the range of crops at risk of infestation. *PloS one* **14**, e0208966, doi:10.1371/journal.pone.0208966 (2019).

45 Kakumani, P. K., Malhotra, P., Mukherjee, S. K. & Bhatnagar, R. K. A draft genome assembly of the army worm, Spodoptera frugiperda. *Genomics* **104**, 134-143, doi:10.1016/j.ygeno.2014.06.005 (2014).

46 Gouin, A. *et al.* Two genomes of highly polyphagous lepidopteran pests (Spodoptera frugiperda, Noctuidae) with different host-plant ranges. *Scientific reports* **7**, 11816, doi:10.1038/s41598-017-10461-4 (2017).

47 Nandakumar, S., Ma, H. & Khan, A. S. Whole-Genome Sequence of the Spodoptera frugiperda Sf9 Insect Cell Line. *Genome Announcements* **5:e00829-17**, doi:https://doi.org/10.1128/genomeA.00829-17 (2017).

48 Cheng, X. *et al.* A simple bead-based method for generating cost-effective co-barcoded sequence reads. *11 October 2018, PROTOCOL (Version 1) available at Protocol Exchange [+https://doi.org/10.1038/protex.2018.116+].*

49 Wang, O. *et al.* Efficient and unique cobarcoding of second-generation sequencing reads from long DNA molecules enabling cost-effective and accurate sequencing, haplotyping, and de novo assembly. *Genome research* **29**, 798-808, doi:10.1101/gr.245126.118 (2019).

50 Huang, J. *et al.* A reference human genome dataset of the BGISEQ-500 sequencer. *GigaScience* **6**, 1-9, doi:10.1093/gigascience/gix024 (2017).

51 Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-293, doi:10.1126/science.1181369 (2009).

52 Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M. & Jaffe, D. B. Direct determination of diploid genome sequences. *Genome research* **27**, 757-767, doi:10.1101/gr.214874.116 (2017).

53 Luo, R. *et al.* SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* **1**, 18, doi:10.1186/2047-217X-1-18 (2012).

54 Dudchenko, O. *et al.* De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92-95, doi:10.1126/science.aal3327 (2017).

55 Li, L., Stoeckert, C. J., Jr. & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome research* **13**, 2178-2189, doi:10.1101/gr.1224503 (2003).

56 Wang, Y. *et al.* MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic acids research* **40**, e49, doi:10.1093/nar/gkr1293 (2012).

57 Nagoshi, R. N., Goergen, G., Plessis, H. D., van den Berg, J. & Meagher, R., Jr. Genetic comparisons of fall armyworm populations from 11 countries spanning sub-Saharan Africa provide insights into strain composition and migratory behaviors. *Scientific reports* **9**, 8311, doi:10.1038/s41598-019-44744-9 (2019).

58 Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology* **59**, 307-321, doi:10.1093/sysbio/syq010 (2010).

59 Chenna, R. *et al.* Multiple sequence alignment with the Clustal series of programs. *Nucleic acids research* **31**, 3497-3500, doi:10.1093/nar/gkg500 (2003).

60 Triant, D. A., Cinel, S. D. & Kawahara, A. Y. Lepidoptera genomes: current knowledge, gaps and future directions. *Current opinion in insect science* **25**, 99-105, doi:10.1016/j.cois.2017.12.004 (2018).

61 Nowell, R. W. *et al.* A high-coverage draft genome of the mycalesine butterfly Bicyclus anynana. *GigaScience* **6**, 1-7, doi:10.1093/gigascience/gix035 (2017).

62 Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210-3212, doi:10.1093/bioinformatics/btv351 (2015).

63 Parra, G., Bradnam, K. & Korf, I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**, 1061-1067, doi:10.1093/bioinformatics/btm071 (2007).

64 Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA* **6**, 11, doi:10.1186/s13100-015-0041-9 (2015).

65 Saha, S., Bridges, S., Magbanua, Z. V. & Peterson, D. G. Empirical comparison of ab initio repeat finding programs. *Nucleic acids research* **36**, 2284-2294, doi:10.1093/nar/gkn064 (2008).

66 Lorenzi, H. *et al.* Genome wide survey, discovery and evolution of repetitive elements in three Entamoeba species. *BMC genomics* **9**, 595, doi:10.1186/1471-2164-9-595 (2008).

67 Thybert, D. *et al.* Repeat associated mechanisms of genome evolution and function revealed by the Mus caroli and Mus pahari genomes. *Genome research* **28**, 448-459, doi:10.1101/gr.234096.117 (2018).

68 Kakumani, P. K. *et al.* De novo transcriptome assembly and analysis of Sf21 cells using illumina paired end sequencing. *Biology direct* **10**, 44, doi:10.1186/s13062-015-0072-7 (2015).

69 Cheng, T. *et al.* Genomic adaptation to polyphagy and insecticides in a major East Asian noctuid pest. *Nature ecology & evolution* **1**, 1747-1756, doi:10.1038/s41559-017-0314-4 (2017).