Whole genome sequencing and bulked segregant analysis reveal a new mechanism of amitraz resistance in the citrus red mite, Panonychus citri (Acari: Tetranychidae)

Shijiang Yu[1], Lin Cong[2], Hao-Qiang Liu[2], and Chun Ran[3]

[1]Citrus Research Institute, Southwest University/Chinese Academy of Agricultural Sciences
[2]Affiliation not available
[3]Citrus Research Institute, Southwest University/Chinese Academy of Agricultural Sciences, National Engineering Research Center for Citrus

April 28, 2020

Abstract

Amitraz is a broad-spectrum insecticide for the control of aphids, psyllids, ticks and mites. Current evidence suggests that ticks and phytophagous mites have developed strong resistance to amitraz. Previous studies have shown that the I61F mutation in the β-adrenergic octopamine receptor is related to amitraz resistance in ticks, but the mutation was not found in Panonychus citri. We therefore used whole genome sequencing and bulked segregant analysis to identify the mechanism by which P. citri is resistant to amitraz. High-quality assembly of the whole P. citri genome was completed, resulting in a genome of approximately 83.97 Mb and a contig N50 of approximately 1.81 Mb. Gene structure predictions revealed 11,577 genes, of which 10,940 genes were annotated. Trait-associated regions in the genome were mapped with bulked segregant analysis and 38 candidate SNPs were obtained, of which T752C had the strongest correlation with the resistant trait, located at the 5' untranslated region (UTR) of the β-2R adrenergic-like octopamine receptor gene. The mutation resulted in the creation of a short hairpin loop structure in mRNA and gene expression was down-regulated by more than 50% in the amitraz-resistant strain. Validation of the T752C mutation in field populations of P. citri found that the correlation between the resistance ratio and the base mutation was 94.40%. Our results suggest that this 5' UTR mutation of β-2R octopamine receptor gene confers amitraz resistance in P. citri and different species may share different mechanisms of resitance to amitraz.

Introduction

The monitoring and management of citrus pest mites is a key issue for citrus production regions across the world each year. Frequent outbreaks of the citrus red mite, *Panonychus citri* (McGregor) are particularly prominent in China, contributing to significant difficulties for the citrus industry at large. The high frequency of outbreaks is partially due to the reproductive and growth characteristics of the *P. citri* , which can reproduce parthenogenetically, potentially leading to a large number of haploid male mites and setting the foundation for a subsequent outbreak (Tuan et al., 2016). *P. citri* can attach to plants through spinning and netting to establish a habitat that protects against rain erosion or predation. *P. citri* can also use spider silk to sway in the wind in the orchard, providing a natural extension of their feeding range (Clotuche et al., 2011). Moreover, because of improper application of acaricide, a broad-spectrum insecticide, mites have gradually developed resistance to acaricides, thus presenting a formidable challenge for citrus farmers. Worldwide reports of the development of resistance have suggested that *P. citri* has developed resistance to many types of chemical acaricides, such as spirodiclofen (Hu et al., 2010; Ouyang et al., 2012), bifenazate

(Van Leeuwen et al., 2011), hexythiazox (Yamamoto et al., 1996), pyridaben, abamectin, fenpropathrin (Pan et al., 2019), and azocyclotin (Ran et al., 2009). As a result, citrus growers have searched for alternatives to acaricides.

At present, the slow development of new minimally-toxic options in the pesticide market has not meet the consistent demands of growers, which has led growers to turn to an older pesticide that is commonly used to control animal parasites—amitraz. As a broad-spectrum insecticide, it has good performance against aphids, psyllids, ticks, and mites. In recent years, it has also been used to control the *P. citri* . According to the Insecticide Resistance Action Committee (IRAC) classification system, amitraz is an octopamine receptor agonist that can stimulate pest mites to increase signaling through multiple excitatory pathways (Van Leeuwen et al., 2015). Octopamine receptors are G protein-coupled receptors (GPCRs), located on the cell membrane. After receiving a stimulus signal, they activate a sequential series of enzyme activity and the production of second messengers (cAMP, Ca^{2+}, IP3, etc.) (Broeck, 2001). The octopamine receptors of the insects are divided into three major categories: α-adrenergic-like, β-adrenergic-like, and octopamine / tyramine (OCT / Tyr) receptors (Evans & Maqueira, 2005). B-adrenergic-like octopamine receptors are present at various growths stages of *Plutella xylostella* , and amitraz, octopamine, and tyramine can each activate this receptor to cause an intracellular increase in cAMP (Huang et al., 2018). Amitraz is easily metabolized and degraded, and produces semi-amitraz chloride (DMPF), N, N-Dimethylformamide (DMF) and 1-Amino-2,4-dimethylbenzene (DMA) upon degradation. Compared to amitraz**,** DPMF is more likely to activate ∝- and β-adrenergic-like octopamine receptors and increase the levels of cAMP and Ca^{2+} in cells (Kita et al., 2017).

For the sensitive *P. citri* in our laboratory, the median lethal dose (LC_{50}) of amitraz was 29.03 mg/L. A 2009 field resistance monitoring report found a resistance ratio of 5.5, suggesting that field populations have moderate resistance to amitraz (Ran et al., 2009). Indeed, given the large-scale use of this pesticide, there is a high risk for the development of resistance by *P. citri* . Therefore, further investigations of the mechanisms by which *P. citri* can become resistant to amitraz may help guide its rational use in the field. However, the focus of research to date has primarily been on ticks, and there have been no reports on *P. citri.* In a previous study, the octopamine receptor gene of*Rhipicephalus microplus* was cloned with the goal of clarifying the target of amitraz. Analysis of those sequencing results revealed that there is a nonsynonymous mutation in the OCT / Tyr receptor in the Brazilian and Mexican strains that are highly resistant to amitraz (T8P and L22S), but there was no subsequent functional analysis or mutation frequency detection in that study (Chen et al., 2007). However, studies that did measure SNP frequency calculations on these two mutation sites confirmed the correlation between the mutation sites and the development of resistance to amitraz (Baron et al., 2015). In addition to the OCT / Tyr receptors, non-synonymous mutations were also found in the β-adrenergic octopamine receptor gene of *R. microplus* . Through a correlation analysis of the mutation frequency and the phenotypic traits, the I61F site mutation was confirmed to be related to amitraz resistance (Corley et al., 2013). Besides the octopamine receptors, ATP-binding cassette transporters (ABC transporters) have also been shown to be involved in the detoxification and transport of amitraz metabolites. Through immunolocalization and RNAi analysis, it was previously found that ABCB10 of *R. microplus* can transport and detoxify amitraz in midgut cells (Lara et al., 2015); subsequent analysis of differential gene expression confirmed the function of this transporter (Koh-Tan et al., 2016). After treating *Ixodes ricinus* cells with amitraz, the stability of the cells was significantly affected, and the expression of ABC transporters was either increased or decreased to varying degrees (Mangia et al., 2018). Follow-up studies have suggested multiple possible mechanisms for the development of tick resistance to amitraz, including mutations in the β-adrenergic octopamine receptor or mutations in the octopamine/tyramine receptor and changes in the activity of enzymes such as monoamine oxidases (MO) or ATP binding cassette transporter (ABC transporters) (Jonsson et al., 2018).

Previously, we cloned octopamine receptor-related genes from the amitraz-sensitive and -resistant strains of *P. citri* but found no nonsynonymous mutations in the open reading frame (data not yet published). Genomic resequencing technologies, such as bulked segregant analysis (BSA), can help screen trait-determining genes step by step at the omics level. This technology has been widely used in plant population genetics. In 2012,

Van Leeuwen et al. first used BSA to screen for mite resistance mutant genes. In that work, it was found that the I1017F mutation in the chitin synthase 1 gene (*TuCHS1*) of*Tetranychus urticae* determines its resistance to etoxazole (Van Leeuwen et al., 2012). In theory, BSA can only be used to locate single gene recessive traits, but the method can also provide clues for the location of candidate sites responsible for complex polygenic traits. BSA can be used on a sample of many mites to obtain better quality candidate sites, but this would first require the acquisition of high-quality whole-genome sequences of species (Van Leeuwen et al., 2016). Therefore, before investigating the genetic determiners of the resistance of *P. citri* to amitraz, we used second- and third-generation sequencing technologies to assemble the genome of *P. citri* . Then, we performed resistance-sensitive mite hybridization and segregated the traits of the offspring to obtain the population DNA pool, used BSA to locate trait sites, and screened for SNPs related to amitraz resistance. As a result, we found a single SNP that was predicative of amitraz resistance, providing a molecular basis for the identification and management of resistance in mites.

Methods

Source of mites

The initial population of *P . citri* was collected from unmanaged wild citrange in 2005 and raised in isolation without contact with pesticides until the present study. This original population was considered the acaricide-sensitive strain (SS); the amitraz resistant strain (RR) was developed from the SS strain by spraying 200 mg/L of amitraz (diluted from 200 g/L, Arysta LifeScience, America) on a regular basis for several years. The resistance of the RR strain was 81.35-fold higher than the SS strain in 2019.

Genomic extraction

P. citri were randomly taken from the sensitive population and placed in four 1.5 mL enzyme-free tubes. The volume of mites in each tube was 0.5 mL. DNA was extracted from the 4 samples using the Mollusc DNA Kit (Omega Bio-Tek, GA, America) and collected using a single adsorption column. The purity, concentration, and integrity of genomic DNA were quantified by NanoDrop2000, Qubit Fluorometric Quantitation, and 0.35% agarose gel electrophoresis.

Library Construction and Sequencing

The genome was sequenced using the Illumina Hiseq and the third generation Oxford Nanopore Technologies (ONT) platforms. The Illumina Hiseq platform method was carried out as follows: (1) Ultrasound treatment reduced genomic DNA to the target fragment size (350 bp), which was then used to construct a small fragment sequencing library through steps such as end repair, adding adenine, adding adapters, selecting target fragments, and PCR; (2) An Agilent 2100 bioanalyzer and qPCR were used to detect library fragment size and library quantification to determine whether the library met the sequencing standards; (3) The library was fixed to the sequencing chip by bridge PCR; (4) A Hiseq sequencer was used for double-end bp (PE 150) sequencing. The data generated by sequencing was used for the next step of information analysis after quality control. Double-end sequencing data was evaluated (GC distribution statistics, quality values Q20, Q30 evaluation) and filtered to obtain high-quality data (clean reads) for genome size assessment, genome assembly, GC content estimation, determination of heterozygosity rate and the evaluation after assembly. The ONT platform method was applied as follows: (1) The genomic DNA was randomly cut after the sample was qualified; magnetic beads were used to enrich and purify large pieces of DNA, which were cut, and then large fragments were recovered to construct a Nanopore library; (2) End repair for both ends of DNA fragments, followed by a ligation reaction with adding adenine; (3) Accurate quantitative detection of DNA libraries with Qubit Fluorometric Quantitation; (4) Addition of DNA libraries of a known concentration and volume into a Flow cell, and transfer of that Flow cell to the PromethION sequencer

for real-time single molecule sequencing to obtain the original sequencing data. ONT sequencing data was evaluated by preliminary analysis (output evaluation, quality evaluation), filtered low-quality reads, and then sublinks were removed to obtain subreads. After further base correction, high-accuracy data was obtained for genome assembly and evaluation of the assembly result.

Assembly and Evaluation

After filtering out the low-quality and short fragments from the Nanopore raw data, the Canu (Koren et al., 2017) software package was used for error correction assembly, and then the WTDGB (https://github.com/ruanjue/wtdbg) software packed was used to assemble the Canu error correction data. Finally, Quickmerge (Chakraborty et al., 2016) software was used to integrate based on the results from Canu and WTDGB assembly. Calibration was performed using Pilon (Walker et al., 2014) software on Hiseq sequencing data. The assembly results were evaluated on the basis of three measures: (1) comparison ratio of second-generation sequencing reads; (2) BUSCO integrity; and (3) CEGMA integrity.

Genome annotations

Annotation of the repeat sequence was carried out with the help of three software packages: LTR FINDER v1.0.5 (Xu & Wang, 2007), RepeatScout v1.0.5 (Price et al., 2005), and PILER-DF v2.4 (Edgar & Myers, 2005). A genomic repeat sequence database was constructed on the basis of the principles of structural prediction and ab initio prediction. The database was classified using PASTEClassifier (Wicker et al., 2007) and then merged with the database of Repbase (Jurka et al., 2005) to create the final repeat sequence database. RepeatMasker v4.0.6 (Tarailo-Graovac et al., 2009) was used to predict the repeat sequence of the constructed repeat sequence database.

The prediction of gene structure in this work was the result of a combination of three different strategies: *de novo* prediction, homologous species prediction, and Unigene prediction. The EVM v1.1.1 (Haas et al., 2008) software package was used to integrate the prediction results. Genscan (Burge & Karlin, 1997), Augustus v2.4 (Stanke & Waack, 2003), GlimmerHMM v3.0.4 (Majoros et al., 2004), GeneID v1.4 (Blanco et al., 2007), and SNAP (version2006-07-28) (Korf, 2004) were used for homology-based gene prediction; GeMoMa v1.3.1 (Keilwagen et al., 2016) was used for same-based prediction of source species; Hisat v2.0.4 and Stringtie v1.2.3 (Pertea et al., 2016) were used for assembly with reference transcripts, and gene prediction was carried out using TransDecoder v2.0 (http://transdecoder.github.io) and GeneMarkS-T v5.1 (Tang et al., 2015). Transcriptome-based Unigene sequences without reference assembly in the data were used to predict the sequence using PASA v2.0.2 (Campbell et al., 2006). Finally, the prediction results obtained by the above methods were integrated using EVM v1.1.1 and modified with PASA v2.0.2. The original transcriptome data was compared with the genome in TopHat (Trapnell et al., 2009), and the number of bases in exon, intron, and intergenic regions were compared to evaluate the gene prediction results.

Pulling data from the Rfam (Griffiths-jones, 2004) and miRBase (Griffiths-jones, 2006) databases, Infenal 1.1 (Nawrocki & Eddy, 2013) was used for rRNA and microRNA prediction. tRNAscan-SE v1.3.1 (Lowe & Eddy, 1997) was used to identify tRNA.

GenBlastA v1.0.4 (She et al., 2008) alignment was used for the pseudogene annotation to find the homologous gene sequence on the genome of the shielded true locus, and then GeneWise v2.4.1 (Birney, 2004) to find the immature termination code and frameshift mutations in the gene sequence.

The predicted gene sequence was compared with functional databases such as NR (Marchler-Bauer et al., 2011), KOG (Tatusov, 2001), GO (Dimmer et al., 2012), KEGG (Kanehisa, 2000), and TrEMBL (Boeckmann, 2003), and then compared with BLAST v2.2.31 (Altschul et al., 1990). Using InterProScan v5.8-49.0 (Zdobnov & Apweiler, 2001) software, through comparing with PROSITE (Bairroch, 1991), HAMAP (Lima et al., 2009), Pfam (Finn, 2006), PRINTS (Attwood & Beck, 1994), ProDom (Bru, 2004), SMART (Letunic, 2004), TIGRFAMs (Haft, 2003), PIRSF (Wu, 2004), SUPERFAMILY (Gough, 2002), CATH -Gene3D (Lees

et al., 2012), and PANTHER (Thomas et al., 2003) database alignment, we produced Motif annotations for the amino acid sequence of the predicted gene.

Bulked segregant analysis

Crossing and character separation

First, we raised ˜200 sensitive *P. citri* (SS) on lemon leaves. The leaves were placed in a petri dish containing sponges and were surrounded by wet cotton threads to prevent mites from escaping. The feeding device was placed in a 28 °C incubator (L14: D10). Mites were reared for 3 days and then eggs were harvested. After the eggs hatched, mites were cultured until the early mating period (resting period). We then selected about 200 amitraz resistant *P. citri*(RR) and obtained a sufficient number of male mites through parthenogenesis. Approximately 500 pre-mating sensitive female mites (SS) and 600 amitraz resistant male mites (RR) were mixed in a breeding device, and naturally crossed to obtain the F1 generation. After the F1 generation hatched, the population was transferred to insect-free Fructus Aurantii and expanded and isolated for rearing. After the F5 generation, the population was large enough that trait segregation had already occurred. On the basis of resistance genetic analysis data (unpublished), 100 mg/L and 2,500 mg/L amitraz solutions were used to distinguish between the sensitive mites (dead) and the resistant mites (live). Each phenotypic sample was placed in a 1.5 ml centrifuge tube and numbered as F (SS) or F (RR). At the same time, phenotypic samples of the parental generations were collected and numbered as P (SS) or P (RR).

Genomic Resequencing

DNA was extracted from P (SS), P (RR), F (SS), and F (RR) samples using the Mollusc DNA Kit. After passing the quality test, the DNA was randomly broken into fragments of approximately 350 bp by ultrasonic disruption. The DNA fragments underwent end repair, adenine was added to the 3' end, a sequencing adapter was added, samples were purified, and PCR amplification was performed to complete the construction of the sequencing library. The library was re-sequenced by Illumina HiSeq after passing quality inspection. Base-calling analysis was performed using the Illumina Casava 1.8 software package. The raw reads obtained by sequencing were filtered, resulting in clean reads for subsequent information analysis. The bwa (Li & Durbin, 2009) software package was used to place the obtained sequencing reads on the reference genome. By comparing the positions of the clean reads to the reference genome, the sequencing depth and genome coverage of each sample were counted, and mutations were detected.

SNP detection and annotation

SNPs were primarily detected using the GATK (Mckenna et al., 2010) software toolkit. According to the positioning results of clean reads in the reference genome, Picard (http://sourceforge.net/projects/picard/) was used to perform mark duplicates, GATK was used to perform local realignment, and base recalibration and other pre-processing steps were used to ensure the accuracy of the detected SNP. Then, we used GATK to detect Single Nucleotide Polymorphisms (SNPs), filtered the results, and thus obtained the final SNP locus set. Small fragments of insertions and deletions between the sample and the reference genome (Small InDel: 1-5bp) were detected according to the location of the sample's clean reads on the reference genome.

SnpEff (Cingolani et al., 2012) software was used to annotate variation (SNP, Small InDel) and predict the effect of variation. On the basis of the position of the mutation site on the reference genome and the gene position information on the reference genome, the region where the mutation site appeared (e.g. in the intergenic region, gene region, CDS region, etc.) and the effect of the mutation (synonymous mutation, non-synonymous mutation, etc.) was determined.

Association Analysis

Prior to association analysis, we filtered out SNP sites with multiple genotypes and then filtered out SNP sites with read support less than 4. Then, we filtered out SNP sites with consistent genotypes and mixed

pooled genes between SNP loci that did not originate from the recessive parents, finally resulting in high-quality credible SNP loci. The association analysis was comprehensively evaluated using two algorithms: Euclidean Distance (ED) and SNP-index.

$$P = 1 - \sum_{x=0}^{y-1} \frac{\binom{K}{x}\binom{M-K}{N-x}}{\binom{M}{N}}$$

$$P = 1 - \sum_{x=0}^{y-1} \frac{\binom{K}{x}\binom{M-K}{N-x}}{\binom{M}{N}}$$

Euclidean distance algorithm: We used sequencing data to find markers with significant differences between DNA pools and evaluated areas associated with different traits (Hill et al., 2013). A larger ED value indicates a greater difference in the markers between the two pools. Differences in the extent of sequencing between pools can lead to biased ED values. In order to eliminate this error, the frequency of each base at each point was used instead of the absolute depth to calculate the ED value. At the same time, to eliminate background noise, the original ED values were processed to the 5^{th} power. We calculated the median value of all sites' ED values plus 3 times of the standard deviation to determine the association threshold. We screened out sites that exceeded the association threshold as candidate association sites and counted the number of SNPs that differed between the two mixed pools on each contig and the number of candidates associated sites. Then, we calculated the probability of associated sites on each contig using the following equation: :; The Benjamini–Hochber method was used to perform multiple test corrections on the probabilities of the associated sites. Finally, the contigs (FDR <0.01) containing the associated sites were screened.

$$P = 1 - \sum_{x=0}^{y-1} \frac{\binom{K}{x}\binom{M-K}{N-x}}{\binom{M}{N}}$$

$$P = 1 - \sum_{x=0}^{y-1} \frac{\binom{K}{x}\binom{M-K}{N-x}}{\binom{M}{N}}$$

SNP-index algorithm: Correlation analysis was performed based on the difference in genotype frequency between mixed pools (Abe). For Δ (SNP-index) statistics, the stronger the correlation between SNP and traits, the closer Δ (SNP-index) is to 1. Because the reference genome was not assembled at the chromosome level, the following strategy was used in association analysis: we first calculated the threshold value by Permutation Test and screened out the sites that exceeded the association threshold. Next, the number of different SNPs between the pools on each contig were counted, as well as the association sites. We then calculated the probability of associated sites on each contig using the follow equation(, again, using Benjamini–Hochber's method to perform multiple test corrections for the probability of associated sites. Finally, we selected contigs that contained associated sites (FDR <0.01).

The results obtained by the above two association analysis methods were used to obtain the intersection and final candidate contigs.

Annotation of candidate regions

We annotated the types of SNP mutations in candidate regions, including SYNONYMOUS_CODING, NON_-SYNONYMOUS_CODING, and so on; BLAST (Altschul, 1997) was used to perform in-depth annotations pulled from multiple databases (NR, Swiss-Prot, GO, KEGG, COG) on the coding genes in the candidate interval.

Analysis of mRNA secondary structure and verification of mutant genes

The RNAfold WebServer (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi) was used to analyze the predicted secondary structure of the mRNA (minimum free energy method, MFE). The expression of the target gene in the sensitive strain (SS) and resistant strain (RS) of *P. citri* was detected by real-time fluorescent quantitative PCR as previously described (Yu et al., 2015). The specific primers were F: TCAAGTGATTCGGGTTACGGG and R: CGGCACAGGCTAACGAGACAA. Candidate SNPs were verified by Sanger sequencing, and PCR amplification primers were designed using Primer Premier 6 software: F: AACCTTCAGAATCAGTGTAATCATC; R: CTTCCATAGATTGTTGTTGTTGTTG. DNA templates were prepared using a Tissue Ex-Amp PCR Kit (Applied Biological Materials Inc, Vancouver, Canada). The method for preparing the DNA template of a single mite was as follows: We isolated ˜20 to 30 *P. citri* , and each mite was placed in a 200 uL PCR tube. The sample was ground with a sterile toothpick, and 20 uL of lysate was added and placed in a PCR instrument, where it was heated at 55 °C for 10 min followed by 95 °C for 5 min. Then, 2 uL of DNA template were aspirated and used for PCR amplification of the β-2R octopamine receptor gene containing the SNP region with specific primers. The genotypes of SNPs were analyzed and counted according to the sequencing results. The mutation frequency was calculated according to the following equation:

$$P = 1 - \sum_{x=0}^{y-1} \frac{\binom{K}{x}\binom{M-K}{N-x}}{\binom{M}{N}}$$

Results

Statistics of genome sequencing results

We constructed a 350 bp library and obtained 23.23 Gb of high-quality data after sequencing and filtering with the Illumina Hiseq platform (SRA accession: PRJNA600040). The total sequencing depth was approximately 301×, the Q20 ratio of sequencing data was above 96.81%, and the Q30 ratio was above 91.25%. The GC content of the genome was about 31.65%, as was roughly expected. (Table 1).

A total of 66.03 Gb of raw data was obtained by the third generation of Nanopore platform sequencer (SRA accession: PRJNA600040). After filtering out short fragments and low-quality data through the use of multiple filter adapters, a total of 25.23 Gb of clean data were obtained, of which the N50Len was 19,718 bp and MeanLen was 18,438 bp (Table 2).

In order to determine whether the extracted sample of DNA was contaminated, 10,000 single-ended reads were randomly selected from the 350 bp library and compared with the NT database by BLAST. Of these, 4.83% of the reads could be matched to the NT library. Among the reads that were successfully matched, reads corresponding to the *P. citri* and *T. urticae* gene data accounted for 36.85% and 35.19%, respectively (Fig.1); 2,000 single-end reads were randomly selected from the three generations of clean reads. A total of 25 (1.25%) reads could be compared, of which 56% were matched to *P. citri* gene data (Fig. 1a). Together, these results suggested relatively few unexpected matches to host plant DNA, so the sample sequencing data was determined to be free of contamination.

In order to evaluate the cytoplasmic DNA content in the sequencing data, a 350 bp library was used to perform a SOAP alignment with the mitochondrial sequence of *P. citri* (GeneBank ID: HM189212.1). The number of reads on the double-ended ratio was 144,768, accounting for 0.09% of the total reads, and the number of reads on the single-ended ratio was 190,764, accounting for 0.12% of the total reads, both of which were lower than the empirical value of 5%. Therefore, these findings indicated that the extranuclear DNA content of the sequencing data was very low and thus was not likely to affect the assembly of the later genome.

The data from the 350 bp library were used to construct a Kmer distribution map of k=19, and the genome size, repeat sequence ratio, and heterozygosity were evaluated. The kmer depth corresponding to the average kmer depth (i.e., the primary peak) was 249 (Fig. 1b). Sequences with a Kmer depth of more than twice the depth of the main peak (i.e. greater than 499) were repeated sequences. It was estimated that the content due to repeated sequences was approximately 13.45%. There was no obvious heterozygous peak, and the estimated heterozygosity was low (about 0.64%). Therefore, the obtained genome of this species belonged to a simple genome which was conducive to the subsequent construction of detailed genomic maps.

Genome assembly and evaluation

The Canu and WTDBG software packages were used to correct and assemble Nanopore raw data, and Quickmerge software was used for integration. Finally, the second-generation sequencing data was correc-

ted by Pilon software. The final genome assembly size was about 83.97 Mb (GeneBank accession number: JAAABK000000000), Contig N50 was about 1.81 Mb, and GC content was 31.33%, with a Gap total length value of 0 (Table 3).

The second-generation sequencing data was used to compare the reads to the reference genome, and the assembly quality was evaluated by comparison. The results showed that the number of clean reads mapped to the reference genome accounted for 96.95% of all Clean Reads, and the value of Properly_mapped (%) was 94.36% (Table 4), indicating that the quality of the genome assembly was intact. The BUSCO v2.0 (Simao et al., 2015) and CEGMA v2.5 (Parra et al., 2007) software packages were used to assess the integrity of the genome assembly, in combination with an arthropoda database containing 1,066 conserved genes from arthropods. Among the genes we assembled, a total of 975 complete BUSCO genes were found, of which 878 were single-copy genes, 24 were Fragmented BUSCO genes, and only 67 genes were not found in the arthropoda library. Completeness of the genome, as assessed by BUSCO, was 91.46% (Table 4). CEGMA contains 458 eukaryotic conserved genes. Among them, 448 sequences (97.82%) were identified in the assembled genome, and 243 of the 248 highly conserved sequences (97.98%) were found in the assembled genome (Table 4), suggesting that the assembled genome was complete.

Genome annotation analysis

Data analysis using repeat sequence prediction software yielded a repeat sequence of 52,069 treaties at 13.93 Mb, accounting for 16.6% of the final genome assembly (Table 5).

The gene structure was determined through a combination of three different strategies: *de novo* prediction, homologous species prediction, and Unigene (RNAseq) prediction. Then, the prediction results were integrated using EVM v1.1.1 software to finally obtain 11,577 genes. The average gene length was 3,942.13 bp, the average exon length was 2,003.09 bp, and the average intron length was 1,921.04 bp (Table 6). The transcriptome database of *P. citrus* was compared with the predicted genes of the genome, and the number of bases in the exon, intron, and intergenic regions were statistically compared. Among them, the number of bases determined as belonging to exons accounted for 82.98%, introns for 3.77%, and intergenic for 13.24%. The total number of genes supported by homology prediction and Unigene prediction was 10,967 (Fig. 1c), accounting for 94.7% of the predicted gene number (11,577). Combined, these data suggest that the quality of genetic prediction was relatively high.

Non-coding RNAs include microRNAs, rRNAs, and tRNAs, which each have a diverse range of known functions. Through prediction, 23 miRNAs, 49 rRNAs, and 117 tRNAs were identified, including 12, 4, and 20 RNA families, respectively. Pseudogenes have a similar sequence to functional genes but have lost their original function due to mutations such as insertions and deletions. Through comparison and search analyses, 238 pseudogenes were predicted with an average length of 2,694.9 bp.

Through gene function annotation analysis, a total of 10,940 genes were annotated on the basis of 7 databases such as NR (Table 7), accounting for 94.50% of the total number of predicted genes (11,577). There were 4,032 genes annotated and classified in the GO database and 5,999 GO terms belonged to the "cellular component" category, 4,611 terms belonged to the "molecular function" category, and 9,208 terms belonged to the "biological process" category. There were 7,849 genes annotated and classified in the KOG database, of which 1,416 genes belonged to the "general function prediction only" category and 1,170 genes belonged to the "signal transduction mechanisms" category. In the TrEMBL database, using the genome of *T. urticae* as a control, 10,905 (94.20%) genes were annotated. A total of 5,668 genes were annotated in the KEGG database representing genes from 268 pathways. The most commonly involved pathways were related to the "lysosome" (ko04142), which had 177 genes. Motif annotation analysis was performed on the amino acid sequences of all predicted genes, and 1,116 motif protein structures and 20,078 domain protein structures were found.

Screening of metabolism-related genes at the genome-wide level

When an exogenous compound enters a mite, it triggers a series of metabolic detoxification processes intended to protect against toxicity. There are many proteins involved in this detoxification process in mites. At present, the well studies detoxification elements include P450 monooxygenases (P450s), glutathione S-transferases (GSTs), the carboxylesterase family, ATP-binding cassette transporters (ABC transporters), major facilitator superfamily (MFS), lipocalins, uridine diphosphate glucuronyl transferases (UGTs), and others, each of which play an important role in various processes of oxidation, hydrolysis, encapsulation, and metabolite transport of heterologous compounds. Among the 10,940 annotated genes, we screened the above-mentioned metabolic-related genes, performed phylogenetic analysis (Neighbor-Joining method, Bootstrap replications = 2,000) using MEGA 7.0 software and performed subfamily classification. P450s are oxido-reductases, and 54 P450s were found in the *P. citri* genome that were related to processing of exogenous compounds. Among them, the representation of members of the CYP4 subfamily was particularly strong (26 genes). In contrast, mitochondrial CYP had only 5 genes (Table 8). Phylogenetic analysis found that genes of the same subfamily tended to be clustered together (Fig. 2a); 19 GSTs genes were retrieved from *P. citri* , and the clusters were grouped into 5 subfamilies of mu, delta, omega, zeta, and kappa, of which the mu and delta subfamily genes appeared to be dominant, with 9 and 7 genes, respectively (Fig. 2b). Screening for carboxylesterase family genes in the Pfam database, 46 gene clusters were scattered, including esterase, acetylcholinesterase, and carboxylate esterase, neuroligin, and others, of which esterase accounted for more than half of the identified clusters. ABC transporters include 8 sub-families of A˜H, and 7 of the A˜G sub-families have been found in the *P. citri* . No sub-family of H was found. The C and G sub-families were the largest, with 32 and 38 genes, respectively. We also found 80 MFS transporters, 16 lipocalins, and 39 UGTs in the genome.

Genomic resequencing analysis

Statistics of Sequencing Data

By mapping the BSA traits to explore the molecular mechanisms of the development of amitraz resistance in *P. citri* , we can accurately locate resistance-correlated SNPs. Genomic resequencing was performed on the 4 DNA samples collected from the parents and offspring, and a total of 76.86 Gbp of data were obtained (SRA accession: PRJNA600137). The clean reads obtained after filtering were a total of 76.12 Gbp, and the Q30 was above 89% (Table 9). The average alignment efficiency between the sample and the reference genome was 91.51%, the average coverage depth was 205.50×, and the genome coverage was 98.78% (with at least one base of coverage).

SNP detection and annotation

The GATK toolkit was used to summarize all the different mutation sites between samples, including SNPs and small fragment insertions and deletions (Small InDel: 1-5 bp). In the parents, there were 66,718 SNP loci between sensitive P (SS) and resistant P (RR) strains, among which 3,775 SNPs caused non-synonymous mutations in the gene; 35,090 Small InDel were also found, including 241 FRAME_SHIFT, 238 CODON_-INSERTION, and 186 CODON_DELETION (Fig. 3a). In the progeny pool, there were 68,285 SNP loci between sensitive F (SS) and resistant F (RR), among which 3,857 SNPs caused non-synonymous mutations in the gene; 35,615 Small InDel were also found, including 231 FRAME_SHIFT, 240 CODON_INSERTION, and 183 CODON_DELETION (Fig. 3b).

Association Analysis

The obtained SNPs were filtered to obtain 32,019 high-quality, reliable SNPs. Based on the Euclidean Distance (ED) algorithm and the probability (P) of enrichment of associated sites, 11 contigs containing significantly enriched associated sites were finally selected. Based on the SNP-index algorithm and the probability value P of enriched association sites (FDR <0.01), 14 contigs containing significantly enriched

association sites were screened. The intersection of the two algorithms was taken, leading to the finding of a total of 2 candidate regions: contig00002 and contig00033. The contig00002 region contained 289 SNPs, of which the ED method was associated with 50 SNPs and the SNP-index method was associated with 182 SNPs. The contig00033 region contained 535 SNPs, of which the ED method was associated with 13 SNPs and the SNP-index method was associated with 269 SNPs (Fig. 3c).

BLAST was used to perform an in-depth annotation using information from multiple databases (NR, Swiss-Prot, GO, KEGG, COG) for the coding genes in the two candidate intervals, and a total of 222 genes were annotated. In the COG annotation, the category with the most gene participation in candidate regions was "R: General function prediction only" (28 genes), followed by "K: Transcription" (9 genes) (Fig. 3d). 86 genes were annotated in the GO database and distributed among 530 GO terms. Among the three GO classifications given to these genes, the number of GO terms included in the category of "biological process" was the largest (259 GO terms). Among the 34 GO subclasses, the top three genes with the most participation in candidate regions were the "metabolic process" (44 genes), the "cellular process" (43 genes), and the" single-organism process" (42 genes) (Fig. 3e). In the KEGG metabolic pathway, 5 genes were predicted to participate in the Wnt signaling pathway and 4 genes each participated in the mTOR signaling pathway, phagosome, inositol phosphate metabolism, and protein processing in the endoplasmic reticulum. The KEGG pathway also predicted that 2 genes were important for ABC transporters (Fig. 3f).

Comparing the SNPs of resistant parents and sensitive parents, there were 91 NON_SYNONYMOUS_CODING, 15 UTR_5_PRIME mutations, 3 STOP_LOST, 1 STOP_GAINED, and 1 START_GAINED in the two candidate regions. There were 78 NON_SYNONYMOUS_CODING, 15 UTR_5_PRIME mutations, 3 STOP_LOST, 1 STOP_GAINED, and 1 START_GAINED mutations between the resistant and sensitive offspring. Through SNP-index and ED association analysis, 38 SNPs (7 at contig00002, 31 at contig00033) were jointly owned in "P (SS) *VS*P (RR)" and "F (SS) *VS* F (RR)", including 29 NON_SYNONYMOUS_CODING and 9 UTR_5_PRIME (Fig. 4a). All the above SNPs mutation types may cause changes in phenotypic species traits.

The ΔSNP-index and ED values were used to evaluate the correlation between SNPs and traits. The larger the ED value, the closer ΔSNP-index is to 1, indicating a stronger correlation between SNPs and traits. In the contig00002 interval, the three sites with the highest evaluation of candidate SNPs were 309075, 309203, and 310326, the ΔSNP-index values of all three were 1, the ED value were 1.41, and the three SNPs were located on the EVM0003193 gene, However, the gene has not been annotated in any of the 7 databases used in this study, so related gene information was not available (Fig. 4b). In the congtig00033 interval, the SNP with the most relevant trait was located at the 797168 site (ΔSNP-index = 1, ED = 1.55), the gene function annotation found that this site (at 752bp) is located in the 5'-untranslated region (5' UTR) of the of the β-2R octopamine receptor gene (GenBank accession numbers: MN928575 and MN928576). Here, the T base mutated to the C base (T752C) (Fig. 4b). The β-2R octopamine receptor gene was not interrupted by introns and had an open reading frame, but no SNPs were found in it; the sequence of the receptor protein was not changed in resistant mites.

μPNA στρυστυρε ανδ εξπρεσσιον λεελ οφ β-2P οστοπαμινε ρεσεπτορ

The 5' UTR base mutation does not affect the structure and function of the protein, but if the mutation causes the secondary structure of the mRNA to change, it may affect the stability of the mRNA or hinder the binding or migration of the ribosome, resulting in the inability to translate the mRNA to protein. Therefore, we compared the predicted secondary structure of mRNA with and without this base mutation using the *RNAfold* WebServer. Because the U base mutated to the C base, two consecutive CC bases formed a complementary pair with two adjacent GG bases. The existence of this interaction forced the local secondary structure of the mRNA to change from a long stem-loop structure to three short hairpin structures (Fig. 4c). The free energy of the thermodynamic ensemble changed from the original -46.97 kcal/mol to a predicted -51.09 kcal/mol; the free energy decreased, and the secondary structure of mRNA was predicted to be stable. After treatment of SS *P. citri* with amitraz (SS-induced), the relative expression

of the β-2R octopamine receptor was 0.79 ± 0.03, significantly lower than the untreated sensitive control group (SS) (P <0.05) (Fig. 4d). Without amitraz treatment, the relative expression in the RS was 0.42 ± 0.04; after amitraz treatment (RS-induced), the relative expression was 0.32 ± 0.02, indicating a significant down-regulation (more than 50%) compared with control group. These findings suggest that the level of gene expression in the amitraz-resistant strain of *P. citri* is lower than that of the sensitive strain.

SNP verification in field mite populations

P. citri were collected from different citrus producing areas in China (SC-MS, YN-NS, YN-XC, GX-YL, JX-GZ) for the evaluations of resistance in these different populations. Virulence was evaluated as previously described (Yu et al., 2016). Among these regions, the field population with the highest resistance to amitraz was GX-YL, with a resistance multiple of 75.64, followed by YN-XC, with a resistance ratio of 51.05. The populations with relatively low resistance were YN-NS and SC-MS, with resistance ratios of 7.55 and 10.56, respectively (Table 10).

A comparison of mutation frequencies showed that, in the RR mites, the candidate sites were all C bases, but in the SS line were all T bases; in the GX-YL population, which had the highest resistance ratio, the T752C mutation frequency was 100%, with all isolated mites having a genotype that was the C/C homozygote. The mutation frequency in the YN-XC population was 96.67%, however the YN-NS population, which had lower resistance ratio, had a T752C mutation frequency of only 36.67%. Three genotypes (T/T, T/C, C/C) were found in the YN-NS population using Sanger sequencing (Fig. 4e), and these three genotypes were also found in the SC-MS population. The T752C mutation frequency was found to be 36.67%. The correlation between the resistance multiplier and the mutation frequency was evaluated by Pearson correlation analysis. The analysis found that the correlation between the two was 94.40%, indicating that the T752C mutation was directly correlated with the development of resistance.

Discussion

In this study, we used the Illumina Hiseq and the Nanopore platforms to complete a high-quality assembly and annotation of the genome of the*P. citri* . The contig N50 was about 1.81 Mb, sufficiently large that it was not necessary to construct a scaffold assembly. The genome size of the *P. citri* was about 83.97 Mb. Multi-dimensional assembly quality assessments showed that the genome assembly was of good quality. A total of 11,577 genes were assembled, of which 10,940 genes were annotated. As of now, in the NCBI datebase, there are 22 entries for genomes of species belonging to the subclass Acari, such as *T. urticae* , *Dermanyssus gallinae, Dermatophagoides farinae* ,*Varroa jacobsoni* , *R. microplus* , etc. (https: //www.ncbi.nlm.nih.gov/genome/?term=Acari), with genome sizes ranging from 40 Mb to 7,400 Mb. Among the Tetranychoidea entries, the genomes of two species have been sequenced: *T. urticae* (Grbic et al., 2011) and *Brevipalpus yothersi* (Navia et al., 2019), with genome sizes of 90.82 Mb and 71.16 Mb, respectively. The genome size of the *P. citri* lies between the two, closest to the two-spotted spider mite, belonging to Tetranychidae. The search for metabolism- and detoxification-related genes in *P. citri* resulted in 54 P450s, 19 GST, 46 carboxylesterase family genes, and 92 ABC transporter genes. The number of these genes in the *T. urticae*genome was 81, 31, 71 and 103, respectively (Van Leeuwen et al., 2016). The number of genes related to metabolic detoxification was higher in*T. urticae* than the *P. citri* , potentially due to the omnivorous nature of *T. urticae* (compared with *P. citri* , which primarily feed from citrus trees). The presence of more detoxification genes suggests that, in the process of the development of resistance against exogenous toxins, *T. urticae* could have an advantage relative to *P. citri.*

Genomic resequencing technologies, such as BSA, QTL, GWAS, etc., can be used to perform correlation analyses on trait-related genes with accurate genomic references. In research on insects, resistance traits are often caused by mutations in target genes. Therefore, screening for high mutation frequency SNP sites related to extreme traits is often the key to studying resistance. Compared with QTL and GWAS, BSA is a more

suitable approach and has been widely used in resistance research. After Van Leeuwen et al. first applied the method to the study resistance mechanisms in *T. urticae* c, BSA has been used to study the target mutations of *T. urticae* against clofentezine, hexythiazox, and other mite growth inhibitors (Demaeght et al., 2014), as well as the mechanism of resistance to phosphine by *Tribolium castaneum* (Jagadeesan et al., 2013) and the mechanism of resistance of *Spodoptera exigua* against Bt toxins (Park et al., 2014). Research on the mechanisms of resistance of *T. urticae* against Mitochondrial Electron Transport Inhibitors of complex I (METI-I) acaricides used a high-resolution QTL mapping method to screen multiple related genes that might have been involved in mite resistance (Snoeck et al., 2019). The method of population genetic analysis, based on genomic resequencing, requires a clear knowledge of the routes of inheritance (e.g. single- or multi-gene inheritance, and recessive or dominant expression). This level of analysis can be achieved through reciprocal cross and backcross experiments with strains of different traits. It has previously been confirmed that the resistance of *T. urticae* to ethizole occurs as a result of the recessive inheritance of a single gene (Van Leeuwen et al., 2012), resistance to chlorfenapyr is an incomplete recessive inheritance controlled by multiple genes (Van Leeuwen et al., 2004), resistance to spirodiclofen is an incomplete dominant inheritance controlled by multiple genes (Van Pottelberge et al., 2009a), and resistance to pyridaben is an incomplete dominant inheritance determined by a single gene on the chromosome (Pottelberge et al., 2009b).

We performed a correlation analysis on the results of BSA through two methods, the ΔSNP-index and the ED method, helped narrow the range of candidate genes. In the end, we screened the SNP with the strongest correlation, which was located at base 752 of 5' UTR, upstream of the start codon of the β-2R octopamine receptor gene (T752C). The populations of *P. citri* from different regions were verified by Sanger sequencing. Of note, there was a strong correlation between the mutation frequency and the extent of resistance of mites sampled from various citrus farms around China. Although 5' UTR base mutations do not cause changes in the protein structure, they may cause changes in the secondary structure of mRNA. Through predictive analysis, it was found that the 5' end of the mRNA of the gene with the mutation does have two short hairpin structures that may affect the efficiency protein translation (Grens & Scheffler, 1990; Gray & Hentze, 1990). Amitraz is an octopamine receptor agonist; if the expression of the octopamine receptor on the cell membrane is decreased, it will reduce the stimulatory signal provided by amitraz. In fact, the expression of this receptor gene was down regulated by more than 50% in resistant strains. We speculate that this is the cause of the resistance of *P. citri*to amitraz. In prokaryotic mRNAs, if base mutations occur in the 5' UTR region, the extent of expression of transcripts and proteins is typically affected. This suggests that UTR has a generally unrecognized but crucial role in transcription at the DNA or RNA level (Berg et al., 2009). Relatedly, a 4 base mutation in the 5' UTR of the whiB7 gene of*Mycobacterium tuberculosis* may be related to its resistance to kanamycin (Chakravorty et al., 2015). In rice, the G98A base mutation in the 5' UTR of the *BML* gene may cause a significant reduction of transcript abundance compared to wild-type (Akhter et al., 2018). As a next step, to test these hypotheses, we will investigate whether this base mutation at the β-2R octopamine receptor leads differential expression of proteins in eukaryotic protein expression systems.

Availability of data and materials

The raw data of the genome sequencing and bulked segregant analysis have been uploaded to the NCBI Sequence Read Archive (SRA) database, and the accession numbers are PRJNA600040 and PRJNA600137, respectively. Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession JAAABK000000000. The version described in this paper is version JAAABK010000000. Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the accession GIIF00000000. The version described in this paper is the first version, GIIF01000000. The gene structures of β-2R octopamine receptor and the sequences alignment between resistant strain and sensitive strain of *P.citri* can be found in GeneBank datebase with the accession number MN928575˜ MN928576.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was supported by the National Key Research and Development Program (2018YFD0201508) ; the National Agricultural Innovation Project (CAAS-XTCX-2016013); the Chongqing Scientific Research Project (cstc2016shms-ztzx80003, cstc2017shms-xdny0443); the Fundamental Research Funds for the Central Universities (XDJK2015C014) and the Research Fund for the Doctoral Program of Southwest University (SWU116061).

Authors' contributions

C.R. conceived the study. C.R. and S.J.Y. designed the study. S.J.Y. and L.C. contributed to the bioinformatic and statistical analyses. S.J.Y., L.C., H.Q.L. and C.R. contributed to the functional analyses. C.R. and H.Q.L. contributed to the supervision. S.J.Y contributed to the writing, with contributions from all authors. All authors read and approved the final manuscript.

References

Abe, A., Kosugi, S., Yoshida, K., Natsume, S., Takagi, H., Kanzaki, H., ... Terauchi, R. (2012). Genome sequencing reveals agronomically important loci in rice using MutMap. *Nature Biotechnology* ,*30* (2), 174–178. doi: 10.1038/nbt.2095

Akhter, D., Qin, R., Nath, U., Alamin, M., Jin, X., & Shi, C. (2018). The brown midrib leaf (bml) mutation in rice (Oryza sativa L.) causes premature leaf senescence and the induction of defense responses. *Genes* , *9* (4), 203. doi: 10.3390/genes9040203

Altschul, S. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* ,*25* (17), 3389–3402. doi: 10.1093/nar/25.17.3389

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

Attwood, T. K., & Beck, M. E. (1994). PRINTS–a protein motif fingerprint database. *"Protein Engineering, Design and Selection,"* *7* (7), 841–848. doi: 10.1093/protein/7.7.841

Bairoch, A. (1991). Prosite: a dictionary of sites and patterns in proteins. *Nucleic Acids Research* , *19* (suppl), 2241–2245. doi: 10.1093/nar/19.suppl.2241

Baron, S., van der Merwe, N. A., Madder, M., & Maritz-Olivier, C. (2015). SNP analysis infers that recombination is involved in the evolution of amitraz resistance in *Rhipicephalus microplus* .*PLOS ONE* , *10* (7), e0131341. doi: 10.1371/journal.pone.0131341

Berg, L., Lale, R., Bakke, I., Burroughs, N., & Valla, S. (2009). The expression of recombinant genes in *Escherichia coli* can be strongly stimulated at the transcript production level by mutating the DNA-region corresponding to the 5'-untranslated part of mRNA.*Microbial Biotechnology* , *2* (3), 379–389. doi: 10.1111/j.1751-7915.2009.00107.x

Birney, E. (2004). GeneWise and Genomewise. *Genome Research* ,*14* (5), 988–995. doi: 10.1101/gr.1865504

Blanco, E., Parra, G., & Guigó, R. (2007). Using geneid to identify genes. In *Current Protocols in Bioinformatics* . doi: 10.1002/0471250953.bi0403s18

Boeckmann, B. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Research* , *31* (1), 365–370. doi: 10.1093/nar/gkg095

Broeck, J. Vanden. (2001). Insect G protein-coupled receptors and signal transduction. *Archives of Insect Biochemistry and Physiology* ,*48* (1), 1–12. doi: 10.1002/arch.1054

Bru, C. (2004). The ProDom database of protein domain families: more emphasis on 3D. *Nucleic Acids Research* , *33* (Database issue), D212–D215. doi: 10.1093/nar/gki034

Burge, C., & Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. *Journal of Molecular Biology* ,*268* (1), 78–94. doi: 10.1006/jmbi.1997.0951

Campbell, M. A., Haas, B. J., Hamilton, J. P., Mount, S. M., & Robin, C. R. (2006). Comprehensive analysis of alternative splicing in rice and comparative analyses with Arabidopsis. *BMC Genomics* , 7, 327. doi: 10.1186/1471-2164-7-327

Chakraborty, M., Baldwin-Brown, J. G., Long, A. D., & Emerson, J. J. (2016). Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. *Nucleic Acids Research* ,*19* , 19. doi: 10.1093/nar/gkw654

Chakravorty, S., Lee, J. S., Cho, E. J., Roh, S. S., Smith, L. E., Lee, J., ... Alland, D. (2015). Genotypic susceptibility testing of*Mycobacterium tuberculosis* isolates for amikacin and kanamycin resistance by use of a rapid sloppy molecular beacon-based assay identifies more cases of low-level drug resistance than phenotypic lowenstein-jensen testin. *Journal of Clinical Microbiology* ,*53* (1), 43–51. doi: 10.1128/JCM.02059-14

Chen, A. C., He, H., & Davey, R. B. (2007). Mutations in a putative octopamine receptor gene in amitraz-resistant cattle ticks. *Veterinary Parasitology* , *148* (3–4), 379–383. doi: 10.1016/j.vetpar.2007.06.026

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., ... Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* ,*6* (2), 80–92. doi: 10.4161/fly.19695

Clotuche, G., Mailleux, A.-C., Astudillo Fernández, A., Deneubourg, J.-L., Detrain, C., & Hance, T. (2011). The formation of collective silk balls in the spider mite *Tetranychus urticae* Koch.*PLoS ONE* , *6* (4), e18854. doi: 10.1371/journal.pone.0018854

Corley, S. W., Jonsson, N. N., Piper, E. K., Cutulle, C., Stear, M. J., & Seddon, J. M. (2013). Mutation in the Rm AOR gene is associated with amitraz resistance in the cattle tick *Rhipicephalus microplus* .*Proceedings of the National Academy of Sciences* , *110* (42), 16772–16777. doi: 10.1073/pnas.1309072110

Demaeght, P., Osborne, E. J., Odman-Naresh, J., Grbić, M., Nauen, R., Merzendorfer, H., ... Van Leeuwen, T. (2014). High resolution genetic mapping uncovers chitin synthase-1 as the target-site of the structurally diverse mite growth inhibitors clofentezine, hexythiazox and etoxazole in *Tetranychus urticae* . *Insect Biochemistry and Molecular Biology* , *51* , 52–61. doi: 10.1016/j.ibmb.2014.05.004

Dimmer, E. C., Huntley, R. P., Alam-Faruque, Y., Sawford, T., O'Donovan, C., Martin, M. J., ... Apweiler, R. (2012). The UniProt-GO Annotation database in 2011. *Nucleic Acids Research* ,*40* (D1), D565–D570. doi: 10.1093/nar/gkr1048

Edgar, R. C., & Myers, E. W. (2005). PILER: identification and classification of genomic repeats. *Bioinformatics* ,*21* (Suppl 1), i152–i158. doi: 10.1093/bioinformatics/bti1003

Evans, P. D., & Maqueira, B. (2005). Insect octopamine receptors: a new classification scheme based on studies of cloned Drosophila G-protein coupled receptors. *Invertebrate Neuroscience* , *5* (3–4), 111–118. doi: 10.1007/s10158-005-0001-z

Finn, R. D. (2006). Pfam: clans, web tools and services. *Nucleic Acids Research* , *34* (90001), D247–D251. doi: 10.1093/nar/gkj149

Gough, J. (2002). SUPERFAMILY: HMMs representing all proteins of known structure. SCOP sequence searches, alignments and genome assignments.*Nucleic Acids Research* , *30* (1), 268–272. doi: 10.1093/nar/30.1.268

Gray, N. K., & Hentze, M. W. (1990). Regulation of protein synthesis by mRNA structure. *Molecular Biology Reports* , 19, 195-200.

Grbić, M., Van Leeuwen, T., Clark, R. M., Rombauts, S., Rouzé, P., Grbić, V., ... Van de Peer, Y. (2011). The genome of*Tetranychus urticae* reveals herbivorous pest adaptations.*Nature* , *479* (7374), 487–492. doi: 10.1038/nature10640

Grens, A., & Scheffler, I. E. (1990). The 5'- and 3'-untranslated regions of ornithine decarboxylase mRNA affect the translational efficiency. *Journal of Biological Chemistry* , 265(20), 11810-11816.

Griffiths-Jones, S. (2004). Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Research* , *33* (Database issue), D121–D124. doi: 10.1093/nar/gki081

Griffiths-Jones, S. (2006). miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Research* , *34* (90001), D140–D144. doi: 10.1093/nar/gkj112

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., ... Wortman, J. R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biology* , *9* (1), R7. doi: 10.1186/gb-2008-9-1-r7

Haft, D. H. (2003). The TIGRFAMs database of protein families.*Nucleic Acids Research* , *31* (1), 371–373. doi: 10.1093/nar/gkg128

Hill, J. T., Demarest, B. L., Bisgrove, B. W., Gorsi, B., Su, Y.-C., & Yost, H. J. (2013). MMAPPR: Mutation Mapping Analysis Pipeline for Pooled RNA-seq. *Genome Research* , *23* (4), 687–697. doi: 10.1101/gr.146936.112

Hu, J., Wang, C., Wang, J., You, Y., & Chen, F. (2010). Monitoring of resistance to spirodiclofen and five other acaricides in*Panonychus citri* collected from Chinese citrus orchards.*Pest Management Science* , *66* (9), 1025–1030. doi: 10.1002/ps.1978

Huang, Q.-T., Ma, H.-H., Deng, X.-L., Zhu, H., Liu, J., Zhou, Y., & Zhou, X.-M. (2018). Pharmacological characterization of a β-adrenergic-like octopamine receptor in *Plutella xylostella* .*Archives of Insect Biochemistry and Physiology* , *98* (4), e21466. doi: 10.1002/arch.21466

Jagadeesan, R., Fotheringham, A., Ebert, P. R., & Schlipalius, D. I. (2013). Rapid genome wide mapping of phosphine resistance loci by a simple regional averaging analysis in the red flour beetle, *Tribolium castaneum* . *BMC Genomics* , *14* (1), 650. doi: 10.1186/1471-2164-14-650

Jonsson, N. N., Klafke, G., Corley, S. W., Tidwell, J., Berry, C. M., & Caline, H. K. T. (2018). Molecular biology of amitraz resistance in cattle ticks of the genus *Rhipicephalus* . *Frontiers in Bioscience – Landmark* , 23(2), 796-810.doi: 10.2741/4617

Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O., & Walichiewicz, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research* ,*110* (1–4), 462–467. doi: 10.1159/000084979

Kanehisa, M. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes.*Nucleic Acids Research* , *28* (1), 27–30. doi: 10.1093/nar/28.1.27

Keilwagen, J., Wenk, M., Erickson, J. L., Schattat, M. H., Grau, J., & Hartung, F. (2016). Using intron position conservation for homology-based gene prediction. *Nucleic Acids Research* ,*44* (9), e89–e89. doi: 10.1093/nar/gkw092

Kita, T., Hayashi, T., Ohtani, T., Takao, H., Takasu, H., Liu, G., ... Ozoe, Y. (2017). Amitraz and its metabolite differentially activate α - and β -adrenergic-like octopamine receptors. *Pest Management Science*

, *73* (5), 984–990. doi: 10.1002/ps.4412

Koh-Tan, H. H. C., Strachan, E., Cooper, K., Bell-Sakyi, L., & Jonsson, N. N. (2016). Identification of a novel β-adrenergic octopamine receptor-like gene (βAOR-like) and increased ATP-binding cassette B10 (ABCB10) expression in a *Rhipicephalus microplus* cell line derived from acaricide-resistant ticks. *Parasites & Vectors* ,*9* (1), 425. doi: 10.1186/s13071-016-1708-x

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 adaptive k -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

Lara, F. A., Pohl, P. C., Gandara, A. C., Ferreira, J. D. S., Nascimento-Silva, M. C., Bechara, G. H., ... Oliveira, P. L. (2015). ATP binding cassette transporter mediates both heme and pesticide detoxification in tick midgut cells. *PLOS ONE* ,*10* (8), e0134779. doi: 10.1371/journal.pone.0134779

Lees, J., Yeats, C., Perkins, J., Sillitoe, I., Rentzsch, R., Dessailly, B. H., & Orengo, C. (2012). Gene3D: a domain-based resource for comparative genomics, functional annotation and protein network analysis. *Nucleic Acids Research* , *40* (D1), D465–D471. doi: 10.1093/nar/gkr1181

Letunic, I. (2004). SMART 4.0: towards genomic data integration.*Nucleic Acids Research* , *32* (90001), 142D – 144. doi: 10.1093/nar/gkh088

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* , *25* (14), 1754–1760. doi: 10.1093/bioinformatics/btp324

Lima, T., Auchincloss, A. H., Coudert, E., Keller, G., Michoud, K., Rivoire, C., ... Bairoch, A. (2009). HAMAP: a database of completely sequenced microbial proteome sets and manually curated microbial protein families in UniProtKB/Swiss-Prot. *Nucleic Acids Research* , *37* (Database), D471–D478. doi: 10.1093/nar/gkn661

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

Majoros, W. H., Pertea, M., & Salzberg, S. L. (2004). TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders.*Bioinformatics* , *20* (16), 2878–2879. doi: 10.1093/bioinformatics/bth315

Mangia, C., Vismarra, A., Genchi, M., Epis, S., Bandi, C., Grandi, G., ... Kramer, L. (2018). Exposure to amitraz, fipronil and permethrin affects cell viability and ABC transporter gene expression in an*Ixodes ricinus* cell line. *Parasites & Vectors* ,*11* (1), 437. doi: 10.1186/s13071-018-3020-4

Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., ... Bryant, S. H. (2011). CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Research* , *39* (Database), D225–D229. doi: 10.1093/nar/gkq1189

Mckenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., ... DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* , *20* (9), 1297–1303. doi: 10.1101/gr.107524.110

Navia, D., Novelli, V. M., Rombauts, S., Freitas-Astúa, J., Santos de Mendonça, R., Nunes, M. A., ... Van de Peer, Y. (2019). Draft genome assembly of the false spider mite *Brevipalpus yothersi* .*Microbiology Resource Announcements* , *8* (6). doi: 10.1128/MRA.01563-18

Nawrocki, E. P., & Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* , *29* (22), 2933–2935. doi: 10.1093/bioinformatics/btt509

Ouyang, Y., Montez, G. H., Liu, L., & Grafton-Cardwell, E. E. (2012). Spirodiclofen and spirotetramat

bioassays for monitoring resistance in citrus red mite, *Panonychus citri* (Acari: Tetranychidae).*Pest Management Science* , *68* (5), 781–787. doi: 10.1002/ps.2326

Pan, D., Dou, W., Yuan, G. R., Zhou, Q. H., & Wang, J. J. (2019). Monitoring the resistance of the citrus red mite (Acari: Tetranychidae) to four acaricides in different citrus orchards in China. *Journal of Economic Entomology* , https://doi.org/10.1093/jee/toz335.

Park, Y., González-Martínez, R. M., Navarro-Cerrillo, G., Chakroun, M., Kim, Y., Ziarsolo, P., ... Herrero, S. (2014). ABCC transporters mediate insect resistance to multiple Bt toxins revealed by bulk segregant analysis. *BMC Biology* , *12* (1), 46. doi: 10.1186/1741-7007-12-46

Parra, G., Bradnam, K., & Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes.*Bioinformatics* , *23* (9), 1061–1067. doi: 10.1093/bioinformatics/btm071

Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., & Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols* , *11* (9), 1650–1667. doi: 10.1038/nprot.2016.095

Price, A. L., Jones, N. C., & Pevzner, P. A. (2005). De novo identification of repeat families in large genomes.*Bioinformatics* , *21* (Suppl 1), i351–i358. doi: 10.1093/bioinformatics/bti1018

Ran, C., Chen, Y., & Wang, J.-J. (2009). Susceptibility and carboxylesterase activity of five field populations of *Panonychus citri* (McGregor) (Acari: Tetranychidae) to four acaricides.*International Journal of Acarology* , *35* (2), 115–121. doi: 10.1080/01647950902917593

She, R., Chu, J. S. C., Wang, K., Pei, J., & Chen, N. (2008). GenBlastA: Enabling BLAST to identify homologous gene sequences.*Genome Research* , *19* (1), 143–149. doi: 10.1101/gr.082081.108

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* ,*31* (19), 3210–3212. doi: 10.1093/bioinformatics/btv351

Snoeck, S., Kurlovs, A. H., Bajda, S., Feyereisen, R., Greenhalgh, R., Villacis-Perez, E., ... Van Leeuwen, T. (2019). High-resolution QTL mapping in *Tetranychus urticae* reveals acaricide-specific responses and common target-site resistance after selection by different METI-I acaricides. *Insect Biochemistry and Molecular Biology* ,*110* , 19–33. doi: 10.1016/j.ibmb.2019.04.011

Stanke, M., & Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* , *19* (Suppl 2), ii215–ii225. doi: 10.1093/bioinformatics/btg1080

Tang, S., Lomsadze, A., & Borodovsky, M. (2015). Identification of protein coding regions in RNA transcripts. *Nucleic Acids Research* , *43* (12), e78–e78. doi: 10.1093/nar/gkv227

Tarailo-Graovac, M., & Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics* , *25* (1), 4-10. doi: 10.1002/0471250953.bi0410s25

Tatusov, R. L. (2001). The COG database: new developments in phylogenetic classification of proteins from complete genomes.*Nucleic Acids Research* , *29* (1), 22–28. doi: 10.1093/nar/29.1.22

Thomas, P. D., Kejariwal, A., Campbell, M. J., Mi, H., Diemer, K., Guo, N., ... Doremieux, O. (2003). PANTHER: A browsable database of gene products organized by biological function, using curated protein family and subfamily classification. *Nucleic Acids Research* ,*31* (1), 334–341. doi: 10.1093/nar/gkg115

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

Tuan, S.-J., Lin, Y.-H., Yang, C.-M., Atlihan, R., Saska, P., & Chi, H. (2016). Survival and reproductive strategies in two-spotted spider mites: demographic analysis of arrhenotokous parthenogenesis

of*Tetranychus urticae* (Acari: Tetranychidae). *Journal of Economic Entomology* , *109* (2), 502–509. doi: 10.1093/jee/tov386

Van Leeuwen, T., & Dermauw, W. (2016). The Molecular evolution of xenobiotic metabolism and resistance in chelicerate mites. *Annual Review of Entomology* , *61* (1), 475–498. doi: 10.1146/annurev-ento-010715-023907

Van Leeuwen, T., Demaeght, P., Osborne, E. J., Dermauw, W., Gohlke, S., Nauen, R., ... Clark, R. M. (2012). Population bulk segregant mapping uncovers resistance mutations and the mode of action of a chitin synthesis inhibitor in arthropods. *Proceedings of the National Academy of Sciences* , *109* (12), 4407–4412. doi: 10.1073/pnas.1200068109

Van Leeuwen, T., Stillatus, V., & Tirry, L. (2004). Genetic analysis and cross-resistance spectrum of a laboratory-selected chlorfenapyr resistant strain of two-spotted spider mite (Acari: Tetranychidae).*Experimental and Applied Acarology* , *32* (4), 249–261. doi: 10.1023/B:APPA.0000023240.01937.6d

Van Leeuwen, T., Tirry, L., Yamamoto, A., Nauen, R., & Dermauw, W. (2015). The economic importance of acaricides in the control of phytophagous mites and an update on recent acaricide mode of action research. *Pesticide Biochemistry and Physiology* , *121* , 12–21. doi: 10.1016/j.pestbp.2014.12.009

Van Leeuwen, T., Van Nieuwenhuyse, P., Vanholme, B., Dermauw, W., Nauen, R., & Tirry, L. (2011). Parallel evolution of cytochrome b mediated bifenazate resistance in the citrus red mite *Panonychus citri* .*Insect Molecular Biology* , *20* (1), 135–140. doi: 10.1111/j.1365-2583.2010.01040.x

Van Pottelberge, S., Van Leeuwen, T., Khajehali, J., & Tirry, L. (2009a). Genetic and biochemical analysis of a laboratory-selected spirodiclofen-resistant strain of *Tetranychus urticae* Koch (Acari: Tetranychidae). *Pest Management Science* , *65* (4), 358–366. doi: 10.1002/ps.1698

Van Pottelberge, S., Van Leeuwen, T., Nauen, R., & Tirry, L. (2009b). Resistance mechanisms to mitochondrial electron transport inhibitors in a field-collected strain of *Tetranychus urticae* Koch (Acari: Tetranychidae). *Bulletin of Entomological Research* , *99* (1), 23–31. doi: 10.1017/S0007485308006081

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

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., ... Schulman, A. H. (2007). A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics* , *8* (12), 973–982. doi: 10.1038/nrg2165

Wu, C. H. (2004). PIRSF: family classification system at the Protein Information Resource. *Nucleic Acids Research* , *32* (90001), 112D – 114. doi: 10.1093/nar/gkh097

Xu, Z., & Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Research* , *35* (Web Server), W265–W268. doi: 10.1093/nar/gkm286

Yamamoto, A., Yoneda, H., Hatano, R., & Asada, M. (1996). Stability of hexythiazox resistance in the citrus red mite, *Panonychus citri*(Mcgregor) under laboratory and field conditions. *Journal of Pesticide Science* , *21* (1), 37–42. doi: 10.1584/jpestics.21.37

Yu, S., Tian, H., Yang, J., Ding, L., Chen, F., Li, X., ... Ran, C. (2015). Cloning of acetyl CoA carboxylase cDNA and the effects of spirodiclofen on the expression of acetyl CoA carboxylase mRNA in*Panonychus citri* . *Entomologia Experimentalis et Applicata* ,*156* (1), 52–58. doi: 10.1111/eea.12314

Zdobnov, E. M., & Apweiler, R. (2001). InterProScan - an integration platform for the signature-recognition methods in InterPro.*Bioinformatics* , *17* (9), 847–848. doi: 10.1093/bioinformatics/17.9.847

Tables

Table 1. Statistics of sample sequencing results. Library: Sequencing library of fine map; Data: Sequencing data volume of the corresponding sequencing library; Depth: Depth of sequencing; Q20: Proportion of bases with sequencing quality value above 20; Q30: Proportion of bases with sequencing quality value above 30.

Sequence data	Library	Depth (×)	Data (Gb)	Q20 (%)	Q30 (%)
Fragment library	350 bp	301.27	23.23	96.81	91.25

Table 2. Filtering statistics from Nanopore sequencing raw data. SeqNum: number of sequences; SumBase: total number of bases; N50Len: length of data N50; N90Len: length of data N90; MeanLen: average length of reads; MaxLen: longest reads length; MeanQual: average reads quality value

Platform	Data type	SeqNum	SumBase	N50Len	N90Len	MeanLen	MaxLen	MeanQual
Nanopore	Clean data	1,368,300	25,229,827,259	19,718	11,389	18,438	302,086	8.63

Table 3. Genome assembly statistics of *P. citri*. Contig number: The number of Contigs above 1 Kb; Contig length: The length of Contig above 1 Kb; Contig N50: The length of Contig N50 above 1 Kb; Contig N90: Contig above 1 Kb N90 length; Contig max: The length of the longest Contig above 1 Kb.

Contig number	Contig length (bp)	Contig N50 (bp)	Contig N90 (bp)	Contig max (bp)	GC content
144	83,967,864	1,807,243	247,900	5,138,740	31

Table 4. Genomic assembly quality assessment. Total_Reads: the number of Clean Reads; Mapped: the number of clean reads mapped to the reference genome as a percentage of all Clean Reads; Properly mapped: the proportion of sequences from two-end sequencing that were located on the reference genome and that also had a distance matching the sequencing fragment Complete length distribution; Complete BUSCOs; Complete and single-copy BUSCOs; Complete and duplicated BUSCOs; Fragmented BUSCOs; Missing BUSCOs; Number of 458 highly conserved CEGs present; % of 458 highly conserved CEGs present; Number of 248 highly conserved CEGs present; % of 248 highly conserved CEGs present.

Panonychus citri	350bp library	Total_reads	156,842,670
		Mapped (%)	96.95%
		Properly_mapped (%)	94.36%
	BUSCO	Complete BUSCOs	975
		Complete and single-copy BUSCOs	878
		Complete and duplicated BUSCOs	97
		Fragmented BUSCOs	24
		Missing BUSCOs	67
	CEGMA	Number of 458 CEGs present in assembly	448
		% of 458 CEGs present in assemblies	97.82%
		Number of 248 highly conserved CEGs present	243
		% of 248 highly conserved CEGs present	97.98%

Table 5. Genomic repeat sequence statistics

Type	Number	Length	Rate(%)
ClassI	31,601	9,606,976	11.44
ClassI/DIRS	223	61,219	0.07
ClassI/LINE	2,089	317,963	0.38
ClassI/LTR	1,708	256,699	0.31
ClassI/LTR/Copia	833	160,582	0.19
ClassI/LTR/Gypsy	4,964	3,778,949	4.5
ClassI/PLE|LARD	21,574	5,115,340	6.09
ClassI/SINE	74	4,623	0.01
ClassI/TRIM	67	64,507	0.08
ClassI/Unknown	69	8,255	0.01
ClassII	16,142	2786052	3.32
ClassII/Crypton	178	14,466	0.02
ClassII/Helitron	1,230	164,584	0.2
ClassII/MITE	432	106,275	0.13
ClassII/Maverick	2,603	1,075,871	1.28
ClassII/TIR	6,254	1,156,490	1.38
ClassII/Unknown	5,445	350,474	0.42
PotentialHostGene	4,288	932,137	1.11
SSR	38	107,733	0.13
Unknown	4,099	1,119,553	1.33
Total	52,069	13,934,963	16.6

Table 6. Statistics of gene structure prediction. GeneLen: total gene length; AveGeneLen: average gene length; ExonLen: total exon length; AveExonLen: average exon length; IntronLen: included total intron length; AveIntronLen: mean intron length.

Soft ware	Gene Num	GeneLen (bp)	AveGeneLen (bp)	ExonLen (bp)	AveExonLen (bp)	IntronLen
EVM	11,577	45,429,646	3924.13	23,189,757	2003.09	22,239,889

Table 7. Functional annotation statistics of genes from comparison to 7 databases.

Database	Annotated_Number	Percentage	100<=Protein length<300	Protein length>=300
GO_Annotation	4,032	34.83%	1,152	2,801
KEGG_Annotation	5,668	48.96%	1,511	4,076
KOG_Annotation	7,849	67.80%	2,094	5,647
Pfam_Annotation	8,591	74.21%	2,346	6,124
Swissprot_Annotation	7,029	60.72%	1,825	5,107
TrEMBL_Annotation	10,905	94.20%	3,257	7,431
nr_Annotation	9,533	82.34%	2,659	6,725
All_Annotated	10,940	94.50%	3,266	7,457

Table 8. Screening of heterologous compound metabolism-related genes at the genome level of *P. citri*

CYP450s	CYP450s	CYP450s	CYP450s	GSTs	GSTs	GSTs	GSTs	GSTs	C
CYP2	CYP3	CYP4	Mitochondrial CYP	mu	delta	omega	zeta	kappa	E
11	12	26	5	9	7	1	1	1	2
Total: 54	Total: 54	Total: 54	Total: 54	Total: 19	Total: 19	Total: 19	Total: 19	Total: 19	T

Table 9. Sequencing and statistical comparison of four DNA samples of parental and offspring. Clean_Reads: the number of reads after filtering; Clean_Base: the number of bases after filtering; Q30: the percentage of bases with a mass value of 30 or more in the total number of bases; GC: G and C bases as a percentage of total bases. Mapped: the number of Clean Reads mapped to the reference genome as a percentage of all Clean Reads; Properly_mapped: the proportion of sequences from two-end sequencing that were located on the reference genome whose distance matched the length distribution of the sequenced fragments; Ave_depth: the average of the sample's coverage depth.

ID	Clean_Reads	Clean_Base	Q30(%)	GC(%)	Mapped(%)	Properly_mapped(%)	Ave_depth
P(SS)	67,466,235	20,165,001,842	92.00	34.42	90.96	88.00	208
F(SS)	52,269,916	15,643,096,538	91.34	32.04	99.51	97.16	192
P(RR)	63,737,350	19,082,853,740	89.89	31.64	93.20	90.37	226
F(RR)	62,575,871	18,736,142,624	90.27	30.68	82.38	79.94	196

Ταβλε 10. Στατιστιςς οφ T752ⁿ μυτατιονς ιν τηε β-2P οςτοπαμινε ρεςεπτορ γενε οφ *Π. ςιτρι* ισολατεδ φρομ διφφερεντ ρεγιονς.Resistance ratio: The ratio of the LC_{50} of amitraz to*P. citri* between the field populations and the sensitive strain in the lab ; n: The number of Sanger sequencing for *P. citri* ; Pearson correlation indicates the degree of correlation between the resistance ratio and the T752C frequency, Sig. < 0.05 indicates a significant difference.

Strains	Coordinate	Slpoe ± SE	LC_{50} (95%CI) (mg L^{-1})	Resistance ratio	n	T/T	T/C	C/C
SS	E106°37'85" N29°76'06"	1.24±0.07	29.03 (20.01-42.11)	1.00	20	20	0	0
RR	E106°37'85" N29°76'06"	1.84±0.24	2361.45 (1982.89-2812.29)	81.35	20	0	0	20
SC-MS	E103°32'20" N29°52'41"	1.60±0.36	306.45 (245.12-383.11)	10.56	30	11	14	5
YN-NS	E103°12'12" N24°20'62"	2.02±0.10	219.17 (188.13-255.33)	7.55	30	13	12	5
YN-XC	E103°11'14" N24°25'15"	1.64±0.15	1481.86 (1223.04-1795.45)	51.05	30	1	0	29
GX-YL	E109°87'52" N22°73'64"	1.42±0.08	2195.85 (1696.22-2842.66)	75.64	25	0	0	25

Strains	Coordinate	Slpoe ± SE	LC_{50} (95%CI) (mg L^{-1})	Resistance ratio	n	T/T	T/C	C/C
JX-GZ	E114°75'71" N24°93'40"	1.9±0.14	1329.54 (1122.12-1575.30)	45.80	20	0	2	18
						χ^2=168.539 *df*=12 Sig.=0	χ^2=168.539 *df*=12 Sig.=0	χ^2=168.539 *df*=12 Sig.=0

Figure Legends

Fig. 1 Evaulation of genome assembly quality . **a**Single-ended reads were randomly selected from the 350 bp library and the Nanopore library and compared with the database of genes from related species. **b** Evaluation of genome size, repeat sequence ratio, and heterozygosity by Kmer analysis. **c** Distribution map of genes derived from 3 prediction methods after integration.

Fig. 2 Phylogenetic analysis of metabolism-related genes from*Panonychus citri* and *Tetranychus urticae* . **a**Neighbor-Joining phylogenetic analysis of P450 protein sequences from*P. citri* and *T. urticae* . XP×××, NP×××, and AWD××× represent the accession number of the *T. urticae* p450 protein in Genebank, EVM××× represents the ID number of p450 protein screening from the genome database of *P.citri* . P450 protein sequences clustered into the four clans: CYP2, CYP3, CYP4, and mitochondrial CYP (CYP M).**b** Neighbor-Joining phylogenetic analysis of glutathione S-transferase(GST) protein sequences from *P. citri* and *T. urticae* . XP×××, NP×××, AFQ×××, and AGE××× represent the accession number of *T. urticae* GST protein in Genebank, EVM××× represents the ID number of GST protein screening from genome database of *P. citri* . GST protein sequences clustered into the five clans: mu, delta, omega, zeta, and kappa. The protein sequences of P450 and GST in*P. citri* can be found in the TSA database (accession number: GIIF00000000)

Fig. 3 SNP detection and annotation through bulked segregant analysis . **a** Statistical differences and types of SNPs between resistant and susceptible samples of *P. citri* . **b**Statistical differences and types of Small InDels between resistant and susceptibile samples of *P. citri* . The ordinate represents the number of SNPs and Small InDels; the x-axis represents the position of SNPs and Small InDels in the gene and the type of mutation. INTERGENIC: intergenic region; INTRAGENIC: SNPs or Small InDel located in the gene; SPLICE_SITE_ACCEPTOR: splice donor mutation; SPLICE_SITE_DONOR: spliced receptor mutation; NON_SYNONYMOUS_CODING: non-synonymous mutations; NON_SYNONYMOUS_START: start codon mutation; START_LOST: start codon loss; STOP_GAINED: new Stop codon; STOP_LOST: stop codon loss; SYNONYMOUS_CODING: synonymous mutation; SYNONYMOUS_STOP: stop codon synonymous mutation; CODON_CHANGE_PLUS_CODON_DELETION: base deletion that is not an integer multiple of 3 on the codon boundary; CODON_CHANGE_PLUS_CODON_INSERTION: an integer that is not 3 on the codon boundary fold base insertion; CODON_DELETION: codon deletion (integer multiple of 3); CODON_INSERTION: codon insertion (integer multiple of 3); EXON_DELETED: entire exon is deleted; FRAME_SHIFT: gene frameshift mutation. The Venn diagram represents the distribution of shared or distinct SNPs and Small InDel among the 4 groups of DNA samples. **c** SNPs related to resistance traits were obtained using ED and SNP-index methods. AssoSNP_SNP-index/ED: the number of candidate SNPs associated with the SNP-index/ED method on the contig.**d** Clusters of Orthologous Groups (COG) classification of genes in the 2 candidate intervals. **e** Functional annotation of genes in the 2 candidate intervals based on gene ontology (GO) categorization.**f** KEGG classification of candidate genes and pathway enrichment.

Fig. 4 Location of target genes and verification of mutation sites . **a** Distribution of SNPs, mutation

types, and gene annotations related to traits**.** **b** ΔSNP-index and ED evaluation results of candidate SNPs. The x-axis represents the position of the SNPs in the corresponding candidate interval, and the y-axis represents the ΔSNP-index value and the original ED value. "*" represents the original ED value of the associated SNPs; "" represents the ΔSNP-index value of the associated SNPs. **c** Comparison of mRNA secondary structure of the β-2R octopamine receptor before (SS) and after (RR) 5' UTR mutation (GenBank accession numbers: MN928575 and MN928576). The presence of U752C forced the local secondary structure of the mRNA to change from a long stem-loop structure to three short hairpin structures. **d** Comparison of mRNA expression of the β-2R octopamine receptor gene between amitraz-resistant (RR) and -sensitive strains (SS) of *P.citri* . Data are shown as means of 3 replicates ± standard deviation; different lowercase letters indicate a significance difference between the 4 samples (P <0.05); SS: sensitive strain of *P. citri* treated with water (control group); RS: amitraz-resistant strain of *P. citri* treated with water; SS/RS-induced: susceptible/amitraz-resistant strain of *P. citri* treated with amitraz. **e** Three genotypes were detected at base pair 752 of the 5' UTR of the β-2R octopamine receptor gene by Sanger sequencing.