RESEARCH
Open Access

ASGDB: a specialised genomic resource for interpreting *Anopheles sinensis* insecticide resistance

Dan Zhou, Yang Xu, Cheng Zhang, Meng-Xue Hu, Yun Huang, Yan Sun, Lei Ma, Bo Shen[*] and Chang-Liang Zhu

Abstract

Background: *Anopheles sinensis* is an important malaria vector in Southeast Asia. The widespread emergence of insecticide resistance in this mosquito species poses a serious threat to the efficacy of malaria control measures, particularly in China. Recently, the whole-genome sequencing and *de novo* assembly of *An. sinensis* (China strain) has been finished. A series of insecticide-resistant studies in *An. sinensis* have also been reported. There is a growing need to integrate these valuable data to provide a comprehensive database for further studies on insecticide-resistant management of *An. sinensis*.

Results: A bioinformatics database named *An. sinensis* genome database (ASGDB) was built. In addition to being a searchable database of published *An. sinensis* genome sequences and annotation, ASGDB provides in-depth analytical platforms for further understanding of the genomic and genetic data, including visualization of genomic data, orthologous relationship analysis, GO analysis, pathway analysis, expression analysis and resistance-related gene analysis. Moreover, ASGDB provides a panoramic view of insecticide resistance studies in *An. sinensis* in China. In total, 551 insecticide-resistant phenotypic and genotypic reports on *An. sinensis* distributed in Chinese malaria-endemic areas since the mid-1980s have been collected, manually edited in the same format and integrated into OpenLayers map-based interface, which allows the international community to assess and exploit the high volume of scattered data much easier. The database has been given the URL: http://www.asgdb.org/.

Conclusions: ASGDB was built to help users mine data from the genome sequence of *An. sinensis* easily and effectively, especially with its advantages in insecticide resistance surveillance and control.

Keywords: Database, *Anopheles sinensis*, Insecticide resistance

Background

Malaria remains one of the most important mosquito-borne infectious parasitic diseases in tropical and subtropical areas [1, 2]. According to the World Health Organization (WHO), an estimated 212 million new cases and 429,000 deaths associated with malaria were reported in 2015 [3]. Malaria has been extensively endemic in China for more than 4000 years, and several large-scale outbreaks of malaria have occurred in Chinese history [4, 5]. Through sustained governmental and international organizational efforts in different control phases, malaria cases were dramatically decreased from approximately 30 million in 1949 to 7855 in

2010 [6, 7]. In the same year, the National Malaria Elimination Programme (2010–2020) was initiated [8]. The mid-term evaluation of this programme was revealed that malaria declined to 3078 cases in 2014, including 25 deaths [9, 10]. Although remarkable progress in malaria control has been achieved, challenges still exist on the path towards malaria elimination. For example, there were continued unpredictable outbreaks of *vivax* malaria in the central region of China along the Huang-Huai River since 2001 [5]. Moreover, the increasing number of imported malaria cases causes new threats to malaria elimination [11].

Without an efficient vaccine, vector control is considered as a key intervention to control and possibly eliminate malaria. Malaria is transmitted by mosquitoes of the genus *Anopheles*. In Southeast Asia, more than 30 *Anopheles* species occur in the domestic environment

* Correspondence: shenbo@njmu.edu.cn
Department of Pathogen Biology, Nanjing Medical University, Nanjing, Jiangsu 210029, People's Republic of China

Zhou *et al. Parasites & Vectors* (2018) 11:32

Page 2 of 11

[12]. *Anopheles dirus*, *An. minimus* and *An. epiroticus* are considered the major vectors which contribute significantly to malaria transmission [13]. In addition, several other *Anopheles* species may also be locally involved [13]. In China, there are four main malaria vectors, *An. sinensis*, *An. dirus*, *An. anthropophagus* (synonymy with *An. lesteri*) and *An. minimus* [14]. Although *An. sinensis* shows zoophilic and exophilic behaviours, it is still considered the most dominant natural vector of *Plasmodium vivax* in China due to several factors: (i) *An. sinensis* is the most widespread species distributed in most parts of China, from the northeast (Liaoning Province) to southwest (Yunnan Province); (ii) *An. sinensis* is susceptible to *P. vivax* and the ability of *An. sinensis* to transmit *P. vivax* has been obviously enhanced in China [15]; (iii) *An. sinensis* can shift from zoophilic (first option) to anthropophilic (second option) behaviour in areas where the number of cattle and pigs decreased [15]. In particular, *An. sinensis* is considered as the most frequently and widely distributed cause of recurring malaria in the central part of China and may represent a major risk for malaria elimination [16, 17].

Vector control is considered as a key intervention to control and possibly eliminate malaria, especially the long-lasting insecticide-treated bednets (LLINs) and indoor residual spraying (IRS), which were used mainly in southern China [18, 19]. Pyrethroids are currently the only class of insecticide approved by the WHO for bednet impregnation, and they are also used for IRS [20]. However, studies have shown that the level of resistance was proportional to the insecticide selection pressure [21]. Thus, constant and intensive insecticide exposure created a high selection pressure and favoured the development of resistance [21–23]. Phenotypic resistance to DDT and pyrethroids in *An. sinensis* could be found throughout China, which has placed the current national efforts of malaria elimination at risk [24]. Resistance to various classes of insecticides in *An. sinensis* has also been reported in South Korea [25].

WHO launched the Global Plan for Insecticide Resistance Management in malaria vectors (GPIRM) in May 2012 [26]. This plan is a collective strategy aimed at collecting baseline information on insecticide resistance at the global scale and further understanding the molecular mechanism of insecticide resistance. Insecticide resistance is a complex phenotype of polygenic inheritance, which is mediated through the interaction of multiple genes. The complete genome of *An. sinensis* (China strain) has been sequenced and fully annotated recently, which provides a credible starting point to determine the roles that multiple genes play in insecticide resistance. Moreover, a series of insecticide resistance-related studies for *An. sinensis* in China have been conducted since the last century, which undoubtedly provides us

with massive and openly accessible information, from the standard WHO susceptibility tests to resistance-related molecules characterization [27–30]. However, these data from different researchers are scattered in various publications. Researchers have to spend a tremendous amount of time to acquire information of interest and to evaluate the current status of insecticide resistance systematically. To host the fast-growing amount of genomic data and aid ongoing insecticide resistance research, a highly integrated information platform is needed for storage, retrieval, visualization and analysis of the *An. sinensis* genomic data and phenotypic data.

Some *An. sinensis* genomic or genetic data sets are available on NCBI (https://www.ncbi.nlm.nih.gov/) [31] or VectorBase (https://www.vectorbase.org/) [32]. NCBI is a database meant for general use and is not specifically designed for people working on vectors and insecticide resistance. Although VectorBase is specially designed for meeting the need for most of the community working on vectors, the accumulated genomic and phenotypic information of *An. sinensis* (China strain) is quite limited. At present, the genomic or genetic information of *An. sinensis* (China strain) in VectorBase mainly focuses on gene sequence, protein information and gene ontology (GO) analysis, which falls short of what will be needed. In this study, we developed an integrated *An. sinensis* genome database (ASGDB; database homepage: http://www.asgdb.org/). Besides providing the published sequences and annotation, ASGDB can also be used as an online analytical platform for visualization of genomic data, orthologous relationship analysis, GO analysis, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis, expression analysis and resistance-related gene analysis. ASGDB also aims to become an important repository for insecticide resistance data for *An. sinensis* in China. Although there are some databases that produce maps of phenotypic insecticide resistance data in malaria vectors (including *An. sinensis*) worldwide such as IR Mapper (http://www.irmapper.com) or VectorBase (https://www.vectorbase.org/popbio), the information of insecticide resistance data for *An. sinensis* in China is limited. One important reason is much valuable information reported in the Chinese medical literature may not be accessible to the malaria research community outside of China because of language barriers. We have collected and processed 551 insecticide-resistant phenotypic and genotypic reports from English or Chinese literature from the mid-1980s or unpublished data provided by the public voluntarily. These data were manually edited in the same format and presented comprehensively and systematically in OpenLayers map-based interface. Users can easily track the dynamics and evolution of

Zhou *et al. Parasites & Vectors* (2018) 11:32

Page 3 of 11

insecticide-resistant information in *An. sinensis* temporally and spatially.

Methods
Data collection
The complete genome sequence and annotation of *An. sinensis* was downloaded from the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/bioproject/PRJNA209295). The orthology relationships between the protein-coding gene sets from the different genomes were generated by the Ensembl Gene Tree method [33]. GO IDs for each gene were obtained from the corresponding InterPro entry. Each GO term is hyperlinked to AmiGO 2 browser at the GO website for details [34]. KEGG was used to perform pathway enrichment analysis of all genes [35]. *Anopheles sinensis* transcriptome data were downloaded from NCBI (BioProject accession numbers: PRJNA293400 and PRJNA339810). In a previous study, the field population of *An. sinensis* was collected in the field in Shifosi (N29.95, E115.62) town of Hubei Province in 2011. After 2 to 3-days post adult emergence, male adult and non-blood female adult mosquitoes were morphological identified and classified as the male-stain and female-stain, respectively. The other parts of non-blood female adult mosquitoes were phenotyped for susceptibility to 0.05% deltamethrin using the standard WHO tube susceptibility bioassay. The mosquitoes which knocked down after one-hour exposure were classified as deltamethrin-susceptible strain (DS-strain) and those survived after the 24-h recovery period were classified as the deltamethrin-resistant strain (DR-strain). Four libraries (male-stain, female-stain, DS-strain and DR-strain) were constructed to provide transcriptomic data to assess the assembly quality of the *An. sinensis* genome [36]. The differential gene expression levels ($P < 0.05$) between female-stain and male-stain or DS-strain and DR-strain were integrated into each gene's basic information. The information for cytochrome P450 (P450), glutathione S-transferase (*gst*), choline/carboxylesterase (*cce*) and cuticular protein (*cp*) genes in several mosquito species and *Drosophila melanogaster* were mainly collected from relevant published papers [37–44]. The insecticide resistance data of *An. sinensis* in China was searched and analysed manually. Most of these data were collected from research articles published in PubMed (http://www.ncbi.nlm.nih.gov/pubmed/), CNKI (http://www.cnki.net/), Wanfang (http://www.wanfangdata.com.cn/) and VIP (http://www.cqvip.com/) since 1988. A comprehensive search of related publications was made by using the following search terms: "*Anopheles sinensis*" AND ("China" or the name of Chinese provinces/ cities/ counties/ autonomous regions/ municipality) AND ("insecticide" or the names of insecticides (shown in Additional file 1: Table S1) or "*kdr*" or "knockdown" or "voltage-gated sodium channel" or "detoxification" or "P450" or "*gst*" or "glutathione-S-transferase" or "*cce*" or "choline/carboxylesterase" or "cuticular"). The full text of each eligible publication was read, and detailed information in the study was extracted and refined manually, including collection details, insecticide resistance monitoring information and potential insecticide resistance mechanisms. The insecticide resistance phenotypic and genotypic data: through deep mining of published papers, 538 insecticide-resistant related events in China were integrated.The vast majority of the events (72%) were collected from Chinese journals. The most common journals were "Journal of Medical Pest Control" (19%, Chinese journal) and "Chinese Journal of Vector Biology and Control" (18%, Chinese journal) after the search. The insecticide resistance phenotypes were determined according to the standard WHO tube bioassay [45], and about 35% of the events lack insecticide resistance phenotypic information.

System construction
ASGDB database was developed on an Apache HTTP server in a Linux (CentOS 6.5) operating system. The web pages were written using PHP, HTML language, Cascading Style Sheets (CSS) and JavaScript. Custom Perl scripts were used to make the database user-friendly with a good interaction interface. The Apache server handles queries from web clients through PHP scripts to perform searches. In the back-end part, all of the data in ASGDB were stored and managed in MySQL relational databases. The JBrowse genome browser is a combination of database and interactive web pages for manipulating and visualizing genes with large-scale sequence datasets [46, 47]. This next generation AJAX-based genome browser, built with JavaScript and HTML5, enables fast and smooth animated genome navigation over the web. ViroBLAST has also been integrated into the ASGDB system [48]. The interface of "Resistance surveillance" utilizes OpenLayers 3 (high-performance, feature-packed web mapping library) to visualize the data. An overview of the ASGDB architecture is given in Fig. 1. The website has been tested on multiple platforms (Linux, Windows and Mac OS) with different web browsers (Firefox, Chrome, IE and Safari). The resistance surveillance map was not compatible with Firefox. We recommend using IE, Chrome, Opera or Safari to open the webpage of the insecticide-resistant map.

Results
Data content
There are three types of data included in ASGDB. Basic data: 19,317 putative protein-coding genes and 2425 non-coding RNAs predicted from 9597 *An. sinensis* assembled scaffolds are obtained. For all protein-coding

Zhou *et al. Parasites & Vectors* (2018) 11:32

Page 4 of 11

Fig. 1 Overview of the ASGDB architecture
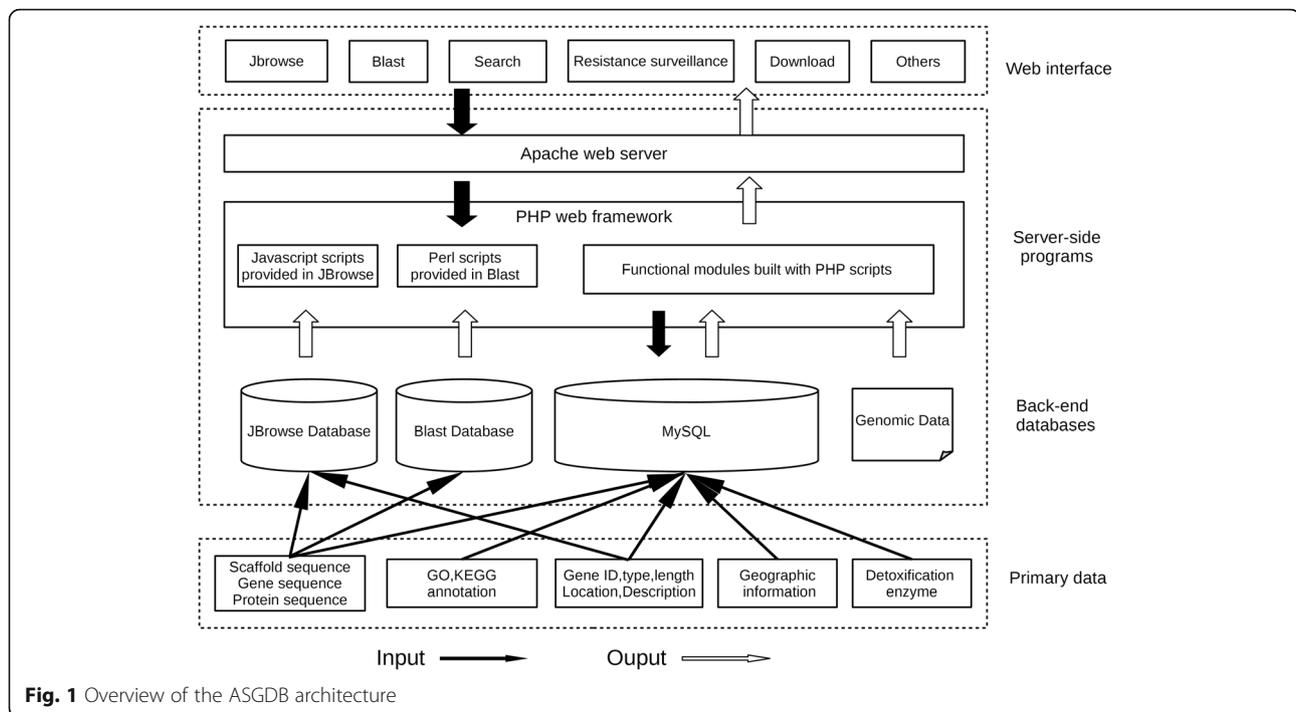
genes, 1424 GO terms, 221 known pathways and 58,207 orthologs are displayed in ASGDB [36]. ASGDB also contains 4047 differentially expressed genes under different conditions from transcriptomic studies, including deltamethrin-resistant/ deltamethrin-susceptible and female/male [36]. Insecticide resistance-related gene data: ASGDB provides the identification and classification information of 93 cytochrome P450s (P450s), 31 glutathione S-transferases (*gsts*), 50 choline/carboxylesterases (*cces*) and 238 cuticular proteins from *An. sinensis*. Related information of these gene families in other species is also provided for comparison. The database also contains 13 unpublished datasets provided by Dr Chang (Bengbu Medical College). Altogether, 551 insecticide-resistant related events are stored and managed in ASGDB, and a user-friendly web interface was developed to help users search and use these data. A summary of data content in ASGDB is shown in Table 1.

Web interface and homepage
ASGDB interface provides direct links to eight individual pages including Home, JBrowse, Search, Download, Resistance-related gene, Resistance surveillance, Contact and Tutorial. All the links are clickable icons. A left mouse click on the icon leads to the respective page.

A brief introduction to ASGDB and photographs of *An. sinensis* taken at different life-cycle stages (egg, larva, pupa and adult) are included at the top part of the homepage. Below these, the latest publications related to *An. sinensis* are displayed. A link to the list of all

relevant publications is also provided, allowing users acquire information of interest quickly and easily. "Resistance surveillance map" provides a map-based interface to access insecticide resistance-related data. All the collection sites are spotted with red pompons on the map. If users are interested in this content, they can click on the map and will be taken to a new page to acquire more detailed information. "Visitor Statistics" shows the number of visitors to the website every week.

JBrowse
Here, we use the JBrowse genome browser to visualize the whole draft genome sequence of *An. sinensis*. The page is divided into two parts. The relatively small panel on the left is a list of different types of factors, on which the icons, from top to bottom, are: "GC content", "gene", "exon", "miRNA", "rRNA", "tRNA" and "reference sequence". All these icons for each scaffold can be tracked to the detailed view. The larger window on the right is the tracks display region. Each track can be turned off by clicking the "cross" in front of the title, which allows users to hide unwanted information for a better user experience. Dragging the tracks up or down can change the positions of tracks to display datasets of interest at the top for convenience. JBrowse also provides efficient panning and zooming of a genomic region in the genome via embedded navigation buttons. With the help of these important and efficient visualization modules, users can easily browse and search on a large scale in a graphic interface.

Zhou *et al. Parasites & Vectors* (2018) 11:32

Page 5 of 11

Table 1 Summary of ASGDB content

Data type	Data statistics
Gene	19,317
None coding RNA	2425
MicroRNAs	56
tRNAs	352
rRNA	2017
GO	1424
KEGG	221
Ortholog	58,207
Differentially expressed genes	4047
Female/Male	1388
Deltamethrin-resistant/ Deltamethrin-susceptible	2659
Insecticide-resistant related events	551
Published articles	538
Unpublished data	13

Users can search for scaffolds to locate regions on the *An. sinensis* genome. These scaffolds are freely selected from the drop-down menu. If the users are interested in a particular region on the scaffold, they can also enter the starting and terminal positions of the region to retrieve detailed information. In the genomic view, rectangular frames with directionality represent the corresponding genes or ncRNAs from the positive or negative strand. A single click on the frame will open an information table, which provides detailed information such as annotation, location, GO, KEGG and sequences. The sequence data for the selected gene can be downloaded in the same page as FASTA files. Other people can see the same region of the *An. sinensis* genome and the collection of open tracks on their screen when the visible URL (accessible either via the browser address bar or the "Share" button) is shared.

Search
ASGDB affords a user-friendly search engine to make it easy to reach specific genes of interest. There are two sub-categories in the search part: simple search and BLAST search. To browse different types of genetic factors, a simple search can be performed using the following parameters: (i) NCBI or ASGDB accession numbers; (ii) Gene name or symbol; (iii) GO ID or GO term; and (iv) KEGG ID or KEGG annotation. Users can enter these parameters to obtain specific gene information from ASGDB, and fuzzy queries are supported. All the matched genes will be linked in the search job when more than one gene is matched with the input keyword. The BLAST search allows searching of genes using the ViroBLAST [48]. Users perform similarity searches against each type of sequences using various BLAST search forms (BLASTn, BLASTp, BLASTx, tBLASTn and tBLASTx). The reference database used for BLAST is all nucleic acid and amino acid sequences of *An. sinensis*. Users can enter nucleotide or protein query sequences or upload a local sequence file in the FASTA format to search against the reference database. The BLAST search tool allows users to set their favourite parameters, such as threshold, Word size, etc., in advanced search.

This multi-functional search module makes it easier to obtain a comprehensive view of each gene. Taking CYP 9 J53 as an example (Fig. 2), users can input many types of keywords, e.g. "KFB49800.1", "CYP 9 J53" or "9 J53" as search content. Pressing on the "GeneID" button will display the detailed information for this gene. At the top part of the gene information page, the users can view some fundamental information about CYP 9 J53, such as description, length and location. Clicking on the right "JBrowse" button enables users to visualize CYP 9 J53 under the background of the scaffold. Below these, the sequence information of CYP 9 J53 is presented. The exon regions are highlighted in red, and the remaining sequences are introns. Clicking on the "show pep" button allows amino acid sequences to be displayed. The lower portion of the information page is the functional feature description of CYP 9 J53, including orthologs, GO and KEGG pathways. In the ortholog part, many-one or many-many orthologous genes to the *An. sinensis* CYP 9 J53 in other mosquito species and fruit fly are displayed. There are links to get the sequences of the orthologs of *An. sinensis* genes in VectorBase and FlyBase. The prediction of GO terms shows that CYP 9 J53 belongs to "iron ion binding" (0005506), "electron carrier activity" (0009055), "heme binding" (0020037) and "oxidation-reduction process" (0055114) categories. Users can click the GO term for detailed term information. CYP 9 J53 participates in "Linoleic acid metabolism" pathway. Clicking the KO (ko00591) will open a new page to show the reference pathway map. Transcriptional results are shown at the bottom of the results page and include technology, comparison, regulation, fold change, published articles, source as a whole. Moreover, users can perform BLAST to find the best hit for the gene of interest via copy-paste sequences or upload sequence FASTA file. Different parameters can be reset to filter and parse the results again. Click one of the links in Score field will locate the pair-wise alignment between the query sequence and subject sequence.

Download
ASGDB provides bulk data downloads, including assembled genome sequences, nucleotide sequences of putative genes, gene annotation and amino acid sequences. All data are hosted and accessible publicly via a browser

Zhou *et al. Parasites & Vectors* (2018) 11:32

Page 6 of 11
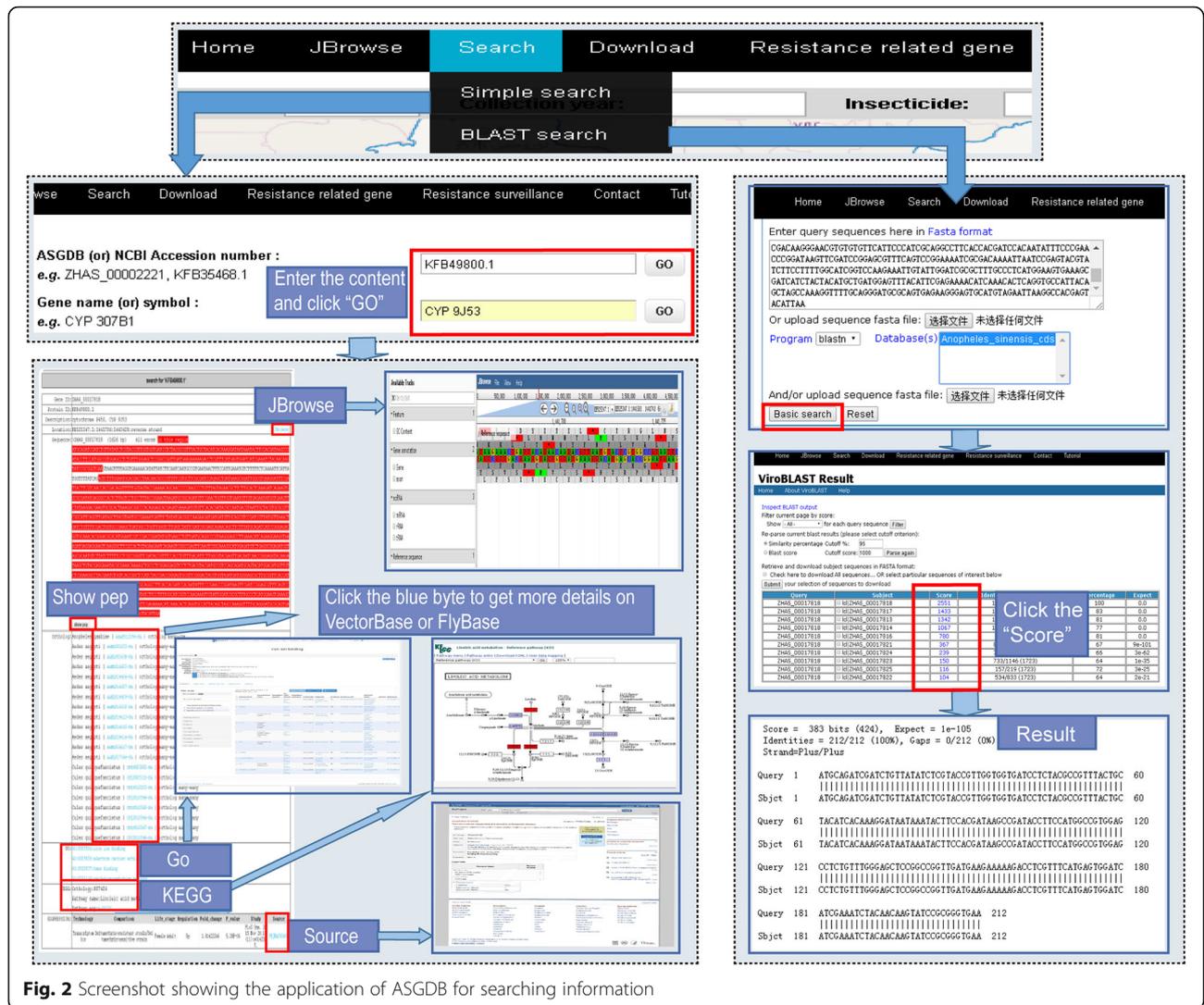
Fig. 2 Screenshot showing the application of ASGDB for searching information

directly. Meanwhile, manual curation of literature related to *An. sinensis* is carried out to fulfil the increasing research demands. We have collected *An. sinensis*-related English literature from PubMed (http://www.ncbi.nlm.nih.gov/pubmed/) and Chinese literature from CNKI (http://www.cnki.net/), Wanfang (http://www.wanfangdata.com.cn/) and VIP (http://www.cqvip.com/). The general information is organized as formatted lists, including title, authors, journal, year and volume. ASGDB also offers some full-text article links of English literature.

Insecticide resistance-related genes
ASGDB provides the identification and classification information of detoxification enzyme superfamilies and cuticular proteins from *An. sinensis*. Three detoxification enzyme superfamilies (P450s, GSTs and CCEs) are primarily responsible for metabolic

resistance in mosquitoes [49]. The cuticle is a major route of insecticide penetration, thickening or changing in the chemical composition of the cuticle serve as another resistance mechanism [50]. Although cuticular resistance has not yet been fully characterized at the molecular level, several examples of putative *cp*s that are the primary components of insect cuticle have been identified as the potential players in insecticide resistance [51–54].

On the basis of a literature review and in-depth data analysis, ASGDB provides detoxification enzyme and *cp* information from several mosquito species and the fruit fly *Drosophila melanogaster*, allowing users to perform rapid and convenient comparative analyses among different Diptera insect species. To aid insecticide resistance research, these genes are further classified into four P450 clans, seven *gst* classes, ten *cce* clades and eleven *cp* families.

Zhou *et al. Parasites & Vectors* (2018) 11:32

Page 7 of 11

Resistance surveillance

The current release of ASGDB has recorded 551 insecticide-resistant phenotypic and genotypic events in *An. sinensis* in China. So far, the geographical distributions of the data cover the majority of *An. sinensis* distribution areas in China.

There are two ways to retrieve insecticide-resistant information in the database (Fig. 3). ASGDB provides an OpenLayers map-based interface for users to obtain data. All the mosquito-sampling sites are identified with small pompons. By clicking on the pompon, a pop-up text box will appear on the map with the most recent insecticide-resistant related record. We choose different colours for pompons to indicate insecticide-resistant levels (Grey: Uncertain; Green: Susceptible; Yellow: Probable resistant; Red: Resistant). Users can also browse all the relevant information in this region on the same page just below the map. It should be noted that when
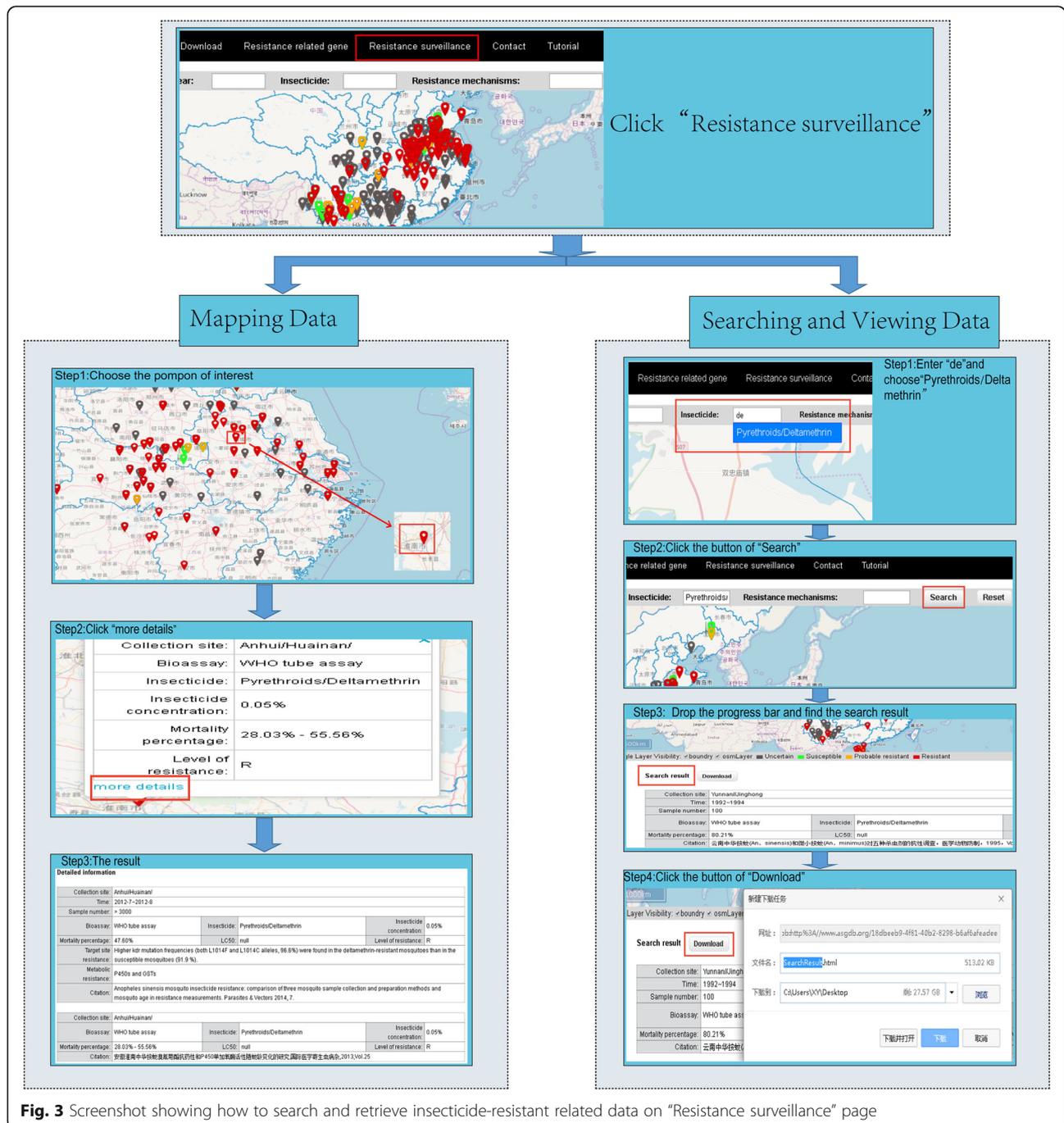
Fig. 3 Screenshot showing how to search and retrieve insecticide-resistant related data on "Resistance surveillance" page

Zhou *et al. Parasites & Vectors* (2018) 11:32

Page 8 of 11

different sampling sites are in proximity, the red pompons might be very near or even superposed on each other. To avoid clicking on the non-target pompon, the users can zoom in to magnify their view of the map by tapping the "plus" button in the upper left corner or by scrolling up with mouse's scroll wheel. Double left click on the map can also simultaneously centre the map and zoom in at the position clicked. Once users think they are done, they can use the "minus" button or scroll down with mouse's scroll wheel to zoom out. To move the map, users can click and drag at any point on the map. ASGDB also provides the users with a search engine to facilitate obtaining the information of interest, such as collection site, collection year, insecticide and resistance mechanism. For example, if the users want to perform an insecticide based search using "deltamethrin", they can type two letters "de" and a list of all possible matching terms will appear. Users can choose the requested term and click on the "Search" button, and only the matching pompons will appear on the map, and all records related to deltamethrin resistance will appear below the map. The retrieved information can be downloaded in bulk.

The data were manually extracted and concisely presented in three aspects: "Collection details", "Insecticide resistance monitoring" and "Insecticide resistance mechanisms". "Collection details" provides temporal and spatial information regarding sample collection. "Insecticide resistance monitoring" surveys the insecticide resistance status in the field population under different insecticide selection pressures. "Insecticide resistance mechanisms" provides further investigations of the resistance mechanisms, including the analysis of gene expression changes, *kdr* or *ace-1* mutations, or elevated enzyme activities of P450s, GSTs and CCEs. We also provided the citation of sources.

Users are encouraged to share their job data related to the insecticide resistance by using the submission procedure, which can ease the process of data collection and sharing, and benefit the dispersal of knowledge. In addition to relevant data, the participants also need to provide valid and open personal information, including institution and e-mail addresses. Although time consuming, it will not only improve content reliability but also increase users' collaborations and communications.

Contact
To establish a platform for community integration of *An. sinensis* data and to aid efficient management of knowledge on *An. sinensis*, ASGDB requires more participation in knowledge curation. Researchers can submit any comments, suggestions, or questions regarding all aspects of ASGDB. To promote researchers' sharing and exchange of knowledge and ideas, the submission

process is simple. No registration requirement is imposed, although users need to provide a valid email address so that our team can contact them in case of any queries.

Tutorial
To help facilitate access and utilize these data, a general tutorial is also available in ASGDB. It provides schematic overview and demonstrates how to get started and navigate through the main features of ASGDB.

Discussion
The basic objective of the ASGDB is to provide an integrative and comparative genomic resource particular to *An. sinensis*. It gathers a wide variety of genetic information, such as the whole draft genome sequences, annotation, pathway, GO terms, orthologous relationships and differentially expressed. The diverse data integration makes it possible to display correlations among various genetic factors and thus to will help users obtain genetic information faster and more accurately, which is a fundamental step in exploring valuable information for further study. As a one-stop resource platform, ASGDB contains a user-friendly interface, convenient search options and enhanced visualization tools, all of which make it easy for researchers to access and analyse whole-genome genomic data and information for *An. sinensis*, even those with little knowledge of bioinformatics.

Another aim of ASGDB is to share insecticide resistance information for *An. sinensis*. We collect resistance-related gene information in *An. sinensis* and other species. This genome-level data-mining strategy of resistance-related genes will also be useful for the follow-up functional study of resistance. For example, researchers could conveniently choose one specific gene of interest or a group of genes (e.g. by classification or gene-expanded clusters) to investigate how resistance phenotypes are generated. The information for genes differentially expressed between deltamethrin-susceptible and deltamethrin-resistant strains should also help identify candidate genes of interest. On the basis of data from the published literature, ASGDB provides a panoramic view of current insecticide-resistant studies of *An. sinensis* in China. These data cover 19 provinces and municipalities in China region, which vary substantially in geography, and economic and social environment. The records were initiated in the mid-1980s and continue to this day. These records systematically and continuously track the dynamics of the insecticide-resistant phenotypic and genotypic information, providing invaluable information to help us understand how insecticide resistance occurs and spreads in *An. sinensis* temporally and spatially. The excessive use of agricultural insecticides should be slowed before the occurrence of high resistance; therefore, the phenotypic

Zhou *et al. Parasites & Vectors* (2018) 11:32

Page 9 of 11

information will be useful to adjust the types and concentrations of insecticides in a rotation scheme to fit local environmental conditions. It could promote the existing resistance management strategies to prolong the effectiveness of insecticides and prevent the occurrence of resistance. The evolution of insecticide resistance is conferred through complex mechanisms, typically requiring the interaction of multiple genes [55–58]. The genotypic data in ASGDB could provide clues to explain the molecular mechanism of insecticide resistance systematically. For example, we could observe whether target-site and metabolic resistance mechanisms occur singly or simultaneously, or judge which mechanism plays more important role in *An. sinensis* at different insecticide-resistance levels. The information of insecticide resistance at the mechanistic level, combined with results of bioassays, could also assist in providing powerful molecular diagnostic tools to aid the monitoring of insecticide resistance in *An. sinensis* at an early stage.

With increasing research on *An. sinensis*, genome re-sequencing, transcriptomic, proteomic and other omics studies are expected to grow continuously, especially with the development of sequencing technologies in the next few years. ASGDB will integrate more types of data and will be updated periodically to fulfil the growing research needs in addressing the genetic complexity of *An. sinensis*. Meanwhile, more insecticide resistance information from neighbouring countries will be integrated into ASGDB in the next step. Also, we encourage researchers to share insecticide-resistant data with the whole scientific community. To attract more participation from the scientific community for ASGDB and to make it an important platform for the insecticide resistance studies of *An. sinensis*, we will develop an incentive system to reward participants according to their contributions.

Conclusions

To further extend our understanding of resistance mechanism and facilitate the implementation of resistance management strategies for mosquito vector control programs, a bioinformatics database named ASGDB was developed. High-quality draft genome sequence integrated with insecticide resistance-related literature, ASGDB provides (i) *An. sinensis* genome database; (ii) insecticide resistance-related gene data; and (iii) the insecticide resistance phenotyping and genotyping data in *An. sinensis* in China region. ASGDB was built to help users to mine data from the genome sequence of *An. sinensis* easily and effectively, especially with its advantages in insecticide resistance surveillance and control. The resistance surveillance related information collected in ASGDB provides dynamic and evolutionary resistance records in whole China region. Detailed records of

insecticide resistance status of *An. sinensis* are useful for adjusting the types of insecticides in a rotation scheme to fit local environmental conditions.

Additional file

Additional file 1: Table S1. The major classes and types of insecticides. (PDF 34 kb)

Abbreviations
ASGDB: *Anopheles sinensis* genome database; BLAST: Basic Local Alignment Search Tool; CCE: Choline/carboxylesterase; CP: Cuticular protein; CSS: Cascading style sheets; GO: Gene ontology; GST: Glutathione S-transferase; ID: Identification; IRS: Indoor residual spraying; KEGG: Kyoto Encyclopedia of Genes and Genomes; KO: KEGG orthology; LLINs: Long-lasting insecticide-treated bednets; NCBI: National Center for Biotechnology Information; P450: Cytochrome P450

Availability of data and materials
All data are fully available without restriction.

Authors' contributions
DZ, YX, CZ, MXH, YH, YS and LM collected, organized and compiled the data. DZ wrote the manuscript and prepared figures. DZ and BS developed the web interface and integrated the tools. BS and CLZ conceived the idea and coordinated the project. All authors read and approved the final manuscript.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

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

References

1. Patarroyo MA, Calderon D, Moreno-Perez DA. Vaccines against *Plasmodium vivax*: a research challenge. Expert Rev Vaccines. 2012;11(10):1249–60.
2. Eastman RT, Pattaradilokrat S, Raj DK, Dixit S, Deng B, Miura K, et al. A class of tricyclic compounds blocking malaria parasite oocyst development and transmission. Antimicrob Agents Chemother. 2013;57(1):425–35.
3. WHO. World malaria report 2016. Geneva: World Health Organization; 2016.
4. Tang L. Achievements in the research on the prevention and treatment of malaria in China. Zhongguo Ji Sheng Chong Xue Yu Ji Sheng Chong Bing Za Zhi. 1999;17(5):257–9.
5. Yin JH, Zhou SS, Xia ZG, Wang RB, Qian YJ, Yang WZ, Zhou XN. Historical patterns of malaria transmission in China. Adv Parasitol. 2014;86:1–19.
6. Yip K. Antimalarial work in China: a historical perspective. Parassitologia. 1998;40(1–2):29–38.

Zhou *et al. Parasites & Vectors* (2018) 11:32

Page 10 of 11

7. Zhou SS, Wang Y, Li Y. Malaria situation in the People's Republic of China in 2010. Zhongguo Ji Sheng Chong Xue Yu Ji Sheng Chong Bing Za Zhi. 2011; 29(6):401–3.

8. WHO: Action plan of China malaria elimination (2010–2020). 2012.

9. Hu T, Liu YB, Zhang SS, Xia ZG, Zhou SS, Yan J, et al. Shrinking the malaria map in China: measuring the progress of the National Malaria Elimination Programme. Infect Dis Poverty. 2016;5(1):52.

10. Zhang L, Zhou SS, Feng J, Fang W, Xia ZG. Malaria situation in the People's Republic of China in 2014. Zhongguo Ji Sheng Chong Xue Yu Ji Sheng Chong Bing Za Zhi. 2015;33(5):319–26.

11. Feng J, Yan H, Feng XY, Zhang L, Li M, Xia ZG, Xiao N. Imported malaria in China, 2012. Emerg Infect Dis. 2014;20(10):1778–80.

12. Trung HD, Bortel WV, Sochantha T, Keokenchanh K, Briet OJ, Coosemans M. Behavioural heterogeneity of *Anopheles* species in ecologically different localities in Southeast Asia: a challenge for vector control. Trop Med Int Health. 2005;10(3):251–62.

13. Trung HD, Van Bortel W, Sochantha T, Keokenchanh K, Quang NT, Cong LD, Coosemans M. Malaria transmission and major malaria vectors in different geographical areas of Southeast Asia. Trop Med Int Health. 2004;9(2):230–7.

14. MOH: Malaria Control Manual. 2007.

15. Pan JY, Zhou SS, Zheng X, Huang F, Wang DQ, Shen YZ, et al. Vector capacity of *Anopheles sinensis* in malaria outbreak areas of central China. Parasit Vectors. 2012;5:136.

16. Zhang HW, Liu Y, Zhang SS, Xu BL, Li WD, Tang JH, et al. Preparation of malaria resurgence in China: case study of vivax malaria re-emergence and outbreak in Huang-Huai plain in 2006. Adv Parasitol. 2014;86:205–30.

17. Zhou SS, Huang F, Wang JJ, Zhang SS, Su YP, Tang LH. Geographical, meteorological and vectorial factors related to malaria re-emergence in Huang-Huai River of central China. Malar J. 2010;9:337.

18. Wang GZ, Wang SQ, Hu XM, Meng F, Li YC, Zeng W, Cai HL. Analysis of malaria epidemic situation in Hainan Province, 2010. Zhongguo Xue Xi Chong Bing Fang Zhi Za Zhi. 2012;24(3):369–70.

19. Zhou SS, Zhang SS, Zhang L, Rietveld AE, Ramsay AR, Zachariah R, et al. China's 1-3-7 surveillance and response strategy for malaria elimination: is case reporting, investigation and foci response happening according to plan? Infect Dis Poverty. 2015;4:55.

20. WHO: Pesticides and their application for the control of vectors and pests of public health importance. 2006.

21. Shi L, Hu H, Ma K, Zhou D, Yu J, Zhong D, et al. Development of resistance to pyrethroid in *Culex pipiens pallens* population under different insecticide selection pressures. PLoS Negl Trop Dis. 2015;9(8):e0003928.

22. Diabate A, Baldet T, Chandre F, Akoobeto M, Guiguemde TR, Darriet F, et al. The role of agricultural use of insecticides in resistance to pyrethroids in *Anopheles gambiae s.l.* in Burkina Faso. Am J Trop Med Hyg. 2002;67(6):617–22.

23. Wilding CS, Smith I, Lynd A, Yawson AE, Weetman D, Paine MJ, Donnelly MJ. A cis-regulatory sequence driving metabolic insecticide resistance in mosquitoes: functional characterisation and signatures of selection. Insect Biochem Mol Biol. 2012;42(9):699–707.

24. Wang DQ, Xia ZG, Zhou SS, Zhou XN, Wang RB, Zhang QF. A potential threat to malaria elimination: extensive deltamethrin and DDT resistance to *Anopheles sinensis* from the malaria-endemic areas in China. Malar J. 2013; 12:164.

25. Chang KS, Yoo DH, Shin EH, Lee WG, Roh JY, Park MY. Susceptibility and resistance of field populations of *Anopheles sinensis* (Diptera: Culicidae) collected from Paju to 13 insecticides. Osong Public Health Res Perspect. 2013;4(2):76–80.

26. WHO: Global plan for insecticide resistance management in malaria vectors. 2012.

27. Wang J. Resistance to two pyrethroids in *Anopheles sinensis* from Zhejiang, China. J Am Mosq Control Assoc. 1999;15(3):308–11.

28. Zhang Z, Yang C. Application of deltamethrin-impregnated bednets for mosquito and malaria control in Yunnan, China. Southeast Asian J Trop Med Public Health. 1996;27(2):367–71.

29. Wang Y, Yu W, Shi H, Yang Z, Xu J, Ma Y. Historical survey of the *kdr* mutations in the populations of *Anopheles sinensis* in China in 1996–2014. Malar J. 2015;14:120.

30. Chang X, Zhong D, Fang Q, Hartsel J, Zhou G, Shi L, et al. Multiple resistances and complex mechanisms of *Anopheles sinensis* mosquito: a major obstacle to mosquito-borne diseases control and elimination in China. PLoS Negl Trop Dis. 2014;8(5):e2889.

31. Jenuth JP. The NCBI. Publicly available tools and resources on the web. Methods Mol Biol. 2000;132:301–12.

32. Giraldo-Calderon GI, Emrich SJ, MacCallum RM, Maslen G, Dialynas E, Topalis P, et al. VectorBase: an updated bioinformatics resource for invertebrate vectors and other organisms related with human diseases. Nucleic Acids Res. 2015;43(Database issue):D707–13.

33. Vilella AJ, Severin J, Ureta-Vidal A, Heng L, Durbin R, Birney E. EnsemblCompara GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates. Genome Res. 2009;19(2):327–35.

34. Gene Ontology C. Gene ontology consortium: going forward. Nucleic Acids Res. 2015;43(Database issue):D1049–56.

35. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, et al. KEGG for linking genomes to life and the environment. Nucleic Acids Res. 2008; 36(Database issue):D480–4.

36. Zhou D, Zhang D, Ding G, Shi L, Hou Q, Ye Y, et al. Genome sequence of *Anopheles sinensis* provides insight into genetics basis of mosquito competence for malaria parasites. BMC Genomics. 2014;15:42.

37. Zhou D, Liu X, Sun Y, Ma L, Shen B, Zhu C. Genomic analysis of detoxification supergene families in the mosquito *Anopheles sinensis*. PLoS One. 2015;10(11):e0143387.

38. Strode C, Wondji CS, David JP, Hawkes NJ, Lumjuan N, Nelson DR, et al. Genomic analysis of detoxification genes in the mosquito *Aedes aegypti*. Insect Biochem Mol Biol. 2008;38(1):113–23.

39. Yan L, Yang P, Jiang F, Cui N, Ma E, Qiao C, Cui F. Transcriptomic and phylogenetic analysis of *Culex pipiens quinquefasciatus* for three detoxification gene families. BMC Genomics. 2012;13:609.

40. Saisawang C, Wongsantichon J, Ketterman AJ. A preliminary characterization of the cytosolic glutathione transferase proteome from *Drosophila melanogaster*. Biochem J. 2012;442(1):181–90.

41. Tijet N, Helvig C, Feyereisen R. The cytochrome P450 gene superfamily in *Drosophila melanogaster*: annotation, intron-exon organization and phylogeny. Gene. 2001;262(1–2):189–98.

42. Zhou D, Duan B, Sun Y, Ma L, Zhu C, Shen B. Preliminary characterization of putative structural cuticular proteins in the malaria vector *Anopheles sinensis*. Pest Manag Sci. 2017;73(12):2519–28.

43. Willis JH. Structural cuticular proteins from arthropods: annotation, nomenclature, and sequence characteristics in the genomics era. Insect Biochem Mol Biol. 2010;40(3):189–204.

44. Ioannidou ZS, Theodoropoulou MC, Papandreou NC, Willis JH, Hamodrakas SJ. CutProtFam-Pred: detection and classification of putative structural cuticular proteins from sequence alone, based on profile hidden Markov models. Insect Biochem Mol Biol. 2014;52:51–9.

45. Organization WH: Test procedures for insecticide resistance monitoring in malaria vector mosquitoes. 2013.

46. Skinner ME, Uzilov AV, Stein LD, Mungall CJ, Holmes IH. JBrowse: a next-generation genome browser. Genome Res. 2009;19(9):1630–8.

47. Westesson O, Skinner M, Holmes I. Visualizing next-generation sequencing data with JBrowse. Brief Bioinform. 2013;14(2):172–7.

48. Deng W, Nickle DC, Learn GH, Maust B, Mullins JI. ViroBLAST: a stand-alone BLAST web server for flexible queries of multiple databases and user's datasets. Bioinformatics. 2007;23(17):2334–6.

49. Hemingway J, Ranson H. Insecticide resistance in insect vectors of human disease. Annu Rev Entomol. 2000;45:371–91.

50. Apperson CS, Georghiou GP. Mechanisms of resistance to organophosphorus insecticides in *Culex tarsalis*. J Econ Entomol. 1975;68(2):153–7.

51. Zhang J, Goyer C, Pelletier Y. Environmental stresses induce the expression of putative glycine-rich insect cuticular protein genes in adult *Leptinotarsa decemlineata* (Say). Insect Mol Biol. 2008;17(3):209–16.

52. Vontas J, David JP, Nikou D, Hemingway J, Christophides GK, Louis C, Ranson H. Transcriptional analysis of insecticide resistance in *Anopheles stephensi* using cross-species microarray hybridization. Insect Mol Biol. 2007; 16(3):315–24.

53. Awolola TS, Oduola OA, Strode C, Koekemoer LL, Brooke B, Ranson H. Evidence of multiple pyrethroid resistance mechanisms in the malaria vector *Anopheles gambiae sensu stricto* from Nigeria. Trans R Soc Trop Med Hyg. 2009;103(11):1139–45.

54. Fang F, Wang W, Zhang D, Lv Y, Zhou D, Ma L, et al. The cuticle proteins: a putative role for deltamethrin resistance in *Culex pipiens pallens*. Parasitol Res. 2015;114(12):4421–9.

55. Joussen N, Heckel DG, Haas M, Schuphan I, Schmidt B. Metabolism of imidacloprid and DDT by P450 CYP6G1 expressed in cell cultures of

Zhou *et al. Parasites & Vectors* (2018) 11:32

Page 11 of 11

Nicotiana tabacum suggests detoxification of these insecticides in Cyp6g1-overexpressing strains of *Drosophila melanogaster,* leading to resistance. Pest Manag Sci. 2008;64(1):65–73.

56. Edi CV, Djogbenou L, Jenkins AM, Regna K, Muskavitch MA, Poupardin R, et al. CYP6 P450 enzymes and ACE-1 duplication produce extreme and multiple insecticide resistance in the malaria mosquito *Anopheles gambiae.* PLoS Genet. 2014;10(3):e1004236.

57. Ranson H, Claudianos C, Ortelli F, Abgrall C, Hemingway J, Sharakhova MV, et al. Evolution of supergene families associated with insecticide resistance. Science. 2002;298(5591):179–81.

58. Kwiatkowska RM, Platt N, Poupardin R, Irving H, Dabire RK, Mitchell S, et al. Dissecting the mechanisms responsible for the multiple insecticide resistance phenotype in *Anopheles gambiae* s.s., M form, from Vallee du Kou, Burkina Faso. Gene. 2013;519(1):98–106.