BMC
Genomics

RESEARCH ARTICLE

Open Access

Transcriptome profiling of pyrethroid resistant and susceptible mosquitoes in the malaria vector, *Anopheles sinensis*

Guoding Zhu[1,2,3†], Daibin Zhong[3†], Jun Cao[2], Huayun Zhou[2], Julin Li[2], Yaobao Liu[2], Liang Bai[2], Sui Xu[2], Mei-Hui Wang[3], Guofa Zhou[3], Xuelian Chang[3], Qi Gao[1,2*] and Guiyun Yan[3*]

Abstract

Background: *Anopheles sinensis* is a major malaria vector in China and other Southeast Asian countries, and it is becoming increasingly resistant to the insecticides used for agriculture, net impregnation, and indoor residual spray. Very limited genomic information on this species is available, which has hindered the development of new tools for resistance surveillance and vector control. We used the 454 GS FLX system and generated expressed sequence tag (EST) databases of various life stages of *An. sinensis*, and we determined the transcriptional differences between deltamethrin resistant and susceptible mosquitoes.

Results: The 454 GS FLX transcriptome sequencing yielded a total of 624,559 reads (average length of 290 bp) with the pooled *An. sinensis* mosquitoes across various development stages. The *de novo* assembly generated 33,411 contigs with average length of 493 bp. A total of 8,057 ESTs were generated with Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) annotation. A total of 2,131 ESTs were differentially expressed between deltamethrin resistant and susceptible mosquitoes collected from the same field site in Jiangsu, China. Among these differentially expressed ESTs, a total of 294 pathways were mapped to the KEGG database, with the predominant ESTs belonging to metabolic pathways. Furthermore, a total of 2,408 microsatellites and 15,496 single nucleotide polymorphisms (SNPs) were identified.

Conclusions: The annotated EST and transcriptome databases provide a valuable genomic resource for further genetic studies of this important malaria vector species. The differentially expressed ESTs associated with insecticide resistance identified in this study lay an important foundation for further functional analysis. The identified microsatellite and SNP markers will provide useful tools for future population genetic and comparative genomic analyses of malaria vectors.

Keywords: Transcriptome, Expressed sequence tag, Pyrethroid resistance, Gene expression, *Anopheles sinensis*

Background

The malaria vector mosquito, *Anopheles sinensis* (Diptera: Culicidae) is widely distributed in China and other Southeast Asian countries [1]. Due to its relatively high zoophilic and exophilic behaviors, it was treated as a secondary vector for many years in the 20th century.

However, the outbreak of *Plasmodium vivax* malaria in central China in which *An. sinensis* was the only vector leads us to reassess its role for malaria transmission [2,3]. Consequently, increased attention has been shifted to *An. sinensis* due to its wide distribution, high abundance and modest susceptibility to malaria parasites. However, little genetic information is available for this species, which has significantly limited the further research and development of new vector control tools.

Vector control is a critical component of all malaria control strategies [4,5]. Currently, it relies primarily on two interventions: long-lasting insecticide nets (LLIN)

* Correspondence: gaoqi54@hotmail.com; guiyuny@uci.edu
†Equal contributors
[1]Department of Parasitology, Medical College of Soochow University, Suzhou 215123, PR China
[3]Program in Public Health, College of Health Sciences, University of California at Irvine, Irvine, CA 92697, USA
Full list of author information is available at the end of the article

and indoor residual spraying (IRS) [6]. Due to their low toxicity and high efficacy, pyrethroids are the only insecticide approved by the World Health Organization (WHO) for bed net impregnation, and they are used often for IRS [7]. The significant increase in insecticide-based malaria vector control in the past decade has resulted in rapid spread of resistance among malaria vectors, which has placed current global efforts in malaria control and elimination at risk [8-10]. In 2012, WHO launched a Global Plan for Insecticide Resistance Management in malaria vectors (GPIRM), which advocates the collection of baseline information on insecticide resistance at the global scale and the development of novel methods to further understand the molecular mechanism of insecticide resistance and to enhance resistance surveillance [11]. In China, deltamethrin and permethrin were the major pyrethroid insecticides for vector control, and pyrethroid resistance was reported in limited number of *An. sinensis* populations in the 2000s [12]. However, deltamethrin resistance has been found in *An. sinensis* from southern and central China where malaria is endemic or epidemic [13-16]. Resistance to various classes of insecticides was also reported in *An. sinensis* in South Korea [17-19].

Two types of pyrethroid insecticide resistance mechanisms have been recognized, including alteration of the target sites in the *para*-type sodium channel that leads to knockdown resistance (*kdr*) and up-regulation of insecticide detoxification enzymes such as P450 monooxygenases, glutathione S-transferases and carboxylesterases [20]. For *An. sinensis*, high *kdr* mutation frequencies have been reported in central China (Hunan, Hubei, Henan and Anhui provinces) whereas Yunan populations in southwestern China completely lacked *kdr* mutations [13,16]. Overall, metabolic detoxification enzyme plays an important role in pyrethroid resistance [13,16]. Recent studies using the microarray hybridization method suggest that cuticle-related genes may also be associated with resistance [21]. The recent development of the RNA-seq method enables us to determine gene expression patterns while providing *de novo* sequencing, assembly and annotation of expressed genes [22]. The RNA-seq method is particularly suitable to species like *An. sinensis* which do not have a complete genome sequence, as it can help with developing expressed sequence tag (EST) databases. Various sequencing platforms (e.g., Roche 454, Solexa/Illumina, ABI SOLiD, etc.) are available, each having its own advantages [22,23]. For example, Roche 454 sequencing technology provides longer sequence reads, and thus it is useful for genome sequence assembly and developing EST databases (e.g., *An. funestus* [24], *Aedes aegypti* [25], *Bactrocera dorsalis* [26], *Melitaea cinxia* [27], *Zygaena filipendulae* [28], *Chrysomela tremulae* [29], *Aphis glycines* [30], *Cimex lectularius* [31], *Manduca sexta*

[32,33], *Laodelphax striatellus* [34], *Stomoxys calcitrans* [35], *Dermacentor variabilis* [35], *Erynnis propertius* and *Papilio zelicaon* [36], *Lygus hesperus* [37] and *Agrilus planipennis* [38]).

The objectives of this study were to enrich the genetic information of *An. sinensis,* an important malaria vector for genetic and comparative genomic studies, and to enhance understanding of insecticide resistance mechanism. We constructed a reference transcriptome of *An. sinensis* using the 454 GS FLX system with mosquito samples from different developmental stages, generated an EST database, and examined the transcriptome profiles between deltamethrin resistant and susceptible mosquitoes.
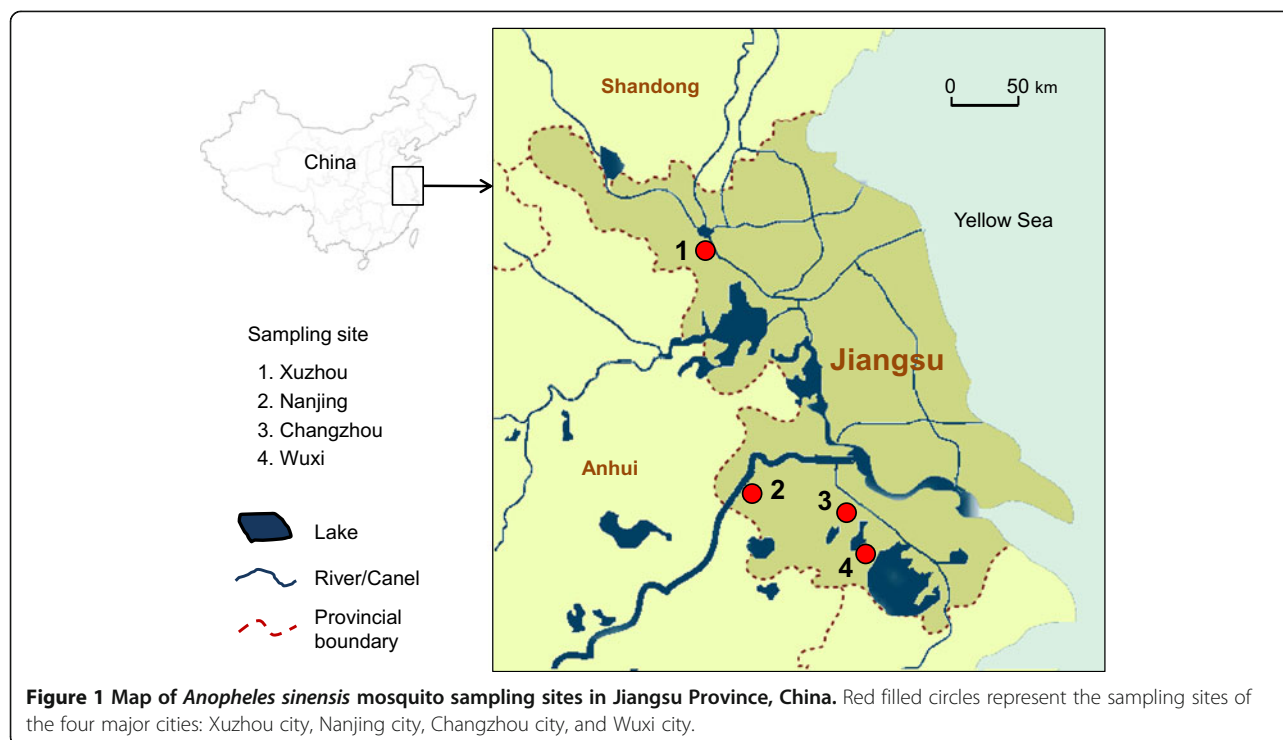
Methods

This study used *An. sinensis* mosquitoes collected from the field and laboratory susceptible colony. Field-collected *An. sinensis* mosquitoes were phenotyped for susceptibility to deltamethrin resistance and used for transcriptome profile determination in the resistant and susceptible populations. The laboratory susceptible colony was primarily used to generate reference expressed sequence tag (EST) databases.

Field mosquito sample collection and insecticide susceptibility bioassay

Anopheline mosquito larvae and pupae were collected from rice fields in four locations: Xuzhou, Nanjing, Changzhou and Wuxi of Jiangsu Province, China, between July and September, 2011 (Figure 1). A total of 30–50 breeding sites were used per site. The study area has been experiencing sporadic *P. vivax* malaria outbreaks. Since the 1980s, deltamethrin impregnated nets (ITNs) and indoor residual spraying (IRS) have been widely used to control mosquitoes. Rice is the major agricultural crop in these study sites. Due to severe insect pest damage to the rice crop, insecticide use for rice pest control has been very intensive, with several rounds of sprays in each growing season. Pyrethroids are commonly used for agricultural pest control in study areas, but other insecticides such as organic phosphates and carbamates are also being used.

The field-collected mosquito larvae and pupae were reared to adulthood in the insectary of the Key Laboratory on Technology for Parasitic Disease Control and Prevention, Ministry of Health, Jiangsu Institute of Parasitic Diseases (JIPD) in Wuxi, Jiangsu Province, China. The average room temperature and relative humidity were $27 \pm 1°C$ and $70 \pm 10\%$, respectively. After the mosquitoes were identified to species morphologically, female adults of *An. sinensis* 3 days post emergence were tested for susceptibility to deltamethrin using the standard WHO tube bioassay with 0.05% deltamethrin test papers [39]. Between 100 and 200

Figure 1 Map of *Anopheles sinensis* mosquito sampling sites in Jiangsu Province, China. Red filled circles represent the sampling sites of the four major cities: Xuzhou city, Nanjing city, Changzhou city, and Wuxi city.

mosquitoes per location were tested with 20 mosquitoes per tube, providing 5–10 biological replicates. The knockdown time was recorded for all tested mosquitoes at 10-minute intervals during the 60 minute exposure time. All mosquitoes alive or dead 24 hours after the recovery period were preserved in 100% ethanol or RNAlater (Qiagen) for subsequent molecular analyses. WHO [39] criteria were used to classify resistance status of the tested mosquito population (resistant if mortality is <90%, probable resistant if mortality is 90–98%, and susceptible if mortality is >98%) [39].

Molecular identification of mosquito species and *kdr* genotyping

Genomic DNA was extracted from single mosquito legs for molecular identification and *kdr* genotyping of *An. sinensis*. A total of 514 field-collected mosquitoes were identified to species using the ribosomal DNA internal transcribed spacer 2 (rDNA-ITS2)-based method [40], and *kdr* genotyping of the field-collected *An. sinensis* mosquitoes followed our previous established allele-specific PCR (AS-PCR) method [16].

Laboratory colony mosquito sample preparation for reference construction

To obtain reference EST information for *An. sinensis*, we used mosquitoes from all development stages (eggs, first to fourth instar larvae, pupae, and female adults) of an *An. sinensis* laboratory colony maintained at

JIPD. A total of 100 eggs, 20 larvae, 20 pupae, and 20 female adult mosquitoes of the *An. sinensis* laboratory colony were used in the study. This laboratory colony was never exposed to any insecticide and was highly sensitive to deltamethrin.

RNA-seq library preparation and 454 sequencing

To maximize transcripts coverage, we constructed 4 RNA libraries from *An. sinensis* mosquitoes of different development stages of JIPD laboratory susceptible colony, and from field collected, deltamethrin-resistant and susceptible mosquitoes. The first library was pooled RNA samples from 100 eggs, 20 larvae and 20 pupae of the *An. sinensis* laboratory colony. Thus this library targeted mosquito immature stages. The second library was RNA samples from 20 female adults, four to five days post emergence, of the *An. sinensis* laboratory colony. The third library was made from 20 female adult mosquitoes reared from field-caught larvae, four to five days post emergence, phenotypically identified as deltamethrin resistant by using the standard WHO tube assay. The fourth library was made from 20 female adult mosquitoes reared from field-caught larvae, four to five days post emergence, phenotypically identified as deltamethrin susceptible. Mosquitoes for the third and fourth libraries were from the same field sites in Jiangsu Province. Data from all four libraries were used to develop the EST database. Comparison of RNA-seq data between the third and fourth libraries

would reveal transcriptome differences associated with deltamethrin resistance.

Field-collected deltamethrin-resistant individuals were confirmed by the WHO insecticide-susceptibility bioassay. The mosquitoes that survived after the 24 hr recovery period were classified as deltamethrin resistant, and those knocked down early during the bioassay were classified as deltamethrin susceptible [41]. This classification is reasonable because the discriminant dosage used in the bioassay (0.05% deltamethrin) kills 99.9% of susceptible mosquitoes [39]. We used deltamethrin resistant and susceptible mosquitoes from the same area under the same growing conditions to reduce the confounding effects of geographic origin of the specimens and larval growing conditions.

Total RNA of *An. sinensis* was extracted from the deltamethrin resistant and susceptible mosquitoes using the PureLink RNA Mini Kit (Life Technologies). Similar procedures were used to extract the total RNA of various stages of the laboratory colony *An. sinensis* mosquitoes used in the EST discovery. Total RNA quantity was assessed using the NanoDrop spectrophotometer (NanoDrop Technologies), and RNA integrity was examined by running on 1.2% agarose gel with the treatment of RNase Zap (Sigma-Aldrich) followed by the Agilent 2100 Bioanalyzer.

The full-length cDNA for each library was synthesized and amplified using the Mint-2 cDNA synthesis kit (Evrogen, Moscow, Russia). Briefly, the first strand of the cDNA was synthesized using a mixture of 0.5 µg total RNA, 1 µl PlugOligo-3 M adapter at 15 µM (5′-AAGCAGTGGTATCAACGCAGAGTGGCCATTACG GCCGGGGG-3′), and 1 µl 3′-end CDS-Gsu adapter at 10 µM (5′-AAGCAGTGGTATCAACGCAGAGT ACTGGAG(T)20VN-3′). Because the first strand cDNA synthesis started from a 3′-end CDS adapter that contained an oligo (dT) 20 sequence which anneals to poly(A) stretches of mRNA, this procedure selected mRNA and helped filtering out rRNA. The mixture was incubated at 70°C for 2 min, and then the incubation temperature was decreased to 42°C. After incubation, each mixture was added to 5 µl reverse transcription solution (2 µl 5X first strand buffer, 1 µl DTT at 20 mM, 1 µl dNTPs at 10 mM each, and 1 µl mint reverse transcriptase), incubated at 42°C for 30 min, and then incubated with IP-solution (solution for incorporation of the PlugOligo sequence) for 1.5 hr at 42°C. Next, the double-stranded cDNA was synthesized in a 50 µl reaction volume containing 40 µl sterile RNase-free water, 5 µl 10X Encyclo buffer, 1 µl dNTP mix (10 mM each), 2 µl PCR primer at 10 µM (5′-AAGCAGTGGTATCAACGCA-GAGT-3′), 1 µl first-strand cDNA, and 1 µl 50X Encyclo polymerase mix. The amplification was performed with the following conditions: one cycle at 95°C for 1 min, followed by 18 cycles at 95°C for 15 sec, 66°C for 20 sec, and 72°C for 3 min.

Our first library was used for EST discovery; therefore, this library was normalized using the Trimmer-2 cDNA normalization kit (Evrogen) to minimize over-presentation of some abundant transcripts. The CDS-Gsu adapter was used for normalization prior to the cDNA library construction. The other three libraries were not normalized because our objective was to determine the transcriptome profiles in field-collected deltamethrin resistant or susceptible mosquitoes and in laboratory-reared susceptible mosquitoes. In this case, the CDS-Gsu adapter was replaced by the CDS-4 M adapter (5′-AAGCAGTGGTATCAACG CAGAGTGGCC GAGGCGGCC(T)4G(T)6C(T)13VN-3′). Each RNA-seq library was sequenced using the 454 GS FLX by the Center for Integrated Biosystems, Utah State University, USA.

RNA-seq data analyses
De novo transcriptome assembly and functional annotation
We first trimmed the adapter sequences from the original reads of the 454 GS FLX and removed the reads with low quality and short sequences (<50 bp). *De novo* assembly was conducted for the individual libraries and for the four libraries pooled to maximize the coverage and to provide a more comprehensive reference transcript. The *de novo* assembly was generated using the CLC Genomics Workbench version 6.0 (CLC Bio) with default parameters (similarity = 0.8, length fraction = 0.5, insertion/deletion cost = 3, and mismatch cost = 3) [42]. The contigs of >200 bp assembled from pooled libraries were BLASTXed against the nr (non-redundant) protein database of the National Center for Biotechnology Information (NCBI), and the taxonomic distribution of hits matching the nr protein database was determined. BLAST output was post-processed and lists of hits with match locations, e-values and bit scores were generated. A cutoff e-value of 1e-5 was used. Gene Ontology (GO) annotations were conducted by using the Blast2GO program [43], and the WEGO software was used to display GO functional classification [44].

Homology-based EST filtering
To determine *An. sinensis* gene homology to other previously sequenced mosquito species, a BLASTN search against *An. gambiae*, *Aedes aegypti* and *Culex quinquefasciatus* transcript-protein sequences in VectorBase database (http://www.vectorbase.org) was conducted. BLASTN results were parsed so that contigs with an E-value <0.001 in bi-directional best hit were retained for each query. Each homolog of the gene pair was then searched with TBLASTX against the three genomes. ESTs matching with an E-value $<1 \times 10^{-5}$ were considered putative homologs. Identical top BLAST hits suggested

putative orthologs, whereas non-identical top BLAST hits suggested the possible presence of paralogs, which were excluded from the analysis. The matched subsequence of the query contig that had the longest aligned region among all queries was used as the candidate transcript. Finally, identical candidate ESTs were collapsed into unique EST sequences.

Identification of differentially expressed genes (DEGs)

To compare the transcriptomes from field mosquitoes and library colony, and the resistant and susceptible mosquitoes, the trimmed reads from the second library (library colony), the third (deltamethrin- resistant) and fourth library (deltamethrin-susceptible) were mapped and aligned to the final EST set using the CLC Bio software. The following mapping parameters were used: mismatch cost = 2, insertion cost = 3, deletion cost = 3, length fraction = 0.5, similarity fraction = 0.8. Gene expression was quantified by RPKM (reads per kilobase per million mapped reads) to minimize the influence of variation in gene length and total number of reads [42]. RPKM values were log2 (RPKM + 0.0001) transformed, and average linkage clustering was used. Genes were classified as differentially expressed if they exhibited two folds or greater changes between the two samples, and statistical significance at P < 0.01 based on the Audic-Claverie method [45-47] and the false discovery rate (FDR) < 0.01 [48].

Microsatellite and SNP discovery

The contigs assembled from the four library sequence reads were searched for microsatellite repeats using the MsatCommander software [49]. A microsatellite was identified if the contig contained a motif that was repeated at least six times for dinucleotides and at least four times for tri-, tetra-, penta-, and hexanucleotides. Candidate SNPs were identified from assembled reads using the CLC Bio software with quality-based variant detection [42]. Parameters were as follows: neighborhood radius = 5, maximum gap and mismatch count = 2, minimum neighborhood quality = 15, minimum central quality = 20, minimum coverage = 10, and minimum variant frequency (%) = 35.0. Putative open reading frames (ORFs) of assembled transcripts with a minimum size of 100 bp were predicted using the GETORF program in the EMBOSS package [50]. The longest putative ORF from any particular sequences were selected. When two similarly long (within 90% of each other) ORFs were found located in opposite strands of the contigs, we followed the method of Pellino [51] by comparing the polarity of the ORFs with the strand orientation from the Blastx analysis. A custom Perl script was used to test whether these SNPs resulted in an amino acid change in the predicted ORF.

Experimental validation of RNA-seq data

Seven differentially expressed transcripts between the deltamethrin resistant and susceptible mosquitoes identified from RNA-seq were selected for independent validation of gene expression using quantitative real-time PCR (qRT-PCR) [52]. We chose these genes because of their important functions in insecticide resistance and mosquito immunity. Twenty An. sinensis female adults reared from field-collected larvae that were phenotyped as deltamethrin resistant or susceptible by the WHO tube bioassay were used. Quantitative real-time PCR reactions were performed in triplicate with LightCycler® 480 SYBR Green I Master. Fold changes in gene expression between resistant and susceptible mosquitoes were derived by the comparative CT method using the 18S ribosomal protein gene as an internal control [52]. Gel electrophoresis of the PCR products was also conducted to verify that a single gene-specific product was produced.

Data deposition

The Roche 454 reads of An. sinensis obtained in this study were submitted to the NCBI Sequence Read Archive under the BioSample accession number of SAMN01892512 - SAMN01892515, and the Transcriptome Shotgun Assembly contigs (length ≥ 200 bp) were deposited to DDBJ/EMBL/GenBank under the accession number GAFE01000001-GAFE01028133 (Bioproject 186896).

Ethics statement

No specific permissions were required for the described field studies. For mosquito collection in rice paddies, oral consent was obtained from field owners in each location. These locations were not protected land, and the field studies did not involve endangered or protected species.

Results

Susceptibility to deltamethrin and *kdr* genotyping

A total of 514 female mosquitoes reared from field-collected larvae were bioassayed for resistance to deltamethrin. We found a high level of resistance in the four mosquito populations because the mortality rate was all <10% whereas the laboratory susceptible strain showed 100% mortality (Additional file 1). PCR analysis with 514 specimens confirmed that all mosquito specimens examined in the bioassay were all *An. sinensis*. Three types of *kdr* mutations at codon position 1014 of the *para*-type sodium channel gene were detected: two mutations (from TTG to TTT and from TTG to TTC) that cause nonsynonymous changes from leucine to phenylalanine (L1014F), and one mutation from TTG to TGT that causes leucine to cysteine substitution (L1014C). The *kdr* mutation frequency was high (88–100%) among the four tested populations and the predominant *kdr*

mutation was L1014F. No *kdr* mutation was detected in the susceptible laboratory strain (Additional file 1).

Assembly of contigs

The Roche 454 GS FLX platform produced a total of 624,559 reads from the 4 libraries. The average raw read length was 290.8 bp. After removal of duplicated, ambiguous and short (<50 bp) sequences, a total of 394,254 reads were used for EST contig assembly (Table 1). The average percentage of reads mapped to contig sets was 71.9%, producing a total of 33,411 EST contigs with an average length of 493 bp. Among them, 31,772 contigs were longer than 50 bp, 28,133 contigs were longer than 200 bp (Figure 2, GenBank accession: GAFE01000001- GAFE01028133).

Homology-based EST filtering and EST function annotation

To determine the taxonomy distribution of the assembled contigs, all 31,772 EST contigs longer than 50 bp were BLASTXed against the NCBI nr (non-redundant protein sequences) database using the expected value of 1e-5. We found that 10,504 unique contigs were matched to the nr database, with the most hits matching to *An. gambiae*, the most important malaria vector in Africa, followed by other mosquito species and insects (Figure 3, Additional file 2).

When we searched 31,772 assembled contigs against transcript databases of three mosquito species (*An. gambiae*, *Ae. aegypti* and *Cx. quinquefasciatus*) with the searching match criteria described in the method by BLASTN, 8,057 ESTs were identified. Among these, 5,545 were from laboratory adults and field-collected mosquitoes (Table 2). These ESTs showed >94% matched hits with AGAP (*An. gambiae*), consistent with BLAST-nr search results in which the dominant match was with *An. gambiae*.

Gene Ontology (GO) classification and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis

The 8,057 ESTs were assigned to GO accession numbers (Additional file 3), and they were classified into 44 function categories under three major domains (biological process, cellular component, and molecular function; Figure 4). To investigate the biological function, all 8,057 ESTs were mapped to the reference pathways in the KEGG. A total of 2,241 ESTs obtained KEGG annotation, and 1,957 ESTs were assigned to 289 KEGG pathways. A large number of ESTs were restricted to single pathways, such as metabolic pathway (390 ESTs), biosynthesis of secondary metabolites (122 ESTs), and microbial metabolism in diverse environment (74 ESTs).

Differentially expressed genes associated with deltamethrin resistance and pathway analyses

Comparisons between field deltamethrin-resistant adult mosquitoes and adult from the susceptible laboratory colony identified a total of 2,088 (25.9%) transcripts that were expressed at 2 folds or greater changes with P-value < 0.01 and FDR < 0.01) (Additional file 4). Among them, 1,197 (14.8%) were up-regulated and 891 (11.1%) down-regulated in the field resistant mosquitoes. Comparison of RNA-seq data with samples between field resistant and field susceptible mosquitoes, we found a total of 2,131 differentially expressed transcripts, 1,079 (13.4%) up-regulated and 1,052 (13.1%) down-regulated in resistant mosquitoes (Figure 5, Additional file 4). Among them, 930 (11.5%) transcripts were expressed only in resistant mosquitoes, while 665 (8.3%) transcripts expressed only in susceptible mosquitoes. Among these significantly differentially expressed genes, 1,286 showed significant matches to the KEGG database, and 916 genes were mapped to 294 pathways. The remaining 370 genes did not have the related genes annotation and KO number when searching with KEGG Mapper (http://www.genome.jp/kegg/tool/map_pathway2.html). The most commonly identified pathway was "metabolic pathway," which was previously found to be involved in insecticide metabolism [53-56]. Among the different transcripts, the majority were in the families of cytochrome P450 monooxygenases (21.7%), NADH dehydrogenase (21.7%), and glutathione S-transferase (8.7%) (Figure 6). The transcripts with the largest increases in expression were also in the metabolic pathway (Additional file 5), especially the genes coding for P450 monooxygenases, GSTs, and the NADH dehydrogenase, a cofactor of P450 monooxygenases. Mitochondrial cytochrome C oxidase and the

Table 1 Summary of reads and assembly from 454 GS FLX sequencing for *Anopheles sinensis*

Library	Raw reads	Average length (bp)	Reads used in assembly*	% reads mapped to contig set	No. contigs	N50 contigs (bp)	Average contig length (bp)	% GC contigs
Immature stage, colony	164,503	341.1	122,661	75.45	17,398	491	447	45.3
Female adults, colony	153,568	299.0	106,268	71.1	7,415	515	453	44.5
Deltamethrin resistant, field	171,679	363.3	124,723	73.0	13,897	524	481	45.5
Deltamethrin susceptible, field	134,809	304.6	91,193	68.3	10,597	464	427	45.2
Total	624,559	290.8	394,254	71.9	33,411	547	493	45.1

*Reads used in assembly is less than the total number of raw reads due to removal of duplicated, ambiguous and short (<50 bp) sequences.

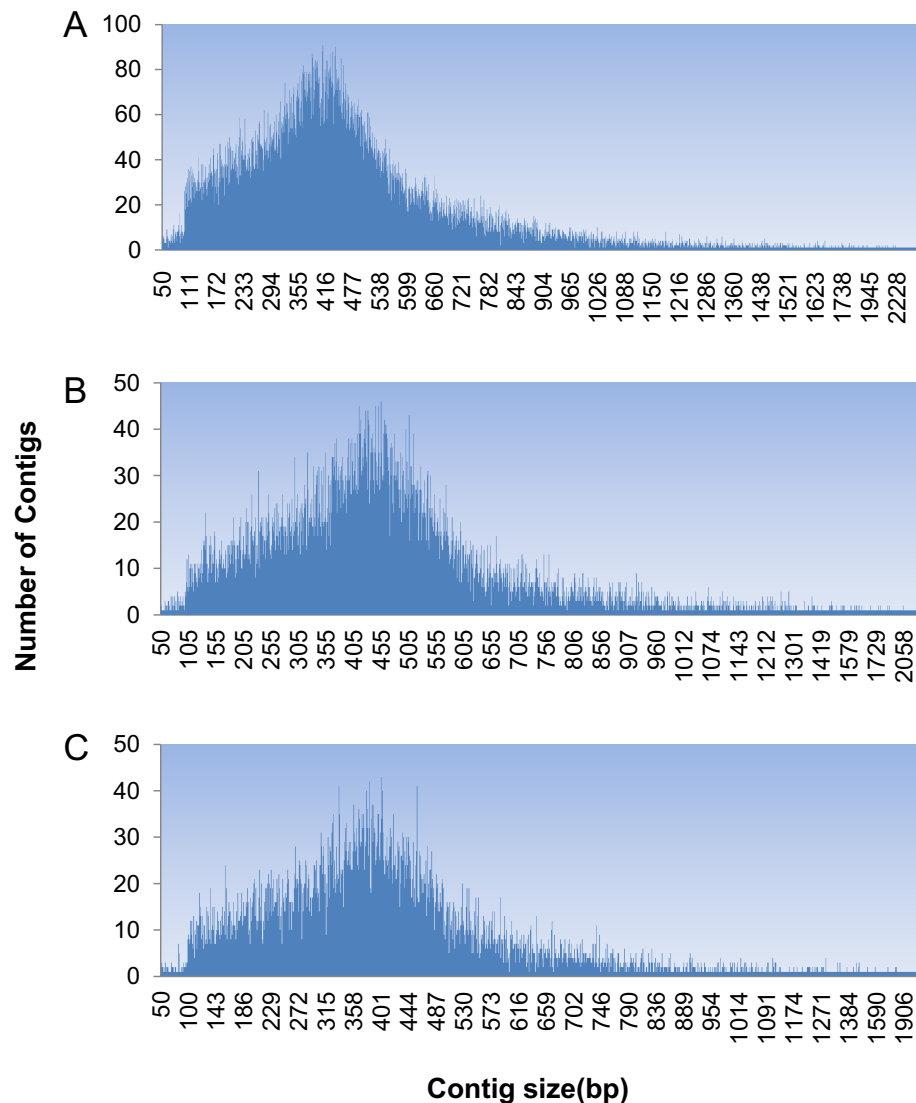**Figure 2 Contig size distributions of *Anopheles sinensis*. A**: laboratory colony and field population (all four groups); **B**: Deltamethrin resistant female adults collected from the field; **C**: Deltamethrin susceptible female adults collected from the field.

cuticular proteins exhibited significantly differential expression between deltamethrin resistant and susceptible mosquitoes.
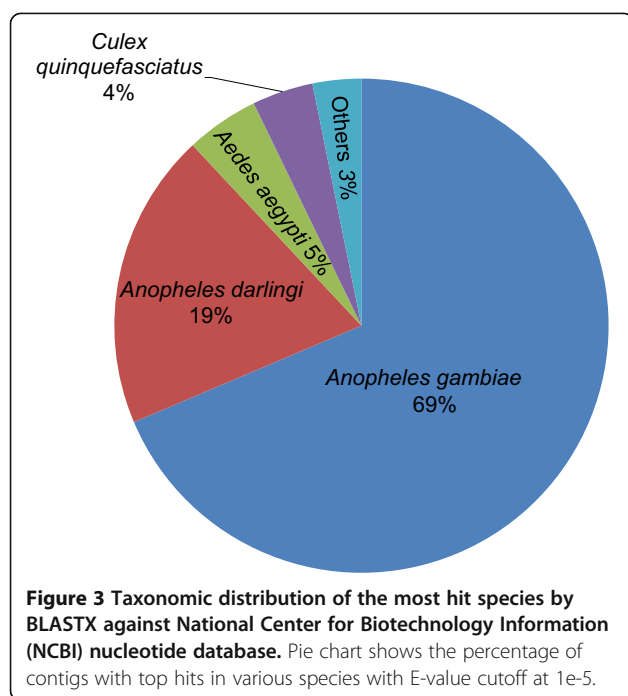
Verification of RNA-seq data by quantitative RT-PCR

The qRT-PCR reactions were carried out with 6 transcripts identified by RNA-seq as differentially expressed between deltamethrin resistant and susceptible mosquitoes. The gene name, function and primer sequences used for qRT-PCR amplification for RNA-seq validation are shown in Additional file 6. When expression data from RNA-seq were compared to qRT-PCR, the direction of fold changes detected by these two methods were consistent for all 6 tested transcripts, although the RNA-seq method

generally identified higher numbers of fold changes (Additional file 7).

Microsatellite discovery

The assembled 31,772 contigs were used for *An. sinensis* microsatellite identification. Overall, a total of 2,408 di-, tri-, tetra-, penta-, and hexanucleotide microsatellites from 2,225 contigs were identified on the conditions that these microsatellites had a minimum 6, 4, 4, 4, and 4 repeat units and each microsatellite was found in at least 5 contigs. The distributions of microsatellites in these assembled contigs are shown in Figure 7. The dinucleotide repeats were the most abundant (51.7%), followed by trinucleotide and tetranucleotide repeats (41.4%). The AC and GT dinucleotide repeats were the most

Figure 3 Taxonomic distribution of the most hit species by BLASTX against National Center for Biotechnology Information (NCBI) nucleotide database. Pie chart shows the percentage of contigs with top hits in various species with E-value cutoff at 1e-5.

abundant (29.7%) classes of repeats, while the CG repeats were the least abundant. Among trinucleotide microsatellites, AGC repeats were the most abundant, while CGG repeats were the least abundant.

SNP discovery

The present study used pooled field mosquito individuals for 454 sequencing, and thus it offered a good opportunity to identify SNPs in the transcribed sequences of *An. sinensis*. The criteria for SNP identification were at least two occurrences of the minority allele at sites covered by at least ten sequences. We found a total of 15,496 SNPs in the pooled transcriptome, providing an estimation of approximately 6 SNPs per 1000 bp with an average coverage of 28. The majority of the SNPs (60%) were transitions (C↔T and A↔G). The percentage of transversions ranged from 7.2% (G↔C) to 14.0% (A↔T). To distinguish synonymous and non-synonymous SNPs, the GETORF program was used to identify open reading frames. A total of 34,918 open reading frames and 11,097 SNPs were identified. Of those SNPs, 2,138 (19.3%) resulted in predicted amino acid substitutions.

To assess the reliability of these SNPs, we examined 15 contigs coding for genes related to metabolism and detoxification (Table 3). A total of 177 SNPs were found in 18,156 bp coding sequences examined. Among them, 65 were transition substitutions, resulting in 27 non-synonymous substitutions, whereas 112 SNPs were transversion substitutions, resulting in 31 non-synonymous substitutions. Overall, transitions resulted in more frequent non-synonymous substitutions (41.5%) than did transversions (27.7%). The average number of SNPs per kb ranged from 2.9 (carboxylesterase) to 33.8 (GSTE6).
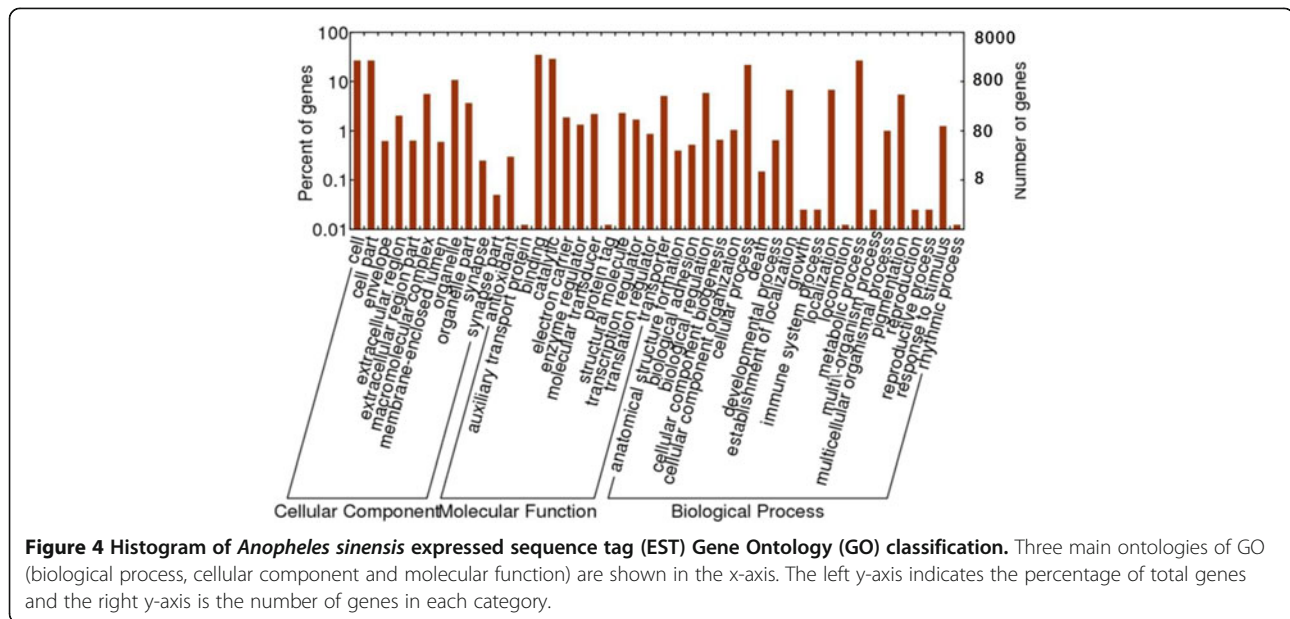
Discussion

Anopheles sinensis is the most important malaria vector in China and other Southeast Asian countries due to its overwhelming dominance in abundance [1,13,57-59]. High levels of insecticide resistances and mechanisms to insecticide have been reported in *An. sinensis* [13,14,16,18,19]. The vector capacity and susceptibility of *An. sinensis* have been examined [3,58,60]. Recently, a preliminary genome sequence of *An. sinensis* was published [61]. However, the lack of transcriptome data has hindered research for molecular mechanisms of malaria transmission and new vector control methods. This study generated an EST database for *An. sinensis* and determined the transcripts associated with deltamethrin resistance. Our *de novo* assembly generated 33,411 contigs with average length of 493 bp. A total of 8,057 ESTs were generated with GO and KEGG annotation. Further, we identified more than 2,400 microsatellite markers. Therefore, this study has enhanced knowledge on the *An. sinensis* genome and transcriptome and has provided new tools for future genetic research

Although *de novo* assembly of 454 reads is more straight forward than Illumina reads, it may not be exhaustive. Published studies on mosquito species that used 454 sequencing include *An. funestus*, *Ae. aegypti* and *Ae. albopictus* [24,62,63]. The 454 sequencing in *An. funestus* generated 375,619 reads, and *De novo* assembly generated 18,103 contigs with average length of 253 bp [24]. In *Ae. aegypti*, 454 pyrosequencing resulted in 0.2 - 0.3 million useable reads and *De novo* assembly generated approximately 15,000 contigs [62]. With >1.1 million quality reads obtained, Poelchau et al. reported 69,474 contigs assembled in *Ae. albopictus* mosquitoes [63]. The present study obtained about 400,000 high-quality reads, and

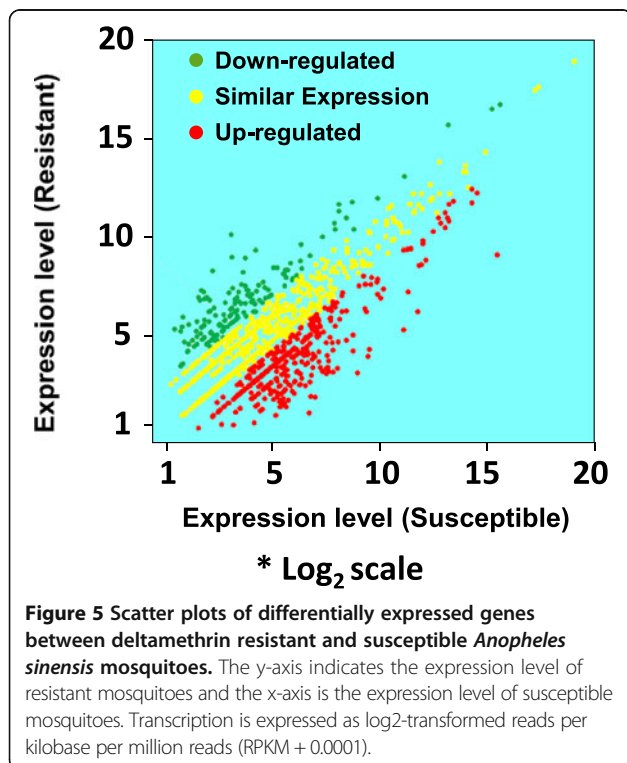**Table 2 Number of expressed sequence tags (ESTs) matching to reference mosquito transcript**

	Number of matches to			
Reference organism	Final EST set	Female adult, colony	Deltamethrin resistance, field	Deltamethrin susceptible, field
Anopheles gambiae	6,781	1,496	2,000	1,743
Aedes aegypti	471	40	53	47
Culex quinquefasciatus	805	52	58	56
Total number of matches	8,057	1,588	2,111	1,846

Figure 4 Histogram of *Anopheles sinensis* expressed sequence tag (EST) Gene Ontology (GO) classification. Three main ontologies of GO (biological process, cellular component and molecular function) are shown in the x-axis. The left y-axis indicates the percentage of total genes and the right y-axis is the number of genes in each category.

De novo assembly generated more than 33,000 contigs with a much higher average length, suggesting that our assembly for *An. sinensis* based 454 sequencing was reasonably good in comparison to the previously published transcript assembly.

RNA-seq allowed a holistic view of the transcriptome and provided absolute rather than relative gene expression measurements. Our RNA-seq analysis identified

Figure 5 Scatter plots of differentially expressed genes between deltamethrin resistant and susceptible *Anopheles sinensis* mosquitoes. The y-axis indicates the expression level of resistant mosquitoes and the x-axis is the expression level of susceptible mosquitoes. Transcription is expressed as log2-transformed reads per kilobase per million reads (RPKM + 0.0001).

more than 2,100 transcripts that may be associated with deltamethrin resistance. Among them, 1,079 transcripts were up-regulated and 1,052 down-regulated. In *An. gambiae*, RNA-seq identified a total of 1,093 transcripts with significantly differential expression between deltamethrin resistant and susceptible mosquitoes from Kenya [41]. These transcripts were distributed over the entire genome, with some transcripts mapped within the previously identified quantitative trait loci (QTLs) linked to permethrin resistance, such as *rtp1* and *rtp2* [41]. The present study with *An. sinensis* detected twice the number of differentially expressed transcripts as in *An. gambiae*. The higher number of differentially expressed transcripts identified in *An. sinensis* may be partially due to the difference in deltamethrin resistance. The study populations used in the present study were extremely resistant, as evidenced by <10% morality in the standard WHO tube test for insecticide susceptibility, in contrast to >66% morality in *An. gambiae*. Further, metabolic detoxification enzymes were found to be very important in *An. sinensis* [16]. The present study confirmed that transcripts with the largest increase in expression were in metabolic pathways, particularly those coding for glutathione S-transferase, P450 monooxygenase, and cytochrome C oxidase detoxification enzymes. This is consistent with previous findings on higher P450 monooxygenase activities in pyrethroid resistant *An. gambiae* [24] and the association between the overexpression of cytochrome C oxidase subunit I gene and pyrethroid resistance in German cockroaches, *Blattella germanica* [64].

Our transcriptome analysis provided insights on the role of specific metabolic detoxification genes in resistance. For example, contig 6816 was annotated to be
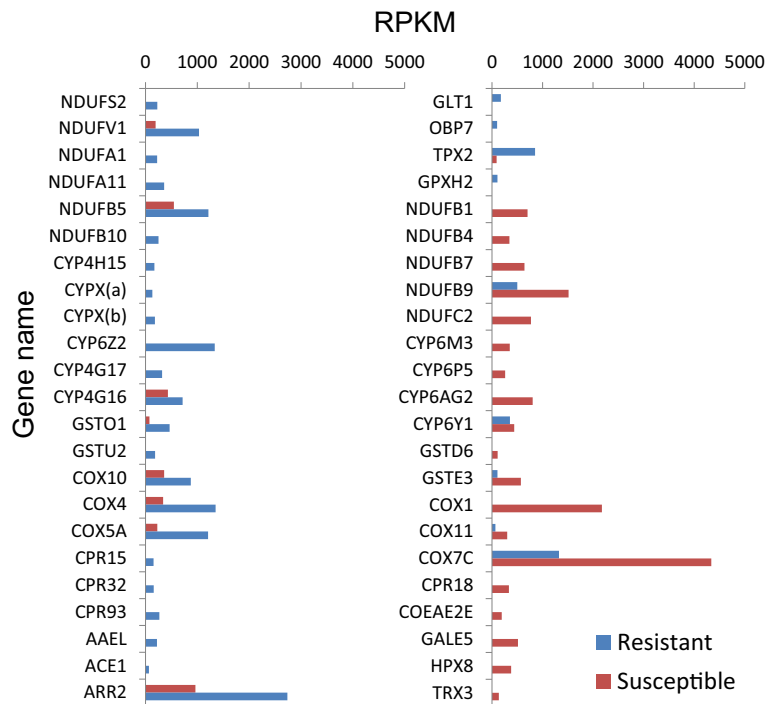
Figure 6 Distribution of expression values in genes involved in metabolic detoxification and insecticide penetration in deltamethrin resistant and susceptible Anopheles sinensis mosquitoes. RPKM refers to the reads per kilobase per million mapped reads.

CYP6Z2 (access no GAFE01005115), and showed 89% similarity in amino acid sequence to AGAP008218 of *An. gambiae*, which was known to be one of the important detoxification genes responsible for pyrethroid resistance in *An. gambiae* [65,66] and *An. arabiensis* [67,68]. The present study found that this gene was strongly up-regulated in resistant mosquitoes and it showed the highest differentiation in expression between resistant and susceptible mosquitoes. The other cytochrome P450 genes that were strongly up-regulated in the resistant *An. sinensis* mosquitoes were CYP4 genes, including CYP4G16, CYP4G17 and CYP4H15. This is consistent

with previous studies in house flies and other mosquito species that showed P450 genes were up-regulated in resistant individuals [69-72]. Similarly, we also found two GST genes (GSTU2 and GSTO1), 3 cytochrome C oxidase genes (COX2, COX5A and COX10), and 6 NADH dehydrogenase genes were significantly up-regulated. It is of interest to note that arrestin gene (ARR2, access no: GAFE01006709) and thioredoxin peroxidase gene (TPX2, access no: GAFE01001198) were found strongly up-regulated in the resistant mosquitoes, consistent with published reports in *Cx. pipiens pallens* [73], *An. arabiensis* and *An. gambiae* [66,68,74].
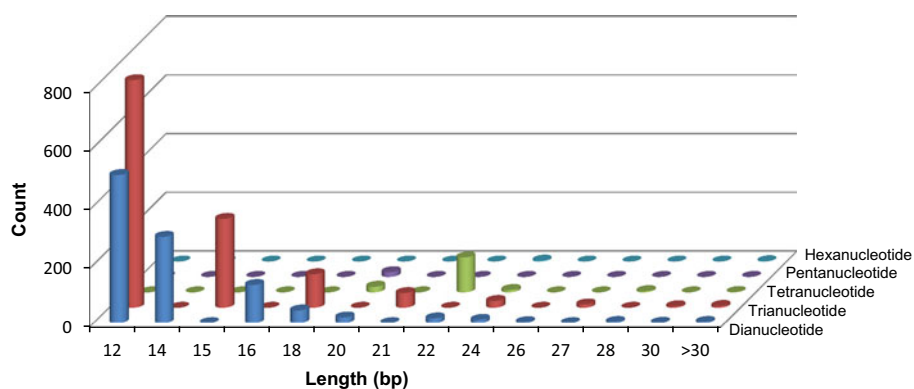

Figure 7 Distribution of microsatellites in Anopheles sinensis expressed sequence tags (ESTs). The y-axis indicates the number of sequences and the x-axis is the length of repeats.

Table 3 Nucleotide polymorphism of genes related to metabolism, detoxification and immunity in *Anopheles sinensis*

Accession no.	Gene	Open reading frame	Chromosome*	Transition	Transversion	Total	Coverage	Average no. SNP per kb
GAFE01006078	CYP6AG2	148–1575	2R	2 (0)	12 (4)	14	11–29	9.8
GAFE01013675	CYP6ZX	193–1587	3R	8 (3)	11 (3)	19	10–25	13.6
GAFE01007218	CYP6Y1	330–1460	3R	2 (0)	4 (0)	6	11–20	5.3
GAFE01004582	GSTT1	63–749	X	0	3 (1)	3	18–34	4.4
GAFE01006360	GSTD5	94–735	2R	0	2 (0)	2	10–10	3.1
GAFE01012264	GSTE6	15–488	3R	10 (6)	6 (2)	16	10–17	33.8
GAFE01006339	Carboxylesterase	515–1204	2R	1 (0)	1 (0)	2	10–10	2.9
GAFE01005990	NYD-GBE	688–2178	3 L	1 (1)	7 (0)	8	12–26	5.4
GAFE01001512	Euk_Ferritin	390–1085	2R	5 (0)	6 (0)	11	25–110	15.8
GAFE01000429	LR1M1	46–1548	2 L	4 (3)	10 (5)	14	16–115	9.3
GAFE01003583	Serpin	116–1564	2R	8 (3)	15 (3)	23	10–35	15.9
GAFE01007678	GlgB	96–2156	3 L	12 (8)	7 (2)	19	11–278	9.2
GAFE01008565	Serine	481–2070	3R	1 (0)	4 (2)	5	11–14	3.1
GAFE01018362	Peroxinectin	364–2040	2R	9 (2)	8 (5)	17	10–24	10.1
GAFE01021559	Tubulin	263–1504	2 L	2 (1)	16 (4)	18	10–15	14.5

The number in parentheses indicates non-synonymous single nucleotide polymorphisms (SNPs).
*Chromosome assignment was based on synteny to *Anopheles gambiae*.

On the other hand, our study also identified some CYP6 genes were down-regulated in the resistant mosquitoes, including CYP6AG2, CYP6M3, CYP6P5 and CYPY1. Other cytochrome C oxidase genes (COX7C and COX1) were also down-regulated. Similar phenomenon was reported in *Cx. quinquefasciatus* mosquitoes [75]. The mechanisms for down-regulation of some P450 genes and their relevance to insecticide resistance are unclear. It has been suggested that reduction in the expression of some metabolic detoxification genes may result from responses to various endogenous and exogenous compounds, or to pathophysiological signals [75-78].

The present study identified cuticular proteins differentially expressed between deltamethrin resistant and susceptible *An. sinensis* mosquitoes. The cuticular proteins may play a role in insecticide resistance by limiting the penetration of insecticides into the mosquito cuticle. This finding is not unique to *An. sinensis*; similar observations were reported in *An. funestus*. It is possible that thicker mosquito cuticles slow insecticide absorption and consequently increase the efficiency of metabolic detoxification [21]. It is important to note that while a transcript may have a large difference in expression between resistant and susceptible mosquitoes, this does not necessarily imply causality in the gene. Therefore, it is important to identify the key candidate genes, and the function of the candidate transcripts on pyrethroid resistance should be confirmed independently.

Molecular markers play an important role in genetic mapping, population genetics and genomics. In this study, we performed a genome-wide scan on the *An. sinensis* transcripts for microsatellites markers. A search for di- to hexa-nucleotide repeats yielded a total of 2,408 potential microsatellite markers. As expected, di-nucleotide repeats were the most frequent microsatellite motifs, followed by trinucleotide repeats. AC/GT (29.7%) and AGC (8.3%) were the most frequent motifs among the di-nucleotide and tri-nucleotide microsatellites respectively. We identified a total of 15,496 potential SNPs in the coding sequencing, or approximately 6 SNPs per 1000 bp. Such a SNP density was comparable to *Aedes aegypti* [79] and *An. gambiae* [80,81] with 5–9 SNPs per 1000 bp, but was substantially lower than *An. funestus* that had an average of 14 SNPs per 1000 bp [24]. The rate of transition (60%) and transversion (40%) substitutions was similar to *An. funestus* [24,82], Copepod [83] and fish [84]. The frequency of non-synonymous substitution SNPs was approximately 20%, similar to human (22%) [85]. These SNPs and microsatellite markers provide an extensive set of genetic markers for *An. sinensis,* and will facilitate future population genetic studies of this important malaria vectors in Southeast Asia.

Conclusion

This study has generated an EST database for *An. sinensis* mosquitoes and enriched the genetic information for this important malaria vector species. RNA-seq analysis identified differentially expressed transcripts, particularly the transcripts coding for metabolic detoxification enzymes and cuticular proteins. The microsatellite and SNP markers identified here provide new tools for future population and evolutionary genetics research.

Additional files

Additional file 1: Mortality rate of deltamethrin resistance bioassay and *kdr* allele frequency in four *Anopheles sinensis* mosquito populations from Jiangsu Province, China.

Additional file 2: The taxonomy distribution of the contigs matched to the nr databases by BLASTX.

Additional file 3: Gene Ontology (GO) accession number assignment of the 8057 ESTs.

Additional file 4: A list of significantly differentially expressed genes between field resistant and laboratory susceptible adult mosquitoes, and between field resistant and field susceptible adult mosquitoes. Relative expression levels and fold changes were log2 (RPKM + 0.0001) transformed. Negative value indicates down-regulation and positive

value indicates up-regulation in the resistant population. Threshold for significant expression: absolute value of Log2 (ratio) > 1, P < 0.01 and FDR corrected value (q) < 0.01.

Additional file 5: KO annotations of the genes with the largest difference in expression between deltamethrin resistant and susceptible *Anopheles sinensis* mosquitoes. Transcripts 1–10 are the genes with increased expression in the resistant mosquitoes, and transcripts 11–20 are the genes with reduced expression in the resistant mosquitoes.

Additional file 6: A list of gene name, function and primers used in qRT-PCR amplification for RNA-seq validation.

Additional file 7: Correlation of expression value measured by RNA-seq and qRT-PCR. The fold change in RNA-seq was measured by the log2 of RPKM (reads per kilobase per million mapped reads). The fold change in qRT-PCR was measured by ΔC_T value.

Authors' contributions
GDZ prepared insect collection, carried out insecticide resistance bioassay, conducted molecular analysis, and drafted the manuscript. JC, HZ, JL, YL, LB, and SX participated in insect collection and insecticide resistance bioassays. GFZ, MHW, and XC helped to analyze the data. GDZ, DZ, QG, and GY conceived and designed the experiments. DZ, QG, and GY drafted the manuscript. All authors read and approved the final manuscript.

Author details
[1]Department of Parasitology, Medical College of Soochow University, Suzhou 215123, PR China. [2]Jiangsu Institute of Parasitic Diseases, Key Laboratory of Parasitic Disease Control and Prevention (Ministry of Health), Jiangsu Provincial Key Laboratory of Parasite Molecular Biology, Wuxi, Jiangsu Province 214064, PR China. [3]Program in Public Health, College of Health Sciences, University of California at Irvine, Irvine, CA 92697, USA.

References
1. Rueda LM, Zhao TY, Ma YJ, Gao Q, Zhu GD, Khuntirat B, Sattabongkot J, Wilkerson RC: Updated distribution records of the *Anopheles (Anopheles) hyrcanus* species-group (Diptera: Culicidae) in China. *Zootaxa* 2007, **1407**:43–55.
2. 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.
3. Zhu G, Xia H, Zhou H, Li J, Lu F, Liu Y, Cao J, Gao Q, Sattabongkot J: Susceptibility of *Anopheles sinensis* to *Plasmodium vivax* in malarial outbreak areas of central China. *Parasit Vectors* 2013, **6**(1):176.
4. Kappe SHI, Vaughan AM, Boddey JA, Cowman AF: That was then but this is now: malaria research in the time of an eradication agenda. *Science* 2010, **328**(5980):862–866.
5. The malERA Consultative Group on Vector Control: A research agenda for malaria eradication: vector control. *PLoS Med* 2011, **8**(1):e1000401.
6. WHO: *Global Insecticide use for Vector-Borne Disease Control.* 5th edition. Geneva, Switzerland: WHO Pesticide Evaluation Scheme (WHOPES), World Health Organization; 2011.
7. WHO: *Pesticides and Their Application for the Control of Vectors and Pests of Public Health Importance.* Geneva, Switzerland: WHO Pesticide Evaluation Scheme (WHOPES), World Health Organization; 2006.
8. Chanda E, Hemingway J, Kleinschmidt I, Rehman AM, Ramdeen V, Phiri FN, Coetzer S, Mthembu D, Shinondo CJ, Chizema-Kawesha E, Kamuliwo M, Mukonka V, Baboo KS, Coleman M: Insecticide resistance and the future of malaria control in Zambia. *PLoS One* 2011, **6**(9):e24336.
9. Githeko AK, Ototo EN, Guiyun Y: Progress towards understanding the ecology and epidemiology of malaria in the western Kenya highlands: opportunities and challenges for control under climate change risk. *Acta Trop* 2012, **121**(1):19–25.
10. Ranson H, N'Guessan R, Lines J, Moiroux N, Nkuni Z, Corbel V: Pyrethroid resistance in African anopheline mosquitoes: what are the implications for malaria control? *Trends Parasitol* 2011, **27**(2):91–98.
11. Maxmen A: Malaria surge feared. *Nature* 2012, **485**(7398):293.
12. Cui F, Raymond M, Qiao C: Insecticide resistance in vector mosquitoes in China. *Pest Manag Sci* 2006, **62**(11):1013–1022.
13. Qin Q, Li Y, Zhong D, Zhou N, Chang X, Li C, Cui L, Yan G, Chen X: Insecticide resistance of *Anopheles sinensis* and *An. vagus* in Hainan Island, a malaria-endemic area of China. *Parasit Vectors* 2014, **7**(1):92.
14. Wang D, Xia Z, Zhou S, Zhou X, Wang R, Zhang Q: 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**(1):164.
15. Xu T, Zhong D, Tang L, Chang X, Fu F, Yan G, Zheng B: *Anopheles sinensis* mosquito insecticide resistance: comparison of three mosquito sample collection and preparation methods and mosquito age in resistance measurements. *Parasit Vectors* 2014, **7**(1):54.
16. Zhong D, Chang X, Zhou G, He Z, Fu F, Yan Z, Zhu G, Xu T, Bonizzoni M, Wang M-H, Cui L, Zheng B, Chen B, Yan G: Relationship between knockdown resistance, metabolic detoxification and organismal resistance to pyrethroids in *Anopheles sinensis*. *PLoS One* 2013, **8**(2):e55475.
17. 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.
18. Kang S, Jung J, Lee S, Hwang H, Kim W: The polymorphism and the geographical distribution of the knockdown resistance (kdr) of *Anopheles sinensis* in the Republic of Korea. *Malar J* 2012, **11**(1):151.
19. Kim H, Baek JH, Lee W-J, Lee SH: Frequency detection of pyrethroid resistance allele in *Anopheles sinensis* populations by real-time PCR amplification of specific allele (rtPASA). *Pestic Biochem Physiol* 2007, **87**(1):54–61.
20. Hemingway J, Hawkes NJ, McCarroll L, Ranson H: The molecular basis of insecticide resistance in mosquitoes. *Insect Biochem Mol Biol* 2004, **34**(7):653–665.
21. Wood O, Hanrahan S, Coetzee M, Koekemoer L, Brooke B: Cuticle thickening associated with pyrethroid resistance in the major malaria vector *Anopheles funestus*. *Parasit Vectors* 2010, **3**:67.
22. Ozsolak F, Milos PM: RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet* 2011, **12**(2):87–98.
23. Gibbons JG, Janson EM, Hittinger CT, Johnston M, Abbot P, Rokas A: Benchmarking next-generation transcriptome sequencing for functional and evolutionary genomics. *Mol Biol Evol* 2009, **26**(12):2731–2744.
24. Gregory R, Darby AC, Irving H, Coulibaly MB, Hughes M, Koekemoer LL, Coetzee M, Ranson H, Hemingway J, Hall N, Wondji CS: A *de novo*

expression profiling of *Anopheles funestus*, malaria vector in Africa, using 454 pyrosequencing. *PLoS One* 2011, **6**(2):e17418.

25. Paris M, Despres L: Identifying insecticide resistance genes in mosquito by combining AFLP genome scans and 454 pyrosequencing. *Mol Ecol* 2012, **21**(7):1672–1686.

26. Zheng W, Peng T, He W, Zhang H: High-throughput sequencing to reveal genes involved in reproduction and development in *Bactrocera dorsalis* (Diptera: Tephritidae). *PLoS One* 2012, **7**(5):e36463.

27. Vera JC, Wheat CW, Fescemyer HW, Frilander MJ, Crawford DL, Hanski I, Marden JH: Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Mol Ecol* 2008, **17**(7):1636–1647.

28. Zagrobelny M, Scheibye-Alsing K, Jensen N, Moller B, Gorodkin J, Bak S: 454 pyrosequencing based transcriptome analysis of Zygaena filipendulae with focus on genes involved in biosynthesis of cyanogenic glucosides. *BMC Genomics* 2009, **10**(1):574.

29. Pauchet Y, Wilkinson P, van Munster M, Augustin S, Pauron D, ffrench-Constant RH: Pyrosequencing of the midgut transcriptome of the poplar leaf beetle *Chrysomela tremulae* reveals new gene families in Coleoptera. *Insect Mol Biol* 2009, **39**(5–6):403–413.

30. Bai X, Zhang W, Orantes L, Jun TH, Mittapalli O, Mian MA, Michel AP: Combining next-generation sequencing strategies for rapid molecular resource development from an invasive aphid species. Aphis glycines. *PLoS One* 2010, **5**(6):e11370.

31. Bai X, Mamidala P, Rajarapu SP, Jones SC, Mittapalli O: Transcriptomics of the Bed Bug (*Cimex lectularius*). *PLoS One* 2011, **6**(1):e16336.

32. Pauchet Y, Wilkinson P, Vogel H, Nelson DR, Reynolds SE, Heckel DG, ffrench-Constant RH: Pyrosequencing the *Manduca sexta* larval midgut transcriptome: messages for digestion, detoxification and defence. *Insect Mol Biol* 2010, **19**(1):61–75.

33. Zou Z, Najar F, Wang Y, Roe B, Jiang H: Pyrosequence analysis of expressed sequence tags for *Manduca sexta* hemolymph proteins involved in immune responses. *Insect Biochem Mol Biol* 2008, **38**(6):677–682.

34. Zhang F, Guo H, Zheng H, Zhou T, Zhou Y, Wang S, Fang R, Qian W, Chen X: Massively parallel pyrosequencing-based transcriptome analyses of small brown planthopper (*Laodelphax striatellus*), a vector insect transmitting rice stripe virus (RSV). *BMC Genomics* 2010, **11**(1):303.

35. Olafson PU, Lohmeyer KH, Dowd SE: Analysis of expressed sequence tags from a significant livestock pest, the stable fly (*Stomoxys calcitrans*), identifies transcripts with a putative role in chemosensation and sex determination. *Arch Insect Biochem Physiol* 2010, **74**(3):179–204.

36. O'Neil S, Dzurisin J, Carmichael R, Lobo N, Emrich S, Hellmann J: Population-level transcriptome sequencing of nonmodel organisms *Erynnis propertius* and *Papilio zelicaon*. *BMC Genomics* 2010, **11**(1):310.

37. Hull JJ, Geib SM, Fabrick JA, Brent CS: Sequencing and *de novo* assembly of the western tarnished plant bug (*Lygus hesperus*) transcriptome. *PLoS One* 2013, **8**(1):e55105.

38. Mittapalli O, Bai X, Mamidala P, Rajarapu SP, Bonello P, Herms DA: Tissue-specific transcriptomics of the exotic invasive insect pest emerald ash borer (*Agrilus planipennis*). *PLoS One* 2010, **5**(10):e13708.

39. WHO: *Test Procedures for Insecticide Resistance Monitoring in Malaria Vector Mosquitoes*. Geneva, Switzerland: World Health Organization; 2013.

40. Gao Q, Beebe NW, Cooper RD: Molecular identification of the malaria vectors *Anopheles anthropophagus* and *Anopheles sinensis* (Diptera: Culicidae) in central China using polymerase chain reaction and appraisal of their position within the Hyrcanus group. *J Med Entomol* 2004, **41**(1):5–11.

41. Bonizzoni M, Afrane Y, Dunn WA, Atieli FK, Zhou G, Zhong D, Li J, Githeko A, Yan G: Comparative transcriptome analyses of deltamethrin-resistant and -susceptible Anopheles gambiae mosquitoes from Kenya by *RNA-Seq*. *PLoS One* 2012, **7**(9):e44607.

42. Bräutigam A, Mullick T, Schliesky S, Weber APM: Critical assessment of assembly strategies for non-model species *mRNA-Seq* data and application of next-generation sequencing to the comparison of C3 and C4 species. *J Exp Bot* 2011, **62**(9):3093–3102.

43. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M: Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 2005, **21**(18):3674–3676.

44. Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, Wang J, Li S, Li R, Bolund L, Wang J: WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res* 2006, **34**(suppl 2):W293–W297.

45. Audic S, Claverie JM: The significance of digital gene expression profiles. *Genome Res* 1997, **7**(10):986–995.

46. Li H, Chen S, Song A, Wang H, Fang W, Guan Z, Jiang J, Chen F: *RNA-Seq* derived identification of differential transcription in the chrysanthemum leaf following inoculation with Alternaria tenuissima. *BMC Genomics* 2014, **15**(1):9.

47. Young ND, Jex AR, Li B, Liu S, Yang L, Xiong Z, Li Y, Cantacessi C, Hall RS, Xu X, Chen F, Wu X, Zerlotini A, Oliveira G, Hofmann A, Zhang G, Fang X, Kang Y, Campbell BE, Loukas A, Ranganathan S, Rollinson D, Rinaldi G, Brindley PJ, Yang H, Wang J, Wang J, Gasser RB: Whole-genome sequence of *Schistosoma haematobium*. *Nat Genet* 2012, **44**(2):221–225.

48. Benjamini Y, Hochberg Y: Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol* 1995, **57**(1):289–300.

49. Faircloth BC: msatcommander: detection of microsatellite repeat arrays and automated, locus-specific primer design. *Mol Ecol Resour* 2008, **8**(1):92–94.

50. Rice P, Longden I, Bleasby A: EMBOSS: the European molecular biology open software suite. *Trends Genet* 2000, **16**(6):276–277.

51. Pellino M, Hojsgaard D, Schmutzer T, Scholz U, Hörandl E, Vogel H, Sharbel TF: Asexual genome evolution in the apomictic *Ranunculus auricomus* complex: examining the effects of hybridization and mutation accumulation. *Mol Ecol* 2013, **22**(23):5908–5921.

52. Schmittgen TD, Livak KJ: Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc* 2008, **3**(6):1101–1108.

53. Copley SD: Evolution of a metabolic pathway for degradation of a toxic xenobiotic: the patchwork approach. *Trends Biochem Sci* 2000, **25**(6):261–265.

54. Copping LG: Metabolic pathways of agrochemicals: part two –insecticides and fungicides. *Pest Manag Sci* 2004, **56**(1):103–104.

55. David JP, Strode C, Vontas J, Nikou D, Vaughan A, Pignatelli PM, Louis C, Hemingway J, Ranson H: The *Anopheles gambiae* detoxification chip: a highly specific microarray to study metabolic-based insecticide resistance in malaria vectors. *Proc Natl Acad Sci USA* 2005, **102**(11):4080–4084.

56. Vontas J, Blass C, Koutsos AC, David JP, Kafatos FC, Louis C, Hemingway J, Christophides GK, Ranson H: Gene expression in insecticide resistant and susceptible *Anopheles gambiae* strains constitutively or after insecticide exposure. *Insect Mol Biol* 2005, **14**(5):509–521.

57. Chow C-Y: Malaria vector in China. *Chinese J Entomology Special Publ* 1991, **6**:67–79.

58. Ree H-I: Studies on *Anopheles sinensis*, the vector species of vivax malaria in Korea. *Korean J Parasitol* 2005, **43**(3):75–92.

59. Sinka M, Bangs M, Manguin S, Chareonviriyaphap T, Patil A, Temperley W, Gething P, Elyazar I, Kabaria C, Harbach R, Hay S: The dominant Anopheles vectors of human malaria in the Asia-Pacific region: occurrence data, distribution maps and bionomic precis. *Parasit Vectors* 2011, **4**(1):89.

60. Pan J-Y, Zhou S-S, Zheng X, Huang F, Wang D-Q, Shen Y-Z, Su Y-P, Zhou G-C, Liu F, Jiang J-J: Vector capacity of *Anopheles sinensis* in malaria outbreak areas of central China. *Parasit Vectors* 2012, **5**(1):136.

61. Zhou D, Zhang D, Ding G, Shi L, Hou Q, Ye Y, Xu Y, Zhou H, Xiong C, Li S, Yu J, Hong S, Yu X, Zou P, Chen C, Chang X, Wang W, Lv Y, Sun Y, Ma L, Shen B, Zhu C: Genome sequence of *Anopheles sinensis* provides insight into genetics basis of mosquito competence for malaria parasites. *BMC Genomics* 2014, **15**(1):42.

62. Price DP, Nagarajan V, Churbanov A, Houde P, Milligan B, Drake LL, Gustafson JE, Hansen IA: The Fat body transcriptomes of the yellow fever mosquito, *Aedes aegypti*, Pre- and post- blood meal. *PLoS One* 2011, **6**(7):e22573.

63. Poelchau M, Reynolds J, Denlinger D, Elsik C, Armbruster P: A *de novo* transcriptome of the Asian tiger mosquito, *Aedes albopictus*, to identify candidate transcripts for diapause preparation. *BMC Genomics* 2011, **12**(1):1–19.

64. Pridgeon JW, Liu N: Overexpression of the cytochrome c oxidase subunit I gene associated with a pyrethroid resistant strain of German cockroaches, *Blattella germanica* (L.). *Insect Biochem Mol Biol* 2003, **33**(10):1043–1048.

65. McLaughlin LA, Niazi U, Bibby J, David JP, Vontas J, Hemingway J, Ranson H, Sutcliffe MJ, Paine MJ: Characterization of inhibitors and substrates of *Anopheles gambiae* CYP6Z2. *Insect Mol Biol* 2008, **17**(2):125–135.

66. Muller P, Donnelly M, Ranson H: Transcription profiling of a recently colonised pyrethroid resistant *Anopheles gambiae* strain from Ghana. *BMC Genomics* 2007, **8**(1):36.

67. Jones C, Haji K, Khatib B, Bagi J, Mcha J, Devine G, Daley M, Kabula B, Ali A, Majambere S, Ranson H: **The dynamics of pyrethroid resistance in** *Anopheles arabiensis* **from Zanzibar and an assessment of the underlying genetic basis.** *Parasit Vectors* 2013, **6**(1):343.

68. Nardini L, Christian R, Coetzer N, Koekemoer L: **DDT and pyrethroid resistance in Anopheles arabiensis from South Africa.** *Parasit Vectors* 2013, **6**(1):229.

69. Hardstone MC, Komagata O, Kasai S, Tomita T, Scott JG: **Use of isogenic strains indicates CYP9M10 is linked to permethrin resistance in** *Culex pipiens quinquefasciatus.* *Insect Mol Biol* 2010, **19**(6):717–726.

70. Liu N, Li T, Reid WR, Yang T, Zhang L: **Multiple cytochrome P450 genes: their constitutive overexpression and permethrin induction in insecticide resistant mosquitoes. Culex quinquefasciatus.** *PLoS One* 2011, **6**(8):e23403.

71. Zhu F, Li T, Zhang L, Liu N: **Co-up-regulation of three P450 genes in response to permethrin exposure in permethrin resistant house flies. Musca domestica.** *BMC Physiol* 2008, **8**(1):18.

72. Zhu F, Liu N: **Differential expression of CYP6A5 and CYP6A5v2 in pyrethroid-resistant house flies. Musca domestica.** *Arch Insect Biochem Physiol* 2008, **67**(3):107–119.

73. Sun Y, Zou P, Yu XY, Chen C, Yu J, Shi LN, Hong SC, Zhou D, Chang XL, Wang WJ, Shen B, Zhang DH, Ma L, Zhu CL: **Functional characterization of an arrestin gene on insecticide resistance of** *Culex pipiens pallens.* *Parasit Vectors* 2012, **5**:134.

74. Nardini L, Christian R, Coetzer N, Ranson H, Coetzee M, Koekemoer L: **Detoxification enzymes associated with insecticide resistance in laboratory strains of Anopheles arabiensis of different geographic origin.** *Parasit Vectors* 2012, **5**(1):113.

75. Yang T, Liu N: **Genome analysis of cytochrome p450s and their expression profiles in insecticide resistant mosquitoes. Culex quinquefasciatus.** *PLoS One* 2011, **6**(12):e29418.

76. Carvalho RA, Azeredo-Espin AM, Torres TT: **Deep sequencing of** *New World screw-worm* **transcripts to discover genes involved in insecticide resistance.** *BMC Genomics* 2010, **11**:695.

77. Cui PH, Lee AC, Zhou F, Murray M: **Impaired transactivation of the human CYP2J2 arachidonic acid epoxygenase gene in HepG2 cells subjected to nitrative stress.** *Br J Pharmacol* 2010, **159**(7):1440–1449.

78. Marinotti O, Nguyen QK, Calvo E, James AA, Ribeiro JM: **Microarray analysis of genes showing variable expression following a blood meal in** *Anopheles gambiae.* *Insect Mol Biol* 2005, **14**(4):365–373.

79. Bonizzoni M, Britton M, Marinotti O, Dunn W, Fass J, James A: **Probing functional polymorphisms in the dengue vector. Aedes aegypti.** *BMC Genomics* 2013, **14**(1):739.

80. Morlais I, Severson DW: **Intraspecific DNA variation in nuclear genes of the mosquito** *Aedes aegypti.* *Insect Mol Biol* 2003, **12**(6):631–639.

81. Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, Wincker P, Clark AG, Ribeiro JC, Wides R, Salzberg SL, Loftus B, Yandell M, Majoros WH, Rusch DB, Lai Z, Kraft CL, Abril JF, Anthouard V, Arensburger P, Atkinson PW, Baden H, de Berardinis V, Baldwin D, Benes V, Biedler J, Blass C, Bolanos R, Boscus D, Barnstead M: **The genome sequence of the malaria mosquito** *Anopheles gambiae.* *Science* 2002, **298**(5591):129–149.

82. Wondji C, Hemingway J, Ranson H: **Identification and analysis of Single Nucleotide Polymorphisms (SNPs) in the mosquito** *Anopheles funestus,* **malaria vector.** *BMC Genomics* 2007, **8**(1):5.

83. Ning J, Wang M, Li C, Sun S: **Transcriptome sequencing and De Novo analysis of the copepod** *Calanus sinicus* **Using 454 GS FLX.** *PLoS One* 2013, **8**(5):e63741.

84. Roberts SB, Hauser L, Seeb LW, Seeb JE: **Development of genomic resources for pacific herring through targeted transcriptome pyrosequencing.** *PLoS One* 2012, **7**(2):e30908.

85. Bustamante CD, Fledel-Alon A, Williamson S, Nielsen R, Todd Hubisz M, Glanowski S, Tanenbaum DM, White TJ, Sninsky JJ, Hernandez RD, Civello D, Adams MD, Cargill M: **Natural selection on protein-coding genes in the human genome.** *Nature* 2005, **437**(7062):1153–1157.