

RESEARCH ARTICLE

Open Access

# Copy number variations among silkworms

Qian Zhao, Min-Jin Han, Wei Sun and Ze Zhang\*

## Abstract

**Background:** Copy number variations (CNVs), which are important source for genetic and phenotypic variation, have been shown to be associated with disease as well as important QTLs, especially in domesticated animals. However, little is known about the CNVs in silkworm.

**Results:** In this study, we have constructed the first CNVs map based on genome-wide analysis of CNVs in domesticated silkworm. Using next-generation sequencing as well as quantitative PCR (qPCR), we identified ~319 CNVs in total and almost half of them (~49%) were distributed on uncharacterized chromosome. The CNVs covered 10.8 Mb, which is about 2.3% of the entire silkworm genome. Furthermore, approximately 61% of CNVs directly overlapped with SDs in silkworm. The genes in CNVs are mainly related to reproduction, immunity, detoxification and signal recognition, which is consistent with the observations in mammals.

**Conclusions:** An initial CNVs map for silkworm has been described in this study. And this map provides new information for genetic variations in silkworm. Furthermore, the silkworm CNVs may play important roles in reproduction, immunity, detoxification and signal recognition. This study provided insight into the evolution of the silkworm genome and an invaluable resource for insect genomics research.

## Background

Copy number variations (CNVs) are defined as DNA sequences ranging from 1 kb to few Mb that have different numbers of repeats among individuals [1,2]. Comparing with single nucleotide polymorphisms (SNPs), CNVs represent a higher percentage of genetic variation and have greater effects on a genome [3,4]. For example, CNVs play roles in determining phenotypic difference among individuals through changing gene structure and dosage, regulating gene expression and function [5-8]. In addition to normal phenotypic variation, CNVs are also related to genetic disease susceptibility [8,9]. And recently, CNV detection is substantially carried out in domesticated animals and these studies revealed that CNVs are associated with several phenotypic traits. For example, duplication of *KIT* gene in pigs determines the *Dominant white* locus [10]; while in sheep, the coat color is related to the duplication of *ASIP* [11]. In ridge-back dogs, hair ridge and predisposition to dermoid sinus are caused by duplication of 4 genes (*FGF3*, *FGF4*, *FGF19* and *ORAOVI*) [12]; and in Shar-Pei dogs, the wrinkled skin phenotype and a periodic fever syndrome are caused

by upstream duplication of *HAS2* [13]. Also, partial deletion of *EDI* gene in bovine caused anhidrotic ecodermal dysplasia [14]. In avian species, CNV in intron 1 of the *SOX5* gene led to the pea-comb phenotype in chicken [15]. Thus, detection of CNVs at a whole-genome level can give a lot of useful information and has been carried out in several domesticated animals, including pigs, sheep, cattle, dogs, horses and chickens [16-28] as well as crops [29]. However, there is no information on CNVs in silkworm.

The domesticated silkworm (*Bombyx mori*), a model of Lepidoptera insects, has great economic value because of its silk production as well as its value as a good bioreactor [30]. It is widely accepted that *B. mori* is domesticated from the wild silkworm, *Bombyx mandarina*, about 5000 years ago [31]. And nowadays, more than 1,000 *Bombyx mori* inbred and mutant strains are kept all over the world [32]. In 2008, an estimated 432 Mb silkworm genome was published [33], with 8.5-fold sequence coverage and N50 size of ~3.7 Mb. And 87% of the scaffold sequences anchored to all 28 chromosomes, which can provide us a reliable genome to analyze the CNVs in silkworm. A previous study showed that the copy number of carotenoid-binding protein (*CBP*), a major determinant of cocoon color, varied greatly

\* Correspondence: zezhang@cqu.edu.cn  
Laboratory of Evolutionary and Functional Genomics, School of Life Sciences, Chongqing University, Chongqing 400044, China

among *B. mori* strains [24]. Thus, the detection of CNVs at a whole-genome level is necessary for understanding phenotypic variations between different silkworms.

As far as we know, comparative genomic hybridization (CGH) and SNP arrays are routinely used for CNV identification [34-37]. However, the power of CNV detection is easily influenced by low probe density. In addition, although a subset of CNVs showed evidence of linkage disequilibrium with flanking SNPs [38], a significant number of CNVs located in the regions are not well recovered by SNP arrays [39,40].

With the development of next-generation sequencing (NGS) and complementary analysis program, there are some better approaches to screen CNVs systematically at a whole-genome level. Generally, NGS employed the read depth (RD) methods to analyze data and previous studies indicated that data with the genome coverage greater than 4 fold are sufficient for RD detection of CNVs [25,41-43]. To date, several methods have exploited sequence data in 1000 Genomes Project Pilot studies to detect CNVs [44,45]. And several programs are developed to analyze CNVs. These programs included CNAnorm (<http://www.precancer.leeds.ac.uk/>), Bayesian information criterion [46], ReadDepth [47], CNV-seq [48], mrsFAST [49] and so on [50]. Specifically, an R package named readDepth can detect CNVs based on sequence depth and then invoke a circular binary segmentation algorithm to call segment boundaries [47]. This program has high sensitivity and specificity and is appropriate for screening CNVs in duplication and repeat-rich regions [47]. In this study, we resequenced 4 silkworms (2 domesticated silkworms and 2 wild silkworms). Then, we first used readDepth to screen the silkworm CNVs at a genome level and second used CNAnorm to recheck the CNVs, which can result in the high-confidence CNVs. Finally we tried to explore the distribution pattern and potential functions of the CNVs.

## Results and discussion

### Resequencing and CNV identification

We resequenced 4 silkworms: 2 domesticated and 2 wild silkworms. The sequencing coverage of these silkworms is greater than 5, indicating that the data are sufficient for CNV identification (Table 1, Additional file 1). The read-Depth was employed to predict CNVs among four silkworms. The initial results of CNVs identified by readDepth were listed in Table 2 and the location information for each

of initial CNVs is shown in Additional file 2. For further analysis, we retained only CNVs obtained by a more stringent criterion (RD differed significantly from the average of genome RD; see Methods). In order to prevent the false positive, we use this conservative filtering way, however, there should be some false negative regions that were abandoned from our analysis, especially regions with lower copy numbers in the genome. The filtration results are also listed in Table 2 (the detail information in Additional file 3). We identified ~348 suggestive CNVs, size ranging from 9.8 kbp to 34.5 kbp. The 348 CNVs covered 11.5 Mb. Then, we used another method CNAnorm to identify the CNV regions in silkworm. The potential CNVs identified by CNAnorm are listed in Additional file 4. Comparison of the results showed that 319 (10.8 Mb) of 348 CNVs by the readDepth were also identified by the CNAnorm (Additional file 4), which is about 2.3% of the silkworm genome. In the following analysis, we focused on these high-confidence CNVs (Additional file 5).

Among four silkworms, the domesticated silkworm N4 contained the largest number of CNVs while wild silkworm NanC contained the fewest. As expected, the “uncharacterized chromosome” (ChrUn), sequences that cannot be mapped to the genome, contains most CNVs (~49%), which is consistent with the observation in cattle [22]. However, the CNVs on ChrUn need to be further investigated since ChrUn contigs are shorter and mapping of ChrUn sequence reads is ambiguous. In our study, CNV detection would be leveraged on the reference genome, thus, copy numbers are reported more like relative copies comparing to the reference genome. A well assembled reference as well as the well-annotated duplications in genome would be important to the CNV detection using this method. Therefore, the correct assemble of the contigs on ChrUn as well as annotations of repeats in the genome may help to improve the identification of CNVs. In order to get the accurate information about the CNVs and excluded false positives, clone-ordered-based approaches for sequence assembly and further annotation of repeats are needed in further study. The remaining CNVs are distributed on the silkworm chromosomes 1–27 and there is no CNV on the chromosome 28.

The positions of CNVs were determined independently within each silkworm and we compared them among different silkworms. Generally, we classified the duplicated

**Table 1** Resequencing data of four silkworms

Sample	Type	Raw bases	Valid bases	Average depth	Read STDEV
N4	Domesticated	7788356400	6222459199	13.31	4.9
NanC	Wild	9567649400	7818574253	7.76	3
XiaF	Domesticated	9176146800	7302579044	14.42	6
AK	Wild	8745576000	7057560414	12.83	5.4

**Table 2 The CNV calls in four silkworms**

Strain	Before filtering			After filtering by RD and CNAnorm		
	Total (bp)	Numbers	Average size (kb)	Total	Numbers	Average size (kb)
N4	96546962	1082	89	4752536	150	32
XiaF	84976008	711	119	3337398	115	29
AK	190353801	640	297	3427899	89	39
NanC	84063418	433	194	2688123	60	45

The initial CNVs calls were kept if the RD values that differ significantly from the average RD and then filtered by CNAnorm.

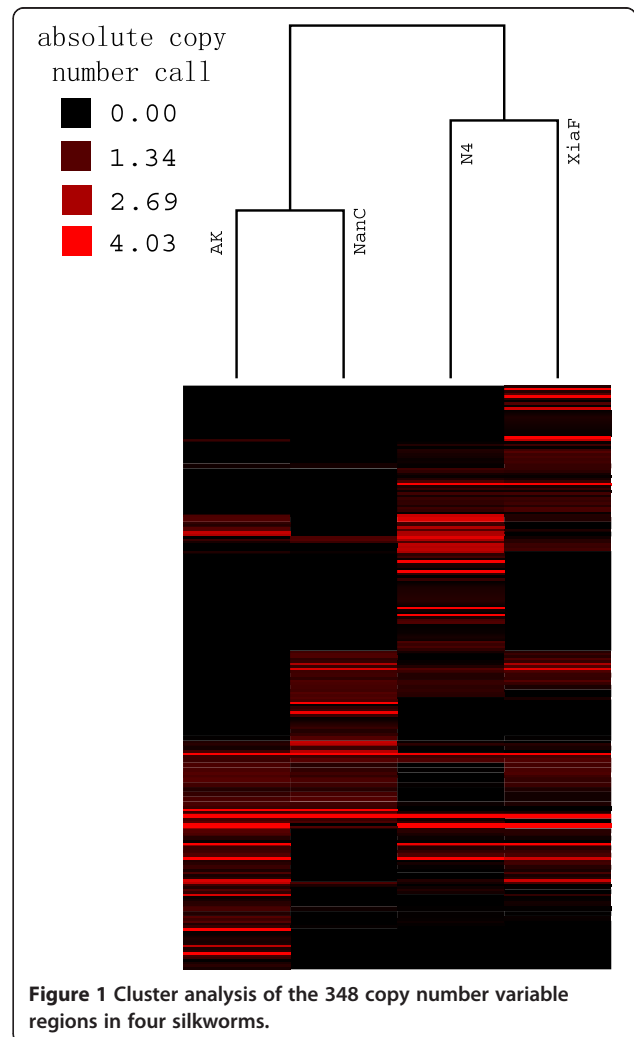
sequences as shared or specific to an individual based on the predicted absolute copy numbers. The results showed that most of the CNVs were shared among two or more silkworms (Additional file 6). Specifically, the domesticated silkworm N4 had the largest number of unique CNVs while wild silkworm NanC contained the smallest number of unique CNVs (Table 2; Additional file 6). In general, a genome is assumed to be more tolerant to duplications than to deletions [51-53], accordingly, CNV gain should be more than loss. However, we found that silkworm had more CNV losses than gains, which is consistent with other species [16,17,19,23]. This result may be due to biological as well as technical reasons. One of the most important mechanisms which may be responsible for CNV formation, named as non-allelic homologous recombination, was proven to generate more deletions than duplications [54]. On the other hand, the detection method may favor the identification of deletions as reported in several other studies [20,44,55]. However, to validate the real status of CNVs, other techniques such as quantitative PCR (qPCR) is necessary.

As previous study showed, the heatmap can also reflect evolutionary relationships among diverse species [25]. Thus, we constructed a heatmap for 4 silkworms using absolute copy numbers in the CNV regions obtained by readDepth (Figure 1). As expected, 2 domesticated silkworms clustered together as other two wild silkworms did. A previous study suggested that a cluster tree constructed by the heatmap of individual-specific CNVs is usually consistent with the individual history [56]. Thus, genomic loci with great agriculture values or QTLs can be identified if there is a larger silkworm sample size and outgroup.

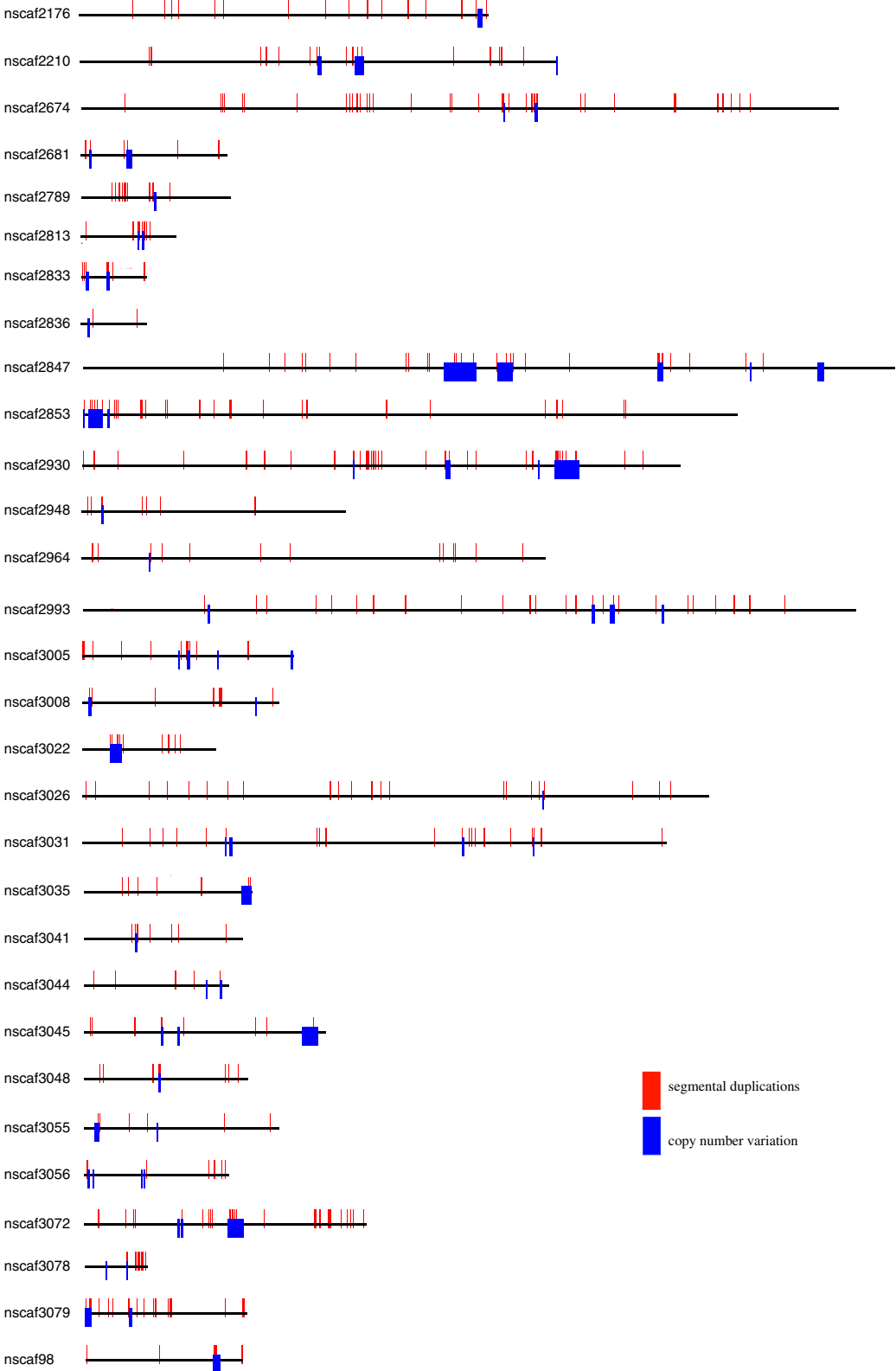
**Overlapping of CNVs with segmental duplications (SDs)**

Previous studies showed that CNVs were enriched in SDs [1,2,57-61]. To test this, we compared the CNVs to the SDs identified by WSSD and WGAC approaches in our previous study [62]. Before filtering the initial CNVs using RD, there were about 94% of SDs exhibiting initial CNVs. And after filtration, approximately 60% of suggestive CNVs directly overlapped with SDs (Figure 2; Additional file 7).

Generally, it is accepted that SDs provide substrates of gene and genome innovation as well as genome rearrangement. SDs are also hotspots of formation of CNVs. Thus, SDs may arise from ancient CNVs fixed in the population [57,63-65]. As observed in other animals (dog, cattle, mouse, rat), there is a consistency (~50%-60%) between large CNVs and SDs (Figure 2) [16,22,60]. Thus, the association of large CNVs with SDs supports the hypothesis that CNV formation is mainly due to nonallelic homologous



**Figure 1 Cluster analysis of the 348 copy number variable regions in four silkworms.**



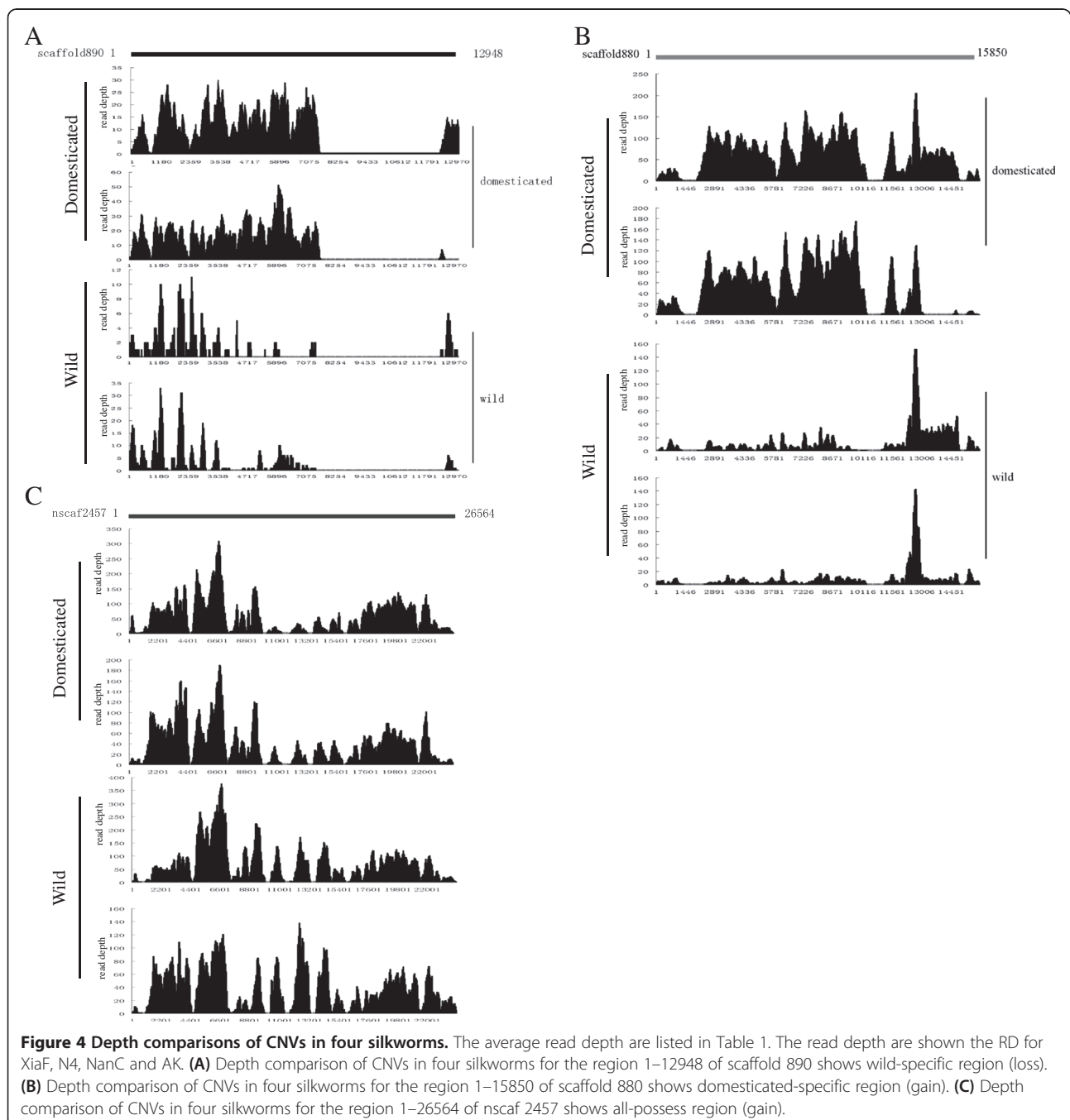
**Figure 2 Silkworm CNVs map.** Only 30 scaffolds were shown and all scaffolds with CNVs information were listed in Additional file 4. The silkworm assembly scaffold is represented as black bars. Larger bars in colors which intersect the scaffold represent the segmental duplications and copy number variation.



should be evolved to adapt to it [67]. For example, cytochrome P450 enzymes are involved in such biological processes in the silkworm [67]. In this study, we identified 10 genes belonged to P450 gene family. We also identified Carboxylesterase (COE), which involved in xenobiotic detoxification as well as pheromone degradation [68], in the CNVs regions. Other genes family related with important functions in lineage-specific evolution included Lipoprotein\_11, heat shock proteins are also identified in our study (Additional file 9).

### Comparative analysis of silkworm CNVs

In order to obtain information related to phenotypic characteristics as much as possible, we classified CNVs as individual-specific, domesticated-specific, wild-specific and all-possessed. Generally, most of the CNVs were shared among two or more silkworms (Additional file 6). However, we identified 80 individual-specific CNVs. Domesticated-specific CNVs are more than wild-specific ones (44 CNVs in domesticated vs. 36 CNVs in wild-specific). Furthermore, the read depth validated this result (Figure 4). Take scaffold



890 as example (Figure 4A), the RD for NanC is less than 4 comparing with the average depth of 7.76. And AK' RD is less than 7 comparing with the average RD of 12.83.

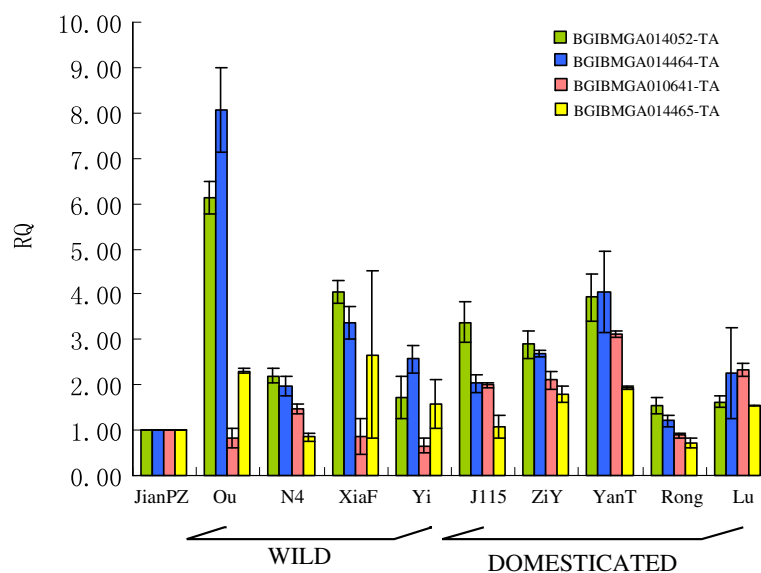
We investigated the genes in the regions of domesticated-specific, wild-specific and all-possessed CNVs. The domesticated-specific CNVs contained 24 functional genes, while wild-specific CNVs contained only 17 genes. We also surveyed the functions and expression patterns of these genes. Most of the genes in these CNV regions are related to detoxification, reproduction and immunity since they were expressed in midgut, testis, ovary and hemocyte, respectively. In domesticated-specific CNV regions, there is an extra gene cluster which was expressed in silkgland (Additional file 10). However, most members of this gene cluster were poorly annotated in the silkworm database, indicating that the functional information on the genes in CNVs has been very limited to date. This deserves further investigation in future.

#### CNV validation by quantitative PCR

We used real time quantitative PCR (qPCR) to validate CNVs in 5 genomic regions as well as 10 genes. Four of five loci (genomic sequences) were validated by this method (Additional file 11). For the exception, the silkworm genome has two copies of Target\_r1 (scaffold984:1...11044) based on the BLASTN searches against *B. mori*. And the qPCR results showed little variation among 4 silkworms (2 domesticated and 2 wild) at this locus. This might be: (1) prediction errors of CNVs, that is, the false positive; (2) polymorphisms such as indels and SNPs that influence binding of the qPCR primers. For four validated regions, we found that there was a big difference

in copy number at the locus of Target\_r3 between domesticated and wild silkworms. That is, domesticated silkworm contained more copies than wild type at this locus based on the qPCR results. Also, this region belongs to domesticated-specific region. Furthermore, we found that only one gene (BGIBMGA014594-TA) is located in this CNV region. However, this gene was poorly annotated so far. A previous study showed that this gene was specifically and highly expressed in testis, indicated that this gene may play important roles in reproduction [69]. Further study is needed to characterize its function.

Besides, we also chose 10 genes to validate the presence of CNVs in different silkworms (Additional file 11). A total of 10 silkworms (4 wild silkworms and 6 domesticated silkworms) were examined: eight of ten genes can be validated by qPCR, except for two genes (BGIBMGA014051, BGIBMGA014594). *F*-test was performed to check whether copy number detected using qPCR showed homogeneity of variance between the reference silkworm and silkworms to be examined. The result suggested that all these 8 loci in silkworms to be examined had greater variance than those in the reference silkworm ( $P < 0.05$ ) (Figure 5, Additional file 11), confirming that the CNVs identified in this study are reliable. For these 8 genes, one (BGIBMGA012385-TA) belonged to P450 gene family, one (BGIBMGA002901-TA) belonged to COesterase and one (BGIBMGA009791-TA) belonged to carotenoid-binding protein. A previous study of microarray expression profiling showed that two (BGIBMGA014464-TA and BGIBMGA014465-TA) of 8 genes were highly expressed in head, integument and hemocyte [69]. Another gene, BGIBMGA014052-TA, was specially and highly expressed in



**Figure 5 qPCR confirmation.** Different bars represent different genes. X-axis shows the different individuals while Y-axis is the value of  $2^{-\Delta\Delta CT}$  that is indicator of duplications (RQ). The domesticated silkworms include JianPZ, Ou, N4, XiaF, Yi, J115, wild silkworms include ZiY, YanT, Rong, Lu.

Malpighian tubule, implying its important role in detoxification in silkworm. BGIBMGA010640-TA, which is involved in lipid metabolic process (GO: 0006629), was highly expressed in midgut. Midgut of silkworm is very important because of its key functions in digesting, resistance and immune response. Genes expressed highly in midgut suggest its important roles in nutrient digestion and absorption, resistance and immune response in silkworm. A previous study used four pathogens to challenge silkworm and investigated the genome-wide gene expression profiles by a microarray [70]. We exploited this dataset to check the expression pattern of BGIBMGA010640-TA as well as expression patterns of another 7 genes that were proven to be resistant to nucleopolyhedrovirus (BmNPV) [71]. Like the above 7 genes, BGIBMGA010640-TA could be induced by 3 pathogens (Additional file 12) [70]. This suggested that BGIBMGA010640-TA may be involved in immune response of silkworm.

The CNVs (86.7%, 12/15) were confirmed to be positive CNVs by qRT-PCR (Figure 5, Additional file 8). It should be emphasized that not all true CNVs could be detected by qPCR, especially some low-copy duplications with less sequence similarities. Thus, 13.3% false positive rate is a conserved estimate in our CNV analysis.

## Conclusion

We have constructed the first CNVs map in silkworm based on next-generation re-sequencing data. A total of ~319 CNVs were identified in the silkworm genome. We presented the frequency, pattern and gene-content of these CNVs. Our results indicated that the genes in CNVs may be involved in specific biological functions such as reproduction, immunity, detoxification and signal recognition. Besides, we identified 80 CNVs that may be individual-specific. Most of genes in these 80 regions were also related to reproduction or detoxification. The data presented in this study provided insight into the evolution of the silkworm genome and an invaluable resource for insect genomics research.

## Methods

### Data sets

#### Genome sequencing and read cleaning

Silkworm genome was obtained from previous studies [33,72]. We prepared libraries for four silkworms (two wild silkworms named as AK and NanC and two domesticated silkworms named as XiaF and N4). We sequenced them using Illumina (HiSeq2000) according to standard manufacturer protocols. The low-quality (Quality < 20) nucleotides were trimmed by sliding a 5 bp window.

#### Read alignment and CNV detection

We used the BWA program to align the paired-end reads to the silkworm genome reference [73], the criteria are the

same as to previous study [47]. For the detection of CNVs among four silkworms, we have applied a program called readDepth [47] using a parameter 0.01 of an FDR rate, which resulted in bins with a size of 1.7 kbp. And readDepth calculates the thresholds for copy number gain and loss for each silkworm (Additional file 13). The readDepth uses a binning procedure to call copy number variants based on sequence depth and then call segment boundaries using a circular binary segmentation algorithm. Our previous results suggested that there are ~1.4% of SDs in the reference genome [62], which can help us to adjust the data in the program. The GC bias was corrected using LOESS method to fit a regression line to the data [41,47].

In order to find the high-confident CNVs, we calculated the read depth (RD) of the regions predicted by the readDepth. And we calculate the average read depth for the unique regions of silkworm identified before [62]. We only kept the regions with RD greater than 3 standard deviations from the mean [25]. Then, these regions whose RD differed significantly from the average of genome RD (Chi square test;  $p < 0.05$ ) were termed as potential CNVs.

Because different algorithms can generate different CNV results [42], we used CNAnorm (<http://www.biocomputor.org/packages/release/bioc/html/CNAnorm.html>) to recheck our CNV regions to reduce the false-positive or false-negative rate. We employed parameters of `-readNum 150, --saveTest, --saveControl` in PERL script of `bam2windows.pl` (a script in the CNAnorm package). The parameter `lambda 7` was used to decrease noise without losing resolution and ploidy (ploidy =  $\text{sugg.ploidy}(\text{CNN4}) + 1$ ) was used to check the potential CNVs in the genome.

#### Heatmap hierarchical cluster analysis

Heatmaps were obtained based on the absolute copy number call generated by readDepth. The `gplots` R package (<http://cran.r-project.org/web/packages/gplots/index.html>) was employed to get the heatmap of the absolute copy number call in four silkworms.

#### Gene content analysis

Gene content of *B. mori* segmental duplications was assessed using the glean consensus gene set (<http://silkworm.genomics.org.cn/>) [74]. We obtained a total of 14,623 silkworm peptides from SilkDB. In addition, using Gene Ontology (GO) [75], we tested the hypothesis that the molecular function, biological process, and pathway terms were under- or overrepresented in CNV regions. Furthermore, we compared the GO results between the genes from SDs and the genes from CNV regions. Pfam [76] was also used to annotate the function of the genes in CNV regions.



### Quantification of CNVs in the silkworm genome by quantitative PCR

Genomics DNAs were extracted from domesticated and wild silkworms, and stored in Tris-EDTA (TE) buffer at 4°C. The primers used in qPCR are designed using Primer 5.0 and listed in Additional file 14. The principle for copy number quantifying using qPCR was described in previous study [77]. According to previous studies, *OR2* was chosen as control because of its highly-conserved sequence and single copy in the silkworm genome [24,78,79]. *Con\_R* is a two-copy region in the silkworm genome according to *B. mori* genome database [71,72,80,81]. We also used this region as control to estimate copy numbers of target regions.

Each PCR reaction was prepared as follows: 10 µl of SYBR-Green PCR master mix, 1 µl of each primer (10 µM), 7 µl of water, and 1 µl of genome template. Quantitative real-time PCR was carried out using the ABI Stepone plus system. The thermocycler program had an initial 95°C denaturation step followed by 40 cycles consisting of a 10-s denaturation at 95°C, a 40-s annealing at 60°C, and a 30-s extension step at 72°C. At the end of each reaction, a dissociation curve was created, which was used to help to detect the presence of primer dimers of other unwanted amplification products that may produce a detectable cycle threshold (Ct) value. Copy number was analyzed according to comparative Ct method. The  $\Delta CT$  and  $\Delta\Delta CT$  were calculated by the formulas  $\Delta CT = CT_{\text{target}} - CT_{\text{control}}$  (single copy) and  $\Delta\Delta CT = \Delta CT_{\text{SD samples}} - \Delta CT_{\text{single copy sample}}$ , respectively. The domesticated silkworm JianPZ was taken as a standard for determining gene copy number.

### Availability of supporting data

Raw sequence reads have been deposited in the ENA database (The European Bioinformatics Institute) with the accession number PRJEB5458 and can also be downloaded from [http://bioinform.cqu.edu.cn/read\\_silkworm/](http://bioinform.cqu.edu.cn/read_silkworm/).

### Additional files

- Additional file 1:** Basic information for RD and reads.
- Additional file 2:** The initial results of CNVs identified by readDepth.
- Additional file 3:** The suggestive ~348 CNVs.
- Additional file 4:** The CNVs identified using CNAnorm.
- Additional file 5:** The CNVs identified both by readDepth and CNAnorm.
- Additional file 6:** Venn diagram showed the comparison of CNV content amongst different silkworms.
- Additional file 7:** Silkworm CNVs map. The silkworm assembly scaffold is represented as black bars. Larger bars in colors which intersect the scaffold represent the segmental duplications and copy number variation.
- Additional file 8:** Expression profiles of the genes located in CNVs based on microarray data. Hierarchical clustering with the average

linkage method was performed. There were as many as 9 tissues used in the gene expression profiling.

**Additional file 9: Functional annotation of genes located in CNVs.** Sheet1 shows the function predictions by BLAST search against nr database. Sheet2 shows the function prediction obtained by Pfam.

**Additional file 10: Comparison of gene expression pattern located in domesticated-specific CNV regions and wild-specific CNVs based on microarray data.** Hierarchical clustering with the average linkage method was performed. There were as many as 9 tissues used in the gene expression profiling. The upper diagram showed the expression profiles of genes in wild-specific CNVs.

**Additional file 11: qPCR validation of predicted CNVs in silkworms.**

**Additional file 12: Expression profiles of 8 genes in silkworm challenged by four pathogens: *Bacillus bombysepticus* (BB, gram-positive bacteria); *Beauveria bassiana* (BJ, fungus); *Escherichia coli* (EC, gram-negative bacteria); *B. mori Nuclear polyhedrosis virus* (NPV, virus).** Data were collected from four time points (3 h, 6 h, 12 h and 24 h; for *Be. bassiana*: 6 h, 12 h, 24 h and 48 h) (Huang, 2010).

**Additional file 13: Thresholds for copy number gain and loss.**

**Additional file 14: A list of primers used in qPCR.**

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

ZZ designed the study. QZ performed the analyses and experiments, and drafted the manuscript. MJH provided help in the data analysis and revised the manuscript. WS provided help in doing experiments and read the manuscript. ZZ supervised the study and revised the manuscript. All authors read and approved the final manuscript.

### Acknowledgements

This work was supported by the Hi-Tech Research and Development (863) Program of China (2013AA102507) and by a grant from National Natural Science Foundation of China (No. 31272363).

Received: 31 July 2013 Accepted: 25 March 2014

Published: 31 March 2014

### References

1. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C: **Detection of large-scale variation in the human genome.** *Nat Genet* 2004, **36**(9):949-951.
2. Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M: **Large-scale copy number polymorphism in the human genome.** *Science* 2004, **305**(5683):525-528.
3. Feuk L, Carson AR, Scherer SW: **Structural variation in the human genome.** *Nat Rev Genet* 2006, **7**(2):85-97.
4. Freeman JL, Perry GH, Feuk L, Redon R, McCarroll SA, Altshuler DM, Aburatani H, Jones KW, Tyler-Smith C, Hurles ME, Carter NP, Scherer SW, Lee C: **Copy number variation: new insights in genome diversity.** *Genome Res* 2006, **16**(8):949-961.
5. Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, Redon R, Bird CP, de Grassi A, Lee C, Tyler-Smith C, Carter N, Scherer SW, Tavaré S, Deloukas P, Hurles ME, Dermitzakis ET: **Relative impact of nucleotide and copy number variation on gene expression phenotypes.** *Science* 2007, **315**(5813):848-853.
6. Cahan P, Li Y, Izumi M, Graubert TA: **The impact of copy number variation on local gene expression in mouse hematopoietic stem and progenitor cells.** *Nat Genet* 2009, **41**(4):430-437.
7. Orozco LD, Cokus SJ, Ghazalpour A, Ingram-Drake L, Wang S, van Nas A, Che N, Araujo JA, Pellegrini M, Lusis AJ: **Copy number variation influences gene expression and metabolic traits in mice.** *Hum Mol Genet* 2009, **18**(21):4118-4129.
8. Zhang F, Gu W, Hurles ME, Lupski JR: **Copy number variation in human health, disease, and evolution.** *Annu Rev Genomics Hum Genet* 2009, **10**:451-481.

9. Stankiewicz P, Lupski JR: **Structural variation in the human genome and its role in disease.** *Annu Rev Med* 2010, **61**:437–455.
10. Pielberg G, Olsson C, Syvanen AC, Andersson L: **Unexpectedly high allelic diversity at the KIT locus causing dominant white color in the domestic pig.** *Genetics* 2002, **160**(1):305–311.
11. Norris BJ, Whan VA: **A gene duplication affecting expression of the ovine ASP gene is responsible for white and black sheep.** *Genome Res* 2008, **18**(8):1282–1293.
12. Salmon Hillbertz NH, Isaksson M, Karlsson EK, Hellmen E, Pielberg GR, Savolainen P, Wade CM, von Euler H, Gustafson U, Hedhammar A, Nilsson M, Lindblad-Toh K, Andersson L, Andersson G: **Duplication of FGF3, FGF4, FGF19 and ORO1 causes hair ridge and predisposition to dermoid sinus in Ridgeback dogs.** *Nat Genet* 2007, **39**(11):1318–1320.
13. Olsson M, Meadows JR, Truve K, Rosengren Pielberg G, Puppo F, Mauceli E, Quilez J, Tonomura N, Zanna G, Docampo MJ, Bassols A, Avery AC, Karlsson EK, Thomas A, Kastner DL, Bongcam-Rudloff E, Webster MT, Sanchez A, Hedhammar A, Remmers EF, Andersson L, Ferrer L, Tintle L, Lindblad-Toh K: **A novel unstable duplication upstream of HAS2 predisposes to a breed-defining skin phenotype and a periodic fever syndrome in Chinese Shar-Pei dogs.** *PLoS Genet* 2011, **7**(3):e1001332.
14. Drogemuller C, Distl O, Leeb T: **Partial deletion of the bovine ED1 gene causes anhidrotic ectodermal dysplasia in cattle.** *Genome Res* 2001, **11**(10):1699–1705.
15. Wright D, Boije H, Meadows JR, Bed'hom B, Gourichon D, Vieaud A, Tixier-Boichard M, Rubin CJ, Imsland F, Hallbook F, Andersson L: **Copy number variation in intron 1 of SOX5 causes the Pea-comb phenotype in chickens.** *PLoS Genet* 2009, **5**(6):e1000512.
16. Nicholas TJ, Cheng Z, Ventura M, Mealey K, Eichler EE, Akey JM: **The genomic architecture of segmental duplications and associated copy number variants in dogs.** *Genome Res* 2009, **19**(3):491–499.
17. Fontanesi L, Beretti F, Martelli PL, Colombo M, Dall'olio S, Occidente M, Portolano B, Casadio R, Matassino D, Russo V: **A first comparative map of copy number variations in the sheep genome.** *Genomics* 2011, **97**(3):158–165.
18. Chen WK, Swartz JD, Rush LJ, Alvarez CE: **Mapping DNA structural variation in dogs.** *Genome Res* 2009, **19**(3):500–509.
19. Bae JS, Cheong HS, Kim LH, NamGung S, Park TJ, Chun JY, Kim JY, Pasaje CF, Lee JS, Shin HD: **Identification of copy number variations and common deletion polymorphisms in cattle.** *BMC Genomics* 2010, **11**:232.
20. Fadista J, Thomsen B, Holm LE, Bendixen C: **Copy number variation in the bovine genome.** *BMC Genomics* 2010, **11**:284.
21. Ramayo-Caldas Y, Castello A, Pena RN, Alves E, Mercade A, Souza CA, Fernandez AI, Perez-Enciso M, Folch JM: **Copy number variation in the porcine genome inferred from a 60 k SNP BeadChip.** *BMC Genomics* 2010, **11**:593.
22. Liu GE, Hou Y, Zhu B, Cardone MF, Jiang L, Cellamare A, Mitra A, Alexander LJ, Coutinho LL, Dell'Aquila ME, Gasbarre LC, Lacalandra G, Li RW, Matukumalli LK, Nonneman D, Regitano LC, Smith TP, Song J, Sonstegard TS, Van Tassell CP, Ventura M, Eichler EE, McDanel TG, Keele JW: **Analysis of copy number variations among diverse cattle breeds.** *Genome Res* 2010, **20**(5):693–703.
23. Kijas JW, Barendse W, Barris W, Harrison B, McCulloch R, McWilliam S, Whan V: **Analysis of copy number variants in the cattle genome.** *Genet* 2010, **482**(1–2):73–77.
24. Sakudoh T, Nakashima T, Kuroki Y, Fujiyama A, Kohara Y, Honda N, Fujimoto H, Shimada T, Nakagaki M, Banno Y, Tsuchida K: **Diversity in copy number and structure of a silkworm morphogenetic gene as a result of domestication.** *Genetics* 2011, **187**(3):965–976.
25. Bickhart DM, Hou Y, Schroeder SG, Alkan C, Cardone MF, Matukumalli LK, Song J, Schnabel RD, Ventura M, Taylor JF, Garcia JF, Van Tassell CP, Sonstegard TS, Eichler EE, Liu GE: **Copy number variation of individual cattle genomes using next-generation sequencing.** *Genome Res* 2012, **22**(4):778–790.
26. Metzger J, Philipp U, Lopes MS, da Camara Machado A, Felicetti M, Silvestrelli M, Distl O: **Analysis of copy number variants by three detection algorithms and their association with body size in horses.** *BMC Genomics* 2013, **14**:487.
27. Doan R, Cohen N, Harrington J, Veazy K, Juras R, Cothran G, McCue ME, Skow L, Dindot SV: **Identification of copy number variants in horses.** *Genome Res* 2012, **22**:899–907.
28. Dupuis MC, Zhang Z, Durkin K, Charlier C, Lekeux P, Georges M: **Detection of copy number variants in the horse genome and examination of their association with recurrent laryngeal neuropathy.** *Anim Genet* 2012, **44**:206–208.
29. Munoz-Amatriain M, Eichten SR, Wicker T, Richmond TA, Mascher M, Steuernagel B, Scholz U, Ariyadasa R, Spannagl M, Nussbaumer T, Mayer KF, Taudien S, Platzer M, Jeddeloh JA, Springer NM, Muehlbauer GJ, Stein N: **Distribution, functional impact, and origin mechanisms of copy number variation in the barley genome.** *Genome Biol* 2013, **14**(6):R58.
30. Chen J, Wu XF, Zhang YZ: **Expression, purification and characterization of human GM-CSF using silkworm pupae (*Bombyx mori*) as a bioreactor.** *J Biotechnol* 2006, **123**(2):236–247.
31. Xiang ZH, et al: *Biology of Sericulture*. Beijing: China Forestry Publishing House; 2005.
32. Banno Y, Shimada T, Kajjura Z, Sezutsu H: **The silkworm-an attractive BioResource supplied by Japan.** *Exp Anim* 2010, **59**(2):139–146.
33. The international silkworm genome consortium: **The genome of a lepidopteran model insect, the silkworm *Bombyx mori*.** *Insect Biochem Mol Biol* 2008, **38**(12):1036–1045.
34. Lai WR, Johnson MD, Kucherlapati R, Park PJ: **Comparative analysis of algorithms for identifying amplifications and deletions in array CGH data.** *Bioinformatics* 2005, **21**(19):3763–3770.
35. LaFramboise T: **Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances.** *Nucleic Acids Res* 2009, **37**(13):4181–4193.
36. Winchester L, Yau C, Ragoussis J: **Comparing CNV detection methods for SNP arrays.** *Brief Funct Genomic Proteomic* 2009, **8**(5):353–366.
37. Pinto D, Darvishi K, Shi X, Rajan D, Rigler D, Fitzgerald T, Lionel AC, Thiruvahindrapuram B, Macdonald JR, Mills R, Prasad A, Noonan K, Gribble S, Prigmore E, Donahoe PK, Smith RS, Park JH, Hurler ME, Carter NP, Lee C, Scherer SW, Feuk L: **Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants.** *Nat Biotechnol* 2011, **29**(6):512–520.
38. McCarroll SA, Kuruvilla FG, Korn JM, Cawley S, Nemes J, Wysoker A, Shapero MH, de Bakker PI, Maller JB, Kirby A, Elliott AL, Parkin M, Hubbell E, Webster T, Mei R, Veitch J, Collins PJ, Handsaker R, Lincoln S, Nizzari M, Blume J, Jones KW, Rava R, Daly MJ, Gabriel SB, Altshuler D: **Integrated detection and population-genetic analysis of SNPs and copy number variation.** *Nat Genet* 2008, **40**(10):1166–1174.
39. Estivill X, Armengol L: **Copy number variants and common disorders: filling the gaps and exploring complexity in genome-wide association studies.** *PLoS Genet* 2007, **3**(10):1787–1799.
40. Campbell CD, Sampas N, Tsalenko A, Sudmant PH, Kidd JM, Malig M, Vu TH, Vives L, Tsang P, Bruhn L, Eichler EE: **Population-genetic properties of differentiated human copy-number polymorphisms.** *Am J Hum Genet* 2011, **88**(3):317–332.
41. Alkan C, Kidd JM, Marques-Bonet T, Aksay G, Antonacci F, Hormozdiani F, Kitzman JO, Baker C, Malig M, Mutlu O, Sahinalp SC, Gibbs RA, Eichler EE: **Personalized copy number and segmental duplication maps using next-generation sequencing.** *Nat Genet* 2009, **41**(10):1061–1067.
42. Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, Abyzov A, Yoon SC, Ye K, Cheetham RK, Chinwalla A, Conrad DF, Fu Y, Grubert F, Hajirasouliha I, Hormozdiani F, Iakoucheva LM, Iqbal Z, Kang S, Kidd JM, Konkel MK, Korn J, Khurana E, Kural D, Lam HY, Leng J, Li R, Li Y, Lin CY, Luo R, et al: **Mapping copy number variation by population-scale genome sequencing.** *Nature* 2011, **470**(7332):59–65.
43. Waszak SM, Hasin Y, Zichner T, Olender T, Keydar I, Khen M, Stutz AM, Schlattl A, Lancet D, Korbel JO: **Systematic inference of copy-number genotypes from personal genome sequencing data reveals extensive olfactory receptor gene content diversity.** *PLoS Comput Biol* 2010, **6**(11):e1000988.
44. Sudmant PH, Kitzman JO, Antonacci F, Alkan C, Malig M, Tsalenko A, Sampas N, Bruhn L, Shendure J, Eichler EE: **Diversity of human copy number variation and multicopy genes.** *Science* 2010, **330**(6004):641–646.
45. Umemori J, Mori A, Ichiyangi K, Uno T, Koide T: **Identification of both copy number variation-type and constant-type core elements in a large segmental duplication region of the mouse genome.** *BMC Genomics* 2013, **14**(1):455.
46. Xi R, Hadipanayis AG, Luquette LJ, Kim TM, Lee E, Zhang J, Johnson MD, Muzny DM, Wheeler DA, Gibbs RA, Kucherlapati R, Park PJ: **Copy number variation detection in whole-genome sequencing data using the Bayesian information criterion.** *Proc Natl Acad Sci U S A* 2011, **108**:E1128–E1136.
47. Miller CA, Hampton O, Coarfa C, Milosavljevic A: **ReadDepth: a parallel R package for detecting copy number alterations from short sequencing reads.** *PLoS One* 2011, **6**(1):e16327.

48. Xie C, Tammi MT: **CNV-seq, a new method to detect copy number variation using high-throughput sequencing.** *BMC Bioinforma* 2009, **10**:80.
49. Hach F, Hormozdiari F, Alkan C, Birol I, Eichler EE, Sahinalp SC: **mrsFAST: a cache-oblivious algorithm for short-read mapping.** *Nat Methods* 2010, **7**:576–577.
50. Zhao M, Wang QG, Wang Q, Jia P, Zhao Z: **Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives.** *BMC Bioinforma* 2013, **14**(suppl 11):S1.
51. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shaperro MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, González JR, Gratacòs M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F: **Global variation in copy number in the human genome.** *Nature* 2006, **444**(7118):444–454.
52. Locke DP, Sharp AJ, McCarroll SA, McGrath SD, Newman TL, Cheng Z, Schwartz S, Albertson DG, Pinkel D, Altschuler DM, Eichler EE: **Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome.** *Am J Hum Genet* 2006, **79**(2):275–290.
53. Brewer C, Holloway S, Zawadzki P, Schinzel A, FitzPatrick D: **A chromosomal duplication map of malformations: regions of suspected haplo- and triplolethality—and tolerance of segmental aneuploidy—in humans.** *Am J Hum Genet* 1999, **64**(6):1702–1708.
54. Turner DJ, Miretti M, Rajan D, Fiegler H, Carter NP, Blayney ML, Beck S, Hurles ME: **Germline rates of de novo meiotic deletions and duplications causing several genomic disorders.** *Nat Genet* 2008, **40**(1):90–95.
55. Fadista J, Nygaard M, Holm LE, Thomsen B, Bendixen C: **A snapshot of CNVs in the pig genome.** *PLoS One* 2008, **3**(12):e3916.
56. Decker JE, Pires JC, Conant GC, McKay SD, Heaton MP, Chen K, Cooper A, Vilkki J, Seabury CM, Caetano AR, Johnson GS, Brenneman RA, Hanotte O, Eggert LS, Wiener P, Kim JJ, Kim KS, Sonstegard TS, Van Tassell CP, Neiberger HL, McEwan JC, Brauning R, Coutinho LL, Babar ME, Wilson GA, McClure MC, Rolf MM, Kim J, Schnabel RD, Taylor JF: **Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics.** *Proc Natl Acad Sci U S A* 2009, **106**(44):18644–18649.
57. Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, Pertz LM, Clark RA, Schwartz S, Segraves R, Oseroff VV, Albertson DG, Pinkel D, Eichler EE: **Segmental duplications and copy-number variation in the human genome.** *Am J Hum Genet* 2005, **77**(1):78–88.
58. She X, Cheng Z, Zollner S, Church DM, Eichler EE: **Mouse segmental duplication and copy number variation.** *Nat Genet* 2008, **40**(7):909–914.
59. Perry GH, Tchinda J, McGrath SD, Zhang J, Picker SR, Caceres AM, lafrate AJ, Tyler-Smith C, Scherer SW, Eichler EE, Stone AC, Lee C: **Hotspots for copy number variation in chimpanzees and humans.** *Proc Natl Acad Sci U S A* 2006, **103**(21):8006–8011.
60. Graubert TA, Cahan P, Edwin D, Selzer RR, Richmond TA, Eis PS, Shannon WD, Li X, McLeod HL, Cheverud JM, Ley TJ: **A high-resolution map of segmental DNA copy number variation in the mouse genome.** *PLoS Genet* 2007, **3**(1):e3.
61. Guryev V, Saar K, Adamovic T, Verheul M, van Heesch SA, Cook S, Pravenec M, Aitman T, Jacob H, Shull JD, Hubner N, Cuppen E: **Distribution and functional impact of DNA copy number variation in the rat.** *Nat Genet* 2008, **40**(5):538–545.
62. Zhao Q, Zhu Z, Kasahara M, Morishita S, Zhang Z: **Segmental duplications in the silkworm genome.** *BMC Genomics* 2013, **14**:521.
63. Emanuel BS, Shaikh TH: **Segmental duplications: an ‘expanding’ role in genomic instability and disease.** *Nat Rev Genet* 2001, **2**(10):791–800.
64. Goidts V, Cooper DN, Armengol L, Schempp W, Conroy J, Estivill X, Nowak N, Hammeister H, Kehrer-Sawatzki H: **Complex patterns of copy number variation at sites of segmental duplications: an important category of structural variation in the human genome.** *Hum Genet* 2006, **120**(2):270–284.
65. Marques-Bonet T, Girirajan S, Eichler EE: **The origins and impact of primate segmental duplications.** *Trends Genet* 2009, **25**(10):443–454.
66. Bulet P, Hetru C, Dimarcq JL, Hoffmann D: **Antimicrobial peptides in insects; structure and function.** *Dev Comp Immunol* 1999, **23**(4–5):329–344.
67. Ai J, Zhu Y, Duan J, Yu Q, Zhang G, Wan F, Xiang ZH: **Genome-wide analysis of cytochrome P450 monooxygenase genes in the silkworm, *Bombyx mori*.** *Gene* 2011, **480**(1–2):42–50.
68. Yu Q, Lu C, Li WL, Xiang ZH, Zhang Z: **Annotation and expression of Carboxylesterases in the silkworm, *Bombyx mori*.** *BMC Genomics* 2009, **10**:553.
69. Xia Q, Cheng D, Duan J, Wang G, Cheng T, Zha X, Liu C, Zhao P, Dai F, Zhang Z, He N, Zhang L, Xiang Z: **Microarray-based gene expression profiles in multiple tissues of the domesticated silkworm, *Bombyx mori*.** *Genome Biol* 2007, **8**(8):R162.
70. Huang L: **A genome-wide analysis of the silkworm host responses to *Bacillus bombyseptius* (Bb) and other pathogens.** Ph.D Thesis, Southwest University, China. 2010.
71. Bao YY, Tang XD, Lv ZY, Wang XY, Tian CH, Xu YP, Zhang CX: **Gene expression profiling of resistant and susceptible *Bombyx mori* strains reveals nucleopolydnavirus-associated variations in host gene transcript levels.** *Genomics* 2009, **94**(2):138–145.
72. Mita K: **Genome of a lepidopteran model insect, the silkworm *Bombyx mori*.** *Seikagaku* 2009, **81**(5):353–360.
73. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform.** *Bioinformatics* 2009, **25**(14):1754–1760.
74. Xia Q, Zhou Z, Lu C, Cheng D, Dai F, Li B, Zhao P, Zha X, Cheng T, Chai C, Pan G, Xu J, Liu C, Lin Y, Qian J, Hou Y, Wu Z, Li G, Pan M, Li C, Shen Y, Lan X, Yuan L, Li T, Xu H, Yang G, Wan Y, Zhu Y, Yu M, Shen W: **A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*).** *Science* 2004, **306**(5703):1937–1940.
75. Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, Wang J, Li S, Li R, Bolund L: **WEGO: a web tool for plotting GO annotations.** *Nucleic Acids Res* 2006, **34**(Web Server issue):W293–W297.
76. Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, Holm L, Sonnhammer EL, Eddy SR, Bateman A: **The Pfam protein families database.** *Nucleic Acids Res* 2010, **38**(Database issue):D211–D222.
77. Andersson DI, Hughes D: **Gene amplification and adaptive evolution in bacteria.** *Annu Rev Genet* 2009, **43**:167–195.
78. Krieger J, Klink O, Mohl C, Raming K, Breer H: **A candidate olfactory receptor subtype highly conserved across different insect orders.** *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 2003, **189**(7):519–526.
79. Nakagawa T, Sakurai T, Nishioka T, Touhara K: **Insect sex-pheromone signals mediated by specific combinations of olfactory receptors.** *Science* 2005, **307**(5715):1638–1642.
80. Shimomura M, Minami H, Suetsugu Y, Ohyanagi H, Satoh C, Antonio B, Nagamura Y, Kadono-Okuda K, Kajiwara H, Sezutsu H, Nagaraju J, Goldsmith MR, Xia Q, Yamamoto K, Mita K: **KAIKObase: an integrated silkworm genome database and data mining tool.** *BMC Genomics* 2009, **10**:486.
81. Duan J, Li R, Cheng D, Fan W, Zha X, Cheng T, Wu Y, Wang J, Mita K, Xiang Z, Xia Q: **SilkDB v2.0: a platform for silkworm (*Bombyx mori*) genome biology.** *Nucleic Acids Res* 2010, **38**(Database issue):D453–D456.

doi:10.1186/1471-2164-15-251

Cite this article as: Zhao et al.: Copy number variations among silkworms. *BMC Genomics* 2014 **15**:251.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

