Insect Biochemistry and Molecular Biology 42 (2012) 935–945

Contents lists available at SciVerse ScienceDirect

Insect Biochemistry and Molecular Biology

journal homepage: www.elsevier.com/locate/ibmb

# Expansion of the silkworm GMC oxidoreductase genes is associated with immunity

Wei Sun [a], Yi-Hong Shen [a], Wen-Juan Yang [a], Yun-Feng Cao [a], Zhong-Huai Xiang [a], Ze Zhang [a,b,*]

[a] The Institute of Sericulture and Systems Biology, Southwest University, Chongqing 400715, China
[b] College of Life Sciences, Chongqing University, Chongqing 400044, China

ARTICLE INFO

*Article history:*
Received 12 August 2012
Received in revised form
13 September 2012
Accepted 17 September 2012

*Keywords:*
GMC oxidoreductases
*Bombyx mori*
Evolution
Expansion
Innate immunity

ABSTRACT

The glucose-methanol-choline (GMC) oxidoreductases constitute a large gene family in insects. Some of these enzymes play roles in developmental or physiological process, such as ecdysteroid metabolism. However, little is known about the functional diversity of the insect GMC family. Here, we identified 43 GMC genes in the silkworm genome, the largest number of GMC genes among all the insect genomes sequenced to date. Similar to the other insects, there is a highly conserved GMC cluster within the second intron of the silkworm *flotillin-2* (*flo-2*) gene. However, the silkworm GMC genes outside of the conserved GMC cluster have experienced a large expansion. Phylogenetic analysis suggested that the silkworm GMCβ subfamily contained 22 copies and made a major contribution to expansion of the silkworm GMC genes. Eighteen of the 22 members of the silkworm GMCβ subfamily are located outside of the conserved GMC cluster, and are known as silkworm expansion genes (SEs). Relative-rate tests showed that SEs evolved significantly faster than the GMCβ genes inside the conserved GMC cluster. Accordingly, the third position GC content (GC3s) and codon bias of SEs are significantly different from those of the GMCβ genes in the conserved GMC cluster. The elevated evolutionary rate of the silkworm GMCβ genes outside of the conserved GMC cluster may reflect the evolution of function diversity. At least 24 of the 43 silkworm GMC genes were differently transcribed and expressed in a tissue- or stage-specific manner during the larval stage. Strikingly, microarray data revealed that four different pathogens upregulated most of the silkworm GMCβ genes. Furthermore, RNA interference of representative upregulated GMCβ genes reduced the survival rate of the silkworm when infected by pathogens. Taken together, the results suggested that expansion of the silkworm GMC oxidoreductase genes is associated with immunity.



## 1. Introduction

Oxidation–reduction may be the most important and basic reaction in all organisms and is accomplished by enzymes that can be classified into different families based on sequence similarity. One of them is the glucose-methanol-choline (GMC) oxidoreductase family (Cavener, 1992). Members of the GMC family share the same domain, an FAD ADP-binding domain located in the N-terminal section (Cavener, 1992). Moreover, they also have five highly conserved blocks with unknown function (Blocks WWW Server: IPB000172 http://blocks.fhcrc.org/). Although enzymes belonging to the GMC family can catalyze at least 11 different reactions (Zámocký et al., 2004), they may be active on a conserved CH–OH group of donors (Cavener, 1992).

As more genome sequences of organisms have become available, more and more GMC genes have been identified, especially in insect genomes. Recently, a study surveyed the evolution of GMC oxidoreductases in several species and revealed that the GMC family experienced a large expansion among insects (*Drosophila melanogaster*, *Anopheles gambiae*, *Apis mellifera*, and *Tribolium castaneum*) compared with other species (Iida et al., 2007). In addition, there is a highly conserved cluster including 10–12 GMC oxidoreductase genes that is located within the second intron of the *flotillin-2* gene in insect genomes. Ecdysone oxidase (EO) is involved in ecdysone metabolism in *Drosophila* and is located in the conserved GMC cluster; therefore, it was speculated that the conserved GMC cluster might have roles in development (Iida et al., 2007). Moreover, some of insect GMC genes have functional information, such as the genes encoding glucose oxidase (GOX), glucose dehydrogenase (GLD), *DmGMCλ*

* Corresponding author. The Institute of Sericulture and Systems Biology, Southwest University, Chongqing 400715, China. Tel.: +86 23 68251029; fax: +86 23 68251128.
*E-mail address:* zezhang@swu.edu.cn (Z. Zhang).

0965-1748/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ibmb.2012.09.006

(CG12398), salicyl alcohol oxidase (SAO) and DmNinaG. Among them, some genes take part in basic physiological processes. For example, both GOX and GLD are active in glucose metabolism (Cavener, 1992; Ohashi et al., 1999), *DmGMCλ* (CG12398) may take part in formation of the eggshell (Fakhouri et al., 2006), and *DmNinaG* may be involved in the biogenesis of visual pigment chromophore in fruit flies (Ahmad et al., 2006; Sarfare et al., 2005). While GMC genes may also be active in some special processes, SAO in the chrysomelid leaf beetle can convert salicyl alcohol into salicylaldehyde, which is used by the larva against its predators (Michalski et al., 2008). Moreover, GOX may suppress the host plant defense responses in some insects (Bede et al., 2006; Diezel et al., 2009; Musser et al., 2002), and both GOX and GLD are important enzymes for insect immunity (Cox-Foster and Stehr, 1994; Yang and Cox-Foster, 2005). Thus, insect GMC genes may have different roles in diverse metabolic processes.

Lepidoptera is a large order of insects that includes moths and butterflies. However, the GMC genes in these insects have not been systematically investigated. The available genome sequences and transcription information of one moth (the silkworm, *Bombyx mori*) and one butterfly (monarch butterfly, *Danaus plexippus*) provide an unprecedented opportunity for investigating the GMC genes in Lepidoptera at the genome-wide scale (The International Silkworm Genome Consortium, 2008; Zhan et al., 2011). Here, a whole genome screen was performed to search for the silkworm and butterfly GMC genes. Forty-three and 33 GMC genes were identified in the silkworm and monarch butterfly, respectively. Comparative and phylogenetic analyses were performed for GMC genes from diverse animals. The number of silkworm GMC genes is the largest among the insect genomes investigated. Expansion of the GMCβ genes is the major cause of the silkworm GMC family expansion. Eighteen of the 22 silkworm GMCβ genes are located outside of the conserved GMC cluster. Most of GMCβ genes are silkworm-specific genes and are clustered together in the phylogenetic tree; therefore, these GMCβ genes were designated as the silkworm expansion genes (SEs). The SEs were shown to have evolved faster than the GMC genes inside of the conserved GMC cluster. Finally, most members of the silkworm GMCβ genes were demonstrated to be upregulated by different pathogens. Moreover, the silkworms whose representative upregulated GMCβ genes were knocked down showed marked susceptibilities to pathogens. Taken together, these results indicate that the expansion of the silkworm GMCβ subfamily may be involved in resistance to various pathogens.

## 2. Materials and methods

### 2.1. Identification of GMC genes in the silkworm genome

The silkworm genome database and predicted protein database were downloaded from SilkDB (http://silkworm.swu.edu.cn/silkdb/) (Duan et al., 2010), and the protein database of the monarch butterfly was downloaded from MonarchBase (http://monarchbase.umassmed.edu/). The protein databases of human *(Homo sapiens*), mouse (*Mus musculus*), chicken (*Gallus gallus*), zebra fish (*Danio rerio*), nematode (*Caenorhabditis elegans*), honey bee (*Apis mellifera*), beetle (*T. castaneum*), and fruit fly (*D. melanogaster*) were all downloaded from National Center for Biotechnology Information (NCBI) FTP site (http://www.ncbi.nlm.nih.gov/Ftp/). The hidden Markov model (HMM) of PF00732 for GMC oxidoreductase domain was downloaded from the Pfam database (http://pfam.sanger.ac.uk/) (Bateman et al., 2004), and was used to search the silkworm and the monarch butterfly predicted protein database using HMMER (Eddy, 1998; Finn et al., 2006). The threshold was set as a score greater than 0.0 and an E-value less than 0.1. The currently assembled genome sequence of the silkworm and the monarch butterfly genome sequence do not cover the entire respective genomes; therefore, the predicted protein databases may be incomplete. Thus, all the results were used as queries to perform TBLASTN searches against the silkworm and butterfly genome sequences, respectively. To eliminate false-positive proteins, we predicted the domain of the candidate sequences using the Pfam online server (http://pfam.sanger.ac.uk/) (Bateman et al., 2004). The same procedure was employed to identify the GMC genes in *D. melanogaster*, *Apis mellifera* and *T. castaneum* to test the efficacy of the method, which was used in a previous study (Iida et al., 2007), and to identify all the GMC genes in other species.

### 2.2. Phylogenetic tree reconstruction

Multiple alignments of protein sequences were made by MUSCLE 3.6 (Edgar, 2004). The alignments were then corrected by eye. Bayesian inference methods (BI) were used to perform the phylogenetic analyses. Three fungal GOX genes (*Aspergillus niger*: Anig GOX; *Penicillium amagasakiense*: Pama GOX; *Aspergillus oryzae*: Aory GOX) were used as the outgroup. Certain insect GMC genes with functional information (e.g., SAO) were also included in this analysis.

Bayesian inferences were performed using MrBayes, V3.1.2 (Ronquist and Huelsenbeck, 2003). The WAG protein model was chosen to perform the Bayesian phylogenetic analysis. Four Markov chains on the data were run for 1,000,000 generations with one cold and three heated chains each, sampling once every 100 trees. To determine the burn-in, the AWTY online program (Nylander et al., 2008) was used to plot the cumulative posterior split probabilities from the Markov chain Monte Carlo (MCMC) runs. The phylogenetic tree was displayed and modified by iTOL (Letunic and Bork, 2007).

### 2.3. Data analysis

A relative-rate test was performed using RRTree software (Robinson-Rechavi and Huchon, 2000). The phylogenetic tree used for the relative-rate test was reconstructed using MrBayes, V3.1.2 (Ronquist and Huelsenbeck, 2003). One *Da. plexippus* GMCβ gene, which is the most divergent from the silkworm GMC genes, was used as the outgroup. Significance of the relative-rate difference was tested using a Bonferroni correction. In addition, the GC3s and codon bias (effective number of codons; ENc) were estimated for the GMCβ genes from inside and outside of the conserved GMC cluster, using the CodonW software (John Peden, Oxford University, available at http://bioweb.pasteur.fr/seqanal/interfaces/codonw.html). The theoretical p*I* (isoelectric point) was also computed for the GMCβ genes on the ExPASy server (http://web.expasy.org/compute_pi/). For comparison with the GMCβ genes, several proteins from the silkworm protein database were chosen as controls. The average length of the GMCβ amino acid sequence is 604.8 ± 57.3. Therefore, the range of lengths of control proteins was set at 547–663. Finally, 1081 proteins were selected from the silkworm predicted protein database. The ENc and p*I* values of the control proteins were estimated. All the statistical analyses were done using R software. In addition, the program MEME-Chip was used to detect the conserved motifs in 2000 bp upstream regions of GMC core orthologs among the five insects (Machanick and Bailey, 2011).

### 2.4. Gene expression analysis

The DaZao strain of silkworm was used to survey the expression profiles of GMC genes. For temporal expression analysis: larvae, pupae or adults were collected at different developmental time

points. For every time point, three individuals were pooled together and then frozen immediately in liquid nitrogen. For spatial expression analysis: eight main tissues were dissected from Day 3 of the fifth instar larvae, and frozen immediately in liquid nitrogen. Every tissue sample was collected from more than three larvae. The samples were homogenized in liquid nitrogen to powders and then added the Trizol reagent (Invitrogen). Total RNA of every sample was extracted according to the manufacturer's instruction. Genomic DNA was digested by RNase-free DNase I (Takara). RNA was quantified by UV spectrum absorbance and reverse-transcribed into first strand cDNA by an M-MLV Reverse Transcriptase Kit (Invitrogen). For expression analysis, the specific amplification primers for GMC genes are shown in Table S1. The PCR products were sequenced to confirm the specificity of the primers.

### 2.5. Microarray analysis

A previous study challenged silkworms using four pathogens and investigated the induced expression profiles of whole silkworm genes using a microarray (Huang, 2010). We retrieved the normalized microarray data from that study. The analysis method was described in a previous study (Huang, 2010). Briefly, four different microorganisms (Gram-positive bacteria: *Bacillus bombyseptieus*; Gram-negative bacteria: *Escherichia coli*; Fungus: *Beauveria bassinan*; and Virus: *B. mori Nuclear polyhedrosis viruses*) were used to infect silkworm larvae (day 3 of the 5th instar). Double distilled water ($ddH_2O$) was used as negative control. Data were collected from four time points (3 h, 6 h, 12 h and 24 h; for *Be. bassinan*: 6 h, 12 h, 24 h and 48 h). Genes were considered to be upregulated if their expression compared with the control was greater than two fold for at least one time point in any of the experiments using the four microorganisms (Huang, 2010). Though it is statistically inefficient to infer differential expression of genes using a fixed threshold cut off (Leung and Cavalieri, 2003), the fold-change (two or higher fold) remains an important feature in microarray analysis (Ambroise et al., 2011). Hierarchical clustering of gene expression patterns was performed using MultiExperiment Viewer (MEV) (Saeed et al., 2006). The nucleotide sequences of the GMC genes were used for BLAST searching against the silkworm probe database (SilkDB, Huang (2010) also used the same probes) to identify a specific probe for each GMC gene.

### 2.6. RNA interference

On the third day of the fifth instar, larvae were used for RNA interference (RNAi) experiments. Specific primers containing the T7 polymerase promoter sequence at their 5′ ends were used to amplify the target genes (Table S1). The PCR products were sequenced to confirm the specificity of the primers. The amplified fragments were then used as templates to generate double-strand RNAs (dsRNAs). The dsRNAs were synthesized *in vitro* by RiboMAX Large Scale RNA Production systems-T7 (Promega) using a manual method. Concentrations of the dsRNAs were quantified by UV spectrum absorbance. Ten microliter solutions containing 30 μg of ds-*BmGMC1*, ds-*BmGMC12* and ds-*BmGMCβ3* were injected into each larva, respectively. The same concentration of ds-*EGFP* or saline was used as controls. In addition, *BmGMCι3*, which was not induced by the pathogens in the microarray analysis, was also knocked down as the negative control. Twelve hours after dsRNA or saline injection, *E. coli* ($10^5$ cells/larvae) or *Ba. bombyseptieus* ($5 \times 10^4$ cells/larvae) or saline were injected into the silkworm. *E. coli* and *Ba. bombyseptieus* were cultured in Luria–Bertani medium at 37 °C. The cells were collected after centrifugation, and the pellets were washed and re-suspended with saline. Finally, the cell numbers were measured and diluted. The survival rate of the silkworm was surveyed after injection. The mRNA levels of the targeted genes were investigated 12 h after microorganism infection, using reverse transcription PCR as described above.

## 3. Results

### 3.1. The GMC genes in silkworm and other species

Searching the *D. melanogaster*, *Apis mellifera* and *T. castaneum* predicted protein databases with the Pfam HMM model PF00732 identified 15, 18 and 23 GMC genes in the respective genomes. The numbers of the GMC genes in these three insects are consistent with previous results (Iida et al., 2007). Therefore, the method used in this study appeared to be reliable. In the silkworm genome, 40 candidate GMC genes were identified (Table 1). Among them, two genes (BGIBMGA012996; BGIBMGA013006) are almost twice as long as the others and both of them have two FAD ADP-binding domains. It was hypothesized that either of these two genes could be two single genes. Indeed, both of them could be divided into two genes when re-analyzed using Softberry (http://linux1.softberry.com/berry.phtml). Therefore, they were designated as BGIBMGA012996-1, BGIBMGA012996-2 and BGIBMGA013006-1, BGIBMGA013006-2, respectively. In addition, ecdysone oxidase (EO), which can covert ecdysone to 3-dehydroecdysone in *Drosophila*, is also the member of the GMC family (Takeuchi et al., 2005). However, the homolog of this gene was not found in the silkworm protein database using the same method. Recently, we identified and cloned the silkworm *EO* gene from an EST library (Sun et al., 2012). Ultimately, 43 GMC genes were identified in the silkworm genome (Table 1). For the monarch butterfly, 33 GMC genes were identified by the same method (Fig. 1). In addition, three fragments that contained part of a GMC oxidoreductase domain were detected in the monarch butterfly genome. Nevertheless, these three fragments are short (less than 200 amino acids), and were not included in subsequent analyses.

Several GMC genes were detected in other species (all listed in Table S2). In summary, the number of GMC genes in lepidopteran insects appeared to have undergone expansion. The silkworm genome has the largest number of GMC genes among the insect genomes sequenced to date.

### 3.2. Phylogeny of GMC genes

Bayesian inference was used to reconstruct the phylogenetic tree of the GMC genes (Fig. 2). All the insect GMC genes clustered together. The phylogenetic tree had three polyphyletic clades (Fig. 2, black arrows). The first clade was comprised nine main subclades, and the topologies of these subclades were similar to those of a previous study (Iida et al., 2007). Thus, the same Greek letters used in the previous study were used to designate these subclades. The phylogenetic tree showed that all the subclades contained the GMC genes from at least four different insects. GMCζ, GMCε, GMCδ, GMCγ and GMCα subclades all have a single copy among the five different insects, except for GMCα, which has two copies in the monarch butterfly. The copy number of GMCλ subfamily varied among the species. The silkworm has five GMCλ genes, and the monarch butterfly has nine copies, six of which are monarch butterfly-specific. In a previous study, the GMCκ subfamily formed a single subclade. However, in this study, this subfamily was clustered with the *Drosophila* GMCι subfamily with a high Bayesian posterior probability value (Bpp = 0.95). Therefore, it was concluded that GMCκ should be a member of GMCι subfamily. In addition, the five silkworm GMCι genes and three monarch butterfly GMCι genes were all grouped together. For

**Table 1**
Summary of the silkworm GMC genes. UN represents the information of the genes cannot be found in SilkDB (http://silkworm.swu.edu.cn/silkdb/). For EST, "+" represents at least one match sequence in SilkDB.

| Gene Name | Gene ID | Scaffold | Chr. | Introns | Protein length | Orientation | EST | Note |
|---|---|---|---|---|---|---|---|---|
| *BmGMC1* | BGIBMGA000068 | nscaf1108 | 24 | 2 | 580 | − | − | SEs |
| *BmGMC2* | BGIBMGA000158 | nscaf1108 | 24 | 2 | 564 | + | + | SEs |
| *BmGMC3* | BGIBMGA005545 | nscaf2829 | 17 | 4 | 579 | − | + | |
| *BmGMC4* | BGIBMGA005608 | nscaf2829 | 17 | 5 | 692 | + | − | SEs |
| *BmGMC5* | BGIBMGA005609 | nscaf2829 | 17 | 0 | 489 | + | + | SEs |
| *BmGMC6* | BGIBMGA005703 | nscaf2830 | UN | 3 | 656 | + | + | SEs |
| *BmGMC7* | BGIBMGA005710 | nscaf2830 | UN | 2 | 585 | + | + | SEs |
| *BmGMC8* | BGIBMGA005711 | nscaf2830 | UN | 2 | 565 | + | + | SEs |
| *BmGMC9* | BGIBMGA009242 | nscaf2943 | 14 | 2 | 612 | − | − | SEs |
| *BmGMC10* | BGIBMGA009924 | nscaf2970 | 8 | 2 | 638 | + | − | SEs |
| *BmGMC11* | BGIBMGA009925 | nscaf2970 | 8 | 2 | 628 | + | − | SEs |
| *BmGMC12* | BGIBMGA010448 | nscaf2993 | 12 | 2 | 608 | − | − | SEs |
| *BmGMC13* | BGIBMGA010461 | nscaf2993 | 12 | 2 | 585 | − | − | SEs |
| *BmGMC14* | BGIBMGA010515 | nscaf2993 | 12 | 0 | 514 | + | − | SEs |
| *BmGMC15* | BGIBMGA010516 | nscaf2993 | 12 | 0 | 465 | + | − | SEs |
| *BmGMC16* | BGIBMGA010517 | nscaf2993 | 12 | 1 | 235 | + | − | SEs |
| *BmGMC17* | BGIBMGA012115 | nscaf3034 | 11 | 1 | 547 | − | − | SEs |
| *BmNinaG* | BGIBMGA012374 | nscaf3041 | 21 | 10 | 574 | − | − | |
| *BmGMC18* | BGIBMGA012586 | nscaf3052 | 19 | 2 | 609 | − | − | |
| *BmGMC19* | BGIBMGA012618 | nscaf3052 | 19 | 3 | 630 | + | + | |
| *BmGMC20* | BGIBMGA012863 | nscaf3058 | 16 | 8 | 538 | − | + | |
| *BmGMC21* | BGIBMGA012872 | nscaf3058 | 16 | 10 | 605 | − | + | |
| *BmGMCα* | BGIBMGA012996-1 | nscaf3058 | 16 | 3 | 631 | + | + | Conserved GMC cluster |
| *BmGMCγ* | BGIBMGA012996-2 | nscaf3058 | 16 | 3 | 613 | + | + | Conserved GMC cluster |
| *BmGMCβ2* | BGIBMGA012997 | nscaf3058 | 16 | 2 | 624 | + | + | Conserved GMC cluster |
| *BmGMCβ3* | BGIBMGA012998 | nscaf3058 | 16 | 2 | 622 | + | + | Conserved GMC cluster |
| *BmGMCβ4* | BGIBMGA012999 | nscaf3058 | 16 | 3 | 636 | + | + | Conserved GMC cluster |
| *BmGMCβ5* | BGIBMGA013000 | nscaf3058 | 16 | 2 | 622 | + | + | Conserved GMC cluster |
| *BmGMCδ1* | BGIBMGA013001 | nscaf3058 | 16 | 3 | 712 | + | + | Conserved GMC cluster |
| *BmGMCε1* | BGIBMGA013002 | nscaf3058 | 16 | 1 | 615 | + | − | Conserved GMC cluster |
| *BmGMCζ1* | BGIBMGA013003 | nscaf3058 | 16 | 3 | 622 | + | − | Conserved GMC cluster |
| *BmGMCθ2* | BGIBMGA013005 | nscaf3058 | 16 | 2 | 603 | + | + | Conserved GMC cluster |
| *BmGMCι1* | BGIBMGA013006-1 | nscaf3058 | 16 | 5 | 602 | + | + | Conserved GMC cluster |
| *BmGMCι2* | BGIBMGA013006-2 | nscaf3058 | 16 | 5 | 620 | + | + | Conserved GMC cluster |
| *BmGMCι3* | BGIBMGA013007 | nscaf3058 | 16 | 4 | 657 | + | + | Conserved GMC cluster |
| *BmGMCι4* | BGIBMGA013008 | nscaf3058 | 16 | 5 | 399 | + | + | Conserved GMC cluster |
| *BmGMCι5* | BGIBMGA013009 | nscaf3058 | 16 | 5 | 657 | + | − | Conserved GMC cluster |
| *BmGLD* | BGIBMGA013215 | nscaf3063 | 16 | 9 | 667 | + | − | |
| *BmGMC22* | BGIBMGA013788 | nscaf3097 | 28 | 3 | 744 | + | − | |
| *BmGMC23* | BGIBMGA013789 | nscaf3097 | 28 | 1 | 467 | + | − | |
| *BmGMC24* | BGIBMGA013951 | nscaf3099 | 28 | 3 | 610 | + | + | |
| *BmGMC25* | BGIBMGA014539 | scaffold782 | UN | 2 | 610 | − | − | SEs |
| *BmEO* | BmEO | nscaf2829 | 17 | 3 | 668 | + | + | SEs |

GMCθ, each insect species has at least two copies, except for the silkworm, which has only one copy.

The second clade was the GLD/GOX clade, which contained the insect glucose dehydrogenase (GLD) and glucose oxidase (GOX) proteins. The GLD proteins from the five insects formed an orthologous subclade, which indicated its conserved function. However, the copy numbers of GOX is different among the insects. The silkworm has two copies (*BmGMC20* and *BmGMC21*), whereas the monarch butterfly has one copy (*DpGMC17*).

All the insect GMCβ genes were clustered into the third clade. This clade contained two subclades: one was the mixed subclade including five insect species, and the other was the Lepidoptera-specific subclade. In the mixed subclade, the GMCβ genes from the same species firstly formed one group, and then clustered with other GMCβ genes, indicating those GMCβ genes that experienced a species-specific expansion. The lepidoptera-specific subclade included 20 GMCβ genes from the silkworm and seven from the monarch butterfly. Among the 20 silkworm GMCβ genes, only four were Lepidoptera-specific orthologs; the other genes were gene specifically expanded in silkworm. The expansion of the GMCβ genes was the main explanation for the silkworm having the largest number of GMC genes among the five insects studied.

### 3.3. Genomic distribution of the silkworm GMC genes

Thirty-nine of the 43 silkworm GMC genes were scattered on 10 chromosomes. The remaining four genes could not be mapped on any chromosomes; however, three of them (*BmGMC6*, *BmGMC7* and *BmGMC8*) are tandem arranged in one scaffold (Table 1; Fig. S1). Among the 39 mapped genes, 36 genes formed seven clusters on chromosomes, each of which contained at least two genes (Fig. S1). In general, tandem duplicated GMC genes were usually grouped together in the phylogenetic tree, except for the *BmGMC3* gene. *BmGMC3* genes, as well as the *BmGMC4*, *BmGMC5* and *BmEO* genes, were located in the same scaffold, nscaf2829, but they did not form a single cluster in the phylogenetic tree (Fig. 2). The genes located in one scaffold often had similar numbers of introns (Table 1; Fig. S1). *BmGMC1*, *BmGMC2*, *BmGMC10* and *BmGMC11* have two introns. In addition, the genes clustered together in the phylogenetic tree also had similar numbers of introns, for example, *BmGMC1*, *BmGMC2*, *BmGMC7* and *BmGMC8* all have two introns (Fig. 1; Fig. S1). *BmGMC20*, *BmGMC21* and *BmGLD*, which belong to the GLD/GOX cluster, have 8–10 introns. However, the genes on the 12th chromosome (*BmGMC12-BmGMC16*) have different numbers of introns, despite being in one group in the phylogenetic tree.

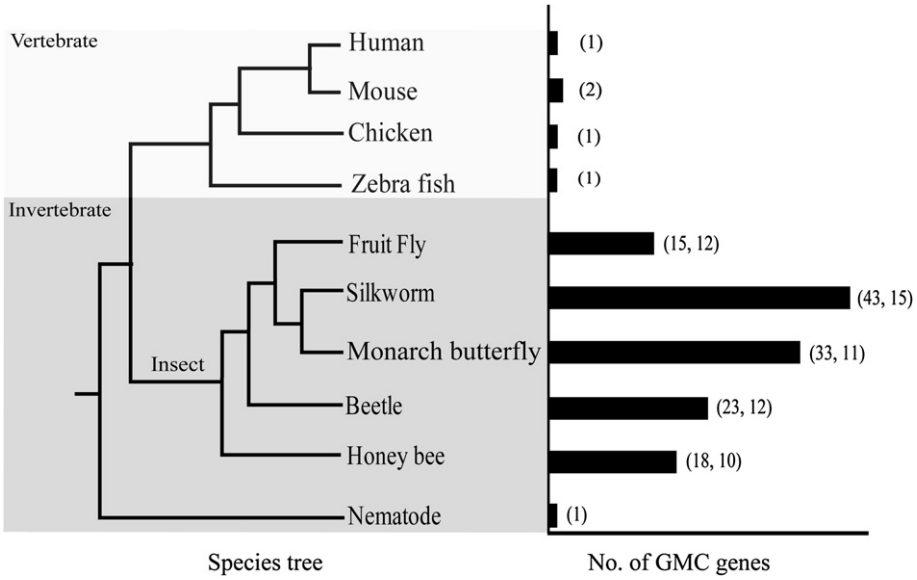


**Fig. 1.** The numbers of GMC genes in the different species used in this study. The left of the figure is the species tree which was modified from Lynch (2007). The right of the figure is the numbers of the GMC genes in different species. The lengths of the black rectangles represent different numbers of the GMC genes. The numbers in the brackets mean the total number of GMC genes and the number of GMC genes inside of the conserved GMC cluster, respectively, see the results.

The largest duplicated group was located on the 16th chromosome (Fig. S1). This group comprised the silkworm GMC genes from two scaffolds (nscaf3058 and nscaf3063); the former containing 17 genes and the later only one gene. Similar to *D. melanogaster* (Iida et al., 2007), there was also a GMC gene cluster within a 379 kb intron of the *flotillin-2* gene, a non-GMC gene encoding a lipid raft-related protein (Fig. S2). The members of the silkworm-conserved GMC cluster had the same transcriptional orientation. These genes were tandemly located in the conserved cluster with an interval comprising three non-GMC genes. The monarch butterfly also had a conserved GMC cluster containing 11 genes that were also interrupted by a non-GMC gene within a 114 kb intron of the *flotillin-2* gene. It should be noted that those non-GMC genes in the silkworm and the monarch butterfly conserved GMC clusters had no sequence similarity. The silkworm-conserved GMC cluster contained 15 genes belonging to eight subfamilies (Fig. S2), and all the subfamilies had single copy, except for *BmGMCι* and *BmGMCβ*, which have five and four copies, respectively. The order of those eight subfamilies within the second intron of *flotillin-2* in the silkworm is the same as in the monarch butterfly and fruit fly (with the exception of GMCγ). Within the conserved GMC cluster, although the copy numbers of GMCι and GMCβ subfamilies were different in the insects investigated, five GMC genes (GMCζ, GMCε, GMCδ, GMCγ and GMCα) were present as single copies. In addition, several conserved motifs were detected in the 2000 bp upstream regions of the five orthologs. Every motif was found in all investigated insects (Fig. S3). This conservation during a long evolutionary history indicates the important regulatory functions of these motifs.

### 3.4. Evolution of the silkworm GMCβ genes

Phylogenetic analysis showed that 22 of the 43 silkworm GMC genes belonged to the GMCβ subfamily. The number of silkworm GMCβ genes was the largest among insects. According to the genomic location of the GMC genes, four GMCβ genes were located in the conserved GMC cluster. As shown above, all five divergent species retained the conserved GMC cluster over a long evolutionary history, indicating the strong evolutionary constraint for the conserved GMC cluster. The remaining 18 GMCβ genes were dispersed on the genome and were located outside of the conserved GMC cluster (these 18 genes were designated as silkworm expansion genes, SEs), and they might have experienced different evolutionary pressures compared with the four clustered GMCβ genes. The program RRtree was used to test for significant differences in evolutionary rate between those silkworm GMCβ genes in the conserved GMC cluster and those outside of the cluster. First, a phylogenetic tree of the silkworm GMCβ genes was reconstructed. The silkworm GMCβ genes in the conserved cluster were firstly grouped together, and then clustered with remaining 18 genes (SEs clade) (Fig. 3). The SEs clade contained three subclades (SE-a, SE-b and SE-c). The results of the relative-rate test revealed that the GMCβ genes outside of the conserved GMC cluster (SEs) had evolved significantly faster than the genes in the conserved GMC cluster (Ka, 0.710 *vs*. 0.619, $P = 0.0013$) (Fig. 3).

The evolutionary rate of proteins may have been influenced by the gene characteristics, for example, the nucleotide composition and codon bias (Lynch and Conery, 2000). The silkworm GMCβ genes in the conserved cluster showed significantly lower GC3s values than SEs (0.397 *vs.* 0.418, *t* test $P = 0.043$) (Fig. 4). In addition, our results also revealed that GMCβ genes in the conserved cluster had significantly higher codon bias compared with SEs and control loci ("GMCβ genes in the cluster" *vs.* SEs: 49.27 *vs*. 54.31, *t* test $P < 0.01$; "GMCβ genes in the cluster" *vs.* control: 49.27 *vs.* 53.54, *t* test $P < 0.01$) (Fig. 4). Therefore, the accelerated evolution of SEs resulted from changes of nucleotide composition and codon bias. The fast evolution may reflect the resultant functional diversification of proteins. The theoretical average p*I* of the SEs was significantly higher than that of the GMCβ genes in the conserved GMC cluster (theoretical p*I*, 7.35 *vs.* 5.88, *t* test $P < 0.01$), but similar to that at control loci (theoretical p*I*, 7.35 *vs.*7.20, *t* test $P = 0.69$).

### 3.5. Spatial and temporal expression patterns of the silkworm GMC genes

The expression profile of a gene family provides a hint of the functions of its members. First, all the silkworm GMC genes were used as queries to search the EST database and only 23 genes had EST evidence (Table 1). RT-PCR was then performed to determine

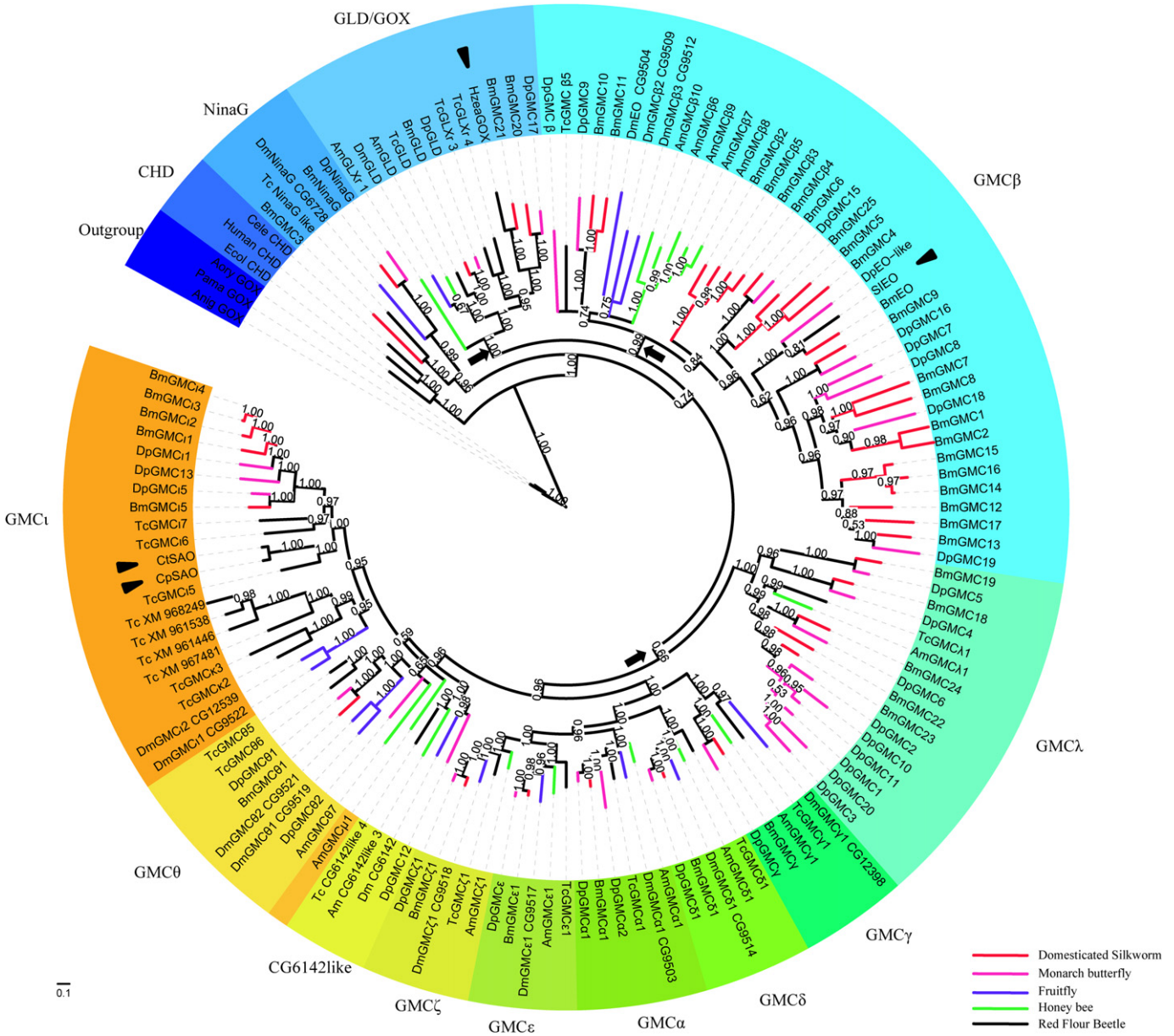


**Fig. 2.** Phylogeny of the GMC genes. The phylogenetic tree was reconstructed using Bayesian inference method in Mrbayes 3.01 software. BPPs less than 50 are not shown. Different colors of branches represent different species. Different colors of leaves represent different GMC subfamilies. Black arrows indicate three main clades. Black arrowheads in leaves mean function known proteins.

the temporal and spatial expression profiles of these silkworm GMC genes.

Based on the results from phylogenetic and genomic distribution analyses, the silkworm GMC genes experienced multiple amplification events, especially in the GMCβ and the GMCι subfamilies. The expression patterns of two duplicated groups from the SEs of the GMCβ subfamily and two duplicated groups from the GMCβ and the GMCι families in the silkworm-conserved GMC cluster were investigated using RT-PCR. The results suggested that most of the genes had transcriptional activity (Fig. 5). *BmGMC6*, *BmGMC16*, *GMCβ3-5* and *BmGMCι2* were not expressed in any tissues from Day 3 of the fifth instar larvae. Among the remaining silkworm GMC genes in Fig. 5, most were detected in at least three tissues. Twelve genes were expressed both in the head and the silk gland. *BmGMC1* and *BmGMC5* were only expressed in the head. However, *BmGMC2* was specifically expressed in the silk gland. Only *BmGMC25* had no expression in those two tissues, but had high expression levels in the integument, fat body and midgut. The expression profiles of these genes were also surveyed during different developmental stages. Similar to the spatial expression patterns, the genes that were not expressed in any tissues had no expression or a very low level of expression from the 5th instar larval through to the adult stage. Most of the remaining genes were mainly expressed in the larval stage, especially in the first two days of the 5th instar larvae. Some genes had a high expression level during the late pupal stage, such as *BmGMC4*, *BmGMC12-15*, *BmGMC17*, *BmGMCι3* and *BmGMι5*.

Overall, each silkworm GMC gene had a distinct expression pattern. Even the silkworm expansion genes (SEs) that were closely related to one another on the phylogenetic tree had different expression patterns. The fruit fly GMC genes also exhibited diverse spatial and developmental expression patterns (Iida et al., 2007).

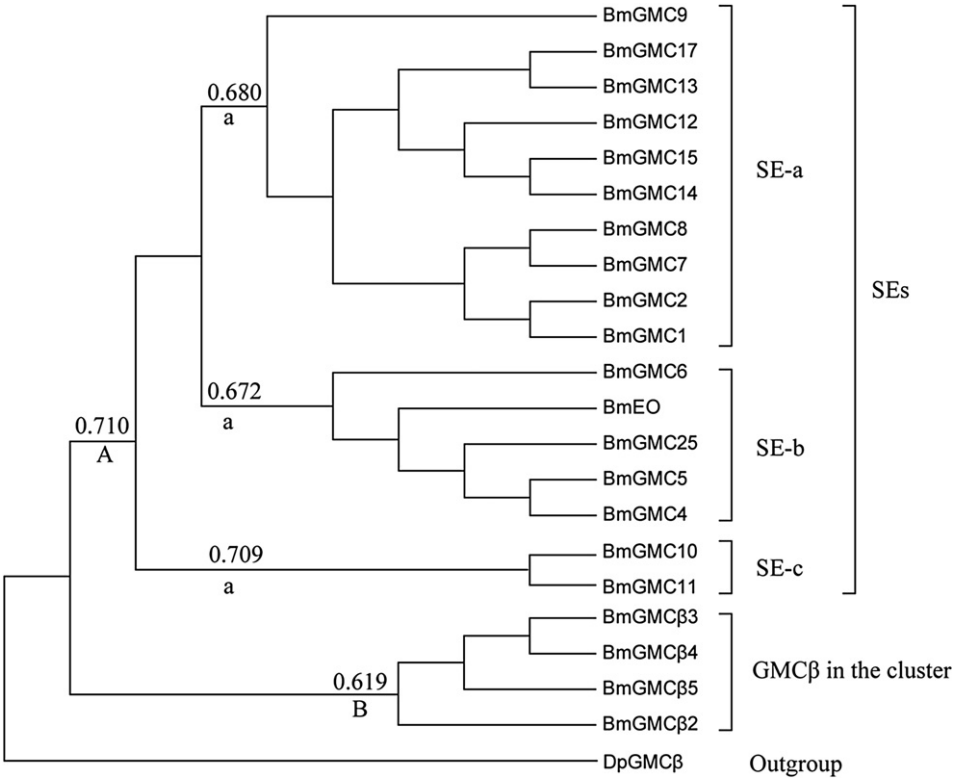


**Fig. 3.** The results for relative-rate test of the silkworm GMCβ genes. The phylogenetic tree was reconstructed using the Bayesian inference method. The monarch butterfly GMCβ gene was used as the outgroup. The number above a branch means the number of non-synonymous substitutions per non-synonymous site. Different alphabet letters below branches mean significantly difference at $P < 0.05$ from the comparison of two clades. SE represents the silkworm expansion GMC genes.

These results indicated that different members of the GMC gene family might serve different biological functions.

### 3.6. Expressions of the silkworm GMCβ genes induced by pathogens

Members of the GMC family can catalyze diverse reactions. Previous studies showed that some of the reactions catalyzed by the GMC family could produce hydrogen peroxide ($H_2O_2$) (Cavener, 1992; Zámocký et al., 2004). Hydrogen peroxide can directly kill pathogenic bacteria and restrain their propagation and pore formation. In addition, $H_2O_2$ can also act as a messenger to induce immune-related genes (Lin et al., 2005). These features of the GMC family were detected in several insect species (Santos et al., 2005; Michalski et al., 2008). In a previous study, four pathogens (*E. coli*, *Ba. bombysepticus*, *Be. bassiana* and *B. mori* nuclear polyhedrosis virus) were used to infect the silkworm, and the induced expression patterns of the silkworm genes were investigated using a microarray (Huang, 2010). We retrieved the microarray information for the silkworm GMC genes from the published data.

The microarray data showed that 16 silkworm GMC genes were upregulated after oral infection with the four microorganisms, compared with the control (Fig. S4) (Huang, 2010). Among these genes, 14 genes belonged to the silkworm GMCβ subfamily. It should be pointed out that 10 of the 14 genes were silkworm expansion genes (SEs). These 10 genes included six genes from the SE-a clade, three genes from the SE-b clade and one gene from the SE-c clade (Fig. 3). The genes in the SE-a clade were always triggered at the late stage after infection (24 h or 48 h). Moreover, *BmGMC12*, *BmGMC13* and *BmGMC14*, which were located on one scaffold, were highly upregulated only by *E. coli* (Fig. S1; Fig. S4). In the previous analysis, *BmGMC6*, *BmGMCβ3*, *BmGMCβ4* and *BmGMCβ5* were not expressed in any tissues and developmental stages. However, the pathogens did induce the expressions of these four genes, especially *BmGMCβ3* and *BmGMCβ4*. Thus, the silkworm GMCβ subfamily experienced a large expansion during evolution, and many silkworm GMCβ genes could be upregulated by different pathogens, which suggested that the expansion of the silkworm GMCβ genes might be associated with immunity.

### 3.7. Knockdown of the GMCβ genes affects the survival rate of infected silkworm

To confirm the roles of the silkworm GMCβ genes in innate immunity, RNA interference (RNAi) was performed. *BmGMC1* and *BmGMCβ3* were highly induced by *Ba. bombyseptieus*. Only *E. coli* significantly triggered *BmGMC12*. Moreover, *BmGMCβ3* is located in the conserved GMC cluster, and *BmGMC1* and *BmGMC12* belong to the silkworm expansion genes outside of the conserved cluster. Therefore, we knocked down these three genes to detect their functions in the silkworm immune system. After successfully reducing mRNAs levels of *BmGMC1*, *BmGMCβ3* and ds-*BmGMC12* by RNAi (24 h after dsRNA injection, Fig. S5), we used *Ba. bombyseptieus* and *E. coli* to infect the corresponding silkworms. In the RNAi experiments, ds-*EGFP* and saline were injected into silkworm and the injected silkworm was used as a control in subsequent infection experiments. We found that the ds-EGFP and saline injected silkworms showed similar rates of mortality after infection (log-rank test: *Ba. bombyseptieus*, $P = 0.758$; *E. coli*, $P = 1$) (Fig. 6). When *E. coli* or *Ba. bombyseptieus* were used to challenge the silkworms, the survival rates of the silkworms whose three representative GMCβ genes were knocked down were significantly lower than that of the control (log-rank test: ds-*BmGMC1 vs.* ds-*EGFP*, $P = 0.002$; ds-*BmGMCβ3 vs.* ds-*EGFP*, $P = 0.010$; ds-*BmGMC12 vs.* ds-*EGFP*, $P = 0.044$) (Fig. 6). The ds-*BmGMC1* and ds-*BmGMCβ3* injected

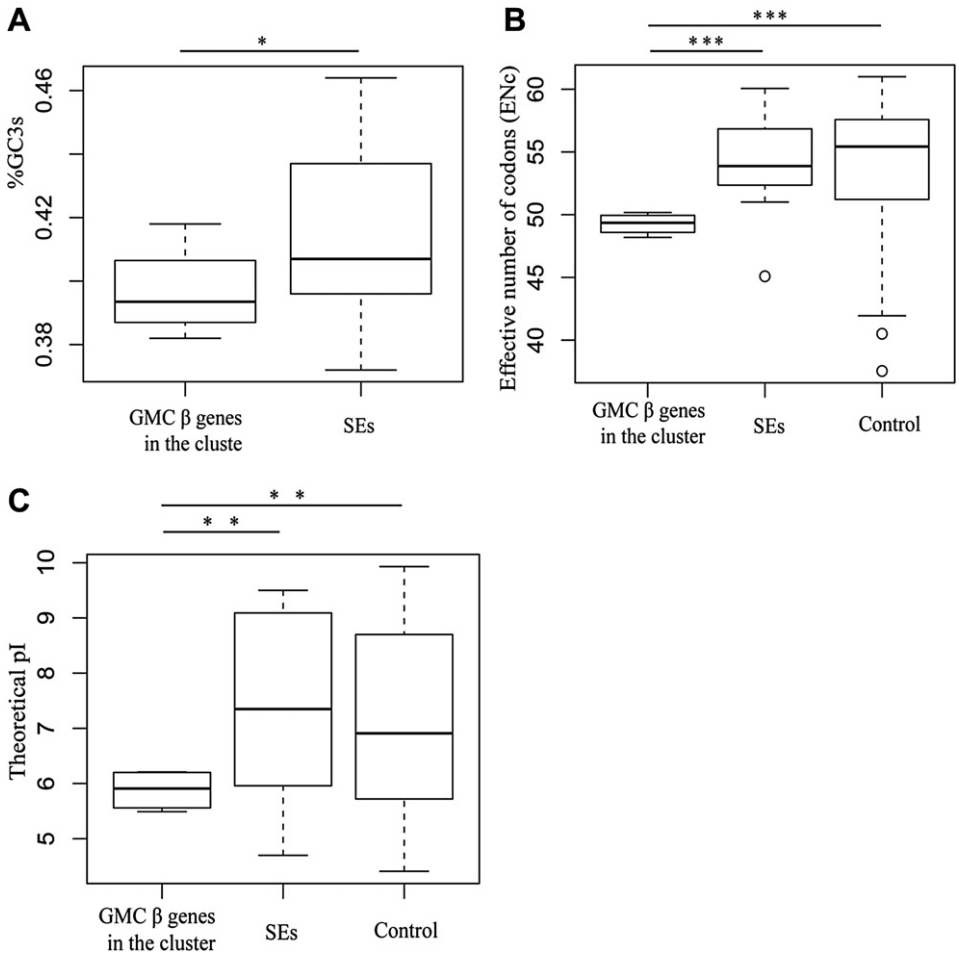


**Fig. 4.** Box-plot comparisons of GC3s content, codon usage bias (ENC) and the theoretical p*I* (isoelectric point) for the silkworm GMCβ genes inside and outside of the conserved GMC cluster. Statistical significance: $^{***}P < 0.001$; $^{**}P < 0.01$; $^{*}P < 0.05$.

silkworms at 18 h after infection by *Ba. bombyseptieus* showed 19.0% and 22.2% survival rate, which was significantly than the control (58.8%). The ds-*BmGMC12* injected silkworm had a low (36.4% *vs*. 66.7%) survival rate compared with the control group at 24 h after infection by *E. coli*. In addition, we also knocked down *BmGMCι3*, which could not be induced by the pathogens in the microarray analysis, as the negative control. The survival rates of *BmGMCι3* knockdown silkworm and the control were similar (log-rank test: *Ba. bombyseptieus,* $P = 0.779$; *E. coli*, $P = 0.397$) (Fig. S6). Some previous studies showed that GMC oxidoreductases might have roles in certain developmental processes (Cavener, 1992; Iida et al., 2007). To exclude the possibility that RNA interference itself may affect larval mortality, the larval weight and survival rate of the RNAi-knockdown silkworm were assessed under normal circumstance. The results showed that all RNAi-knockdown silkworms grew and developed normally (Fig. S7). Taken together, these results strongly indicated that the *BmGMC1*, *BmGMC12* and *BmGMCβ3* genes might have important roles in defense against pathogens.

## 4. Discussion

In the present study, we identified the 43 GMC genes in the silkworm genome. Similar to a previous study (Iida et al., 2007), 15 GMC genes of eight GMC subfamilies formed a GMC cluster in the intron of the *flo-2* gene. Synteny analysis showed that the GMC cluster is highly conserved across different insects (Fig. S2). In addition, it was suggested that five single copy genes (GMCζ, GMCε, GMCδ, GMCγ and GMCα) were the core genes of the conserved GMC cluster (Iida et al., 2007). We also found some conserved motifs upstream of every core gene in all five insect species (Fig. S3). The conserved order, orientation and motifs of the GMC cluster indicated their important functions among different species.

The remaining 28 silkworm GMC genes outside of the conserved GMC cluster are dispersed on different chromosomes. Most of these genes belong to the GMCβ subfamily, indicating that this subfamily has experienced a large expansion in the silkworm. In addition, GMC genes in other insect species have also experienced lineage-specific duplication events in some GMC subfamilies (Fig. 1). Chien et al. (2004) suggested that the expansion of a paralogous gene family might be associated with adaptation to specific environments and the development of novel life strategies. Yamanaka et al. (1998) also showed that the large CspA family is beneficial for *E. coli* to respond to different environment stresses. Therefore, these lineage-specific expansions of the GMC genes and subsequent functional divergences might help insects to adapt to diverse environments, because insects have the broadest habitats.

Gene duplication may be the result of unequal crossing over, transposon-mediated or chromosome rearrangement events. Unequal crossing over often produces tandem duplicated genes, whereas a transposon-mediated duplicated gene is random distributed at any location in a genome (Zhang, 2003). Compared

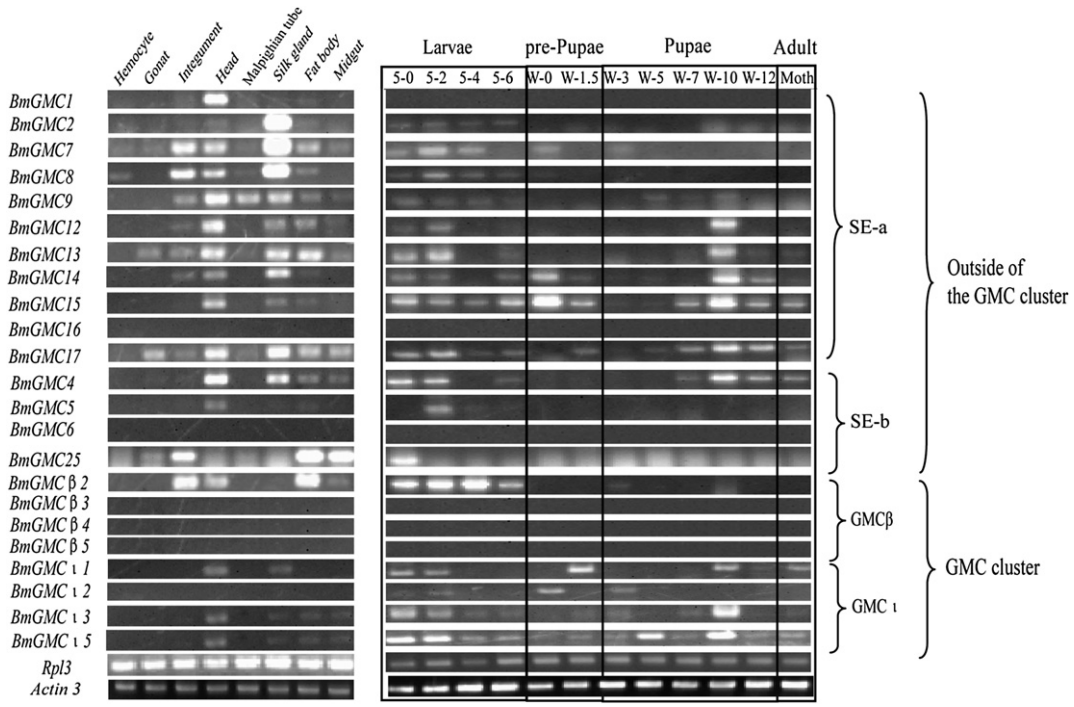


**Fig. 5.** Expression patterns of the silkworm GMC genes. (A) The spatial expression profiles of GMC genes in the silkworm, the gonad includes ovary and testis. (B) The temporal expression profiles of GMC genes in the silkworm.

with other species, the most obvious characteristic is that the silkworm GMC genes outside of the conserved GMC cluster experienced a burst expansion. Although there are also lineage-specific expansions in some insects (e.g. nine GMCλ genes in the monarch butterfly and nine GMCι genes in the red flour beetle), the number of silkworm GMC genes (28 copies) outside of the cluster is the largest among the insects investigated. The silkworm GMC genes outside of the conserved GMC cluster formed several small groups, each containing 2–5 genes in the phylogenetic tree. Moreover, most of the tandemly arranged silkworm GMC genes on one scaffold have similar numbers of introns (Fig. S1). This indicated that the tandemly arranged GMC genes might be the result of unequal

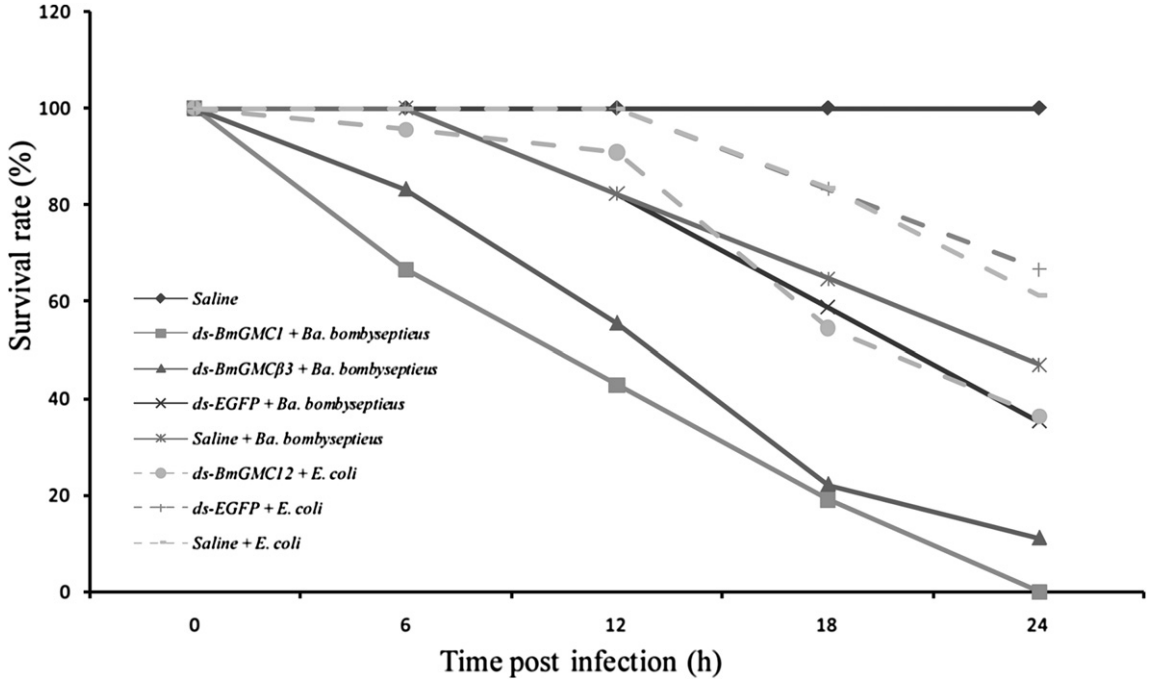


**Fig. 6.** Survival rates of the pathogen infected silkworms after dsRNA injection. Thirty micrograms of the dsRNA were injected into each silkworm. Twelve hours after dsRNA injection, the pathogens were used to infect the silkworm. The dotted lines mean that the dsRNA-injected silkworm was challenged with *E. coli*. The solid lines mean that the dsRNA injected silkworm was challenged with *Ba. bombyseptieus*.

crossing over. Moreover, the silkworm GMC groups outside of the conserved GMC cluster are randomly located on10 different chromosomes (Table 1; Fig. S1). Transposons make up ~35% of the silkworm genome (Osanai-Futahashi et al., 2008). To investigate whether the dispersed silkworm GMC groups were mediated by transposons, the average density of repetitive sequences of the 28 GMC genes outside of the conserved GMC cluster was compared with that of whole silkworm predicted genes (14,623 genes). A bootstrap method was used for this comparison (Efron and Tibshirani, 1986). The detailed method is described in Fig. S8. The result showed that the mean density of repetitive sequences of the 28 silkworm GMC genes was significantly higher than that of 28 random selected genes ($P = 0.0116$; 100,000 samplings; Fig. S8). Therefore, it is likely that the silkworm GMC groups dispersed on different scaffolds or chromosomes were created by transposon-mediated gene duplication. Taken together, both unequal crossing over and transposon-mediated events contributed to the expansion of the silkworm GMC genes outside of the conserved GMC cluster.

Among the 28 silkworm genes outside of the conserved GMC cluster, 18 belonged to the GMCβ subfamily (SEs). The relative evolutionary rates of the GMCβ genes inside and outside of the cluster were compared. The results showed that the silkworm expansion GMCβ genes had accelerated evolution rates (Fig. 3). The rapid evolution of SEs led to an increase of the GC3s content and a decrease of the codon usage bias of the GMCβ genes outside of the conserved GMC cluster. Moreover, the rapid evolution influenced the protein characteristics. The theoretical p*I*s of the SEs were significantly higher than those of the GMCβ genes in the conserved cluster. Previous analysis showed that changes in theoretical p*I*s of the members of murine β-Defensin reflected changes in antibacterial function (Morrison et al., 2003). Therefore, the different theoretical p*I*s among members of GMCβ genes may reflect their diverse biological functions.

Previous studies showed that several GMC genes, including *GLD* and *GOX*, respond to pathogen infection (Cox-Foster and Stehr, 1994; Yang and Cox-Foster, 2005). The former was speculated to take part in the melanized encapsulation reaction via oxidative free radicals (Cox-Foster and Stehr, 1994), and the latter may be secreted by worker bees to sterilize food when they feed on the larvae (Santos et al., 2005). These results indicated that some insect GMC genes might be immune-related. Recently, Huang (2010) used four different microorganisms to infect silkworm and surveyed the expressions of silkworm genes at a genomic level using a microarray. The expression patterns of the silkworm GMC genes were examined after infection by four microorganisms and 16 silkworm GMC genes were upregulated by the different pathogens. Fourteen of these 16 (64%) GMC genes belonged to the GMCβ subfamily. Interestingly, four silkworm GMCβ genes (*BmGMC6*, *BmGMCβ3*, *BmGMCβ4* and *BmGMCβ5*) were not expressed under normal conditions; however, they were induced by the pathogens. Among the 22 silkworm GMCβ genes, 18 were silkworm expansion genes (SEs). Ten of the 18 genes were triggered by at least one microorganism. Moreover, silkworms whose three representative GMCβ genes (*BmGMC1*, *BmGMC12* and *BmGMCβ3*) were knocked down exhibited higher susceptibilities to pathogens. This indicated that these genes are essential for silkworm to defense against the pathogens. Thus, the expansion of the silkworm GMCβ subfamily may be involved in resistance to pathogenic microorganisms.

In insect species, there is little functional information on GMC genes. Iida et al. (2007) proposed that GMC genes might be involved in developmental and immune process. However, few studies showed that GMC genes have a role in immunity. The present study demonstrated that the silkworm GMCβ subfamily has experienced an expansion and the expanded genes respond to pathogenic infection. Previous studies showed that some immune-related genes experienced major family expansions in insects. A large expansion of serpin genes was observed in *Tribolium* (Zou et al., 2007). A burst expansion of the immunity-related fibrinogen-domain (FBN) family was detected in *A. gambiae* (Christophides et al., 2002). In Lepidoptera, the silkworm genome has many antimicrobial peptide genes (Tanaka et al., 2008). These species-specific expansions of immunity-related genes may reflect different responses to different pathogenic microorganisms in different habitats of insects (Zou et al., 2007). Similar to these immune-related genes, an expansion of the silkworm GMCβ subfamily may also be important for innate immunity in the silkworm.

## Acknowledgments

We thank Dr. Tian Li (Southwest University) for writing the Perl program. This research was supported by a grant from National Natural Science Foundation of China (No. 30970409), by Natural Science Foundation Project of CQ CSTC (cstc2012jjB80007) and the Doctorial Innovation Fund of Southwest University (Kb2009019).

## Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.ibmb.2012.09.006.

## References

Ahmad, S.T., Joyce, M.V., Boggess, B., et al., 2006. The role of *Drosophila* ninaG oxidoreductase in visual pigment chromophore biogenesis. Journal of Biological Chemistry 281, 9205.
Ambroise, J., Bearzatto, B., Robert, A., et al., 2011. Impact of the spotted microarray preprocessing method on fold-change compression and variance stability. BMC Bioinformatics 12, 413.
Bateman, A., Coin, L., Durbin, R., et al., 2004. The Pfam protein families database. Nucleic Acids Research 32, D138–D141.
Bede, J.C., Musser, R.O., Felton, G.W., et al., 2006. Caterpillar herbivory and salivary enzymes decrease transcript levels of *Medicago truncatula* genes encoding early enzymes in terpenoid biosynthesis. Plant Molecular Biology 60, 519–531.
Cavener, D.R., 1992. GMC oxidoreductases: a newly defined family of homologous proteins with diverse catalytic activities. Journal of Molecular Biology 223, 811–814.
Chien, M., Morozova, I., Shi, S.D., et al., 2004. The genomic sequence of the accidental pathogen *Legionella pneumophila*. Science 305, 1966–1968.
Christophides, G.K., Zdobnov, E., Barillas-Mury, C., et al., 2002. Immunity-related genes and gene families in *Anopheles gambiae*. Science 298, 159–165.
Cox-Foster, D.L., Stehr, J.E., 1994. Induction and localization of FAD-glucose dehydrogenase (GLD) during encapsulation of abiotic implants in *Manduca sexta* larvae. Journal of Insect Physiology 40, 235–249.
Diezel, C., von Dahl, C.C., Gaquerel, E., et al., 2009. Different lepidopteran elicitors account for cross-talk in herbivory-induced phytohormone signaling. Plant Physiology 150, 1576–1586.
Duan, J., Li, R., Cheng, D., et al., 2010. SilkDB v2. 0: a platform for silkworm (*Bombyx mori*) genome biology. Nucleic Acids Research 38, D453–D456.
Eddy, S.R., 1998. Profile hidden Markov models. Bioinformatics 14, 755–763.
Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32, 1792.
Efron, B., Tibshirani, R., 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Statistical Science 1, 54–75.
Fakhouri, M., Elalayli, M., Sherling, D., et al., 2006. Minor proteins and enzymes of the *Drosophila* eggshell matrix. Developmental Biology 293, 127–141.
Finn, R.D., Mistry, J., Schuster-Böckler, B., et al., 2006. Pfam: clans, web tools and services. Nucleic Acids Research 34, D247–D251.
Huang, L., 2010. A Genome-wide Analysis of the Silkworm Host Responses to *Bacillus bombyseptieus* (Bb) and Other Pathogens. Southwest University.
Iida, K., Cox-Foster, D.L., Yang, X., et al., 2007. Expansion and evolution of insect GMC oxidoreductases. BMC Evolutionary Biology 7, 75.
Letunic, I., Bork, P., 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics 23, 127–128.
Leung, Y.F., Cavalieri, D., 2003. Fundamentals of cDNA microarray data analysis. Trends in Genetics 19, 649–659.
Lin, W., Hu, X., Zhang, W., et al., 2005. Hydrogen peroxide mediates defence responses induced by chitosans of different molecular weights in rice. Journal of Plant Physiology 162, 937–944.
Lynch, M., 2007. The Origins of Genome Architecture. Sinauer Associates.

Lynch, M., Conery, J.S., 2000. The evolutionary fate and consequences of duplicate genes. Science 290, 1151.
Machanick, P., Bailey, T.L., 2011. MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics 27, 1696–1697.
Michalski, C., Mohagheghi, H., Nimtz, M., et al., 2008. Salicyl alcohol oxidase of the chemical defense secretion of two *Chrysomelid* leaf beetles. Journal of Biological Chemistry 283, 19219–19228.
Morrison, G.M., Semple, C.A., Kilanowski, F.M., et al., 2003. Signal sequence conservation and mature peptide divergence within subgroups of the murine β–defensin gene family. Molecular Biology and Evolution 20, 460–470.
Musser, R.O., Hum-Musser, S.M., Eichenseer, H., et al., 2002. Herbivory: caterpillar saliva beats plant defences. Nature 416, 599–600.
Nylander, J.A.A., Wilgenbusch, J.C., Warren, D.L., et al., 2008. AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. Bioinformatics 24, 581.
Ohashi, K., Natori, S., Kubo, T., 1999. Expression of amylase and glucose oxidase in the hypopharyngeal gland with an age-dependent role change of the worker honeybee (*Apis mellifera* L.). European Journal of Biochemistry 265, 127–133.
Osanai-Futahashi, M., Suetsugu, Y., Mita, K., et al., 2008. Genome-wide screening and characterization of transposable elements and their distribution analysis in the silkworm, *Bombyx mori*. Insect Biochemistry and Molecular Biology 38, 1046–1057.
Robinson-Rechavi, M., Huchon, D., 2000. RRTree: relative-rate tests between groups of sequences on a phylogenetic tree. Bioinformatics 16, 296–297.
Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572.
Saeed, A.I., Bhagabati, N.K., Braisted, J.C., et al., 2006. TM4 microarray software Suite. Methods in Enzymology 411, 134–193.
Santos, K.S., Delazari dos Santos, L., Anita Mendes, M., et al., 2005. Profiling the proteome complement of the secretion from hypopharyngeal gland of Africanized nurse-honeybees (*Apis mellifera* L.). Insect Biochemistry and Molecular Biology 35, 85–91.
Sarfare, S., Ahmad, S.T., Joyce, M.V., et al., 2005. The *Drosophila* ninaG oxidoreductase acts in visual pigment chromophore production. Journal of Biological Chemistry 280, 11895–11901.
Sun, W., Shen, Y.H., Qi, D.W., et al., 2012. Molecular cloning and characterization of ecdysone oxidase and 3-dehydroecdysone-3α-reductase involved in the ecdysone inactivation pathway of Silkworm, *Bombyx mori*. International Journal of Biological Science 8, 125.
Takeuchi, H., Rigden, D.J., Ebrahimi, B., et al., 2005. Regulation of ecdysteroid signalling during *Drosophila* development: identification, characterization and modelling of ecdysone oxidase, an enzyme involved in control of ligand concentration. Biochemistry Journal 389, 637.
Tanaka, H., Ishibashi, J., Fujita, K., Nakajima, Y., Sagisaka, A., Tomimoto, K., Suzuki, N., Yoshiyama, M., Kaneko, Y., Iwasaki, T., Sunagawa, T., Yamaji, K., Asaoka, A., Mita, K., Yamakawa, M., 2008. A genome-wide analysis of genes and gene families involved in innate immunity of *Bombyx mori*. Insect Biochemistry Molecular Biology 38 (12), 1087–1110.
The International Silkworm Genome Consortium, 2008. The genome of a lepidopteran model insect, the silkworm *Bombyx mori*. Insect Biochemistry Molecular Biology 38, 1036–1045.
Yamanaka, K., Fang, L., Inouye, M., 1998. The CspA family in *Escherichia coli*: multiple gene duplication for stress adaptation. Molecular Microbiology 27, 247–255.
Yang, X., Cox-Foster, D.L., 2005. Impact of an ectoparasite on the immunity and pathology of an invertebrate: evidence for host immunosuppression and viral amplification. Proceedings of the National Academy of Sciences of the United States of America 102, 7470–7475.
Zámocký, M., Hallberg, M., Ludwig, R., et al., 2004. Ancestral gene fusion in cellobiose dehydrogenases reflects a specific evolution of GMC oxidoreductases in fungi. Gene 338, 1–14.
Zhan, S., Merlin, C., Boore, J.L., et al., 2011. The monarch butterfly genome yields insights into long-distance migration. Cell 147, 1171–1185.
Zhang, J., 2003. Evolution by gene duplication: an update. Trends in Ecology and Evolution 18, 292–298.
Zou, Z., Evans, J.D., Lu, Z., et al., 2007. Comparative genomic analysis of the *Tribolium* immune system. Genome Biology 8, R177.