2017

Expression map of a complete set of gustatory receptor genes in chemosensory organs of *Bombyx mori*

Huizhen Guo
Tingcai Cheng

See next page for additional authors

Follow this and additional works at: https://digitalcommons.uri.edu/bio_facpubs

The University of Rhode Island Faculty have made this article openly available. Please let us know how Open Access to this research benefits you.

This is a pre-publication author manuscript of the final, published article.

Terms of Use
This article is made available under the terms and conditions applicable towards Open Access Policy Articles, as set forth in our Terms of Use.

Citation/Publisher Attribution
Available at: http://dx.doi.org/10.1016/j.ibmb.2017.02.001

This Article is brought to you for free and open access by the Biological Sciences at DigitalCommons@URI. It has been accepted for inclusion in Biological Sciences Faculty Publications by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons@etal.uri.edu.
Authors

This article is available at DigitalCommons@URI: https://digitalcommons.uri.edu/bio_facpubs/84
Accepted Manuscript

Expression map of a complete set of gustatory receptor genes in chemosensory organs of Bombyx mori


PII: S0965-1748(17)30021-8
DOI: 10.1016/j.ibmb.2017.02.001
Reference: IB 2925

To appear in: Insect Biochemistry and Molecular Biology

Received Date: 25 August 2016
Revised Date: 30 January 2017
Accepted Date: 2 February 2017


This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Expression map of a complete set of gustatory receptor genes in chemosensory organs of

*Bombyx mori*

Huizhen Guo¹, Tingcai Cheng¹, Zhiwei Chen¹, Liang Jiang¹, Youbing Guo¹, Jianqiu Liu¹,
Shenglong Li¹, Kiyoko Taniai², Kiyoshi Asaoka², Keiko Kadono-Okuda², Kallare P Arunkumar³,
Jiaqi Wu⁴, Hirohisa Kishino⁴, Huijie Zhang⁵, Rakesh K Seth⁶, Karumathil P. Gopinathan⁷, Nicolas
Montagné⁸, Emmanuelle Jacquin-Joly⁹*, Marian R Goldsmith¹⁰*, Qingyou Xia¹, Kazuei Mita¹*

¹State Key Laboratory of Silkworm Genome Biology, Southwest University, Chongqing 400715, P.
R. China.

²National Institute of Agrobiological Sciences, 1-2 Owashi, Tsukuba, Ibaraki 305-8634, Japan.

³Centre for DNA Fingerprinting and Diagnostics, Hyderabad 500001, India.

⁴Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi,
Bunkyo-ku Tokyo 113-8657, Japan.

⁵Ministry of Education Key Laboratory of Diagnostic Medicine, College of Laboratory Medicine,
Chongqing Medical University, Chongqing 400016, China.

⁶Department of Zoology, University of Delhi, Delhi 110007, India

⁷Indian Institute of Science, Bangalore, India.

⁸Sorbonne Universités, UPMC Univ Paris 06, Institute of Ecology and Environmental Sciences
iEES-Paris, 4 place Jussieu, F-75005 Paris, France.
Abstract

Most lepidopteran species are herbivores, and interaction with host plants affects their gene expression and behavior as well as their genome evolution. Gustatory receptors (Grs) are expected to mediate host plant selection, feeding, oviposition and courtship behavior. However, due to their high diversity, sequence divergence and extremely low level of expression it has been difficult to identify precisely a complete set of Grs in Lepidoptera. By manual annotation and BAC sequencing, we improved annotation of 43 gene sequences compared with previously reported Grs in the most studied lepidopteran model, the silkworm, *Bombyx mori*, and identified 7 new tandem copies of *BmGr30* on chromosome 7, bringing the total number of *BmGrs* to 76. Among these, we mapped 68 genes to chromosomes in a newly constructed chromosome distribution map and 8 genes to scaffolds;
we also found new evidence for large clusters of BmGrs, especially from the bitter receptor family. RNA-seq analysis of diverse BmGr expression patterns in chemosensory organs of larvae and adults enabled us to draw a precise organ specific map of BmGr expression. Interestingly, most of the clustered genes were expressed in the same tissues and more than half of the genes were expressed in larval maxillae, larval thoracic legs and adult legs. For example, BmGr63 showed high expression levels in all organs in both larval and adult stages. By contrast, some genes showed expression limited to specific developmental stages or organs and tissues. BmGr19 was highly expressed in larval chemosensory organs (especially antennae and thoracic legs), the single exon genes BmGr53 and BmGr67, were expressed exclusively in larval tissues, the BmGr27–BmGr31 gene cluster on chr7 displayed a high expression level limited to adult legs and the candidate CO₂ receptor BmGr2 was highly expressed in adult antennae, where few other Grs were expressed. Transcriptional analysis of the Grs in B. mori provides a valuable new reference for finding genes involved in plant-insect interactions in Lepidoptera and establishing correlations between these genes and vital insect behaviors like host plant selection and courtship for mating.

**Keywords:** insect-plant interactions, annotation, taste, gustatory receptor, Bombyx mori, RNA-seq.

1. Introduction

Insects, especially phytophagous insects, have formed specific relationships with their host plants for a long period of evolution. The interaction and co-evolution between insects and host plants laid the foundation for insects to survive and expand into genetic races (Ehrlich and Raven,
1964; Jiang et al., 2015), and for processes in which chemical senses play critical roles, such as detection of food, oviposition sites, predators, and mates. Chemical reception is mediated by specialized sensory neurons located in appendages such as antennae, mouthparts, legs, and ovipositors. While olfactory neurons recognize volatile cues, gustatory neurons sense soluble chemicals by contact. Both mechanisms involve expression of chemosensory genes whose genomic organization, expression and evolution participate in shaping the process of insect-plant interactions (Engsontia et al., 2014; Gardiner et al., 2008; McBride et al., 2007; Vieira and Rozas, 2011; Xu et al., 2016).

At present, the signal transduction process of insect olfactory sensory cells is relatively clear (Benton et al., 2006; Sato et al., 2008; Smart et al., 2008; Wicher et al., 2008), while less is known about the gustatory process (Sato et al., 2011; Zhang et al., 2011). Both involve transmembrane chemoreceptors encoded by olfactory receptor genes (Ors), gustatory receptor genes (Grs) and ionotropic receptor genes (Irs) (Cande et al., 2013; Hansson and Stensmyr, 2011).

Comparative studies of chemosensory gene families among lepidopteran species can lead to new insights into understanding the mechanisms of host plant specialization and insect adaptation. Whereas much effort has been made to identify Or genes in Lepidoptera (Montagne et al., 2015), only fragmentary data are available about Grs, limited to species with extensive published genome sequences such as Bombyx mori (Wanner and Robertson, 2008; Sato et al., 2011; Zhang et al., 2011), Danaus plexippus (Zhan et al., 2011), Heliconius melpomene (Briscoe et al., 2013), Plutella xylostella (You et al., 2013), Manduca sexta (Koenig et al., 2015) and Hyphantria cunea (Zhang et al., 2016).
Depending on their host plant ranges, phytophagous insects can be categorized as monophagous, oligophagous or polyphagous. The domesticated silkworm, *B. mori*, is a typical oligophagous insect that feeds mainly on leaves of mulberry tree and its close taxonomic relatives. *B. mori* genome sequence data published in 2004 (Mita et al., 2004; Xia et al., 2004) led to the identification of 65 candidate *Gr* genes (further referred as *BmGrs*) (Wanner and Robertson, 2008). Four additional partial *BmGrs* (Sato et al., 2011; Zhang et al., 2011) were identified in the highly improved *B. mori* genome sequence published in 2008 (International Silkworm Genome, 2008). However, due to the very low expression level and high divergence observed in insect chemoreceptor sequences as well as short contigs in their genome assemblies, the precise composition and structure of the *BmGr* family have not been characterized completely, nor has the nomenclature of the chemosensory genes been unified, notably for the *Grs*. Subsequently, RNA-seq technology became available as a highly sensitive and low cost method to detect weakly expressed genes, alternative splice variants and novel transcripts using ultra-high-throughput sequencing technology (Buermans and den Dunnen, 2014). This situation gave us an opportunity to carry out a more comprehensive analysis of the silkworm *Gr* gene family.

Here we identified and characterized a complete set of *BmGrs* from the improved silkworm genome assembly and sequences of additional bacterial artificial chromosome (BAC) clones encompassing *Gr* genes. Subsequently, RNA-seq analysis provided a precise expression map of *BmGrs* in various chemosensory organs. The results of this study yield a valuable new reference for comparative studies of plant-insect interactions in Lepidoptera.
2. Results

2.1 Identification and nomenclature of a complete set of BmGr genes

Based on the BmGr sequences in the KAIKObase (Shimomura et al., 2009), SilkDB (Duan et al., 2010) and the NCBI reference database, especially the reported BmGr amino acid sequences (Wanner and Robertson, 2008), we performed BLASTp (cutoff e-value: 1e-05) and tBLASTn using the highly improved silkworm genome assembly (International Silkworm Genome, 2008). We identified a total of 76 BmGrs, among which 26 gene sequences (BmGr9, 11, 13, 14, 15, 16, 17, 18, 24, 25, 26, 27, 30, 33, 34, 41, 50, 51, 53, 56, 57, 58, 60, 62, 63 and 64) were the same as previously reported. For genes initially reported to lack a 5’ end (BmGr1, 2 and 5), we determined the first exon and completed the full-length sequences. We made minor revisions of the remaining sequences except for BmGr44 and BmGr66, which differed completely from those previously published.

Zhang et al. (2011) reported BmGr66-69 as new Gr genes which we checked initially with a BLAST search and subsequently reannotated as follows. We chose H. Robertson’s Gr66 (379aa) as BmGr66 (Wanner and Robertson, 2008) because BmGr66_Zhang (341aa) aligned with BmGr67 (acc. # NM_001246287.1, 351aa) [1-314]. BmGr67_Zhang, encoding 275aa, was identical to a portion of BmGr68 (encoding 418aa), which was derived from a BmGr66 sequence reported by Sato et al. (2011) (acc. # AB600835.1, 344aa, partial). BmGr68_Zhang encoding 177aa was a part of BmGr44 (306aa) [130-306]. tBLASTn search of BmGr69_Zhang (188aa) in KAIKObase showed that it hit sequences between BmGr17 and BmGr14 on chr7. A subsequent BLASTp search in NCBI showed homology with BmGr14 (370aa) with an amino acid identity of 93/202 (46%). Since BmGr17, 14, 15 and 16 formed a gene cluster, we checked whether any new Grs were predicted in the domain chr7:
3,547,001–3,552,000 using fgenesh and genescan gene prediction programs and KAIKObase, but could not find any *Gr* genes in this region. Therefore we concluded that *BmGr69_Zhang* was not a genuine *Gr*.

We found one gene, *BmGr31*, located in a gap on chr7: 3,442,653–3,534,450. Since we found several other *Gr* genes flanking the gap (91kb), we investigated whether it might harbor other copies of *Gr* in a large gene cluster. Sequencing of BAC clone 092J04 [chr7: 3,376,7764–3,551,821] (http://sgp.dna.affrc.go.jp/KAIKObase/), which covered the gap, revealed 7 additional tandemly aligned copies of *BmGr30*. We also corrected published sequences for *BmGr29* (acc. #: BK006604) and *BmGr30* (acc. #: BK006605), which were located just at the edges of the gap. Comparison of the *BmGr29–BmGr31* sequences showed high sequence identity. Thus, in total we found 10 copies of *BmGr30* in this gap (Figure 1).

For the nomenclature of these genes, we basically followed the previously published name order (Sato et al., 2011; Wanner and Robertson, 2008; Zhang et al., 2011). For the newly identified genes on chr7, we listed them as *BmGr30-1* to *BmGr30-8* and adjusted the gene numbering based on their chromosomal location. The 154,052 bp sequence of BAC 092J04 is available in DDBJ under accession number LC056060. We improved 43 genes with suffixes marked “XX” or previously published partial *Gr* gene structures by this work. As noted above, *BmGr66* was based on the *Gr66* sequence kindly provided by H. Robertson, while *BmGr66_AB600835* (Sato et al., 2011) was changed to *BmGr68*, and *BmGr68_NM_001246288* (Sato et al., 2011) was changed to *BmGr69*. This resulted in an increase of the total number of *BmGr* from 69 to 76 and significant improvement of previous annotations. All gene sequences and detailed gene information for *BmGr* are shown in
Table S1.

2.2 BmGrs are distributed on 16 chromosomes and most are arranged in clusters

We improved a previously constructed map of BmGrs (Engsontia et al., 2014) to localize 68 genes to chromosomes (Figure 2). The remaining 8 Grs were mapped to scaffolds. BmGr65, composed of four exons, had been problematic, because the first two exons were on chr12 and the other two exons were on chr23 in the previous annotation. Here we corrected the location of BmGr65 and determined that it was close to BmGr64 on chr12 (Document S1). BmGr41, BmGr42 and BmGr43 were located on unmapped Bm_scaf444 with a size of 21,432bp. KAIKObase showed a gap of 22kb between Bm_scaf1_contig631 and contig630 [chr13: 11,532,372–11,554,690]. Here, we determined that BmGr41–43 were localized in this gap by PCR using genomic DNA (Document S2).

We also confirmed the gene structure of BmGr46 and the order of the genes in the corresponding cluster which should be BmGr46–BmGr41–BmGr42–BmGr43 (Document S3).

From the chromosome distribution map, we found that most BmGrs were in clusters. Putative CO₂ receptors (BmGr1, 2 and 3) (Wanner and Robertson, 2008) were located separately on chr7, chr8 and chr23, whereas putative sugar receptors, BmGr4, BmGr5 and BmGr6 (Mang et al., 2016; Wanner and Robertson, 2008), formed a cluster on chr15; BmGr7 and BmGr8 remained unmapped. Most of the remaining Grs, provisionally considered as putative bitter receptors (Wanner and Robertson, 2008), formed clusters, especially on chr7 (16 genes) and chr13 (16 genes). BmGr14–BmGr17, located in the 3’ region of the chr7 gene cluster, and 12 Grs, BmGr27–BmGr31, in the 5’ portion, presented a high level of sequence identity. BmGr49–BmGr52 formed a cluster on chr6, expansion of which is proposed to have occurred by recent gene duplication events (Figures 2 and 3).
A phylogenetic analysis (Figure 3) was carried out using the updated BmGr repertoire together with Grs identified in M. sexta and H. melpomene. As observed previously (Engsontia et al., 2014), candidate CO₂ receptors (BmGr1–3) grouped within a single clade and exhibited a high conservation level among the three lepidopteran species. The same applies for candidate sugar receptors (BmGr4–8) and for the Drosophila DmGr43a orthologs, which include D-fructose receptor BmGr9 (Sato et al., 2011), and the inositol receptor BmGr10 (Kikuta et al., 2016). The remaining 66 BmGrs, including all the newly identified genes, belong to highly divergent clades of candidate bitter receptors, with very few one-to-one orthology relationships evidenced among the three species (e.g., BmGr54 and 63). Genes clustered on the same chromosome also belonged to the same clade in the phylogenetic tree. Most bitter Grs shared a similar gene structure composed of 3–5 exons, except for BmGr53 and BmGr67 which had a single exon each, whereas CO₂ and sugar receptor genes had more complex gene structures composed of 5–9 exons and 8–12 exons, respectively. The CDS sequence of bitter receptor genes within each gene cluster shared more than 70% identity, whereas intron identity decreased to around 45%. Notably, we observed 95%–99% homology for CDS sequences and 62%–94% for introns among the newly identified copies of the chr7 gene cluster, suggesting they expanded through very recent gene duplication events. Thus, not only the exons but also the introns were congruent with the phylogeny (Figure 3).

2.3. Transcriptional analysis of BmGrs in silkworm larval and moth chemosensory organs

Determining whether identified Gr genes are indeed expressed and in which tissue they are expressed is essential for understanding their role in insect interactions with the environment. Here we used an RNA-seq analysis based on the more precise gene annotation to determine the
spatio-temporal expression profiles of BmGrs in various chemosensory organs. RNA-seq detected
most transcripts in one or more of the following tissues: larval antennae, larval maxillary galea +
palps, larval legs, adult antennae and adult legs, all collected from both males and females (Figures
S1 and S2). Only two genes (BmGr25 and 61) had no evidence of expression in any of the tissues
investigated. By contrast, BmGr63 was highly expressed in every tissue, and BmGr18 and 19 were
expressed in every tissue except adult antennae. It is notable that 46 Gr genes (Figure S1) were
expressed in larval maxillae, 44 Gr genes expressed in larval thoracic legs and 52 Gr genes (Figure
S2) expressed in adult legs, indicating that these pairs of appendages are important gustatory tissues
in B. mori. Relatively fewer Gr genes were expressed in larval or adult antennae. Among these,
BmGr19 showed an extremely high expression level in larval antennae, whereas the candidate CO₂
receptor BmGr2 was highly expressed in adult antennae. BmGr53 and BmGr67, which shared the
same single exon gene structure, were expressed highly in the larval stage, but showed no expression
in adult organs. BmGr63 seemed to show a high expression level in all chemosensory organs
examined in both larval and adult stages.

Considering the putative bitter receptor gene cluster found on chr7, the 12 highly homologous
Grs constituting the 5’ part of the gene cluster (BmGr27–31) were expressed exclusively in adult
legs. In contrast, BmGr14–17, found on the same chromosome, were weakly expressed in larval
maxillae, but were not expressed in other tissues. Four genes that formed a cluster on chr6 (BmGr49–
52) were highly expressed in larval organs, but presented almost no expression in adult organs. The
large Gr gene cluster on chr13 (Figure 2 and 4) was split into three parts [BmGr41/42/43/45/46/48],
[BmGr32–37/40/47] and [BmGr39]. Grs in the first and third parts were highly expressed in larval
organs, with almost no expression in adult tissues, whereas the seven Grs in the central part were mainly expressed in adult legs (Figures 4 and S1). These expression patterns were consistent with the phylogenetic tree (Figure 3).

Several of these Grs showed sex-biased expression. BmGr10, 16, 28, 29, 37, 38, 45, 49, 58 and 66 were male-biased, whereas BmGr6, 15, 18, 20, 24, 30-1–30-8, 33–36, 40, 55, 62, 67 and 69 were female-biased. Comparison of Gr gene expression between larval legs and moth legs called attention to two sets of genes with complementary patterns (Figure S2). Notably, BmGr19 (chromosome site unknown) was expressed at an extremely high level in larval thoracic legs, whereas expression of the BmGr27–31 gene cluster (chr7) was strictly limited to moth legs.

Larval mouthparts contain many sensilla located on maxillary galea, maxillary palps and epipharynx, which are the most important chemosensory organs for larval food selection (Ishikawa and Hirao, 1961). In order to understand which genes could be responsible for silkworm larval food selection, we dissected out larval maxillary palps, maxillary galea, epipharynx and thoracic legs separately to perform RNA-seq (Figure S3). More BmGrs were expressed in the maxillary palps than maxillary galea or epipharynx. The inositol receptor BmGr8 (Zhang et al., 2011) was mainly expressed in the maxillary galea, whereas the D-fructose receptor BmGr9 was expressed highly in maxillary palps. The three candidate CO₂ receptors (BmGr1, 2 and 3) were highly expressed in the maxillary palps of both sexes, and sugar receptors were mainly expressed in the maxillary galea and palps; however, bitter receptors could be detected not only in maxilla, but also in epipharynx. Genes highly expressed at the larval stage (BmGr18 and BmGr19) had much higher expression in epipharynx and maxillary palps than maxillary galea in both sexes. The recently expanded BmGrs
identified on chr7 (BmGr27 to 31) showed almost no expression in these organs.

3. Discussion

Thanks to genome sequencing and comparative genomics, in recent years data have accumulated on the size of Gr gene families in different species and on their evolution (Engsontia et al., 2014; Lavagnino et al., 2012; Zhang et al., 2016). Information on the expression patterns of Grs in different developmental stages, different sexes and different chemosensory organs is much sparser, and functional data have been mostly limited to Drosophila Grs (Isono and Morita, 2010). Such data are needed for understanding how the Gr repertoires of various species have evolved for discrimination of suitable host plants. In Lepidoptera, only two large scale Gr expression analyses have been reported: one focused only on adult tissues (Briscoe et al., 2013) and the other on adult and larval tissues (Xu et al., 2016). At the functional level, only three Grs have been functionally characterized in B. mori, all responding to sugars (Mang et al., 2016; Sato et al., 2011; Zhang et al., 2011).

In the present work, we re-annotated the previously reported Grs of B. mori with precise gene structures and a unified nomenclature based on the improved silkworm genome assembly. Furthermore, we identified 7 additional copies of BmGr30 by sequencing. The BmGrs annotated here are located on 16 chromosomes, mostly in clusters in the same chromosomal regions. Based on their high level of sequence identity and similar exon structure, the gene clusters of chr7 and chr13 were formed by recent gene duplication and expansions events. The phylogenetic analysis shows that candidate bitter receptor gene duplications generally occurred after M. sexta, H. melpomene and B. mori separated. In other words, bitter receptors may have evolved faster than
other chemosensory genes. This strongly suggests that bitter receptors made a significant contribution to adaptation of Lepidoptera to novel ecological niches. Although the total number of Gr genes we identified is in the range of what has been described in most other lepidopteran species, a recent study revealed significant expansion of the Gr gene family in the armyworm, *H. armigera* (Lepidoptera: Noctuidae), with a description of up to 197 genes (Xu et al., 2016). Our study confirms the hypothesis that Gr expansion may be linked to polyphagy, since a much lower number of Grs could be identified in the oligophagous *B. mori*.

We also performed the first comprehensive and detailed transcriptional analysis of *BmGr* genes by RNA-seq of a wide range of chemosensory organs including larval antennae, epipharynx, maxillary galea, maxillary palps, and thoracic legs, and moth antennae and legs, each separated by sex. Interestingly, this analysis revealed that most genes located in the same cluster (e.g., genes in the different sections of chr7 and chr13, and *BmGr49–52* on chr6 (Figure 5)) shared the same expression pattern, suggesting their expression is controlled by the same upstream regulatory element(s). It will be interesting to clarify if the clustered Grs are expressed in the same GRN (gustatory receptor neurons) or in a subset of GRNs, as it is known that GRNs can express multiple Gr genes (Scott et al., 2001), contrary to olfactory receptor neurons which usually express only one Or gene in addition to the co-receptor Orco. This also calls attention to the need for understanding how the stage- and tissue-limited expression of gustatory genes is coordinated with the regulatory neuronal network required for food choice and other sensory perception.

Several genes showed sex-biased, developmental stage or tissue specific expression and thus may be involved in stage or sex specific behaviors, such as food choice (larvae), oviposition site
selection (females) or contact behaviors (males and females). These distinctive expression patterns
suggest prioritization of future Gr functional studies: BmGr67 was not only female-biased, but also
highly expressed in larval antennae, with no expression in adult tissues. Two candidate CO₂
receptors contrasted in relatively specific expression patterns: BmGr1 in larval maxilla and BmGr2,
which was highly expressed in moth antennae. BmGr63, 18 and 19 exhibited high expression levels
in nearly all organs examined in both larval and adult stages, suggesting an important role in
gustation.

Specific expression of Gr(s) ensures the functional diversity of GRNs that together shape the
way the insect senses its external environment. In caterpillars, these Grs are expressed in a pair of
styleconic sensilla located on the maxillary galea that each house 4 GRNs in basiconic sensilla
located at the top of the maxillary palps, as well as in sensilla on the epipharynx (Dethier, 1937;
Ishikawa and Hirao, 1961; Schoonhoven.LM, 1969; Shields, 2009). In B. mori larvae, one GRN in
the lateral styleconic sensillum is sensitive to sucrose, and the other three respond to myo-inositol,
glucose, or salts (Ishikawa, 1963, 1966; Ishikawa and Hirao, 1963). One GRN in the medial
sensillum is sensitive to bitter compounds and the other GRNs respond to water, salts, acids,
ecdysone, or 20-hydroxyecdysone (Descoins and Marion-Poll, 1999; Tanaka et al., 1994). Our
results are consistent with these physiological data since we found evidence for the expression of
sugar receptors, especially the inositol sensitive BmGr8 (Zhang et al., 2011), as well as candidate
bitter receptors, in the galea. The gustatory sensilla found in the maxillary palps are involved in
food detection and selection (Ishikawa et al., 1969) and we found many Grs expressed in this
tissue.
In this report, we found that not only maxillae but also antennae and legs expressed gustatory receptors in silkworm, as observed in other Lepidoptera (Briscoe et al., 2013; Legeai et al., 2011; Xu et al., 2016; Zhang et al., 2016), which is in agreement with the observation that these organs carry taste sensilla. In particular, Gr expression in female moth legs has been proposed to be a determinant of oviposition site choice in *H. melpomene* (Briscoe et al., 2013). In accordance with this hypothesis, we found overexpression of many *Grs* in female moth legs compared to males. In addition, data in Figures S1 and S2, clearly show that a set of *BmGrs* was over-expressed in adult legs compared to larval legs and another subset was enriched in larval legs compared to adults, suggesting they play different roles in these differing life stages. It has been reported that during mating, a male fruit fly taps the female with its tarsal leg sensilla to make contact with the female abdomen or to detect the cuticular hydrocarbons (Bray and Amrein, 2003; Ling et al., 2014). In several lepidopteran species, it has been reported that scales act as a releaser of the copulation attempt, or play a key role in the recognition of the correct object for copulation (Ono, 1977). Although not yet described in *B. mori*, such a behavior may exist in this species. *Grs* with high expression in silkworm legs, for example, *BmGr27–31*, are good candidates to mediate such interactions, possibly via recognition of cuticular hydrocarbons during courtship behavior. Those expressed in larval legs would be responsible for food recognition, such as *BmGr18, BmGr19, BmGr53* and so on. Thus, the precise *Gr* expression map in chemosensory organs provides clues for identifying candidate *Grs* involved in critical insect behaviors.
4. Materials and methods

4.1 Bioinformatics, re-annotation, and nomenclature

We referenced sequence data from published papers on the identification of silkworm *BmGrs* (Wanner and Robertson, 2008; Zhang et al., 2011) and combined it with the silkworm genome data (http://sgp.dna.affrc.go.jp/KAIKObase/ (Shimomura et al., 2009) and http://www.silkdb.org/silkdb/ (Duan et al., 2010)) and NCBI reference data (http://www.ncbi.nlm.nih.gov/nucleotide/), followed by checking manually all members of the gene family individually. In addition, H. Robertson kindly provided us with all BmGr amino acid sequences that he identified (Wanner and Robertson, 2008). We performed tBLASTn search (cutoff e-value: 1e-05) using amino acid sequences of reported *Grs* to identify all possible candidate genes. For each identified gene, each exon/intron boundary was checked manually. Each identified gene was also evaluated by BLASTp search in public protein databases, and examined by HMMER3 search (cutoff e-value: 1e-03) using the Pfam database as well as ExPASy Prosite Release 20.120 <prosite.expasy.org>. We determined a unified nomenclature by following that of the *Grs* in the papers cited (Sato et al., 2011; Wanner and Robertson, 2008; Zhang et al., 2011).

4.2 BAC Sequencing

In order to complete the *BmGr* gene cluster on chr7, we sequenced BAC 092J04 by the BAC shotgun method as follows. We constructed two shotgun libraries of 2kb and 5kb from which we picked 590 clones randomly for each library, followed by pair-end sequencing with an ABI3730 DNA Analyzer (Applied Biosystems). After vector-trimming and removal of low quality reads (QV<20), we assembled all pair-end reads with the programs Phrap 1.08081222 (de la Bastide and...
McCombie, 2007) and Consed 16.0 (Gordon et al., 1998). The sequence of BAC092J04 (154,052 bp) is available in DDBJ under accession number LC056060.

4.3 Chromosomal distribution of chemosensory gene families in the silkworm genome

Using the physical map of *B. mori* provided in SilkDB we imported the *BmGr* nucleotide sequence data into the online SilkMap tool (http://www.silkdb.org/silksoft/silkmap.html) to output a gene-distribution map automatically.

4.4 Phylogenetic analysis of *BmGrs*

We conducted phylogenetic analysis with silkworm (Table S1) and other lepidopteran Grs from *H. melpomene* (Briscoe et al., 2013) and *M. sexta* (Koenig et al., 2015). Amino acid sequences were automatically aligned by the Mafft program version 7 (http://mafft.cbrc.jp/alignment/software/algorithms/algorithms.html), using E-INS-i strategy (Katoh and Standley, 2013). As the alignment showed highly conservative and non-conservative regions, only the conservative regions were retained for further analysis, and sequences with lengths of 340 aa were used for tree inference. Model selection was conducted by Mega version 6 (Tamura et al., 2013) and LG+Gamma+I mode (Hasegawa et al., 1985; Le and Gascuel, 2008; Yang, 1994) was found best for our dataset. The maximum likelihood tree was inferred by RaxML version 8 (Stamatakis, 2014) using the LG+Gamma+I model. To evaluate the confidence of the tree topology, the bootstrap method (Sanderson and Wojciechowski, 2000) was applied with 1000 replications using the rapid bootstrap algorithm (Stamatakis et al., 2008).
4.5 Silkworm strains and sample preparation

The silkworm strain, *Dazao*, was maintained in the Silkworm Gene Resource Library, Southwest University, China, by rearing on fresh mulberry leaves in standard conditions of 12 h light and 12 h dark cycle at 25°C. We collected the maxillae and antennae from about 500 2 to 3-day-old fifth instar larvae of each sex, and antennae, forelegs, midlegs and hindlegs from moths, also separated by sex. We washed all tissues with Phosphate-buffered saline (PBS) buffer (NaCl 137mmol/L, KCl 2.7mmol/L, Na₂HPO₄ 4.3mmol/L, KH₂PO₄ 1.4mmol/L, pH 7.4), and then put them directly into Trizol reagent (Invitrogen, USA) to avoid RNA degradation followed by storage at −80°C until use.

4.6 RNA extraction and Illumina sequencing

We extracted total RNA from the tissues prepared above using Trizol reagent according to the manufacturer’s instructions (Invitrogen, USA), and digested contaminating genomic DNA with RNase-free DNase I (Takara, China). We suspended purified RNA in 20mM sodium acetate buffer (pH 5.2) and quantified samples using a Qubit 2.0 Fluorometer (Life Technologies, Grand Island, NY). We assessed the integrity and quality of the mRNA samples using an Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA). We used 1 μg total RNA to make cDNA libraries using a TruSeq RNA sample preparation kit (Illumina, San Diego, CA). In total we prepared 24 individual cDNA libraries by ligating sequencing adaptors to cDNA fragments by PCR amplification and synthesized cDNA products using random hexamer primers, yielding an average length of 260bp. We generated raw sequencing data using an Illumina HiSeq2000 system (Illumina, USA).
4.7 Statistical analysis

We manually removed the polyA using fqtrim (v0.93) ([http://ccb.jhu.edu/software/fqtrim/index.shtml](http://ccb.jhu.edu/software/fqtrim/index.shtml)), rRNA and tRNA with Bowtie2 (v2.2.3) (Langmead and Salzberg, 2012) and low quality reads (QV<20) with Trimmomatic (v0.32) (Bolger et al., 2014), then evaluated and calculated the gene expression level using RSEM software (Li and Dewey, 2011) with the fragments per kb per million reads (FPKM) method (Mortazavi et al., 2008).

We directly used the gene sequences as an alignment reference. We integrated each sample into an expression matrix and illustrated the data with heatmaps using R (Logiciel) ([https://www.r-project.org/](https://www.r-project.org/)).

Acknowledgements

This work was supported by the Chinese government fund “One thousand plan for the foreign experts recruitment program (No. WQ20125500074)” and the National Basic Research Program of China (No. 2012CB114600).

Author contributions

HG and KM designed research, MRG, HK, KA, KT, KK, KPA, HZ and QX provided suggestions for research. HG and KM performed most of experiments with the assistance of LJ, ZC, JL and SL. HG, KM, TC, YG and JW analyzed data. HG and KM wrote the primary manuscript. MRG, KK, HK, KPG, RKS, KT, KPA, NM, EJJ and KA revised the manuscript.

Competing financial interests

The authors declare no competing financial interests.
References


Ishikawa, S., 1963. Responses of maxillary receptors in the larva of the silkworm, Bombyx mori, to


Mang, D., Shu, M., Endo, H., Yoshizawa, Y., Nagata, S., Kikuta, S., Sato, R., 2016. Expression of a sugar clade gustatory receptor, BmGr6, in the oral sensory organs, midgut, and central nervous system of larvae of the silkworm Bombyx mori. Insect biochemistry and molecular biology 70,


Figure legends

Figure 1. Detailed gene distribution of *BmGrs* in the Chromosome 7 gene cluster. BAC clone 092J04 [chr7:3,376,776…3,551,821] was shown to encompass a gap found in the organization of *BmGr* genes. Sequencing of this BAC revealed 7 additional copies of *BmGr30* tandemly aligned in this gap. The 154,052 bp sequence of BAC092J04 is available in DDBJ under accession number LC056060. Red arrows indicate direction of gene transcription.

Figure 2. Chromosomal distribution map of *BmGr* genes. The current status of gene cluster distribution is indicated. Newly identified *Gr* genes are marked with dark dots.

Figure 3. Evolutionary relationships of the *Grs* identified in three lepidopteran insect genomes. Amino acid sequences were automatically aligned by the Maftt program version 7, using the E-INS-i strategy. The evolutionary history was inferred using a maximum likelihood tree with RaxML version 8 using the LG+Gamma+I model. Model selection was conducted by Mega version 6 and LG+Gamma+I mode. Bootstrap support was 1000 replicates. Putative CO₂ and sugar gustatory receptors show conserved relationships among the three Lepidoptera species, while the remaining bitter receptors are more divergent. Bm, *Bombyx mori*; Hm, *Heliconius melpomene*; Ms, *Manduca sexta*.

Figure 4. Distribution map of *BmGr* genes in the Chromosome 13 gene cluster.

Figure 5. Expression map of *BmGr* genes in larval and adult chemosensory organs. Colored dots on larval and adult chemosensory organs correspond to positions of high expression of *BmGr* genes clustered in chr 7 and chr 13. Other genes mentioned in the text are listed near tissues where they are also highly expressed. Bar graphs are based on the FPKM values of the genes which are shown on
the Y axis. The X axis stands for tissues as follows: white, LA; black, LM; horizontal stripes, LTL; vertical stripes, MA; and grey, ML. LA, larval antenna; LM, larval maxilla; LTL, larval thoracic legs; MA, moth antenna; ML, moth legs.

Supplementary information

Supplementary Material

Document S1

Document S2

Document S3

Table S1
Graphical abstract

By manual annotation and BAC sequencing, we improved 43 gene sequences compared with previously reported Grs in the most studied Lepidoptera model, the silkworm Bombyx mori, and identified 7 new tandem copies of BmGr30 on chromosome 7, bringing the total number of BmGrs to 76.

RNA-seq analysis of diverse BmGr expression patterns in chemosensory organs of larvae and adults enabled us to draw a precise organ specific map of BmGr expression.

Body map of Gr expression provides a valuable new reference for finding the genes involved in plant-insect interactions in Lepidoptera and establishing correlations between these genes and vital insect behaviors like host plant selection and courtship for mating.
Highlights

1) By manual annotation and BAC sequencing, we improved 43 gene sequences compared with previously reported Grs in the most studied Lepidoptera model, the silkworm *Bombyx mori*, and identified 7 new tandem copies of *BmGr30* on chromosome 7, bringing the total number of *BmGrs* to 76.

2) Sequencing of BAC clone 092J04 revealed 7 additional tandemly aligned copies of *BmGr30*, resulted in 16 Gr gene cluster on chr7. In addition, we found 16 Gr gene cluster on chr13 using BAC clone.

3) RNA-seq analysis of diverse *BmGr* expression patterns in chemosensory organs of larvae and adults enabled us to draw a precise organ specific map of *BmGr* expression. Body map of Gr expression provides a valuable new reference for finding the genes involved in plant-insect interactions in Lepidoptera and establishing correlations between these genes and vital insect behaviors like host plant selection and courtship for mating.