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

Citation/Publisher Attribution
Guo, H., Cheng, T., Chen, Z., Jiang, L., Guo, Y., Liu, J., Li, S., Taniai, K., Asaoka, K., Kadono-Okuda, K., Arunkumar, K.P., Wu, J., Kishino, H., Zhang, H., Seth, R.K., Gopinathan, K.P., Montagné, N., Jacquin-Joly, E., Goldsmith, M.R., Xia, Q., Mita, K., Expression map of a complete set of gustatory receptor genes in chemosensory organs of *Bombyx mori, Insect Biochemistry and Molecular Biology* (2017).
Available at: http://dx.doi.org/10.1016/j.ibmb.2017.02.001

Accepted Manuscript

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

Huizhen Guo, Tingcai Cheng, Zhiwei Chen, Liang Jiang, Youbing Guo, Jianqiu Liu, Shenglong Li, Kiyoko Taniai, Kiyoshi Asaoka, Keiko Kadono-Okuda, Kallare P. Arunkumar, Jiaqi Wu, Hirohisa Kishino, Huijie Zhang, Rakesh K. Seth, Karumathil P. Gopinathan, Nicolas Montagné, Emmanuelle Jacquin-Joly, Marian R. Goldsmith, Qingyou Xia, Kazuei Mita

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

Please cite this article as: Guo, H., Cheng, T., Chen, Z., Jiang, L., Guo, Y., Liu, J., Li, S., Taniai, K., Asaoka, K., Kadono-Okuda, K., Arunkumar, K.P., Wu, J., Kishino, H., Zhang, H., Seth, R.K., Gopinathan, K.P., Montagné, N., Jacquin-Joly, E., Goldsmith, M.R., Xia, Q., Mita, K., Expression map of a complete set of gustatory receptor genes in chemosensory organs of *Bombyx mori*, *Insect Biochemistry and Molecular Biology* (2017), doi: 10.1016/j.ibmb.2017.02.001.

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

Huizhen Guo[1], Tingcai Cheng[1], Zhiwei Chen[1], Liang Jiang[1], Youbing Guo[1], Jianqiu Liu[1], Shenglong Li[1], Kiyoko Taniai[2], Kiyoshi Asaoka[2], Keiko Kadono-Okuda[2], Kallare P Arunkumar[3], Jiaqi Wu[4], Hirohisa Kishino[4], Huijie Zhang[5], Rakesh K Seth[6], Karumathil P. Gopinathan[7], Nicolas Montagné[8], Emmanuelle Jacquin-Joly[9*], Marian R Goldsmith[10*], Qingyou Xia[1], Kazuei Mita[1*]

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

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

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

[4]*Graduate 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.

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

[7]*Indian Institute of Science, Bangalore, India.*

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

[9]*INRA, Institute of Ecology and Environmental Sciences iEES-Paris, Route de Saint-Cyr, F-78000 Versailles, France.*

[10]*University of Rhode Island, Kingston 02881, USA.*

*Correspondence authors:

State Key Laboratory of Silkworm Genome Biology, Chongqing 400715, China. E-mail: mitakazuei@gmail.com. Tel: +86-18983794244. Fax: +86-023-68250892; Department of Biological Sciences, University of Rhode Island, Kingston, Rhode Island 02881, USA. E-mail: mki101@uri.edu. Tel: 1-401-874-2637. Fax: 1-401-874-2065; and INRA, Institute of Ecology and Environmental Sciences iEES-Paris, Route de Saint-Cyr, F-78000 Versailles, France. E-mail: emmanuelle.joly@inra.fr.

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 *Gr*s 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 CO_2 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:

125 3,547,001–3,552,000 using fgenesh and genescan gene prediction programs and KAIKObase, but
126 could not find any *Gr* genes in this region. Therefore we concluded that *BmGr69_Zhang* was not a
127 genuine *Gr*.

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

136 For the nomenclature of these genes, we basically followed the previously published name order
137 (Sato et al., 2011; Wanner and Robertson, 2008; Zhang et al., 2011). For the newly identified genes
138 on chr7, we listed them as *BmGr30-1* to *BmGr30-8* and adjusted the gene numbering based on their
139 chromosomal location. The 154,052 bp sequence of BAC 092J04 is available in DDBJ under
140 accession number LC056060. We improved 43 genes with suffixes marked "XX" or previously
141 published partial *Gr* gene structures by this work. As noted above, *BmGr66* was based on the *Gr66*
142 sequence kindly provided by H. Robertson, while *BmGr66*_AB600835 (Sato et al., 2011) was
143 changed to *BmGr68*, and *BmGr68*_NM_001246288 (Sato et al., 2011) was changed to *BmGr69*. This
144 resulted in an increase of the total number of *BmGrs* from 69 to 76 and significant improvement of
145 previous annotations. All gene sequences and detailed gene information for *BmGrs* 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 CO_2 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 CO_2 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 CO_2 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

188 spatio-temporal expression profiles of *BmGrs* in various chemosensory organs. RNA-seq detected
189 most transcripts in one or more of the following tissues: larval antennae, larval maxillary galea +
190 palps, larval legs, adult antennae and adult legs, all collected from both males and females (Figures
191 S1 and S2). Only two genes (*BmGr25 and 61)* had no evidence of expression in any of the tissues
192 investigated. By contrast, *BmGr63* was highly expressed in every tissue, and *BmGr18* and *19* were
193 expressed in every tissue except adult antennae. It is notable that 46 *Gr* genes (Figure S1) were
194 expressed in larval maxillae, 44 *Gr* genes expressed in larval thoracic legs and 52 *Gr* genes (Figure
195 S2) expressed in adult legs, indicating that these pairs of appendages are important gustatory tissues
196 in *B. mori*. Relatively fewer *Gr* genes were expressed in larval or adult antennae. Among these,
197 *BmGr19* showed an extremely high expression level in larval antennae, whereas the candidate CO_2
198 receptor *BmGr2* was highly expressed in adult antennae. *BmGr53* and *BmGr67*, which shared the
199 same single exon gene structure, were expressed highly in the larval stage, but showed no expression
200 in adult organs. *BmGr63* seemed to show a high expression level in all chemosensory organs
201 examined in both larval and adult stages.

202 Considering the putative bitter receptor gene cluster found on chr7, the 12 highly homologous
203 *Grs* constituting the 5' part of the gene cluster (*BmGr27–31*) were expressed exclusively in adult
204 legs. In contrast, *BmGr14–17*, found on the same chromosome, were weakly expressed in larval
205 maxillae, but were not expressed in other tissues. Four genes that formed a cluster on chr6 (*BmGr49–*
206 *52*) were highly expressed in larval organs, but presented almost no expression in adult organs. The
207 large *Gr* gene cluster on chr13 (Figure 2 and 4) was split into three parts [*BmGr41*/*42*/*43*/*45*/*46*/*48*],
208 [*BmGr32–37*/*40*/*47*] and [*BmGr39*]. *Grs* in the first and third parts were highly expressed in larval

209 organs, with almost no expression in adult tissues, whereas the seven *Grs* in the central part were
210 mainly expressed in adult legs (Figures 4 and S1). These expression patterns were consistent with the
211 phylogenetic tree (Figure 3).

212 Several of these *Grs* showed sex-biased expression. *BmGr10, 16, 28, 29, 37, 38, 45, 49, 58* and
213 *66* were male-biased, whereas *BmGr6, 15, 18, 20, 24, 30-1–30-8, 33–36, 40, 55, 62, 67* and *69* were
214 female-biased. Comparison of *Gr* gene expression between larval legs and moth legs called attention
215 to two sets of genes with complementary patterns (Figure S2). Notably, *BmGr19* (chromosome site
216 unknown) was expressed at an extremely high level in larval thoracic legs, whereas expression of the
217 *BmGr27–31* gene cluster (chr7) was strictly limited to moth legs.

218 Larval mouthparts contain many sensilla located on maxillary galea, maxillary palps and
219 epipharynx, which are the most important chemosensory organs for larval food selection (Ishikawa
220 and Hirao, 1961). In order to understand which genes could be responsible for silkworm larval food
221 selection, we dissected out larval maxillary palps, maxillary galea, epipharynx and thoracic legs
222 separately to perform RNA-seq (Figure S3). More *BmGrs* were expressed in the maxillary palps than
223 maxillary galea or epipharynx. The inositol receptor *BmGr8* (Zhang et al., 2011) was mainly
224 expressed in the maxillary galea, whereas the D-fructose receptor *BmGr9* was expressed highly in
225 maxillary palps. The three candidate CO_2 receptors (*BmGr1, 2* and *3*) were highly expressed in the
226 maxillary palps of both sexes, and sugar receptors were mainly expressed in the maxillary galea and
227 palps; however, bitter receptors could be detected not only in maxilla, but also in epipharynx. Genes
228 highly expressed at the larval stage (*BmGr18* and *BmGr19*) had much higher expression in
229 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 CO_2 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 styloconic 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 styloconic 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

335 McCombie, 2007) and Consed 16.0 (Gordon et al., 1998). The sequence of BAC092J04 (154,052
336 bp) is available in DDBJ under accession number LC056060.

337 *4.3 Chromosomal distribution of chemosensory gene families in the silkworm genome*

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

341 *4.4 Phylogenetic analysis of BmGrs*

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

354

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℃. 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, Na_2HPO_4 4.3mmol/L, KH_2PO_4 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^{o}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 1μg 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), 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/).

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.

397 **References**

398

399 Benton, R., Sachse, S., Michnick, S.W., Vosshall, L.B., 2006. Atypical membrane topology and
400 heteromeric function of *Drosophila* odorant receptors in vivo. PLoS biology 4, e20.
401 Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence
402 data. Bioinformatics 30, 2114-2120.
403 Bray, S., Amrein, H., 2003. A putative *Drosophila* pheromone receptor expressed in male-specific
404 taste neurons is required for efficient courtship. Neuron 39, 1019-1029.
405 Briscoe, A.D., Macias-Munoz, A., Kozak, K.M., Walters, J.R., Yuan, F., Jamie, G.A., Martin, S.H.,
406 Dasmahapatra, K.K., Ferguson, L.C., Mallet, J., Jacquin-Joly, E., Jiggins, C.D., 2013. Female
407 behaviour drives expression and evolution of gustatory receptors in butterflies. PLoS genetics 9,
408 e1003620.
409 Buermans, H.P., den Dunnen, J.T., 2014. Next generation sequencing technology: Advances and
410 applications. Biochimica et biophysica acta 1842, 1932-1941.
411 Cande, J., Prud'homme, B., Gompel, N., 2013. Smells like evolution: the role of chemoreceptor
412 evolution in behavioral change. Current opinion in neurobiology 23, 152-158.
413 de la Bastide, M., McCombie, W.R., 2007. Assembling genomic DNA sequences with PHRAP.
414 Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.] Chapter 11,
415 Unit11 14.
416 Descoins, C., Marion-Poll, F., 1999. Electrophysiological responses of gustatory sensilla of
417 *Mamestra brassicae* (Lepidoptera, Noctuidae) larvae to three ecdysteroids: ecdysone,
418 20-hydroxyecdysone and ponasterone A. Journal of insect physiology 45, 871-876.
419 Dethier, V.G., 1937. Gustation and olfaction in lepidopterous larv ae. Biol Bull-Us 72, 7-23.
420 Duan, J., Li, R., Cheng, D., Fan, W., Zha, X., Cheng, T., Wu, Y., Wang, J., Mita, K., Xiang, Z., Xia,
421 Q., 2010. SilkDB v2.0: a platform for silkworm (*Bombyx mori*) genome biology. Nucleic acids
422 research 38, D453-456.
423 Ehrlich, P.R., Raven, P.H., 1964. Butterflies and Plants - a Study in Coevolution. Evolution 18,
424 586-608.
425 Engsontia, P., Sangket, U., Chotigeat, W., Satasook, C., 2014. Molecular evolution of the odorant and
426 gustatory receptor genes in lepidopteran insects: implications for their adaptation and speciation.
427 Journal of molecular evolution 79, 21-39.
428 Gardiner, A., Barker, D., Butlin, R.K., Jordan, W.C., Ritchie, M.G., 2008. *Drosophila* chemoreceptor
429 gene evolution: selection, specialization and genome size. Molecular ecology 17, 1648-1657.
430 Gordon, D., Abajian, C., Green, P., 1998. Consed: a graphical tool for sequence finishing. Genome
431 research 8, 195-202.
432 Hansson, B.S., Stensmyr, M.C., 2011. Evolution of insect olfaction. Neuron 72, 698-711.
433 Hasegawa, M., Kishino, H., Yano, T., 1985. Dating of the human-ape splitting by a molecular clock
434 of mitochondrial DNA. Journal of molecular evolution 22, 160-174.
435 International Silkworm Genome, C., 2008. The genome of a lepidopteran model insect, the silkworm
436 *Bombyx mori*. Insect biochemistry and molecular biology 38, 1036-1045.
437 Ishikawa, S., 1963. Responses of maxillary receptors in the larva of the silkworm, *Bombyx mori*, to

stimulation by carbohydrates. J. Cell Comp. Physiol. 61, 99–107.
Ishikawa, s., 1966. Electrical response and function of a bitter substance receptor associated with the maxillary sensilla of the larva of the silkworm, *Bombyx mori.* J. Cell Comp. Physiol. 67.
Ishikawa, S., Hirao, T., 1961. Tactile sense organ on the head of silkworm larva (*Bombyx mori*). J. Seric Sci Jpn. 30.
Ishikawa, S., Hirao, T., 1963. Electrophysiological studies of taste sensation in the larvae of the silkworm, *Bombyx mori.* Responsiveness of sensilla styloconica on the maxilla. Bull. Seric. Exp. Sta. Jpn. 19, 297-357.
Ishikawa, S., Hirao, T., Arai, N., 1969. Chemosensory Basis of Hostplant Selection in Silkworm. Entomol Exp Appl 12, 544-&.
Isono, K., Morita, H., 2010. Molecular and cellular designs of insect taste receptor system. Frontiers in cellular neuroscience 4, 20.
Jiang, L., Huang, C.L., Sun, Q., Guo, H.Z., Peng, Z.W., Dang, Y.H., Liu, W.Q., Xing, D.X., Xu, G.W., Zhao, P., Xia, Q.Y., 2015. Overexpression of host plant urease in transgenic silkworms. Molecular Genetics and Genomics 290, 1117-1123.
Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular biology and evolution 30, 772-780.
Kikuta, S., Endo, H., Tomita, N., Takada, T., Morita, C., Asaoka, K., Sato, R., 2016. Characterization of a ligand-gated cation channel based on an inositol receptor in the silkworm, *Bombyx mori*. Insect biochemistry and molecular biology 74, 12-20.
Koenig, C., Hirsh, A., Bucks, S., Klinner, C., Vogel, H., Shukla, A., Mansfield, J.H., Morton, B., Hansson, B.S., Grosse-Wilde, E., 2015. A reference gene set for chemosensory receptor genes of *Manduca sexta*. Insect biochemistry and molecular biology 66, 51-63.
Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nature methods 9, 357-359.
Lavagnino, N., Serra, F., Arbiza, L., Dopazo, H., Hasson, E., 2012. Evolutionary Genomics of Genes Involved in Olfactory Behavior in the *Drosophila melanogaster* Species Group. Evolutionary bioinformatics online 8, 89-104.
Le, S.Q., Gascuel, O., 2008. An improved general amino acid replacement matrix. Molecular biology and evolution 25, 1307-1320.
Legeai, F., Malpel, S., Montagne, N., Monsempes, C., Cousserans, F., Merlin, C., Francois, M.C., Maibeche-Coisne, M., Gavory, F., Poulain, J., Jacquin-Joly, E., 2011. An Expressed Sequence Tag collection from the male antennae of the Noctuid moth *Spodoptera littoralis*: a resource for olfactory and pheromone detection research. BMC genomics 12, 86.
Li, B., Dewey, C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC bioinformatics 12, 323.
Ling, F., Dahanukar, A., Weiss, L.A., Kwon, J.Y., Carlson, J.R., 2014. The molecular and cellular basis of taste coding in the legs of *Drosophila*. The Journal of neuroscience : the official journal of the Society for Neuroscience 34, 7148-7164.
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,

85-98.
McBride, C.S., Arguello, J.R., O'Meara, B.C., 2007. Five *Drosophila* genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. Genetics 177, 1395-1416.
Mita, K., Kasahara, M., Sasaki, S., Nagayasu, Y., Yamada, T., Kanamori, H., Namiki, N., Kitagawa, M., Yamashita, H., Yasukochi, Y., Kadono-Okuda, K., Yamamoto, K., Ajimura, M., Ravikumar, G., Shimomura, M., Nagamura, Y., Shin, I.T., Abe, H., Shimada, T., Morishita, S., Sasaki, T., 2004. The genome sequence of silkworm, *Bombyx mori*. DNA research : an international journal for rapid publication of reports on genes and genomes 11, 27-35.
Montagne, N., de Fouchier, A., Newcomb, R.D., Jacquin-Joly, E., 2015. Advances in the identification and characterization of olfactory receptors in insects. Progress in molecular biology and translational science 130, 55-80.
Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., Wold, B., 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature methods 5, 621-628.
Ono, T., 1977. The Scales as a Releaser of the Copulation Attempt in Lepidoptera. Namrwissenschaften 1997, 386-387.
Sanderson, M.J., Wojciechowski, M.F., 2000. Improved bootstrap confidence limits in large-scale phylogenies, with an example from *Neo-Astragalus* (Leguminosae). Systematic biology 49, 671-685.
Sato, K., Pellegrino, M., Nakagawa, T., Nakagawa, T., Vosshall, L.B., Touhara, K., 2008. Insect olfactory receptors are heteromeric ligand-gated ion channels. Nature 452, 1002-1006.
Sato, K., Tanaka, K., Touhara, K., 2011. Sugar-regulated cation channel formed by an insect gustatory receptor. Proceedings of the National Academy of Sciences of the United States of America 108, 11680-11685.
Schoonhoven.LM, 1969. Gustation and Foodplant Selection in Some Lepidopterous Larvae. Entomol Exp Appl 12, 555-&.
Scott, K., Brady, R., Jr., Cravchik, A., Morozov, P., Rzhetsky, A., Zuker, C., Axel, R., 2001. A chemosensory gene family encoding candidate gustatory and olfactory receptors in *Drosophila*. Cell 104, 661-673.
Shields, V.D., 2009. Fine structure of the galeal styloconic sensilla of larval *Lymantria dispar* (Lepidoptera: Lymantriidae). Annals of the Entomological Society of America 102, 1116-1125.
Shimomura, M., Minami, H., Suetsugu, Y., Ohyanagi, H., Satoh, C., Antonio, B., Nagamura, Y., Kadono-Okuda, K., Kajiwara, H., Sezutsu, H., Nagaraju, J., Goldsmith, M.R., Xia, Q., Yamamoto, K., Mita, K., 2009. KAIKObase: an integrated silkworm genome database and data mining tool. BMC genomics 10, 486.
Smart, R., Kiely, A., Beale, M., Vargas, E., Carraher, C., Kralicek, A.V., Christie, D.L., Chen, C., Newcomb, R.D., Warr, C.G., 2008. *Drosophila* odorant receptors are novel seven transmembrane domain proteins that can signal independently of heterotrimeric G proteins. Insect biochemistry and molecular biology 38, 770-780.
Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312-1313.
Stamatakis, A., Hoover, P., Rougemont, J., 2008. A rapid bootstrap algorithm for the RAxML Web servers. Systematic biology 57, 758-771.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Molecular biology and evolution 30, 2725-2729.

Tanaka, Y., Asaoka, K., Takeda, S., 1994. Different Feeding and Gustatory Responses to Ecdysone and 20-Hydroxyecdysone by Larvae of the Silkworm, *Bombyx-Mori*. Journal of chemical ecology 20, 125-133.

Vieira, F.G., Rozas, J., 2011. Comparative genomics of the odorant-binding and chemosensory protein gene families across the Arthropoda: origin and evolutionary history of the chemosensory system. Genome biology and evolution 3, 476-490.

Wanner, K.W., Robertson, H.M., 2008. The gustatory receptor family in the silkworm moth *Bombyx mori* is characterized by a large expansion of a single lineage of putative bitter receptors. Insect molecular biology 17, 621-629.

Wicher, D., Schafer, R., Bauernfeind, R., Stensmyr, M.C., Heller, R., Heinemann, S.H., Hansson, B.S., 2008. *Drosophila* odorant receptors are both ligand-gated and cyclic-nucleotide-activated cation channels. Nature 452, 1007-1011.

Xia, Q., Zhou, Z., Lu, C., Cheng, D., Dai, F., Li, B., Zhao, P., Zha, X., Cheng, T., Chai, C., Pan, G., Xu, J., Liu, C., Lin, Y., Qian, J., Hou, Y., Wu, Z., Li, G., Pan, M., Li, C., Shen, Y., Lan, X., Yuan, L., Li, T., Xu, H., Yang, G., Wan, Y., Zhu, Y., Yu, M., Shen, W., Wu, D., Xiang, Z., Yu, J., Wang, J., Li, R., Shi, J., Li, H., Li, G., Su, J., Wang, X., Li, G., Zhang, Z., Wu, Q., Li, J., Zhang, Q., Wei, N., Xu, J., Sun, H., Dong, L., Liu, D., Zhao, S., Zhao, X., Meng, Q., Lan, F., Huang, X., Li, Y., Fang, L., Li, C., Li, D., Sun, Y., Zhang, Z., Yang, Z., Huang, Y., Xi, Y., Qi, Q., He, D., Huang, H., Zhang, X., Wang, Z., Li, W., Cao, Y., Yu, Y., Yu, H., Li, J., Ye, J., Chen, H., Zhou, Y., Liu, B., Wang, J., Ye, J., Ji, H., Li, S., Ni, P., Zhang, J., Zhang, Y., Zheng, H., Mao, B., Wang, W., Ye, C., Li, S., Wang, J., Wong, G.K., Yang, H., Biology Analysis, G., 2004. A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). Science 306, 1937-1940.

Xu, W., Papanicolaou, A., Zhang, H.J., Anderson, A., 2016. Expansion of a bitter taste receptor family in a polyphagous insect herbivore. Scientific reports 6, 23666.

Yang, Z., 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. Journal of molecular evolution 39, 306-314.

You, M., Yue, Z., He, W., Yang, X., Yang, G., Xie, M., Zhan, D., Baxter, S.W., Vasseur, L., Gurr, G.M., Douglas, C.J., Bai, J., Wang, P., Cui, K., Huang, S., Li, X., Zhou, Q., Wu, Z., Chen, Q., Liu, C., Wang, B., Li, X., Xu, X., Lu, C., Hu, M., Davey, J.W., Smith, S.M., Chen, M., Xia, X., Tang, W., Ke, F., Zheng, D., Hu, Y., Song, F., You, Y., Ma, X., Peng, L., Zheng, Y., Liang, Y., Chen, Y., Yu, L., Zhang, Y., Liu, Y., Li, G., Fang, L., Li, J., Zhou, X., Luo, Y., Gou, C., Wang, J., Wang, J., Yang, H., Wang, J., 2013. A heterozygous moth genome provides insights into herbivory and detoxification. Nature genetics 45, 220-225.

Zhan, S., Merlin, C., Boore, J.L., Reppert, S.M., 2011. The monarch butterfly genome yields insights into long-distance migration. Cell 147, 1171-1185.

Zhang, H.J., Anderson, A.R., Trowell, S.C., Luo, A.R., Xiang, Z.H., Xia, Q.Y., 2011. Topological and functional characterization of an insect gustatory receptor. PloS one 6, e24111.

Zhang L.W., Kang K., Jiang S.C., Zhang Y.N., Wang T.T., Zhang J., Sun L., Yang Y.Q., Huang C.C., Jiang L.Y., Ding D.G. 2016. Analysis of the antennal transcriptome and insights into olfactory genes in *Hyphantria cunea* (Drury). PloS one 11, e0164729

564 **Figure legends**

565 **Figure 1. Detailed gene distribution of *BmGrs* in the Chromosome 7 gene cluster.** BAC clone
566 092J04 [chr7:3,376,776…3,551,821] was shown to encompass a gap found in the organization of
567 *BmGr* genes. Sequencing of this BAC revealed 7 additional copies of *BmGr30* tandemly aligned in
568 this gap. The 154,052 bp sequence of BAC092J04 is available in DDBJ under accession number
569 LC056060. Red arrows indicate direction of gene transcription.

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

572 **Figure 3. Evolutionary relationships of the *Grs* identified in three lepidopteran insect genomes.**
573 Amino acid sequences were automatically aligned by the Mafft program version 7, using the E-INS-i
574 strategy. The evolutionary history was inferred using a maximum likelihood tree with RaxML
575 version 8 using the LG+Gamma+I model. Model selection was conducted by Mega version 6 and
576 LG+Gamma+I mode. Bootstrap support was 1000 replicates. Putative CO_2 and sugar gustatory
577 receptors show conserved relationships among the three Lepidoptera species, while the remaining
578 bitter receptors are more divergent. Bm, *Bombyx mori*; Hm, *Heliconius melpomene*; Ms, *Manduca*
579 *sexta*.

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

581 **Figure 5. Expression map of *BmGr* genes in larval and adult chemosensory organs.** Colored dots
582 on larval and adult chemosensory organs correspond to positions of high expression of *BmGr* genes
583 clustered in chr 7 and chr 13. Other genes mentioned in the text are listed near tissues where they are
584 also highly expressed. Bar graphs are based on the FPKM values of the genes which are shown on

585 the Y axis. The X axis stands for tissues as follows: white, LA; black, LM; horizontal stripes, LTL;
586 vertical stripes, MA; and grey, ML. LA, larval antenna; LM, larval maxilla; LTL, larval thoracic legs;
587 MA, moth antenna; ML, moth legs.

588

589 **Supplementary information**

590 Supplementary Material

591 Document S1

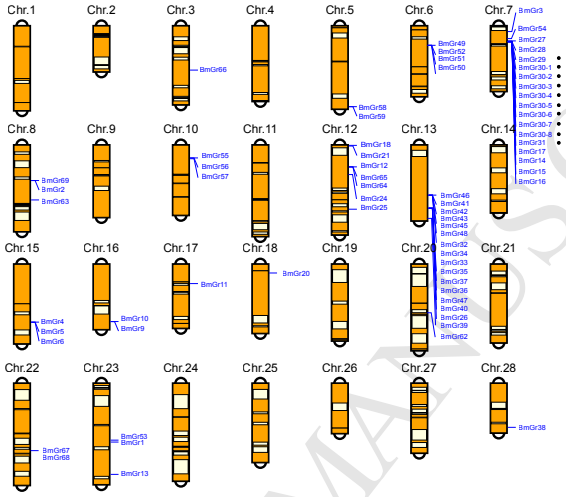
592 Document S2

593 Document S3

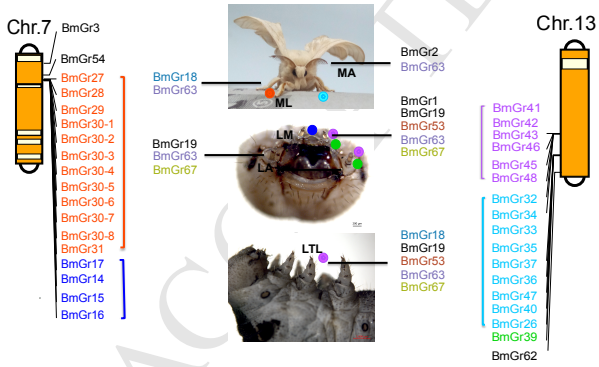
594 Table S1

Graphical abstract

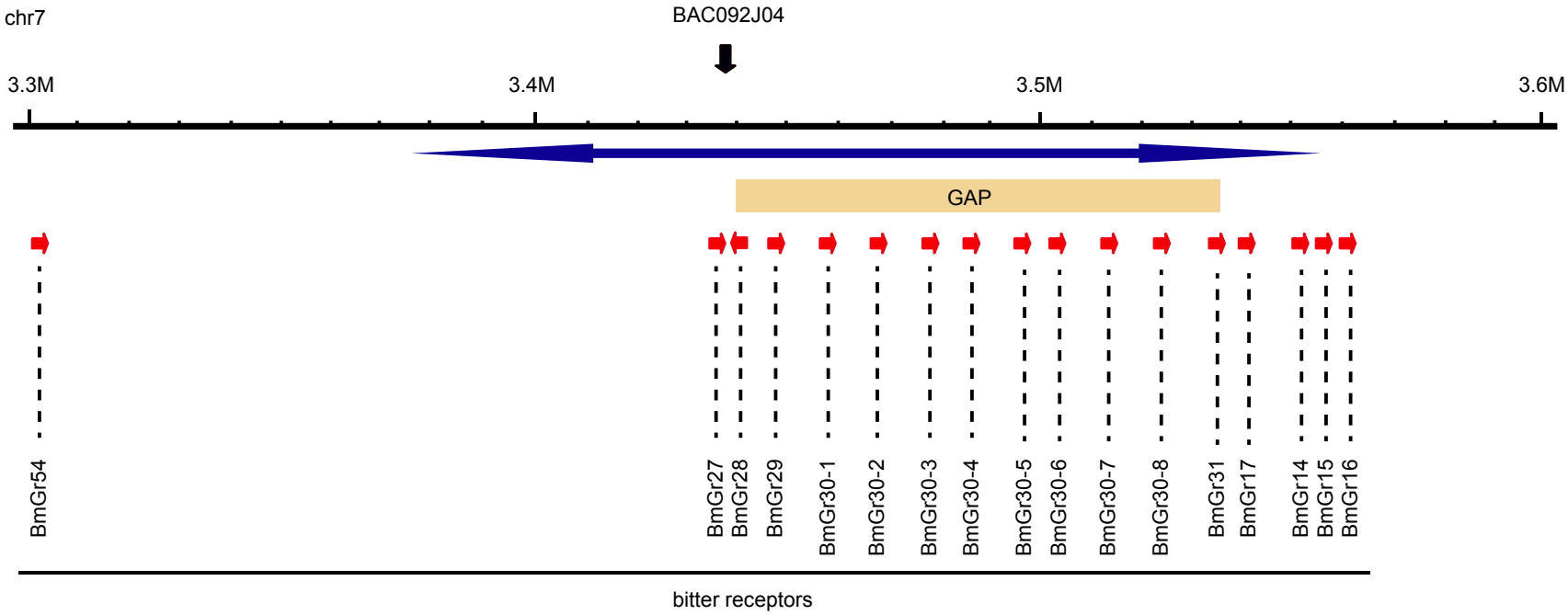
By manual annotation and BAC sequencing, we improved 43 gene sequences compared with previously reported *Gr*s in the most studied Lepidoptera model, the silkworm *Bombyx mori*, and identified 7 new tandem copies of *BmGr30* on chromosome 7, bringing the total number of *BmGrs* to 76.

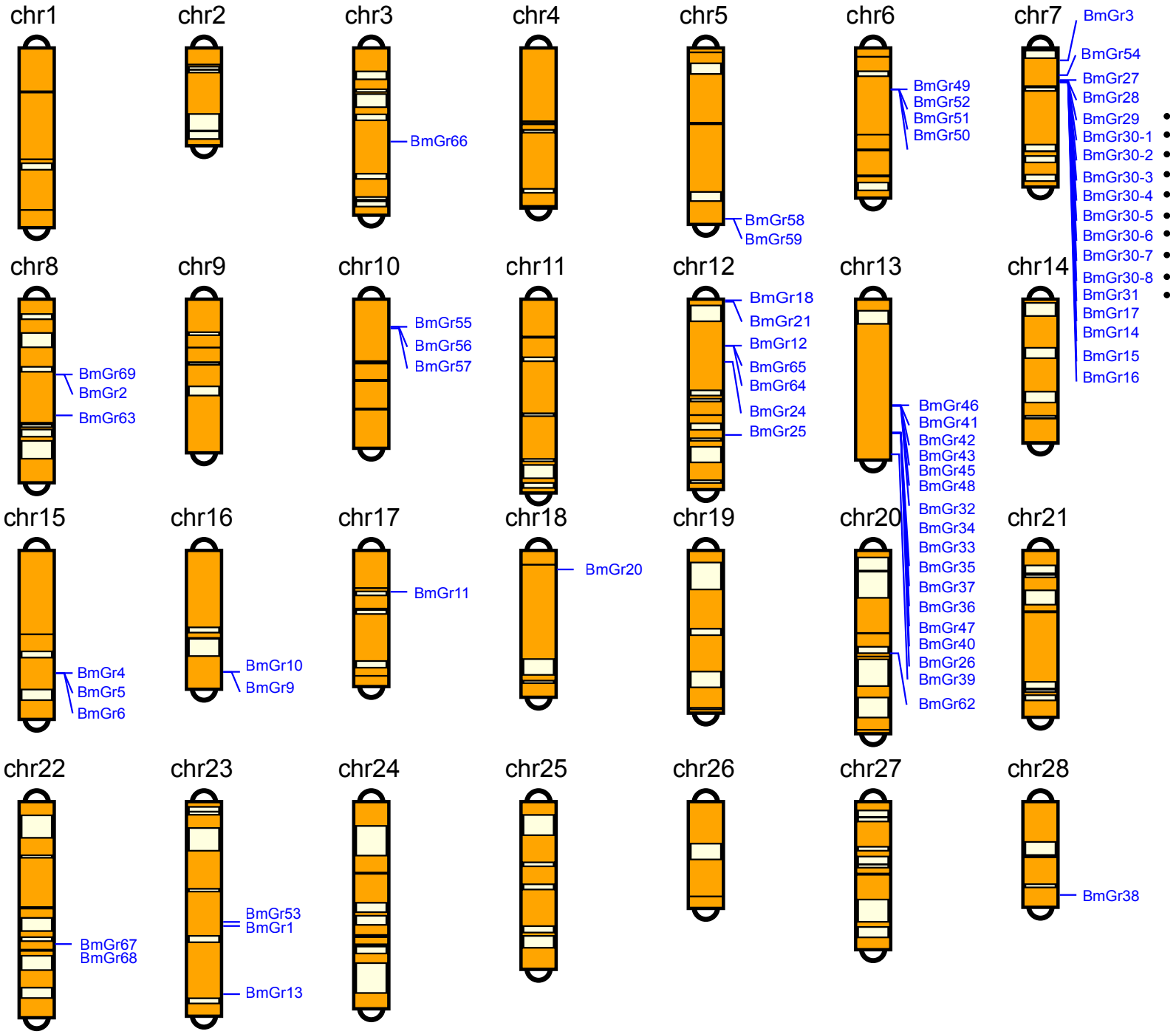


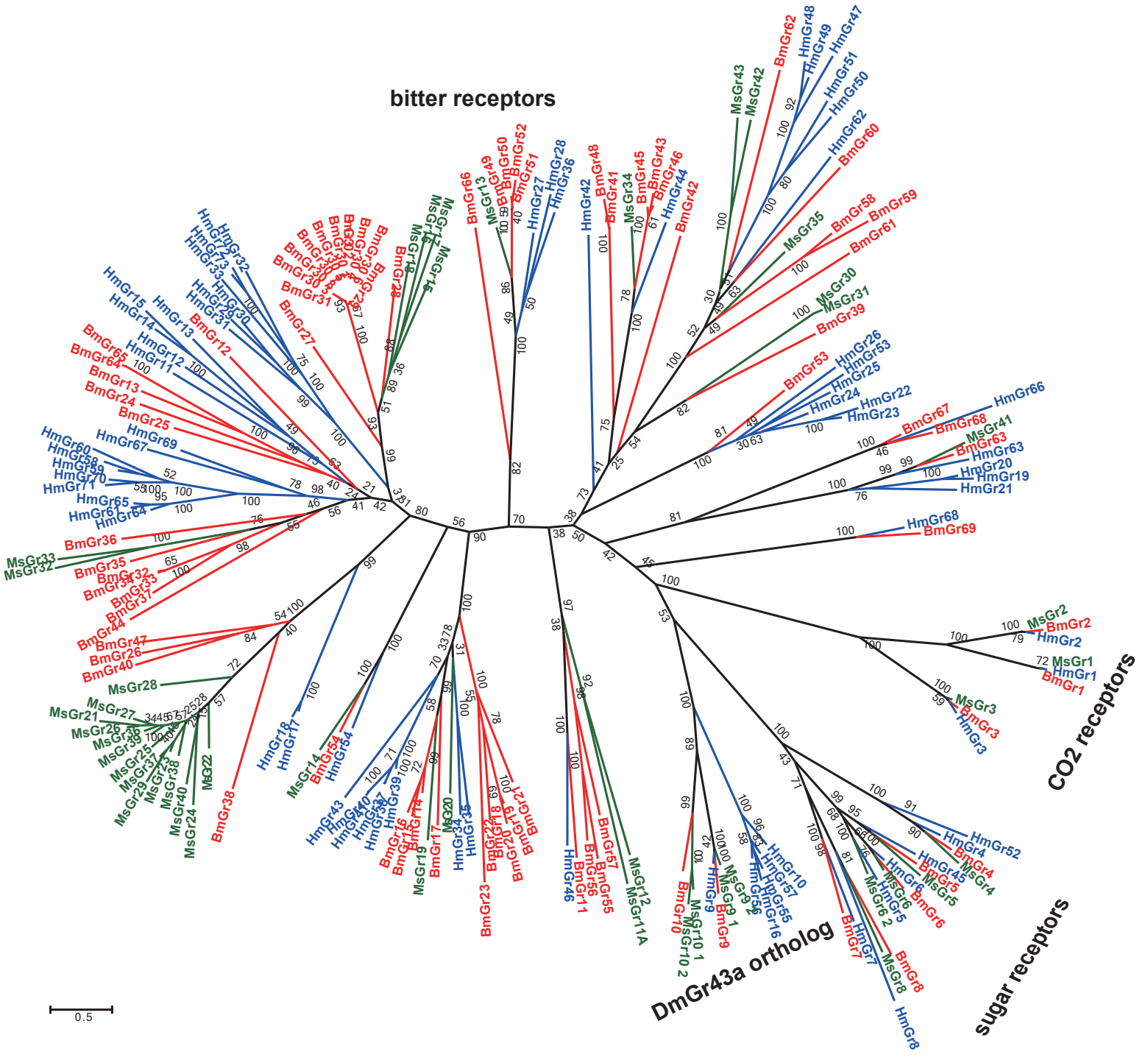
RNA-seq analysis of diverse *BmGr* expression patterns in chemosensory organs of larvae and adults enabled us to draw a precise organ specific map of *BmGr* expression.

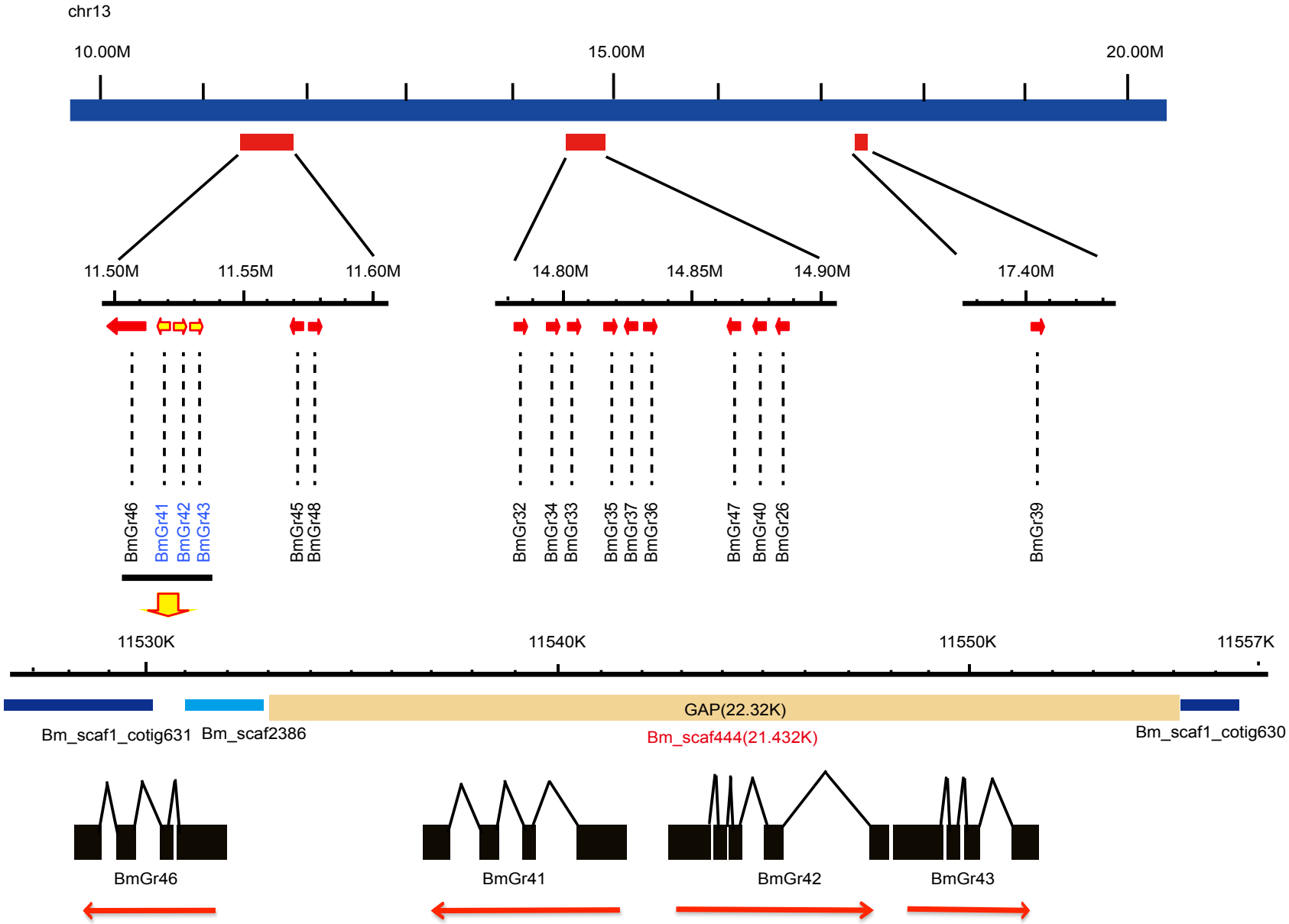


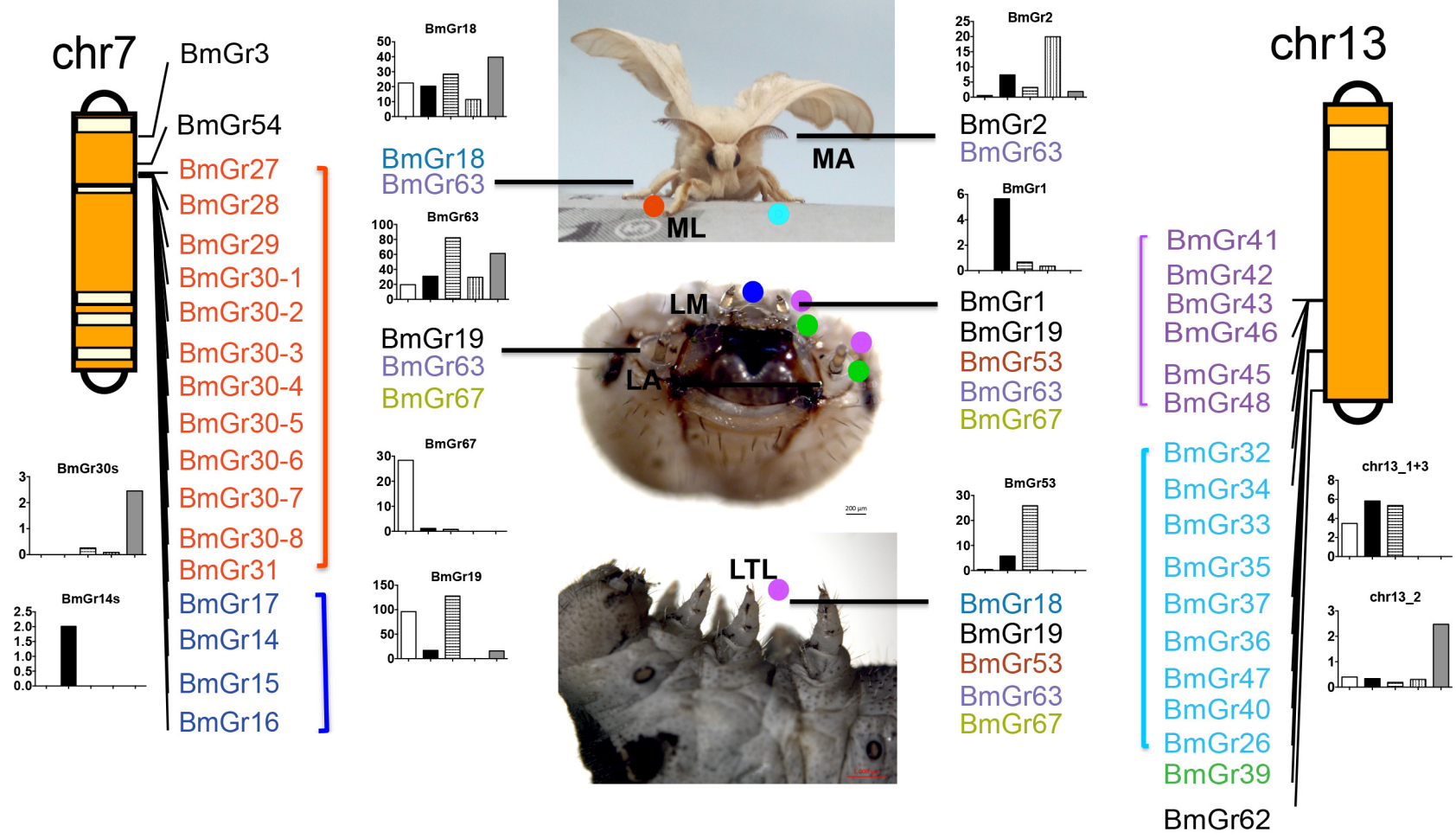
Body map of *Gr* expression provides a valuable new reference for finding the genes involved in plant-insect interactions in Lepidoptera and establishing correlations between these genes and vital insect behaviors like host plant selection and courtship for mating.


chr7
BAC092J04
3.3M
3.4M
3.5M
3.6M
GAP
BmGr54
BmGr27
BmGr28
BmGr29
BmGr30-1
BmGr30-2
BmGr30-3
BmGr30-4
BmGr30-5
BmGr30-6
BmGr30-7
BmGr30-8
BmGr31
BmGr17
BmGr14
BmGr15
BmGr16
bitter receptors


chr1
chr2
chr3
BmGr66
chr4
chr5
BmGr58
BmGr59
chr6
BmGr49
BmGr52
BmGr51
BmGr50
chr7
BmGr3
BmGr54
BmGr27
BmGr28
BmGr29
BmGr30-1
BmGr30-2
BmGr30-3
BmGr30-4
BmGr30-5
BmGr30-6
BmGr30-7
BmGr30-8
BmGr31
BmGr17
BmGr14
BmGr15
BmGr16
chr8
BmGr69
BmGr2
BmGr63
chr9
chr10
BmGr55
BmGr56
BmGr57
chr11
chr12
BmGr18
BmGr21
BmGr12
BmGr65
BmGr64
BmGr24
BmGr25
chr13
BmGr46
BmGr41
BmGr42
BmGr43
BmGr45
BmGr48
BmGr32
BmGr34
BmGr33
BmGr35
BmGr37
BmGr36
BmGr47
BmGr40
BmGr26
BmGr39
BmGr62
chr14
chr15
BmGr4
BmGr5
BmGr6
chr16
BmGr10
BmGr9
chr17
BmGr11
chr18
BmGr20
chr19
chr20
chr21
chr22
BmGr67
BmGr68
chr23
BmGr53
BmGr1
BmGr13
chr24
chr25
chr26
chr27
chr28
BmGr38


bitter receptors
CO2 receptors
DmGr43a ortholog
sugar receptors
0.5


chr13
10.00M
15.00M
20.00M
11.50M
11.55M
11.60M
14.80M
14.85M
14.90M
17.40M
BmGr46
BmGr41
BmGr42
BmGr43
BmGr45
BmGr48
BmGr32
BmGr34
BmGr33
BmGr35
BmGr37
BmGr36
BmGr47
BmGr40
BmGr26
BmGr39
11530K
11540K
11550K
11557K
GAP(22.32K)
Bm_scaf1_cotig631
Bm_scaf2386
Bm_scaf444(21.432K)
Bm_scaf1_cotig630
BmGr46
BmGr41
BmGr42
BmGr43


chr7
BmGr3
BmGr54
BmGr27
BmGr28
BmGr29
BmGr30-1
BmGr30-2
BmGr30-3
BmGr30-4
BmGr30-5
BmGr30-6
BmGr30-7
BmGr30-8
BmGr31
BmGr17
BmGr14
BmGr15
BmGr16
BmGr30s
BmGr14s
BmGr18
BmGr63
BmGr19
BmGr63
BmGr67
BmGr18
BmGr63
BmGr67
BmGr19
MA
ML
LM
LA
LTL
BmGr2
BmGr63
BmGr1
BmGr19
BmGr53
BmGr63
BmGr67
BmGr53
BmGr18
BmGr19
BmGr53
BmGr63
BmGr67
BmGr2
BmGr1
BmGr53
chr13
BmGr41
BmGr42
BmGr43
BmGr46
BmGr45
BmGr48
BmGr32
BmGr34
BmGr33
BmGr35
BmGr37
BmGr36
BmGr47
BmGr40
BmGr26
BmGr39
BmGr62
chr13_1+3
chr13_2

Highlights

1) By manual annotation and BAC sequencing, we improved 43 gene sequences compared with previously reported *Gr*s 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.