Molecular Characterization of Sugar Taste Receptors in Cotton Bollworm *Helicoverpa armigera*

Journal:	<i>Genome</i>
Manuscript ID	gen-2017-0086.R1
Manuscript Type:	Article
Date Submitted by the Author:	06-Jul-2017
Complete List of Authors:	Xu, Wei; Murdoch University, School of Veterinary and Life Sciences Liu, Nai-Yong; Southwest Forestry University Liao, Yalin; Murdoch University Anderson, Alisha; CSIRO Ecosystem Sciences
Is the invited manuscript for consideration in a Special Issue? :	This submission is not invited
Keyword:	Gustatory receptor, sugar receptor, <i>Helicoverpa armigera</i> , gene expression

SCHOLARONE™
Manuscripts

1 **Molecular Characterization of Sugar Taste Receptors in Cotton**
2 **Bollworm *Helicoverpa armigera***

3

4 Wei Xu^{1*}, Naiyong Liu², Yalin Liao¹, Alisha Anderson³

5 WX: W.Xu@murdoch.edu.au

6 NYL: Naiyong_2013@163.com

7 YL: Y.Liao@murdoch.edu.au

8 AA: Alisha.Anderson@csiro.au

9

10

11 ¹School of Veterinary and Life Sciences, Murdoch University, Murdoch, WA 6150,
12 Australia

13 ²Key Laboratory of Forest Disaster Warning and Control of Yunnan Province, Southwest
14 Forestry University, Kunming, 650224, China.

15 ³CSIRO Ecosystem Sciences, Black Mountain, ACT 2601, Australia

16

17

18 *Corresponding author: Dr. Wei Xu, Email: W.Xu@Murdoch.edu.au.

19

20

21

Abstract

Insects utilize sugars as their essential energy and nutrient sources; therefore, the sense of sugar detection plays a critical role in insect behaviours. Previously using genomic and transcriptomic approaches, we identified eight putative sugar gustatory receptor (GR) genes from the cotton bollworm *Helicoverpa armigera* (Hübner) (Lepidoptera: Noctuidae). Here we further validated these annotated sugar receptor genes (*HarmGr4-8* and *10-12*) and found *HarmGr10* may be a pseudogene carrying a stop codon in the open reading frame (ORF). Sequence alignment revealed *H. armigera* sugar GR sequences are conserved at C-terminus and phylogenetic analysis showed that insect sugar GRs have evolved in a family-specific manner. Interestingly, all eight *H. armigera* sugar GRs are localized in a tandem array on the same scaffold of the genome. *In silico* gene expression and reverse transcription (RT)-PCR analysis showed that *HarmGr10* is specifically expressed in male adult testes while *HarmGr11* is specifically expressed in female adult ovaries, suggesting *H. armigera* sugar GRs may be involved in reproduction-related functions. This study improves our knowledge on insect sugar receptors and gustatory systems.

Key words: Gustatory receptor, sugar receptor, *Helicoverpa armigera*, gene expression

41 Introduction

42 Sugars play critical roles in insect life as valuable energy and food resources. For example,
 43 bees collect nectar from blossoms in the field and convert them into honey, which contains a
 44 high level of sugars (Pham-Delegue et al. 1990); mosquitoes utilize sugars from flowers and
 45 plants as an energy source (Foster 1995); aphids feed on plant sap and produce sugar-rich
 46 honeydew, attracting ants to collect and milk (Perkins et al. 2013); populations of the German
 47 cockroach have evolved an adaptive behavioural aversion to glucose, a phagostimulant
 48 component of toxic baits (Wada-Katsumata et al. 2013). The perception of sugars is also
 49 utilized by insects to assess and evaluate the nutritional value of foods (Kent and Robertson
 50 2009; Slone et al. 2007). Therefore, insects require sensitive, accurate and robust sensory
 51 systems to detect sugars and regulate these diverse behaviours across different insect species
 52 (Kent and Robertson 2009; Slone et al. 2007).

53 Sugar detection primarily occurs at taste sensilla distributed throughout the insect body,
 54 especially on the major chemosensory tissues such as antennae, mouthparts, tarsi and
 55 ovipositors (van der Goes van Naters and Carlson 2006). Gustatory receptors (GRs) are
 56 localized on the dendrites inside the gustatory sensilla and act as an interface between insects
 57 and their environmental taste chemicals (van der Goes van Naters and Carlson 2006). GR
 58 genes were firstly identified from the *Drosophila melanogaster* genome using sequence
 59 similarity and phylogenetic clustering (Clyne et al. 2000). Insect GRs have been classified
 60 into the following 4 clades: “GR43a-like” (Abdel-Latif 2007; Sato et al. 2011; Xu et al.
 61 2012), “CO₂” (Jones et al. 2007; Kirkness et al. 2010; Ning et al. 2016; Xu and Anderson
 62 2015), “bitter” (Xu et al. 2016) and “sugar” (Chyb et al. 2003; Dahanukar et al. 2001;
 63 Dahanukar et al. 2007; Jiao et al. 2007; Jiao et al. 2008a; Slone et al. 2007; Xu et al. 2015;
 64 Xu et al. 2016; Zhang et al. 2011). Although the “GR43a-like” receptors have been shown to
 65 have active responses to sugars e.g. fructose, they are not the “sugar” receptors we discuss

here (Sato et al. 2011; Xu et al. 2012). Eight GR genes have been mapped to *Drosophila* sweet taste neurons, defining a distinct clade of sugar receptors (Dahanukar et al. 2007; Jiao et al. 2007; Jiao et al. 2008b; Slone et al. 2007), whose orthologous genes are found across various insect species (Kent and Robertson 2009). Recently published genome and transcriptome sequences from various lepidopteran insect species like *Danaus plexippus*, *Heliconius melpomene*, *Helicoverpa armigera* and *Helicoverpa assulta*, revealed that putative sugar gustatory receptors are conserved among many insect species (Briscoe et al. 2013; Engsontia et al. 2014; Xu et al. 2016). Insect GRs were originally considered as a large group of G protein-coupled receptors (GPCRs) (Slone et al. 2007), however, subsequent *in vitro* studies have shown that insect GRs have an inverted GPCR transmembrane structure *in vitro* (Zhang et al. 2011) with an intracellular N-terminus and an extracellular C-terminus, similar to insect odorant receptors (ORs).

Currently most studies on insect sugar receptors have been primarily focused on *Drosophila* (Bredendiek et al. 2011; Chyb et al. 2003; Dahanukar et al. 2001; Dahanukar et al. 2007; Freeman et al. 2014; Jiao et al. 2007; Jiao et al. 2008b; Mishra et al. 2013; Robertson et al. 2003; Slone et al. 2007). Much less attention has been paid to the sugar receptors of lepidopteran insects, which consists a large group of herbivore insect species as well as serious agricultural pests. Five sugar GRs were firstly identified from the model insect of Lepidoptera, *Bombyx mori* (Wanner and Robertson 2008). Since then, a number of lepidopteran sugar GRs were identified and studied (Pitts et al. 2014; Zhang et al. 2011). *H. armigera* (Hübner) (Lepidoptera: Noctuidae), one of the most destructive agricultural pests, feed on a large number of cultivated crops including cotton, tomato, soybeans and corn. In our previous studies, we identified eight candidate sugar GRs from the *H. armigera* genome and transcriptome (Liu et al. 2014; Xu et al. 2016). Here we further performed gene cloning, sequence validation, alignment, phylogenetic, gene structural, expression analyses and *in*

91 *silico* expression profiles of these sugar receptors. This study improves our understanding of
92 insect sugar receptor family, insect-plant interaction and evolution.

Draft

93 **Materials and methods**

94 **Insect rearing, tissue collection and RNA purification**

95 *H. armigera* pupae were kindly provided by Professor Myron Zalucki of University of
96 Queensland (Perkins et al. 2013) and reared at 26 °C and 70% humidity. Ten testes, ten
97 ovaries, twenty male antennae, twenty female antennae, twenty male tarsi and twenty female
98 tarsi were collected from adult aged 1-3 days. All collected tissues were immediately stored
99 in *RNAlater* (Invitrogen, USA). Total RNA was purified using Qiagen RNeasy mini kit
100 (Qiagen, USA) kits according to the manufacturer's protocol. The purified RNA was treated
101 with *DNaseI* (Ambion, USA) at 37 °C for 30 min, then quantified and qualified using
102 NanoDrop ND-2000 (Thermo Scientific, USA).

104 ***RT-PCR and gene cloning***

105 The cDNA templates were prepared from purified RNA samples using the SuperScript®
106 VILO™ cDNA Synthesis Kit (Invitrogen, USA), according to the manufacturer's manual.
107 RT-PCR was performed using gene-specific primers (Table 2) for the full length *H. armigera*
108 sugar GR ORF sequences (Xu et al. 2016). The PCR program was performed by using
109 Phusion High-Fidelity DNA Polymerase (New England BioLabs, USA) as follows: 95 °C for
110 3 min; 50 cycles at 95 °C for 30 s, 60 °C for 30 s and 72 °C for 90 s; and final extension at
111 72 °C for 10 min. PCR products were purified using QIAquick gel extraction reagents
112 (Qiagen, USA), cloned into the pBluescript SKII (-) vector (Stratagene, USA) and
113 transformed to competent *Escherichia coli* cells DH5α (New England BioLabs, USA). Three
114 randomly selected colonies were picked for plasmid purification and DNA sequencing. The
115 sequencing results were compared to previously annotated sequences and exon/intron
116 analysis (<http://wormweb.org/exonintron>). Multiple transmembrane domains of receptors

were predicted by TMpred (http://www.ch.embnet.org/software/TMPRED_form.html) (Hofmann and Stoffel 1993) HMMTOP (<http://www.enzim.hu/hmmtop/>) (Tusnady and Simon 1998) and TMHMM (<http://www.cbs.dtu.dk/services/TMHMM/>) (Krogh et al. 2001).

Sugar GR sequence alignment and phylogenetic analysis

Sequence alignment analysis of *H. armigera* sugar GRs was performed by using multiple alignment program ClustalW (Cost matrix, BLOSUM) implemented in Geneious software 8.0.5 (<http://www.geneious.com>) (Kearse et al. 2012). Then HarmGR4-8 and 10-12 sequences were used for phylogenetic analysis with other known insect sugar receptors from *D. melanogaster*, *Aedes aegypti*, *Culex quinquefasciatus*, *Anopheles. gambiae*, *B. mori*, *D. plexippus*, *H. melpomene*, *Tribolium castaneum*, *Zootermopsis nevadensis*, *Pogonomyrmex barbatus*, *Apis mellifera*, *Nasonia vitripennis*, *Plutella xylostella* and *Acyrtosiphon pisum* (Arensburger et al. 2010; Briscoe et al. 2013; Kent and Robertson 2009; Kent et al. 2008; Kirkness et al. 2010; Robertson et al. 2010; Robertson and Wanner 2006; Wang et al. 2014; Wanner and Robertson 2008; Xu et al. 2016; Zhang et al. 2011). A Neighbour-Joining tree was built using the default settings based on Jukes-Cantor Model with 1000 bootstrap replications by Geneious Tree Builder (Geneious 8.0.5).

Gene expression profile

In our previous study, total RNA samples from antennae, mouthparts, epidermis, fat body, foreguts, midguts, hindguts, Malpighian tubules, haemocytes, hearts, trachea, ventral nerve, silk glands, salivary glands and muscle of 5th instar larvae; male antennae, female antennae, male heads (with antennae), female heads (with antennae), male tarsi, female tarsi, male thorax, female thorax, male abdomens (with testes), female abdomens (with ovaries), male

141 testes and female ovaries from day 0 to day 5 adults; embryo, 3rd instar larvae (whole body),
142 post-feeding stage larvae (whole body) and pupae were sequenced and the sequenced data
143 were used here for expression studies (Xu et al. 2016).

144 *In silico* expression profiles were generated using DEW (<http://dew.sourceforge.net/>)
145 which is an automated pipeline that: 1) used Bowtie2 (Xu et al. 2016) to align the reads of
146 each library against a sequence file comprised of our assembly and the chemosensory genes;
147 2) post-processed the alignments with eXpress to account for isoforms and paralogues; 3)
148 performed Trimmed Mean Normalization using edgeR and estimated Fragments Per Kilobase
149 Per Million reads (FPKM) (Xu et al. 2016); 4) and visualized the results as described (Xu et
150 al. 2016).

151

Draft

Results and discussion

Sugar GRs in various insect species

First, we summarized the number of insect sugar GRs identified from selected insect species which have available genome sequences (Fig. 1). From *H. armigera*, eight sugar GR genes were identified from the genome (Xu et al. 2016) and transcriptome (Liu et al. 2014), more than *B. mori* (five sugar GRs), *H. melpomene* (seven sugar GRs) and *P. xylostella* (five sugar GRs) (Xu et al. 2016) but less than *D. plexippus* (nine sugar GRs) (Briscoe et al. 2013). Sixteen orthologous sugar GR genes were identified from *T. castaneum*, which is the highest number in all insect species studied until now. Six of *T. castaneum* sugar GRs were found to be localized in the lacinia of the *Tribolium* larvae (Abdel-Latief 2007). In *D. melanogaster* and malaria mosquito *A. gambiae*, eight sugar GRs were identified respectively (Kent and Robertson 2009). Ten sugar GRs were identified from *A. aegypti* (Kent et al. 2008) while fourteen sugar GRs were identified from *C. quinquefasciatus* (Arensburger et al. 2010). In both pea aphid (*A. pisum*) and termite (*Z. nevadensis*), six sugar GRs were detected. Two orthologous sugar GR genes have been identified from honeybee (*A. mellifera*), ant (*P. barbatus*) and wasp (*N. vitripennis*) (Fig. 1). The low number of sugar GRs in these species may be attributed to required nurturing life in the hive. However, in locust (*Locusta migratoria*) (Wang et al. 2014) and human louse (*Pediculus humanus*) (Kirkness et al. 2010) (Fig. 1), not a single orthologous sugar GR gene was detected, suggesting they may use other mechanisms for sugar sensation. For example, ionotropic receptors (IRs) are a recently characterized family of insect chemosensory receptors and some of them show specific expression in insect taste neurons (Koh et al. 2014). Previously five *H. assulta* sugar GRs were identified by using transcriptomic analysis (Xu et al. 2015). We also successfully obtained eight full-length ortholog sugar GR genes from *Helicoverpa zea* genome assembly and named HzeaGR4-8 and HzeaGR10-12 (Supplementary data), which show high (95-99%)

identities to their ortholog genes in *H. armigera* (Supplementary Table 1). *H. assulta*, *H. zea* and *H. armigera* are very close species with different host plants. *H. assulta* is a specialist insect with larvae that mainly feed on the plants of the Solanaceae whereas *H. armigera* is a generalist pest and its larvae can feed on over 200 different plants. *H. zea* hosts have been recorded from over 123 species in 29 families, more than *H. assulta* but less than *H. armigera* (Cunningham and Zalucki 2014).

Sequence cloning, alignment and phylogenetic analysis

We reanalysed the annotated *H. armigera* sugar GR genes, designed primers, cloned these *Gr* genes and performed sequence alignment (Fig. 2). Compared to the sequences annotated using genome/transcriptome data previously (Xu et al. 2016), a number of differences were identified and revised. The revised HarmGR gene sequences were submitted to GenBank (*HarmGr4*-KY806282, *HarmGr5*-KY806283, *HarmGr6*-KY806284, *HarmGr7*-KY806285, *HarmGr8*-KY806286, *HarmGr10*-KY806287, *HarmGr11*-KY806288 and *HarmGr12*-KY806289). *HarmGr4* is shorter than previously annotated at the N-terminus. Sequence alignment results suggest the N-terminus is not conserved as C-terminus, which may inspire difficulties in the annotation at the N-terminus (Fig. 2). *HarmGr10* may be a pseudogene because a stop codon was detected in the cloned ORF sequence, leading to only 274 amino acids. By comparison between HarmGR10 PCR products and annotated sequence, a fragment of four nucleic acids (AAAA) from 521 to 524 replaced the genome annotated sequences (TAC), resulting in a premature stop codon as well as a truncated GR10 protein. Pseudogenes have been found in other insect sugar receptors, for example, *Gr64e* is a pseudogene in both *Drosophila pseudoobscura* and *Drosophila persimilis* (Kent and Robertson 2009). *Gr5a* is a severely damaged pseudogene in the Hawaiian *Drosophila grimshawi* (Kent and Robertson 2009). Pseudogenes have long been believed as “junk”

sequences in the genome during the evolution. However, recent results showed that some pseudogenes seem to regulate their protein-coding cousins. Many pseudogenes can be transcribed into RNA and exhibit a tissue-specific pattern of activation, which may be the case for *HarmGR10* in this study. Recently, we found a new expanded group of *H. armigera* “bitter” GRs, which show truncated amino acids but are still functional to detect plant compounds in insect cells (Xu et al. 2016). For the other sugar receptor genes (*HarmGR5*, 6, 7, 8, 11 and 12), the cloned sequences showed 94-98% identities to the previously annotated sequences. Our results demonstrated that sequences annotated from genome or transcriptome data need validation in the wet lab before functional study. High-throughput genome projects may have mis-assemblies or mis-annotations, requiring substantial manual curation before biological study. Another possibility is that the strain used for GR cloning in this study is different from the genome sequencing strain, leading to different sequences.

We performed sequence alignment analysis of *H. armigera* sugar GRs (Fig. 2) and the result showed that *H. armigera* sugar GRs are conserved at the C-terminal. Further study on 33 lepidopteran sugar receptors (Supplementary Fig. 1) also showed C-terminal are conserved across the family, akin to olfactory receptors studied previously (Ray et al. 2014) (Fig. 3). Previous studies on *Drosophila* have shown that insect GRs form functional heteromultimers *in vivo* (Jiao et al. 2008a; Jones et al. 2007; Lee et al. 2009), with genetic studies indicating that co-expression of multiple GRs is essential for the detection of compounds such as CO₂, sucrose, D-glucose and trehalose (Dahanukar et al. 2001; Dahanukar et al. 2007; Jiao et al. 2008a; Jones et al. 2007; Moon et al. 2006). The conserved C-terminal of *H. armigera* sugar GRs may be involved in the interaction to the co-receptors or the cell membrane.

The phylogenetic analysis of the sugar GRs from various insect species showed insect sugar receptors were divided into species-specific groups (Fig. 3). The Diptera-specific group

houses all sugar GRs of *D. melanogaster*, *A. gambiae*, *A. aegypti* and *C. quinquefasciatus*. Coleoptera-specific group houses all 16 sugar GR candidates of *T. castaneum*. Four Lepidoptera-specific groups were formed in the tree (Fig. 3). Lepidoptera-specific group 1 contains HarmGR5, 6, 10; BmorGR5, 6; HmelGR5, 6, 45; DpleGR2, 32, 38, 42; PxylGR1, 2 and 3. Lepidoptera-specific group 2 contains HarmGR4, 7, 8; BmorGR4; HmelGR4, 52; DpleGR3, 4, 37; PxylGR4 and 44. Lepidoptera-specific group 3 contains HarmGR11, 12; BmorGR7, 8; HmelGR7, and DpleGR40. Lepidoptera-specific group 4 contains HmelGR8 and BmorGR8. These results suggest Lepidopteran sugar receptors may come from four ancestral genes. There are also two specific groups of sugar receptors in Hemiptera and Hymenoptera families respectively. Termite sugar GRs formed a Blattodea-specific group.

GR gene localization and structure

All eight *H. armigera* sugar GR genes are clustered in a 200,000-bp fragment on scaffold 0 (Fig. 4A) in the genome. Similar results were also observed in other insect genomes. In *H. melpomene*, *Gr4*, 5, 7 and 52 were mapped on chromosome 6 but *Gr6*, 8 and 45 have not been mapped so far (Briscoe et al. 2013). In honeybee and wasp, the two sugar GRs were clustered on the same chromosome too (Kent and Robertson 2009). In *T. castaneum*, all the sugar GR candidates are on the same chromosome. In *D. melanogaster*, *Gr64a*, *b*, *c*, *d*, *e*, *f* and *61a* were clustered on the chromosome 3L while *Gr5a* is on chromosome X (Kent and Robertson 2009). *HarmGr4*, 6, 7, 8, 10 and 12 are transcribed in opposite directions to *HarmGr5* and 11. *H. armigera* GR genes also showed different exon/intron structures (Fig. 4B). For example, HarmGR4, 7, 8 and 10 consist of nine exons. HarmGR12 consists of 10 exons. HarmGR5 and 6 show 11 exons while HarmGR11 showed 12 exons (Fig. 4B).

Predicted topologies

Previous studies showed that both insect olfactory receptors (ORs) and GRs encode seven TMDs with an intracellular N-terminus and an extracellular C-terminus by using *in silico* or cellular biology methods (Kent and Robertson 2009; Smart et al. 2008; Zhang et al. 2011). Whether this topology is present throughout the insect sugar GR family is unknown. In *H. armigera*, two new topologies from “bitter receptor” family were reported: N-terminus and C-terminus both are intracellular or both extracellular (Xu et al. 2016). To investigate the membrane topology of *H. armigera* sugar GRs, we first applied algorithms (TMPred, HMMTOP and TMHMM) to predict transmembrane domains (TMDs). Results showed very diverse TMD topologies (Table 1). For HarmGR7 and 8, all three programs predicted seven conserved TMDs (Table 1). For others, the predicted TMD results are variable, from six to ten TMDs. Therefore, we cannot exclude other possible topologies are present in insect sugar receptors.

Expression profile

To characterize the potential function of these candidate sugar receptors, we also built an expression profile of these identified *H. armigera* sugar GRs from the 31 transcriptomic libraries (Fig. 5). All eight *H. armigera* sugar GR genes were detected in at least one library (Fig. 5). In general, sugar GR genes have a broader expression in adult tissues than in larval tissues (Fig. 5), especially in adult antennae, tarsi, testes and ovaries. Four sugar GRs (HarmGR5, 6, 10 and 12) are detected from larval mouthpart, which is a major gustatory tissue at larvae stage. Taken into account that in *H. armigera* fifth instar larvae, sucrose was sensed by a gustatory receptor neuron (GRN) in the lateral sensillum, and *myo*-inositol by a GRN in the medial sensillum (Zhang et al. 2013), HarmGR5, 6 10 and 12 may be the

275 candidate receptors for sucrose or myo-inositol. At adult stage, HarmGR4 was detected in
276 female adult ovaries, tarsi, antennae, both male and female heads (with antennae) and male
277 testes. HarmGR5 was expressed in adult tarsi and heads (with antennae) but not the antennae,
278 suggesting they may be expressed in other tissues on heads, for example, the proboscis, an
279 organ used to suck sugar-rich nectar, plant sap, tree sap, and pollen or complementary food.
280 HarmGR6 and 7 were both expressed in adult antennae and tarsi. Previous studies showed
281 that in *H. armigera*, the tarsal gustatory sensilla play an essential role in perceiving sugars
282 available in floral nectar and provide chemosensory information determining feeding
283 behaviour (Zhang et al. 2010). HarmGR8 was expressed in post-feeding larvae stage, pupae
284 stage and adult antennae (Fig. 5). HarmGR10 was expressed in male adult testes while
285 HarmGR11 was expressed in the female ovaries (Fig. 5). This result was further confirmed
286 by RT-PCR (Fig. 5B). The RT-PCR analyses of other HarmGR genes have been performed
287 previously (Liu et al. 2014). Interestingly HarmGR10 showed specific expression in male
288 testes while HarmGR11 showed specific expression in female ovaries. This is the first report
289 of specific sugar receptors in lepidopteran reproductive systems. Previous studies showed by
290 using Gr-GAL4 drivers in *D. melanogaster*, *Gr28b.b*, *Gr28b.c*, *Gr32a*, and *Gr64c*, were
291 detected in neurons that appear to innervate both the male and female reproductive organs,
292 suggesting these GRs possibly function in regulation of accessory gland secretion in males or
293 regulation of ovulation in females (Park and Kwon 2011). Another study showed that GR43a,
294 a fructose gustatory receptor, is expressed in the uterus of *D. melanogaster*, raising the
295 possibility that it plays a role in reproduction or mating (Miyamoto et al. 2012). Furthermore,
296 odorant receptor, Orco, was detected in the flagella of malaria mosquito, *A. gambiae*,
297 spermatozoa. Orco-specific agonists, antagonists as well as other odorant compounds could
298 robustly activate flagella beating. Therefore, chemosensory receptors may mediate sperm
299 activation in insects (Pitts et al. 2014). Another study reported that sugar detection is related

to ovarian maturation in female *Phormia regina* (Solari et al. 2015). Thus, it is reasonable to propose that HarmGR10 and 11 are involved in nutrient sensation and regulates reproduction behaviours. The expression levels of HarmGRs in lab colonised *H. armigera* in this study may be different from wild insects because this colony has been fed on artificial diet for many years, possibly resulting in lower GR expression compared with wild insects which feed on plants.

In summary, eight *H. armigera* sugar GR genes were cloned, validated and analysed by using alignment, phylogenetic analysis, RT-PCR and *in silico* gene expression profile, which have shown high, selective and specific expressions in major gustatory tissues. HarmGR10 may be a pseudogene, but highly expressed in male adult testes. HarmGR11 is highly expressed in female ovaries. We propose *H. armigera* sugar GRs play multi-functional roles in detecting a broad range of sugars. Our study provides new knowledge about insect sugar gustatory receptors and may help better understand insect chemosensory system and insect-plant interactions.

316 Acknowledgments

317 We would like to thank all the members of *Helicoverpa* genome consortium for permission
318 to use the genome sequences ahead of the publication. We also thank Professor Myron
319 Zalucki of University of Queensland providing us *H. armigera* pupae and Dr Alexie
320 Papanicolaou of Hawkesbury Institute for the Environment in Western Sydney University
321 helping us in the *in silico* expression analysis. We thank Western Australia State Agricultural
322 Biotechnology Centre (SABC, [http://www.murdoch.edu.au/Research-](http://www.murdoch.edu.au/Research-capabilities/SABC/Facilities/DNA-Sequencing/)
323 [capabilities/SABC/Facilities/DNA-Sequencing/](http://www.murdoch.edu.au/Research-capabilities/SABC/Facilities/DNA-Sequencing/)) for technical support. Dr Wei Xu is the
324 recipient of an Australian Research Council Discovery Early Career Researcher Award
325 (DECRA) (DE160100382).

Draft

Reference

Abdel-Latif, M. 2007. A Family of Chemoreceptors in *Tribolium castaneum* (Tenebrionidae: Coleoptera). PLoS One **2**(12). doi:10.1371/journal.pone.0001319.

Arensburger, P., Megy, K., Waterhouse, R.M., Abrudan, J., Amedeo, P., Antelo, B., Bartholomay, L., Bidwell, S., Caler, E., Camara, F., Campbell, C.L., Campbell, K.S., Casola, C., Castro, M.T., Chandramouliswaran, I., Chapman, S.B., Christley, S., Costas, J., Eisenstadt, E., Feschotte, C., Fraser-Liggett, C., Guigo, R., Haas, B., Hammond, M., Hansson, B.S., Hemingway, J., Hill, S.R., Howarth, C., Ignell, R., Kennedy, R.C., Kodira, C.D., Lobo, N.F., Mao, C., Mayhew, G., Michel, K., Mori, A., Liu, N., Naveira, H., Nene, V., Nguyen, N., Pearson, M.D., Pritham, E.J., Puiu, D., Qi, Y., Ranson, H., Ribeiro, J.M., Roberston, H.M., Severson, D.W., Shumway, M., Stanke, M., Strausberg, R.L., Sun, C., Sutton, G., Tu, Z.J., Tubio, J.M., Unger, M.F., Vanlandingham, D.L., Vilella, A.J., White, O., White, J.R., Wondji, C.S., Wortman, J., Zdobnov, E.M., Birren, B., Christensen, B.M., Collins, F.H., Cornel, A., Dimopoulos, G., Hannick, L.I., Higgs, S., Lanzaro, G.C., Lawson, D., Lee, N.H., Muskavitch, M.A., Raikhel, A.S., and Atkinson, P.W. 2010. Sequencing of *Culex quinquefasciatus* establishes a platform for mosquito comparative genomics. Science **330**(6000): 86-88. doi: 10.1126/science.1191864.

Bredendiek, N., Hutte, J., Steingraber, A., Hatt, H., Gisselmann, G., and Neuhaus, E.M. 2011. Go alpha is involved in sugar perception in *Drosophila*. Chem. Senses **36**(1): 69-81. doi: 10.1093/chemse/bjq100.

Briscoe, A.D., Macias-Munoz, A., Kozak, K.M., Walters, J.R., Yuan, F., Jamie, G.A., Martin, S.H., Dasmahapatra, K.K., Ferguson, L.C., Mallet, J., Jacquin-Joly, E., and Jiggins, C.D. 2013. Female behaviour drives expression and evolution of gustatory receptors in butterflies. PLoS Genet. **9**(7): e1003620. doi: 10.1371/journal.pgen.1003620.

Chyb, S., Dahanukar, A., Wickens, A., and Carlson, J.R. 2003. *Drosophila* Gr5a encodes a taste receptor tuned to trehalose. Proc. Natl. Acad. Sci. U S A **100** Suppl 2: 14526-14530.

Clyne, P.J., Warr, C.G., and Carlson, J.R. 2000. Candidate taste receptors in *Drosophila*. Science **287**(5459): 1830-1834.

Cunningham, J.P., and Zalucki, M.P. 2014. Understanding *Heliothine* (Lepidoptera: Heliothinae) Pests: What is a Host Plant? J. Econ. Entomol. **107**(3): 881-896. doi: 10.1603/EC14036.

Dahanukar, A., Foster, K., van der Goes van Naters, W.M., and Carlson, J.R. 2001. A Gr receptor is required for response to the sugar trehalose in taste neurons of *Drosophila*. Nat. Neurosci. **4**(12): 1182-1186.

Dahanukar, A., Lei, Y.T., Kwon, J.Y., and Carlson, J.R. 2007. Two Gr genes underlie sugar reception in *Drosophila*. Neuron **56**(3): 503-516. doi: 10.1016/j.neuron.2007.10.024.

Engsontia, P., Sangket, U., Chotigeat, W., and Satasook, C. 2014. Molecular evolution of the odorant and gustatory receptor genes in lepidopteran insects: implications for their adaptation and speciation. J. Mol. Evol. **79**(1-2): 21-39. doi: 10.1007/s00239-014-9633-0.

- 372 Foster, W.A. 1995. Mosquito sugar feeding and reproductive energetics. *Annu. Rev.*
 373 *Entomol.* **40**: 443-474. doi: 10.1146/annurev.en.40.010195.002303.
- 374
- 375 Freeman, E.G., Wisotsky, Z., and Dahanukar, A. 2014. Detection of sweet tastants by a
 376 conserved group of insect gustatory receptors. *Proc. Natl. Acad. Sci. U S A* **111**(4): 1598-
 377 1603. doi: 10.1073/pnas.1311724111.
- 378
- 379 Hofmann, K. and Stoffel, W.. 1993. TMBASE - A database of membrane spanning protein
 380 segments. *Biol. Chem. Hoppe-Seyler* **374**(166).
- 381
- 382 Jiao, Y., Moon, S.J., and Montell, C. 2007. A *Drosophila* gustatory receptor required for the
 383 responses to sucrose, glucose, and maltose identified by mRNA tagging. *Proc. Natl. Acad.*
 384 *Sci. U S A* **104**(35): 14110-14115.
- 385
- 386 Jiao, Y., Moon, S.J., Wang, X., Ren, Q., and Montell, C. 2008a. Gr64f is required in
 387 combination with other gustatory receptors for sugar detection in *Drosophila*. *Curr. Biol.*
 388 **18**(22): 1797-1801. doi: 10.1016/j.cub.2008.10.009.
- 389
- 390 Jiao, Y., Moon, S.J., Wang, X., Ren, Q., and Montell, C. 2008b. Gr64f is required in
 391 combination with other gustatory receptors for sugar detection in *Drosophila*. *Curr. Biol.* **18**:
 392 1797-1801.
- 393
- 394 Jones, W.D., Cayirlioglu, P., Kadow, I.G., and Vosshall, L.B. 2007. Two chemosensory
 395 receptors together mediate carbon dioxide detection in *Drosophila*. *Nature* **445**(7123): 86-90.
 396 doi: 10.1038/nature05466.
- 397
- 398 Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S.,
 399 Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., & Drummond, A.
 400 (2012). Geneious Basic: an integrated and extendable desktop software platform for the
 401 organization and analysis of sequence data. *Bioinformatics*, **28**(12), 1647-1649.
- 402
- 403 Kent, L.B., and Robertson, H.M. 2009. Evolution of the sugar receptors in insects. *BMC*
 404 *Evol. Biol.* **9**: 41. doi: 10.1186/1471-2148-9-41.
- 405
- 406 Kent, L.B., Walden, K.K., and Robertson, H.M. 2008. The Gr family of candidate gustatory
 407 and olfactory receptors in the yellow-fever mosquito *Aedes aegypti*. *Chem. Senses* **33**(1): 79-
 408 93.
- 409
- 410 Kirkness, E.F., Haas, B.J., Sun, W., Braig, H.R., Perotti, M.A., Clark, J.M., Lee, S.H.,
 411 Robertson, H.M., Kennedy, R.C., Elhaik, E., Gerlach, D., Kriventseva, E.V., Elsik, C.G.,
 412 Gaur, D., Hill, C.A., Veenstra, J.A., Walenz, B., Tubio, J.M., Ribeiro, J.M., Rozas, J.,
 413 Johnston, J.S., Reese, J.T., Popadic, A., Tojo, M., Raoult, D., Reed, D.L., Tomoyasu, Y.,
 414 Kraus, E., Mittapalli, O., Margam, V.M., Li, H.M., Meyer, J.M., Johnson, R.M., Romero-
 415 Severson, J., Vanzee, J.P., Alvarez-Ponce, D., Vieira, F.G., Aguade, M., Guirao-Rico, S.,
 416 Anzola, J.M., Yoon, K.S., Strycharz, J.P., Unger, M.F., Christley, S., Lobo, N.F.,
 417 Seufferheld, M.J., Wang, N., Dasch, G.A., Struchiner, C.J., Madey, G., Hannick, L.I.,
 418 Bidwell, S., Joardar, V., Caler, E., Shao, R., Barker, S.C., Cameron, S., Bruggner, R.V.,
 419 Regier, A., Johnson, J., Viswanathan, L., Utterback, T.R., Sutton, G.G., Lawson, D.,
 420 Waterhouse, R.M., Venter, J.C., Strausberg, R.L., Berenbaum, M.R., Collins, F.H., Zdobnov,

- 421 E.M., and Pittendrigh, B.R. 2010. Genome sequences of the human body louse and its
 422 primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proc. Natl.*
 423 *Acad. Sci. U S A* **107**(27): 12168-12173. doi: 10.1073/pnas.1003379107.
- 424
- 425 Koh, T.W., He, Z., Gorur-Shandilya, S., Menuz, K., Larter, N.K., Stewart, S., and Carlson,
 426 J.R. 2014. The *Drosophila* IR20a Clade of Ionotropic Receptors Are Candidate Taste and
 427 Pheromone Receptors. *Neuron* **83**(4): 850-865. doi: 10.1016/j.neuron.2014.07.012.
- 428
- 429 Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. 2001. Predicting
 430 transmembrane protein topology with a hidden Markov model: application to complete
 431 genomes. *J. Mol. Biol.* **305**(3): 567-580. doi: 10.1006/jmbi.2000.4315.
- 432
- 433 Lee, Y., Moon, S.J., and Montell, C. 2009. Multiple gustatory receptors required for the
 434 caffeine response in *Drosophila*. *Proc. Natl. Acad. Sci. U S A* **106**(11): 4495-4500. doi:
 435 10.1073/pnas.0811744106.
- 436
- 437 Liu, N.Y., Xu, W., Papanicolaou, A., Dong, S.L., and Anderson, A. 2014. Identification and
 438 characterization of three chemosensory receptor families in the cotton bollworm *Helicoverpa*
armigera. *BMC Genomics* **15**(1): 597. doi: 10.1186/1471-2164-15-597.
- 439
- 440 Mishra, D., Miyamoto, T., Rezenom, Y.H., Broussard, A., Yavuz, A., Slone, J., Russell,
 441 D.H., and Amrein, H. 2013. The molecular basis of sugar sensing in *Drosophila* larvae. *Curr.*
 442 *Biol.* **23**(15): 1466-1471. doi: 10.1016/j.cub.2013.06.028.
- 443
- 444 Miyamoto, T., Slone, J., Song, X.Y., and Amrein, H. 2012. A Fructose Receptor Functions as
 445 a Nutrient Sensor in the *Drosophila* Brain. *Cell* **151**(5): 1113-1125. doi:
 446 10.1016/j.cell.2012.10.024.
- 447
- 448 Moon, S.J., Kottgen, M., Jiao, Y., Xu, H., and Montell, C. 2006. A taste receptor required for
 449 the caffeine response in vivo. *Curr. Biol.* **16**(18): 1812-1817. doi: 10.1016/j.cub.2006.07.024.
- 450
- 451 Ning, C., Yang, K., Xu, M., Huang, L.Q., and Wang, C.Z. 2016. Functional validation of the
 452 carbon dioxide receptor in labial palps of *Helicoverpa armigera* moths. *Insect Biochem. Mol.*
Biol. **73**: 12-19. doi: 10.1016/j.ibmb.2016.04.002.
- 453
- 454 Park, J.H., and Kwon, J.Y. 2011. A Systematic Analysis of *Drosophila* Gustatory Receptor
 455 Gene Expression in Abdominal Neurons which Project to the Central Nervous System. *Mol.*
 456 *Cells* **32**(4): 375-381. doi: 10.1007/s10059-011-0128-1.
- 457
- 458 Perkins, L.E., Cribb, B.W., Brewer, P.B., Hanan, J., Grant, M., de Torres, M., and Zalucki,
 459 M.P. 2013. Generalist insects behave in a jasmonate-dependent manner on their host plants,
 460 leaving induced areas quickly and staying longer on distant parts. *P. Roy. Soc. B-Biol. Sci.*
 461 **280**(1756). doi: 10.1098/Rspb.2012.2646.
- 462
- 463 Pham-Delegue, M.H., Etievant, P., Guichard, E., Marilleau, R., Douault, P., Chauffaille, J.,
 464 and Masson, C. 1990. Chemicals involved in honeybee-sunflower relationship. *J. Chem.*
Ecol. **16**(11): 3053-3065. doi: 10.1007/BF00979612.
- 465

- Pitts, R.J., Liu, C., Zhou, X.F., Malpartida, J.C., and Zwiebel, L.J. 2014. Odorant receptor-mediated sperm activation in disease vector mosquitoes. *Proc. Natl. Acad. Sci. U S A* **111**(7): 2566-2571. doi: 10.1073/pnas.1322923111.
- Ray, A., van Naters, W.G., and Carlson, J.R. 2014. Molecular determinants of odorant receptor function in insects. *J. Biosci.* **39**(4): 555-563.
- Robertson, H.M., Gadau, J., and Wanner, K.W. 2010. The insect chemoreceptor superfamily of the parasitoid jewel wasp *Nasonia vitripennis*. *Insect Mol. Biol.* **19 Suppl 1**: 121-136. doi: 10.1111/j.1365-2583.2009.00979.x.
- Robertson, H.M., and Wanner, K.W. 2006. The chemoreceptor superfamily in the honey bee, *Apis mellifera*: expansion of the odorant, but not gustatory, receptor family. *Genome Res.* **16**(11): 1395-1403.
- Robertson, H.M., Warr, C.G., and Carlson, J.R. 2003. Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U S A* **100 Suppl 2**: 14537-14542.
- Sato, K., Tanaka, K., and Touhara, K. 2011. Sugar-regulated cation channel formed by an insect gustatory receptor. *Proc. Natl. Acad. Sci. U S A* **108**(28): 11680-11685. doi: 10.1073/pnas.1019622108.
- Slone, J., Daniels, J., and Amrein, H. 2007. Sugar receptors in *Drosophila*. *Curr. Biol.* **17**(20): 1809-1816. doi: 10.1016/j.cub.2007.09.027.
- Smart, R., Kiely, A., Beale, M., Vargas, E., Carraher, C., Kralicek, A.V., Christie, D.L., Chen, C., Newcomb, R.D., and Warr, C.G. 2008. *Drosophila* odorant receptors are novel seven transmembrane domain proteins that can signal independently of heterotrimeric G proteins. *Insect Biochem. Mol. Biol.* **38**(8): 770-780. doi: 10.1016/j.ibmb.2008.05.002.
- Solari, P., Stoffolano, J.G., Jr., De Rose, F., Barbarossa, I.T., and Liscia, A. 2015. The chemosensitivity of labellar sugar receptor in female *Phormia regina* is paralleled with ovary maturation: Effects of serotonin. *J. Insect Physiol.* **82**: 38-45. doi: 10.1016/j.jinsphys.2015.08.007.
- Tusnady, G.E., and Simon, I. 1998. Principles governing amino acid composition of integral membrane proteins: application to topology prediction. *J. Mol. Biol.* **283**(2): 489-506. doi: 10.1006/jmbi.1998.2107.
- van der Goes van Naters, W., and Carlson, J.R. 2006. Insects as chemosensors of humans and crops. *Nature* **444**(7117): 302-307. doi: 10.1038/nature05403.
- Wada-Katsumata, A., Silverman, J., and Schal, C. 2013. Changes in taste neurons support the emergence of an adaptive behavior in cockroaches. *Science* **340**(6135): 972-975. doi: 10.1126/science.1234854.
- Wang, X., Fang, X., Yang, P., Jiang, X., Jiang, F., Zhao, D., Li, B., Cui, F., Wei, J., Ma, C., Wang, Y., He, J., Luo, Y., Wang, Z., Guo, X., Guo, W., Zhang, Y., Yang, M., Hao, S., Chen, B., Ma, Z., Yu, D., Xiong, Z., Zhu, Y., Fan, D., Han, L., Wang, B., Chen, Y., Wang, J., Yang, L., Zhao, W., Feng, Y., Chen, G., Lian, J., Li, Q., Huang, Z., Yao, X., Lv, N., Zhang, G., Li,

- Y., Zhu, B., and Kang, L. 2014. The locust genome provides insight into swarm formation and long-distance flight. *Nat. Commun.* **5**(2957): 1-9. doi: 10.1038/ncomms3957.
- Wanner, K.W., and 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 Mol. Biol.* **17**(6): 621-629. doi: 10.1111/j.1365-2583.2008.00836.x.
- Xu, W., and Anderson, A. 2015. Carbon dioxide receptor genes in cotton bollworm *Helicoverpa armigera*. *Die Naturwissenschaften* **102**(3-4): 11. doi: 10.1007/s00114-015-1260-0.
- Xu, W., Papanicolaou, A., Liu, N.Y., Dong, S.L., and Anderson, A. 2015. Chemosensory receptor genes in the Oriental tobacco budworm *Helicoverpa assulta*. *Insect Mol. Biol.* **24**(2): 253-263. doi: 10.1111/imb.12153.
- Xu, W., Papanicolaou, A., Zhang, H.J., and Anderson, A. 2016. Expansion of a bitter taste receptor family in a polyphagous insect herbivore. *Sci. Rep.* **6**: 23666. doi: 10.1038/srep23666.
- Xu, W., Zhang, H.J., and Anderson, A. 2012. A sugar gustatory receptor identified from the foregut of cotton bollworm *Helicoverpa armigera*. *J. Chem. Ecol.* **38**(12): 1513-1520. doi: 10.1007/s10886-012-0221-8.
- Zhang, H.J., Anderson, A.R., Trowell, S.C., Luo, A.R., Xiang, Z.H., and Xia, Q.Y. 2011. Topological and functional characterization of an insect gustatory receptor. *PLoS One* **6**(8): e24111. doi: 10.1371/journal.pone.0024111
- Zhang, H.J., Faucher, C.P., Anderson, A., Berna, A.Z., Trowell, S., Chen, Q.M., Xia, Q.Y., and Chyb, S. 2013. Comparisons of Contact Chemoreception and Food Acceptance by Larvae of Polyphagous *Helicoverpa armigera* and Oligophagous *Bombyx mori*. *J. Chem. Ecol.* **39**(8): 1070-1080. doi: 10.1007/s10886-013-0303-2.
- Zhang, Y.F., van Loon, J.J., and Wang, C.Z. 2010. Tarsal taste neuron activity and proboscis extension reflex in response to sugars and amino acids in *Helicoverpa armigera* (Hubner). *J. Exp. Biol.* **213**(Pt 16): 2889-2895. doi: 10.1242/jeb.042705.

Figure Legend

Figure 1. Number of sugar GRs identified from various insect species. Lepidoptera insects (*H. armigera*, *B. mori*, *H. melpomene* and *D. plexippus*, *P. xylostella*), Diptera insects (*A. gambiae*, *A. aegypti*, *C. quinquefasciatus* and *D. melanogaster*); Coleoptera insect (*T. castaneum*); Hymenoptera insects (*P. barbatus*, *A. mellifera* and *N. vitripennis*); Hemiptera insect (*A. pisum*); Phthiraptera insect (*P. humanus*); Orthoptera (*L. migratoria*) and Blattodea (*Z. nevadensis*) (Arensburger et al. 2010; Briscoe et al. 2013; Kent and Robertson 2009; Kent et al. 2008; Kirkness et al. 2010; Robertson et al. 2010; Robertson and Wanner 2006; Wang et al. 2014; Wanner and Robertson 2008; Xu et al. 2016; Zhang et al. 2011). The number of candidate sugar GRs differ across species and fall into discrete groups.

Figure 2. Amino acid sequence alignment analysis of *H. armigera* sugar GRs (HarmGR4-8, 11 and 12). The conserved amino acids are labelled in black colour. C-terminus of these sugar GRs are more conserved than the N-terminus.

Figure 3. Phylogenetic analysis of insect sugar GRs. Diptera-specific group (black), Coleoptera-specific group (blue), Lepidoptera-specific (green), Hymenoptera-specific (purple), Hemiptera (orange) and Blattodea (grey). Lepidoptera-specific sugar GRs were further divided into four subgroups. Hemiptera-specific sugar GRs were further divided into two subgroups.

Figure 4. Schematic representation of *H. armigera* sugar GRs genomic organization. The localization on scaffold (A) and exon/intron structures (B) of *H. armigera* sugar receptor genes are shown. Solid blocks represent exons.

Figure 5. Expression profiles of *H. armigera* sugar GRs in various tissues, stages of larvae, male and female adults as revealed by read mapping. (A) *In Silico* expression map of eight HarmGRs in various samples. The expression of GRs in tissues include: antennae, mouthparts, epidermis, fat body, foreguts, midguts, hindguts, Malpighian tubules, haemocytes, hearts, trachea, ventral nerve, silk glands, salivary glands and muscle of 5th instar larvae; male antennae, female antennae, male heads, female heads, male tarsi, female tarsi, male thorax, female thorax, male abdomens, female abdomens, male testes and female ovaries from day 0 to day 5 adults; embryo, 3rd instar larvae (whole body), post-feeding stage larvae (whole body) and pupae. The number in the box is the normalized FPKM value (Dark red, max. value; yellow, mid. value; and white, min. value). **(B) RT-PCR analysis of HarmGR10 and 11 in ♂An, ♂Ta, Testes, ♀An, ♀Ta and Ovaries.** Beta-actin gene was used as a control.

Draft

Table 1. The numbers of predicted trans-membrane domains of *H. armigera* sugar GRs by using TMpred, HMMTOP and TMHMM.

	Predicted transmembrane configuration					
	TMpred		HMMTOP		TMHMM	
	N-*	TMD	N-*	TMD	N-*	TMD
HarmGR4	+	7	+	7	+	6
HarmGR5		10		8	+	7
HarmGR6		9		8	+	9
HarmGR7	+	7	+	7	+	7
HarmGR8	+	7	+	7	+	7
HarmGR11		8		8		7
HarmGR12	+	8	+	7	+	7

* N-terminus locates intracellularly.

Transmembrane domain (TMD).

Table 2. The designed gene-specific primers for cloning and sequencing *H. armigera* sugar GRs by using RT-PCR.

HarmGR4F	ATGAAAGGGAATTCGGTCAAG
HarmGR4R	CTACTTATCATATTGTATTAAACTAATTC
HarmGR5F	ATGCAAAATGGATGGAACA
HarmGR5R	CTACTTAATAACACCACTGACAGACTC
HarmGR6F	ATGGGGCAAACAAGTTTTAGA
HarmGR6R	TCAAGTAGCTAGTGTTGTAATAGTGTGG
HarmGR7F	ATGTCAAGTCGAGGGTTC
HarmGR7R	CTATACTCCAATAGCATTTTTAGTCA
HarmGR8F	ATGTCGAGTAAGGAATT
HarmGR8R	TTAGATGTTAGTAGCA
HarmGR10F	AATAAATTGGTTGGTAAGAAAAGAAATG
HarmGR10R	CTAGTTCTTGTTCAATTCGTCGCTGCTG
HarmGR11F	ATGCCATCGAACTATTTCTCAAGAC
HarmGR11R	TTATCTGTTGAACTGTAGTAATACC
HarmGR12F	ATGAAAGTCCACCAACTGA
HarmGR12R	ACTGTAAGTGTGCGAAG

112x149mm (300 x 300 DPI)

246x565mm (300 x 300 DPI)

238x650mm (300 x 300 DPI)

129x93mm (300 x 300 DPI)

129x94mm (300 x 300 DPI)

Supplementary Figure 1. Amino acid sequence alignment analysis of *H. armigera* sugar GRs and other Lepidoptera sugar receptors. The conserved amino acids are labelled in black colour. C-terminus of these sugar GRs are more conserved than the N-terminus.

Draft

A

1 Supplementary Table 1, Percent Identities (%) of *H. armigera* and *H. zea* sugar gustatory receptors at amino acid level - created by Clustal2.1

2 <http://www.ebi.ac.uk/Tools/services/rest/clustalo/result/clustalo-l20170623-063359-0938-93626133-oy/pim>

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1:	HzeaGR4	100.00	98.01	40.34	40.59	40.83	40.83	22.97	22.83	28.95	29.21	23.64	24.87	21.78	22.13	26.27	26.27
2:	HarmGR4	98.01	100.00	40.50	40.75	41.50	41.50	23.96	23.81	29.11	29.38	23.43	24.73	22.04	22.40	26.01	26.01
3:	HzeaGR8	40.34	40.50	100.00	98.59	65.33	64.15	26.02	26.43	26.14	26.14	25.45	26.15	24.61	24.20	23.58	23.32
4:	HarmGR8	40.59	40.75	98.59	100.00	66.04	64.86	26.56	26.98	26.40	26.40	25.45	26.15	25.13	24.73	24.35	24.09
5:	HzeaGR7	40.83	41.50	65.33	66.04	100.00	98.13	22.49	23.16	26.52	26.52	24.68	25.64	23.30	23.14	23.39	23.91
6:	HarmGR7	40.83	41.50	64.15	64.86	98.13	100.00	22.49	23.16	26.77	26.77	24.68	25.64	23.04	22.87	23.39	23.91
7:	HzeaGR10	22.97	23.96	26.02	26.56	22.49	22.49	100.00	96.14	39.32	38.83	35.35	35.63	23.00	23.23	26.53	26.79
8:	HarmGR10	22.83	23.81	26.43	26.98	23.16	23.16	96.14	100.00	38.78	38.29	35.98	36.26	23.62	23.60	26.41	26.67
9:	HzeaGR6	28.95	29.11	26.14	26.40	26.52	26.77	39.32	38.78	100.00	98.89	47.25	46.71	32.45	32.52	30.52	30.75
10:	HarmGR6	29.21	29.38	26.14	26.40	26.52	26.77	38.83	38.29	98.89	100.00	47.25	46.71	32.45	32.52	30.28	30.52
11:	HzeaGR5	23.64	23.43	25.45	25.45	24.68	24.68	35.35	35.98	47.25	47.25	100.00	97.69	31.20	31.76	29.93	30.41
12:	HarmGR5	24.87	24.73	26.15	26.15	25.64	25.64	35.63	36.26	46.71	46.71	97.69	100.00	30.83	31.37	30.05	30.53
13:	HzeaGR11	21.78	22.04	24.61	25.13	23.30	23.04	23.00	23.62	32.45	32.45	31.20	30.83	100.00	95.63	34.07	33.83
14:	HarmGR11	22.13	22.40	24.20	24.73	23.14	22.87	23.23	23.60	32.52	32.52	31.76	31.37	95.63	100.00	34.16	33.92
15:	HzeaGR12	26.27	26.01	23.58	24.35	23.39	23.39	26.53	26.41	30.52	30.28	29.93	30.05	34.07	34.16	100.00	97.24
16:	HarmGR12	26.27	26.01	23.32	24.09	23.91	23.91	26.79	26.67	30.75	30.52	30.41	30.53	33.83	33.92	97.24	100.00

Supplementary data, *Helicoverpa zea* sugar receptor DNA and Amino acid sequences

>HzeaGR4

ATGGAAATTAAGCTGTGTAACACTATTTGTAGTTTTGACAAAGAGAATTGAATATATGAAAGGGAATTCGGTCA
AGAATTTAAATGAAAAGAAACGTGACGACTTCCTACCAACATTAAACAGTGTGTTCTCAAGGCTCGATTCTTC
GGTATATCTGGGTATGGGCTCACAATATCATTTTGTGGAGTCTCATATTGTTCTCAATGCTGGTTGTCATGGA
ATCCGTGGCTATATGGAGAGTTGTGACTCTGCTTGGAGAATGGCTTATGAGTGCTTCTAATAATGGTCTAATC
GGTCGTCTCTCTGGCGCAATATTTTACGGTAACGCATTAATTTCTCTGTTCTCTATCATCGAAGTTTGTGCACTCC
TGGCGGAGTCTATCAAATTACTGGTTGAGGATGGAAACAGTACTGCTCTGGACTTCCCTCCTGATGTCAGAA
TAAGGAAGAGAACTATTTATATAACTGCTTTTGTGGTTTCTGTTGCTGTAGTTGAGCATATATTGAGCATGATA
TCTGCTACTGGCGTTGGATTTCCTCCAGAAGAGTTCTTGATAGATACGTCACACTGTCCCATGGATTTATTTTG
AAAGCTCAAGACTACACCATATGGAAGGCCATACCAATCGTCGTGCTCAGCAAATTGGCGACAGCTCTGTGGA
ACTTTCAGGACCTGATCATCATTCTGATAAGCATGGGCCTATCCTCGCGATACGATAGGCTCAATTTGTACGTG
AGGCATGTCGTCTCTGTGGAGAAGCAATTTGAATCTAAACAGAGGTTTGGTACTGAATTATACTTACAAATTC
AAGTATGGCGTCGTTTGGCGGAAGCGTATGTGCGACAATCGACACTCGTGCGCATGGTGGACCGAAATCTAG
GCTCTTTGGTGTTACTGTCCAATATAAATAATTTGTACTTCATATGTTTGCAGATCTATCTTGGGATACACAAAT
CCTCCGGCTCAACAATCAGTCGGTGTTACTTCTTATTTTCTCTCGGCTGGTTGATATTCGCGCCTGCAGTGATG
TGTTGGCGGCATCAGACGTACATCTTCACTCGCAGAGAGCTCTCAAGAGTTTGCACGCGTGTCCCAGTGCCGC
TTATAATGTGGAGATCAAACGTCTGCAATACCAGTTGGCTCACGACTTTGTGGCTCTAACTGGCATGGGTTTCT
TCTCATTACGCAGGGAGCTACTACTAGAAGTAGCAGCTGCTATATTAAGTATGAATTAGTTTTAATACAGTAT
GATAAG

>HzeaGR5

ATGCAAAATGGATGGAACAACGTTATTTTGAACATCAGTGTTGGCAGTGTGAACACAGTGAATTATCTCTTCA
GAACATGGGAGCGGTTGGCACCGAATCGTAACATGGATTTGTATTCTTTGGAAAAGTTAAGAAGTAAAA
ATGATTGGAATTACCCAGTACATGTCCGCTATCAGGATCAAGTGATGGCTGAAAAAGAGAAACCATGTACAAC
CTTCAAACAGCCATGAGGGTGACCTTGACTATAGGCCAGTGTTTTGGCCTGAACCCAGTCCAGGGTATCAGA
GAGAAAGATGCTTCGAAATTGAGATTCAAATTACTATCAGGGCGTTGCCTATTCACGTTTTTATCTCTGATCGG
TCAGTTCATCATGGCGTTTGTATTATTTCTAAGTTTGTCAAAGAACTTCATCTACAGTAGATACAGCAAGTG
ATTTTGGCTTCACAACAACTATACTGTTTTTCCGCATTGCAACAAATTGGCCGAAGCTATGCATGCATATTGCA
AAAGTGGAGAGTGTAGATCCTAATACAGACAACAAATTAGGAAAGAAGTTCAATATTGCTTGTTCGATAC
TGTTCTTGGCGTTGATGGAGCACTTATTTTCAAGACTGCATGGAATATCCATAGCCTTGGACTGCTTTCCTGAT
ACTCTGTGTATGAGAGCTTCATGAAACACAGTTTTTCAGTGGCTCTTGGTTTCATACCGTACAGTGACTTCGC
AGGAGGTATGGCTCATTTTTCTAACCTACAATGCACATTTAACTGGAACCTTGGCGATGTATTGTCATGTGCA
TGAGTATGTACTTGACCGCGAGGCTCGAACAAGTCAATCAAAGGATTATTGCTGTAAAGATAAAGAAATCCCC
ATCGTCATTCTGGCGACAATGCGAGAAGACTACAACAGATCAGTGCACTTAGTCCGGCAGGTGGACAAGAT
CATCGGAGGCGTGGTCTTCATATCTTTCGCTAGCAACCTGTTCTTGTATGCTCACAACCTGCTGCATACACTCGC
AGGAGGTATAAAGGCGTCTCAGAGATGCAAGCCAGAAATTGGAGCAGATCGAAGATTCTTCTACGGCTACGA
ACACTCGATATACTTCGTGTTCTCCTTCTCTTCTGTTGATCAGGTCTCTGGCTGTCTCCCTCACTGCAGCCAA
GGTTCATGCAGCTAGCTTGGAGCCAGCCTACTCACTGTATGATGTGTCTTCTGCCAACTATTGTGTTGAGGTGG
AGAGGTTCTTGATCAGATACATGGGGACACGGTGGCGTTGAGCGGCCTTCAATTCTCCACGTCAAGAGAG
GACTTGATTGACGATTGCGGGTACCATCGTAACGTATGAGTTGGTGCTGATGCAGTTCACAGGAATCACTCC
AACGGCATCGCCAGAGTCCGTTAGTGGTGTTATTAAG

>HzeaGR6

ATGGGGCAAACAAGTTTCAGAAGAAATATGTCTTTCTGGATTCTGTAAAAAAGAACAAGGTAGACGTGGCTA
AGCCTAAAGTGAAGAACATCACATCATTCCAGGATGCTCTGAGAGTTACCGTCATCATAGGACAGGTCTTCTC
CTTACTCCCTTCGTGCGAGTATTTACCAATGTGGCTAGTAATGTCAAGTTTGTGAAAACTTCATGGAAATGTG
TATACTCGTTGCTGTCTCTCATTGGTCAAATGTTTCATGGCAGTACTGTGTATCAATAAGCTGGCAAAGACTACT
GTTTCCCTTAATGGCACATCGCCAGTAATATTCTACGTAACCACCTGCGTGACGATGATGCTGTTCTTCCAAGT
GGCTCGGCGCTGGCCCGCGCTGGTGCAGCACATATCTAAAGCCGAGGACATGGATCCCAACTTCGATTGCAG
CCTCACTAGGAAGTGAATATTACTTGCGCTGTGTCCTGATACTGGCTTTGTGTGAACACATACTATCGTTAT
TATCAGCATTGCGGGAGCCTCGGCCTGCTATTAGGAATGGATACTTACGAAGGTTTCGTCACCCATTTCTAT
CCTTGGGTGTTACCTATCTACCTATTCCGTCGTATTAGGTGTTTTAACCAGTTTCTCCACTTCCAGTCGACGT
TCATATGGAAGTTCTCCGACTTGTTTGTGATCTGCATGAGCTACTATCTGACGTCACGACTGGAGCAGGTCAAC
AGGAACTGCTGGCGGCTCAGGGCAAGTATTTACCAGAGATTTTCTGGAGGGCTACTCGTGAGGACTACTGC
AGAGCAACTCAGATAGTGCGGAAAGTAGACGAGGTGATCAGTGGAGTCGTCTTCATATCATTGCTAACAAC
CTCTTCTCATCTGTCTGCAACTCTTCAATACTTTGGAAGACGGCTTAAAAGGCACAGGAGAATGCACTCCGAA
GTTGAAAAAATCGTAGTATCAAAAACCGGGCCACTAGGGGGTCACGAGGCCGAGCATACTTCTCTTCTCC
CTGGTTTACTTGCTGTCTCGGTCTGTAGCAGTGTCTCTCATAGCTTCTCAAGTGAACAGTGCTTCGTAGTACCT
GCGCTGTGCTGTATGATGTTCTTACCTGTGTATTGTGTTGAGGTGCAGAGGTTTTTGGACCAGGTCAACG
GGGACAAAGTGGCGTTGAGTGGCCTCCAGTTCTTCTCAGTAACCAGAGGATTACTTTAACTGTGGCGGGCAC
CATAGTGACGTATGAGCTAGTAATGTTCCAGTTCAACTCATCGACACCCACCCTCAATATAACTTCTCCCACTGT
TGTTACTCACACTATTACAACACTAGCTACT

>HzeaGR7

ATGTCAAGTCGAGGGTTCGGACAATTTCTCGGGATGGAAATCTACTTCTGCCTGAACAACCAATCATGACG
ACTTTCTAACAGTCATGGAAAAGGTGTTCAAATGGTCTTGTCTAATCGGAGTGCTGGGTCCAAGCGACACAT
CAACTATGCCTGGAGTGGATTCACTCTTAGTGCTACTATTCATGGAGAGTCAAGCTATATGGAAGGTCATA
AAAGCACTCGCTGGATGGGCTATCGATACTGCTGGACAAAGGAGCGTCACAGCCCGTTTAGCAGGGACCATC
TTCTACACCATAGCAATTTATCCCTGGTCTTGTATCAAGGCTGTACAGATCTTGGGGAGAGCTATCAGCGTT
ATGGGCCAGGGTTGAGAGGATCATGGCTGTCAAAGCCCCGCTGATAAGACCTTGAAGAGAAGAATGTATTT
CTTCTTAGGATTTATGACCGTGTGTTCAATTGTTGGAACATCTTATGAGTGTTGTTTCTGCAATAGGACTGGACT
GTCCGCCAGCTTTGATAATAAAGAGATACGTTTTGATATCACATGGATTTATGATATTAAGACATGAGTACTCC
GACTGGTACGCCCTGCCTCTAATCTTCATGAGCACATTAGCTTCTCTGCTATGGAAGTTCCAAGACGTACTAAT
AGTACTGATCAGTATGGGTCTAACCTCAAGATATAGCAGACTAAACCAGTGCTTGGCGAAAATCTGTGCACTA
GAGAGGAAGCAGATGGATTCTGATAAAAAAACGAAGCAACAAAAGTATACGCATGGCGCAAATTAAGAGA
AGCCTACGTGAAACAAGCTATGTTAGTTCGAAAGTGGACGATGCCATCGGCGGCATCATTATACTGTCATGC
TTTTGCAACTTTTACTTTATATGTTTGCAGTTATTTTGGGAATTACTCAGAGCAAAGCCAGCGAGCCTATCAG
GACAGCCTACTACTTCATGTCATTAGGCTGGATTGCTTTCAGAGTCATCTGCGTTGTGCTAGCTGCCTCCGACA
TCAACGTCCACTCCCATTAGCTCTCAAACACATCTATACTCACGACAGTCATAGTTACAATATTGAGATGGGA
CGTCTACAGGACCAGCTGTCTAAAGACTACGTAGCTCTGAGCGGCAAGGGATTTTCTATTTGAGCAAAAGTA
TTTTGCTTCAGATGGCTGGCGGATAATCACTTACGAACCTATGTTGATACAATTTGATGACCAGGGGACTGAT
GATGTTCAACTGAATTTGACAAAAATGCTATTGGAGTA

>HzeaGR8

ATGTGAGCAAGGAATTCAAACAATTTTTCGACAAAACAAATTACTTTTACCTCAGCAACCGATTTATGATGA
CTTCTTAGACGTTATAGAAAAGGTGTTCCACTGGTCTTGTCTACGGTGTATTGCGATCCAAAAGGTTTCATCA
GTTTAATATGGAGCACTCTGATACTGGGATCTTTGGTCATCATCGAAGTCTTGGCTATCTGGAAGGTTATAAG

AGCACTCGCTGGCGTTGCTAGAGACATGAGTGGTCATCGAAGCGTAACAGCCCGCTAGCAGGAACCATCTT
CTATTCCATTTCAATACTATCCCTGGTGCTGATCTCAAACTTTACTACAATTGGAGAACTGATATAGCTGGAGT
ATGGGGTAAAGTGGAACGCTCTGTTGGTGTTAAATCCGGTGGATAAGACATTGAAATGTCGCATGTCTTTT
GTCGCTGGTCTTATGACGTTCTGTTCACTTTGAACATGCGCTGAGCATACTAGCATCAGTAGGATTTGACTG
TCCGCCATCTTTGATACTCAAGCGATATGTTCTAGTTTACATGGATTCATTTTATGGGACAGGACTATTCGG
AATGGTTTGCCATGCCTCTAGTCATTATAAGTACCATAGCAACTCTTCTATGGAACTTTCAAGACCAGCTGATA
GTACTCATCAGTATGGGCTTAACCTCAAGGTATAGAAGACTCAACGAATGTCTGGCCAAATTTTGTGAATTGG
AAAAGCACCATATGGATTCTGATAAGAAAGTGGAAGCAGTAAAAGTGTAACCTTGGCGCAAGATAAGAGAA
GCGTACGTGAAGCAAGCGATGCTGGTCCGCAAAATTGATGTCGCTCTCGGCGGTATTATCATCCTGTCTTGT
CCTGCAACTTCTATTTTATATGCTTGCAAATGTTTCTGGGAATTACGCAAGGGATGTCCACAGATTTTCTGACA
GGGGTATATTACATGGTATCTTTAGCTTGGCTATGTATCAGAGTGTTAAGCGTGGTATTAGCTGCATCTGGTGT
AAATACTCACTCAAAATTAGCGCTCAATCATCTTTATACTTATGAAACTCATTGTTATAATGTTGAGGTAGAGA
GACTCCAAGACCAACTTACTAAAGATTACATCGCTCTAAGCGGGATGGGTTTCTTCATCTAAACAAAACCTATT
TTGTTACAGATGGCAGGAGCAATAATAACTTATGAACCTGTGCTAATTCAGTTTGTATGATCAAGGCAGTGATG
GGATTGCATTAAACGCTACTAACATC

>HzeaGR10

ATGGAATATGGGTTAGATGCTAAAATTTGAAGGAATTAGAAAGTATCAATTTAGAGTATTGCAAGAACTC
AATTGGACGATGAAGCTAAAAACGAAGTACTAAGCCGTGGATTGTGGATGACAATCAATTGGTTGTTAAGA
AAAGAAATGATGTTAAAGATCAATTCACAGCTTTCCAAAAGCAATGAAGATTTGCTCGTATGGGGTCAAAC
CATAGGCCTCAATCCCGTGAAGTACTACTACAAAAGATCCTTCGAAAATGAGATTCACAGTGTACACGTGG
AAATTCCTCTTCTCGTTAACAATGGCGATCGCTCAGACCTTGAAGTACTTTATGCATCTATAAATTGTTTGA
GAACCTACCTCTATCAGTGCTTTGGGTTTTGTAACTTCTTACATCGACGTGTTTACCACATTTTGTTCATAC
TGATCGCTTCAAAATGGCCCGCTTGATGCAAGATATTGTTAGATCCAACTTGATGAATATGTTGATAAGAA
AATTATTACAAAGTGCAGGATAACTTGCTGTATCTTTATTGGTATGGCTTTGATGGAACACTTCTTGTGATATT
ATCAAGAGTGGCTAGAGTGATAGAATGTTCAAAAATGAAACAGACCATGGAGAGGTATTGTGAAGGTGAC
ATCTCCGTGGCTGTATGATTGAATGTACCATACGTTGTTGCAGTCGCTGTCATGGTCCAGTACGTGAATTTAA
TAACGACAGCGAACTGGAAGTACTCGTACATATTCATCGTATGTGTTAGCACGTATCTGTCTTCGATATTGAAC
CAAATTAATAAGAAGATCGAGTTGGAAGCTCAGAAGACTCACGTGCCAGCAAAAACCTGGATCAATCTAAGG
GAGGATTACACACGCGCGACGCACTTAGTCAAGCGATTGATGACGTCATCAGCGGGATTGTGCTCGTGACG
TACGCAAAATGATCTGTTCTTCATATGTCTGCAGCTTTATAATGTGCTCTCGAACATGTCGAAAGCGGCTCACT
CGTCAATAAATTGTGCTGATCAAGACGGGACTTTTCGAGCCTATTCTACCCTGCGTACCTGATATACTCAG
TACTCTATCTTTAGTAAGATTTCTGACTGTATCCATAGTGGCATCGGGAGTGAATACAGCCTCCTTGCTACCC
GCTCCATTCTCTATGGTATACCGACCACAGCGTATACTAAAGAGGTGGAAAGGTTCCAGAACCAAGTGAACG
GTGACGTCGTTGCACTCAGCGGTCTGCACTTTTTTACATCACCAGAGATCTCGTTCTTACTGTAAGTGTACCTA
GT

>HzeaGR11

ATGACATCGAAGCTATTCCTCAAGACATTCAACATGGCGACTAGACATCGGTTAGACAAGCGAGAGATATGTG
GCCTGCATAGTACTGTCAGAGGAACCTGTTCTGCAGCCGAGTGATGGGGTACTCCCTGTCTCTGGTCTCACT
TGTCCTACTAGCCGGAGGTTGAGGTTACATTCGGGTCTCCATACACGGTCCTGTACGTGGCGAGTTTGTTCG
GCCAGTACTGATGTTCTGTCATGACCTTCTGTTGGCTGATGATGAACGGCATCTCCTAGCTAATATAACAAAC
GCAGTATTCTACACATCAAGCCTGATTTTCATTAATTCTACTACACATCGGTCGCTGCTGGCCCTCGCTGGT
GGGACGCTGGAGACGCTGGAGCGGAGCTGCCTCCCTTCCACAGGAACGTGGCCTCTATCAGCAACACGCT

GGAGCGCGAGCTGCCTCCCTTCCACAGGAACGTGGCCTCCATCAGCAACGTTGTGGAACACCTCCTTTCAGTG
TTTTATGGGTTAAAAGTAGCATGCGCTTGTGACTCTAACAATGTGGCTGAAAACTATTTTCGCTTCAATATGCC
TTGGATCTTTGATTATACTCCTTTACTATATGGAAGGGGGCTCTTAGTGAGTTGTTCAATATTCAGTCGACGTT
TGTGTGGAGTTTGAACGACCTACTCATCATGGTGATCAGCATCTACCTGACGGAACATCTGCTCATACACAATG
AGCTCCTCAAGAAGGCTGCAGAACAGGAGCACTTTTCTGCCTAGAATTCCGGACGCAATACTTGAAGATAGT
CCGCCTAGTGAAGCTGATTAATGGTCAGTTCGGCATCTATATACTGACTTCGTTCCGGCAGCAACCTGTATTGGA
TATGTACGCAACTTTTTATAGCTTGAGTCGGACGCAAACGGGACATTTTATTACCTGTACGTTCAAGGATTCA
CAGACTAAAGTGCCTCCTGTCAATGAATACGAGAATCCCCTTTGTCCGTGGATGTTAGGGGAAAAAGGGGCCT
TGAATGGCGTAGAACACTCGATATATTTACATATTCGTTCTCGTTTCTGTTGCTGAGAACTCTGCTCGTATTGT
TGCTGGCTGCCAGGATACATTCCAATTCTGTTGCCCCCTTTATGTGCTGTACGGAATCCCAGTTCCTCGTTTC
ATATTGAGGTGGAGCGATTTATTGCACAAATAAATAACTTGAAAGTCGCCATGAGCGGGTTGGACTTCTTCTA
TGTTACAAGGACTATGATTTAACTCTCCTGGGAACTATCGTCACATATGAACTAGTATTGCTACAATTCAACA
GA

>HzeaGR12

ATGAAAGTCCACGAACTGACAATGAGAAGATCCAACAAGCGTTGTGGTTTACATGTATGCTTACGTACACGCTA
TGCGGCTAGCCCGATGGACAGGATTCTTTCCTCTTCAGGGTCTTGGACAAGCCTACGCTGATGGGACCAGGTA
CAAAATTTTGTCTCTCTACTTCAATTACAATTCGCCACCTTATTGGGACAAGTGTGATGCTTCTGCTTCGCTATT
CTACTCTTCTTCAAAGTCAAGTCACTCTGAATTCTATATCAAACGTCATATTTTACGTGACGTCACTAATTTTCG
GCGGTATTATTTCTGAAGTTAGCGAAACAATGGCCGAGGTTGATGGCGAGGGCAACTGAAACCGAGCAGGG
GTTAACTGAACTGAAGTTGCAAAACAAAGTCATTGTTAAATGCTGCGTGATCGCTTACGTTGCTATGGCACTG
GCTTTGGTGGAACACATCCTCTGCACGAGTTTCAACTTGACATTTGTCATGCATTGCCTCAAAGAAGCCGGGAT
TACTACCAATGTGATGGAGAAGTATGTGGTGCACAGGATGCCATATGTCTTCAAATATGTACCTTACAGCCTTT
TCTGGGCTTTCTTGTGTTGAGTACCTGTGCTTTCAGTCTACCTTCGTTTGGAGCTTCAATGATGTCCTGATCACCT
GTTTCAGTATTTATATTACAGCTTACTTTTCGGAGTTTGAATCAAGTCGTTACTGCTAATTTAAAAAGGACAAG
GATAATATGATACCTTGGTCCACTCTCAGGGTGCCTACTCCAAGTTAGTGAGACTGGTAAAAGAAATTGATA
ATCAGATCAGTTCCTTTATTCTGCTGGCGTTCTTCACTGATCTTTTCTACATCTGCTTGCAGCTATTTAACTCTTT
GCATCGAAATTACGTGAGCTTCAAATTTTGCAATGAACTACAGACCAAAACAAGCTCTAACGTCGCCCTCATACC
TCCTTTACTACTTATATTCCTTTATATTCCTGGTACTCCGAGCCACGATGCTATCACTCTTCGCTTCTAACGTACA
CTGTGCTGCTTTGGAACCTGTGCACGCCGTGTATGATGTGCCATCTACATTTATGATAATGAGGTTTCGTCGAT
TTCAGTTGCAACTTCATCATACTCAAGTAGGATTGACCGGAAAGTTCTTTTATGTCACAAGGAATATGGTTCTG
AAAGTAATCGGCACAATAATACTTACGAAATAGTTTTGTTGTCAGTACACCATAACGCCAAACCCTTATTATAA
TGGTACAAAAGTAATACTGAATATTTCTTCGCACAGTTACAGT

>HzeaGR4

MEIKLCKLFFVLTTRIEYMKGNSVKNLNEKKRDDFLPTLNSVFLKARFFGISGYGLTISF
CWSLILFMSLVVMESVAIWRVVTLGWLMSASNNGLIGRLSGAIFYGNALISLFLSSKF
VHSWRSLSNYWLRMETSTALDFPPDVIRKRTIYITAFVVSVAVVEHILSMISATGVGF
PEEFLYRYVTLSHGFILKAQDYTIWKAIPVVL SKLATALWNFQDLIIILISMGLSSRYD
RLNLYVRHVVSVEKQFESKQRFGTLYLQIQVWRRLREAYVRQSTLVRMVDRLGSLVLL
SNINNLYFICLQIYLG I HKSSGSTISRCYFLFSLGWLIFRACSVVLAASDVHLHSQRALK
SLHACPSAAYNVEIKRLQYQLAHDFVALTGMGFFSLRRELLLEVAAA I LKYEVLVLIQYDK

>HzeaGR5

MQNGWNNVISNISVGSVNTVNYLFRTWERLAPNRNMDLYSLEKFKKYKNDWNYPVHVRYQ
DQVMAEKEKPCTTFTQAMRVTLTIGQCFGLNPVQGIREKDASKLRFKLLSGRCLFTFLSL
IGQFIMAFVLFLSLFKETSSSTVDTASDFGFTTTILFFRIATNWPKLCMHIAKVESVDPNT
DNKLGKKFNIAFCSILFLALMEHLFSELHGISIALDCFPDTPVYESFMKHSFQWLFGFIP
YSDFAGGMAHFSNLQCTFNWNFADVFVMCMSMYLTARLEQVNQRIIAAKDKNSPSSFWRT
MREDYNRSVHLVRQVDKIIGGVVFISFASNLFVCSQLLHTLAGGIKASQRCKPEIGADR
RFFYGYEHSIYFVFSFSLVIRSLAVSLTAAKVHAASLEPAYSLEYDVSSANYCVEVERFL
DQIHGDTVALSGLQFFHVKRGVLVLT I AGTIVTYELVLMQFTGITPTASPESVSGVIK

>HzeaGR6

MGQTSFRRNMSFWIPVKKNKVDVAKPKVKNITSFQDALRVTVIIGQVFSLLPFVGVFTNV
ASNVKFVKTSWKCVYSLLSLIGQMFAVLCINKLAKTTVSLNGTSPVIFYVTTCVTMMLF
FQVARRWPALVQHISKAEDMDPNFDCSLTRKCNITCAVVLILALCEHILSLLSAFAGASA
CYSGMDTYEGFVTHFYPPWVFTYLPYSVVLGVLTQFLHFQSTFIWNFSDLFVICMSYYLTS
RLEQVNRKLLAAQGYLPEIFWRATREDYCRATQIVRKVDEVISGVVFISFANNLFFICL
QLFNTLEDGLKGTGECTPKLKKIVVSKTGPLGGHEAAAYFLFSLVYLLSRVAVSLIASQ
VNSASSVPAPVLYDVSPVYCEVQRFLDQVNGDKVALSGLQFFSVTRGLLLTVAGTIVT
YELVMFQFNSSTPTLNITSPTVVTHTITTLAT

>HzeaGR7

MSSRGFGQFLRDGNLLLPEQPNHDDFLTVMKEVFKWSC LIGVLGSKRHINYAWSGFILLV
LLFMESQAIWKVICALAGWAIDTAGQRSVTARLAGTIFYTIAILSLVLSSRLYRSWGELS
ALWARVERIMAVKAPDPKTLKRRMYFFLGFM TVCSLLEHLMSVVSAIGLDCPPALI I KRY
VLISHGFMI LRHEYS DWYALPLIFMSTLASLLWNFQDVLIVLISMGLTSRYSRNLQCLAK
ICALERKQMDSDKKNEATKVYAWRKLREAYVKQAMLVRKVDDAIGGIIILSCFCNFYFIC
LQLFLGITQSKASEPIRTAYYFMSLGWICFRVICVLAASDINVHSRLALKHIYTHDSHS
YNIEMGR LQDQLSKDYVALSGKGFFYLSKSILLQ MAGAIITYELMLIQFDDQGTDDVQLN
LTKNAIGV

>HzeaGR8

MSSKEFKQFLRQNKLLLPQQPIYDDFLDVIEKVFWHSCFYGVFGSKRFISLIWSTLILGS
LVIIIEVLAIWKVIRALAGVARDMSGHRSVTARLAGTIFYSISILSLVLISKLYYNWRTDI
AGVWGKVERS VGVKIPVDKTLKCRMSFVAGLMTFC SFFEHALSILASVGFD CPPSLILKR
YVLVSHGFIFMGQDYSEWFAMPLV I ISTIATLLWNFQDQLIVLISMGLTSRYRRLNECLA
KFCELEKHHMDSDDKKEAVKVYTWKIREAYVKQAMLVRKIDVALGGIIILSCSCNFYFI
CLQMFLGITQGMSTDFLTGVYYMVSLAWLCIRVLSVLAASGVNTHSKLALNHLYTYETH
CYNVEVERLQDQLTKDYIALSGMGFFHLNKTILLQ MAGAIITYELVLIQFDDQGS DGIAL
NATNI

>HzeaGR10

MEYGLDAKISKELESINFRVLQETQLDDEAKKRSTKPWIVDDNQLVVKKRNDVKDQFTAF
QKAMKILLVWGQTIGLNPVTGILQKDPSKMRFTVYTWKFLFSLTMAIAQTIGTTLCIYKL
FREPTSISALGFVTFFTSTCFTTFLFILIASKWPALMQDIVRSKLDEYVDKKIITKCRIT
CCIFIGMALMEHFLSILSRVARVIECSQNETDHGEVFVKVTS PWLYDLNVPYVVAVAVMV
QYVNLITTANWNYSYIFIVCVSTYLSSILNQINKKIELEAQKTHVPAKTWINLREDYTRA
THLVKRFDDVISGIVLVTYANDLFFICLQLYNVLSNMSKAAQLVNKLCPDQDGTFRAYS
PAYLIYSVLYLLVRFLT VSI VASGVNTASLLPAPILYGIPTTAYTKEVERFQNQVNGD
VVALSGLHFFYITRDLVLT VSVPS

>HzeaGR11

MTSKLFLKTFNMATRHRLDKREICGLHSTVRGTLFCSRVMGLLPVSGLTCPTSRRRLRFTF
RSPYTVLYVASLFGQLLMFVMTFCWLMNGISLANITNAV FYTSSLISSLILLHIGRCWP
SLVGSVETLERELPPFHRN VASISNTLERELPPFHRN VASISNVVEHLLSVFYGLKVACA
CDSNNVAENYFRFNMPWIFDYTPFTIWK GALSELFNIQSTFVWSLNDLLIMVISIYLTEH
LLIHNELLKKA AEQE HFSCL EFR TQYLKIVRLVKLINGQFGIYILTSFGSNLYWICTQLF
YSLSR TQTGHFITCTFKDSQTKVPPVNEYENPLCPWMLGEKGALNGVEHSIYFTYSFSFL
LLR TLLVLLLAARIHSNSVAPLYVLYGIPSSRFHIEVERFIAQINN LKVAMSGLDFFYVT
RTMILTLLGTIVTYELVLLQFNR

>HzeaGR12

MKVHELTMRRSNKRCGLHVCLRHAMRLARWTGFFPLQGLGQAYADGTRYKILSLYFIYNF
ATLLGQLVMSCFAILLFFQTEVTLNSISNVIFYVTSLISAVLFLKLAKQWPRLMARATET
EQGLTELKLQNKVIVKCCVIAYVAMALALVEHILCTSFNLTFVMHCLKEAGITTNVMENY
VVHRMPYVFKYVPYSLFWAFLFEYLC LQSTFVWSFNDVLITCF SIYITAYFRSLNQVVT
NSKKDKDNMIPWSTLRVHYSKLVRLVKEIDNQISSFILLAFFTDLFYICLQLFNSLHRNY
VSFKFCNELQTKQALTSPSYLLYYLYSFI FLVLRATMLS LFASNVHCAALEPVHAVYDVP
STLYDNEVRRFQLQLHHTQVGLTGKFFYVTRNMVLKVIGTIITYEIVLLQYTITPNPYYN
GTKVILNISSHSYS