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

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Genome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>gen-2017-0086.R1</td>
</tr>
<tr>
<td>Manuscript Type</td>
<td>Article</td>
</tr>
<tr>
<td>Date Submitted</td>
<td>06-Jul-2017</td>
</tr>
</tbody>
</table>
| 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 |
| Special Issue?  | This submission is not invited              |
| Keyword:        | Gustatory receptor, sugar receptor, Helicoverpa armigera, gene expression |

[https://mc06.manuscriptcentral.com/genome-pubs](https://mc06.manuscriptcentral.com/genome-pubs)
Molecular Characterization of Sugar Taste Receptors in Cotton

Bollworm *Helicoverpa armigera*

Wei Xu¹*, Naiyong Liu², Yalin Liao¹, Alisha Anderson³

WX: W.Xu@murdoch.edu.au
NYL: Naiyong_2013@163.com
YL: Y.Liao@murdoch.edu.au
AA: Alisha.Anderson@csiro.au

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

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

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

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

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

Sugar detection primarily occurs at taste sensilla distributed throughout the insect body, especially on the major chemosensory tissues such as antennae, mouthparts, tarsi and ovipositors (van der Goes van Naters and Carlson 2006). Gustatory receptors (GRs) are localized on the dendrites inside the gustatory sensilla and act as an interface between insects and their environmental taste chemicals (van der Goes van Naters and Carlson 2006). GR genes were firstly identified from the Drosophila melanogaster genome using sequence similarity and phylogenetic clustering (Clyne et al. 2000). Insect GRs have been classified into the following 4 clades: “GR43a-like” (Abdel-Latif 2007; Sato et al. 2011; Xu et al. 2012), “CO₂” (Jones et al. 2007; Kirkness et al. 2010; Ning et al. 2016; Xu and Anderson 2015), “bitter” (Xu et al. 2016) and “sugar” (Chyb et al. 2003; Dahanukar et al. 2001; Dahanukar et al. 2007; Jiao et al. 2007; Jiao et al. 2008a; Slone et al. 2007; Xu et al. 2015; Xu et al. 2016; Zhang et al. 2011). Although the “GR43a-like” receptors have been shown to 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
silico expression profiles of these sugar receptors. This study improves our understanding of insect sugar receptor family, insect-plant interaction and evolution.
Materials and methods

Insect rearing, tissue collection and RNA purification

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

RT-PCR and gene cloning

The cDNA templates were prepared from purified RNA samples using the SuperScript® VILO™ cDNA Synthesis Kit (Invitrogen, USA), according to the manufacturer's manual. RT-PCR was performed using gene-specific primers (Table 2) for the full length *H. armigera* sugar GR ORF sequences (Xu et al. 2016). The PCR program was performed by using Phusion High-Fidelity DNA Polymerase (New England BioLabs, USA) as follows: 95 °C for 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 72 °C for 10 min. PCR products were purified using QIAquick gel extraction reagents (Qiagen, USA), cloned into the pBluescript SKII (-) vector (Stratagene, USA) and transformed to competent *Escherichia coli* cells DH5α (New England BioLabs, USA). Three randomly selected colonies were picked for plasmid purification and DNA sequencing. The sequencing results were compared to previously annotated sequences and exon/intron analysis ([http://wormweb.org/exonintron](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 *Acyrthosiphon 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
testes and female ovaries from day 0 to day 5 adults; embryo, 3\textsuperscript{rd} instar larvae (whole body),
post-feeding stage larvae (whole body) and pupae were sequenced and the sequenced data
were used here for expression studies (Xu et al. 2016).

\textit{In silico} expression profiles were generated using DEW (http://dew.sourceforge.net/)
which is an automated pipeline that: 1) used Bowtie 2 (Xu et al. 2016) to align the reads of
each library against a sequence file comprised of our assembly and the chemosensory genes;
2) post-processed the alignments with eXpress to account for isoforms and paralogues; 3)
performed Trimmed Mean Normalization using edger and estimated Fragments Per Kilobase
Per Million reads (FPKM) (Xu et al. 2016); 4) and visualized the results as described (Xu et
al. 2016).
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-Latif 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 (*HarmGr*4-KY806282, *HarmGr*5-KY806283, *HarmGr*6-KY806284, *HarmGr*7-KY806285, *HarmGr*8-KY806286, *HarmGr*10-KY806287, *HarmGr*11-KY806288 and *HarmGr*12-KY806289). *HarmGr*4 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 CO$_2$, 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; Pxy1GR1, 2 and 3. Lepidoptera-specific group 2 contains HarmGR4, 7, 8; BmorGR4; HmelGR4, 52; DpleGR3, 4, 37; Pxy1GR4 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 ancestoral genes. There are also two specific groups of sugar receptors in Hemiptera and Hymenoptra 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
candidate receptors for sucrose or myo-inositol. At adult stage, HarmGR4 was detected in female adult ovaries, tarsi, antennae, both male and female heads (with antennae) and male testes. HarmGR5 was expressed in adult tarsi and heads (with antennae) but not the antennae, suggesting they may be expressed in other tissues on heads, for example, the proboscis, an organ used to suck sugar-rich nectar, plant sap, tree sap, and pollen or complementary food.

HarmGR6 and 7 were both expressed in adult antennae and tarsi. Previous studies showed that in *H. armigera*, the tarsal gustatory sensilla play an essential role in perceiving sugars available in floral nectar and provide chemosensory information determining feeding behaviour (Zhang et al. 2010). HarmGR8 was expressed in post-feeding larvae stage, pupae stage and adult antennae (Fig. 5). HarmGR10 was expressed in male adult testes while HarmGR11 was expressed in the female ovaries (Fig. 5). This result was further confirmed by RT-PCR (Fig. 5B). The RT-PCR analyses of other HarmGR genes have been performed previously (Liu et al. 2014). Interestingly HarmGR10 showed specific expression in male testes while HarmGR11 showed specific expression in female ovaries. This is the first report of specific sugar receptors in lepidopteran reproductive systems. Previous studies showed by using Gr-GAL4 drivers in *D. melanogaster*, Gr28b.b, Gr28b.c, Gr32a, and Gr64c, were detected in neurons that appear to innervate both the male and female reproductive organs, suggesting these GRs possibly function in regulation of accessory gland secretion in males or regulation of ovulation in females (Park and Kwon 2011). Another study showed that GR43a, a fructose gustatory receptor, is expressed in the uterus of *D. melanogaster*, raising the possibility that it plays a role in reproduction or mating (Miyamoto et al. 2012). Furthermore, odorant receptor, Orco, was detected in the flagella of malaria mosquito, *A. gambiae*, spermatozoa. Ocro-specific agonists, antagonists as well as other odorant compounds could robustly activate flagella beating. Therefore, chemosensory receptors may mediate sperm 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.
Acknowledgments

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


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

<table>
<thead>
<tr>
<th></th>
<th>Predicted transmembrane configuration</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TMpred</td>
<td>HMMTOP</td>
<td>TMHMM</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>N-*</td>
<td>TMD</td>
<td>N-*</td>
<td>TMD</td>
<td>N-*</td>
</tr>
<tr>
<td>HarmGR4</td>
<td>+</td>
<td>7</td>
<td>+</td>
<td>7</td>
<td>+</td>
</tr>
<tr>
<td>HarmGR5</td>
<td></td>
<td></td>
<td>10</td>
<td>8</td>
<td>+</td>
</tr>
<tr>
<td>HarmGR6</td>
<td></td>
<td></td>
<td>9</td>
<td>8</td>
<td>+</td>
</tr>
<tr>
<td>HarmGR7</td>
<td>+</td>
<td>7</td>
<td>+</td>
<td>7</td>
<td>+</td>
</tr>
<tr>
<td>HarmGR8</td>
<td>+</td>
<td>7</td>
<td>+</td>
<td>7</td>
<td>+</td>
</tr>
<tr>
<td>HarmGR11</td>
<td></td>
<td></td>
<td>8</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>HarmGR12</td>
<td>+</td>
<td>8</td>
<td>+</td>
<td>7</td>
<td>+</td>
</tr>
</tbody>
</table>

* N-terminus locates intracellularly.

Transmembrane domain (TMD).
<table>
<thead>
<tr>
<th>HarmGR4F</th>
<th>ATGAAAGGGAATTCCGTCGAAG</th>
</tr>
</thead>
<tbody>
<tr>
<td>HarmGR4R</td>
<td>CTACTTATCATATTGTATTTAAAATAATTC</td>
</tr>
<tr>
<td>HarmGR5F</td>
<td>ATGCAAAATGGATGGAACA</td>
</tr>
<tr>
<td>HarmGR5R</td>
<td>CTACTTAATAACACCACGTGACAGACTC</td>
</tr>
<tr>
<td>HarmGR6F</td>
<td>ATGGGGCAAAACAAGTTTTAGA</td>
</tr>
<tr>
<td>HarmGR6R</td>
<td>TCAAGTAGCTAGTGGTATAGTGTGG</td>
</tr>
<tr>
<td>HarmGR7F</td>
<td>ATGTCAAGTCGAGGGTTC</td>
</tr>
<tr>
<td>HarmGR7R</td>
<td>CTATCTCCAATAGCATTTTTAGCA</td>
</tr>
<tr>
<td>HarmGR8F</td>
<td>ATGTCGAGTAAGGAATT</td>
</tr>
<tr>
<td>HarmGR8R</td>
<td>TTAGATGTTAGTAGCA</td>
</tr>
<tr>
<td>HarmGR10F</td>
<td>AATAAATTGGTTGGTAAGAAAAGAAAATG</td>
</tr>
<tr>
<td>HarmGR10R</td>
<td>CTAGTTCTTGTTCAATTCGTCGCTGCTG</td>
</tr>
<tr>
<td>HarmGR11F</td>
<td>ATGCCATCGAAACTATTTCTCAAGAC</td>
</tr>
<tr>
<td>HarmGR11R</td>
<td>TTATCTGGGACTGTAATACC</td>
</tr>
<tr>
<td>HarmGR12F</td>
<td>ATGAAAGTCCACCAAACCTGA</td>
</tr>
<tr>
<td>HarmGR12R</td>
<td>ACTGAAACTGTGCGAAG</td>
</tr>
<tr>
<td></td>
<td>GR4</td>
</tr>
<tr>
<td>----------------</td>
<td>-----</td>
</tr>
<tr>
<td>Thorax</td>
<td>0.01</td>
</tr>
<tr>
<td>Head</td>
<td></td>
</tr>
<tr>
<td>Antennae</td>
<td>0.12</td>
</tr>
<tr>
<td>Testes</td>
<td>0.22</td>
</tr>
<tr>
<td>Ovary</td>
<td>0.47</td>
</tr>
<tr>
<td>Foregut</td>
<td></td>
</tr>
<tr>
<td>Midgut</td>
<td></td>
</tr>
<tr>
<td>Malpighian tubules</td>
<td></td>
</tr>
<tr>
<td>Silk gland</td>
<td></td>
</tr>
<tr>
<td>Fat body</td>
<td></td>
</tr>
<tr>
<td>Ventral nerve</td>
<td></td>
</tr>
<tr>
<td>Hemocytes</td>
<td></td>
</tr>
<tr>
<td>Salivary gland</td>
<td></td>
</tr>
<tr>
<td>Heart</td>
<td></td>
</tr>
<tr>
<td>Trachea</td>
<td></td>
</tr>
<tr>
<td>Muscle</td>
<td></td>
</tr>
<tr>
<td>Epidennis cuticle</td>
<td></td>
</tr>
<tr>
<td>Antennae</td>
<td></td>
</tr>
<tr>
<td>Mouthpart</td>
<td></td>
</tr>
<tr>
<td>Embryo</td>
<td></td>
</tr>
<tr>
<td>3rd instar whole</td>
<td></td>
</tr>
<tr>
<td>Post_feeding</td>
<td></td>
</tr>
<tr>
<td>Pupa</td>
<td></td>
</tr>
</tbody>
</table>

* values in blank cell are all 0.

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

http://www.ebi.ac.uk/Tools/services/rest/clustalo/result/clustalo-I20170623-063359-0938-93626133-oy/pim

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>HzeaGR4</td>
<td>100.00</td>
<td>98.01</td>
<td>40.34</td>
<td>40.59</td>
<td>40.83</td>
<td>40.83</td>
<td>22.97</td>
<td>22.83</td>
<td>28.95</td>
<td>29.21</td>
<td>23.64</td>
<td>24.87</td>
<td>21.78</td>
<td>22.13</td>
<td>26.27</td>
</tr>
<tr>
<td>2</td>
<td>HarmGR4</td>
<td>98.01</td>
<td>100.00</td>
<td>40.50</td>
<td>40.75</td>
<td>41.50</td>
<td>41.50</td>
<td>23.96</td>
<td>23.81</td>
<td>29.11</td>
<td>29.38</td>
<td>23.43</td>
<td>24.73</td>
<td>22.04</td>
<td>22.40</td>
<td>26.01</td>
</tr>
<tr>
<td>3</td>
<td>HzeaGR8</td>
<td>40.34</td>
<td>40.50</td>
<td>100.00</td>
<td>98.59</td>
<td>65.33</td>
<td>64.15</td>
<td>26.02</td>
<td>26.43</td>
<td>26.14</td>
<td>25.45</td>
<td>26.15</td>
<td>24.61</td>
<td>24.20</td>
<td>23.58</td>
<td>23.32</td>
</tr>
<tr>
<td>4</td>
<td>HarmGR8</td>
<td>40.59</td>
<td>40.75</td>
<td>98.59</td>
<td>100.00</td>
<td>66.04</td>
<td>64.86</td>
<td>26.56</td>
<td>26.98</td>
<td>26.40</td>
<td>26.40</td>
<td>25.45</td>
<td>26.15</td>
<td>21.13</td>
<td>24.73</td>
<td>24.35</td>
</tr>
<tr>
<td>5</td>
<td>HzeaGR7</td>
<td>40.83</td>
<td>41.50</td>
<td>65.33</td>
<td>66.04</td>
<td>100.00</td>
<td>98.13</td>
<td>22.49</td>
<td>23.16</td>
<td>26.52</td>
<td>26.52</td>
<td>24.68</td>
<td>25.64</td>
<td>23.30</td>
<td>23.14</td>
<td>23.39</td>
</tr>
<tr>
<td>6</td>
<td>HarmGR7</td>
<td>40.83</td>
<td>41.50</td>
<td>64.15</td>
<td>64.86</td>
<td>98.13</td>
<td>100.00</td>
<td>22.49</td>
<td>23.16</td>
<td>26.77</td>
<td>26.77</td>
<td>24.68</td>
<td>25.64</td>
<td>23.04</td>
<td>22.87</td>
<td>23.39</td>
</tr>
<tr>
<td>7</td>
<td>HzeaGR10</td>
<td>22.97</td>
<td>23.96</td>
<td>26.02</td>
<td>26.56</td>
<td>22.49</td>
<td>22.49</td>
<td>100.00</td>
<td>96.14</td>
<td>39.32</td>
<td>38.83</td>
<td>35.35</td>
<td>35.63</td>
<td>35.00</td>
<td>23.00</td>
<td>23.23</td>
</tr>
<tr>
<td>8</td>
<td>HarmGR10</td>
<td>22.83</td>
<td>23.81</td>
<td>26.43</td>
<td>26.98</td>
<td>23.16</td>
<td>23.16</td>
<td>96.14</td>
<td>100.00</td>
<td>38.78</td>
<td>38.29</td>
<td>35.98</td>
<td>36.26</td>
<td>36.12</td>
<td>23.62</td>
<td>23.60</td>
</tr>
<tr>
<td>9</td>
<td>HzeaGR6</td>
<td>28.95</td>
<td>29.11</td>
<td>26.14</td>
<td>26.40</td>
<td>26.52</td>
<td>26.77</td>
<td>39.32</td>
<td>39.78</td>
<td>100.00</td>
<td>98.89</td>
<td>47.25</td>
<td>46.71</td>
<td>32.45</td>
<td>32.52</td>
<td>30.52</td>
</tr>
<tr>
<td>10</td>
<td>HarmGR6</td>
<td>29.21</td>
<td>29.38</td>
<td>26.14</td>
<td>26.40</td>
<td>26.52</td>
<td>26.77</td>
<td>38.83</td>
<td>38.29</td>
<td>98.89</td>
<td>100.00</td>
<td>47.25</td>
<td>46.71</td>
<td>32.45</td>
<td>32.52</td>
<td>30.28</td>
</tr>
<tr>
<td>11</td>
<td>HzeaGR5</td>
<td>23.64</td>
<td>23.43</td>
<td>25.45</td>
<td>25.45</td>
<td>24.68</td>
<td>24.68</td>
<td>35.35</td>
<td>35.98</td>
<td>47.25</td>
<td>47.25</td>
<td>100.00</td>
<td>97.69</td>
<td>31.20</td>
<td>31.76</td>
<td>29.93</td>
</tr>
<tr>
<td>12</td>
<td>HarmGR5</td>
<td>24.87</td>
<td>24.73</td>
<td>26.15</td>
<td>26.15</td>
<td>25.64</td>
<td>25.64</td>
<td>35.63</td>
<td>35.63</td>
<td>46.71</td>
<td>46.71</td>
<td>97.69</td>
<td>100.00</td>
<td>30.83</td>
<td>31.37</td>
<td>30.05</td>
</tr>
<tr>
<td>13</td>
<td>HzeaGR11</td>
<td>21.78</td>
<td>22.04</td>
<td>24.61</td>
<td>25.13</td>
<td>23.30</td>
<td>23.04</td>
<td>23.00</td>
<td>23.62</td>
<td>32.45</td>
<td>32.45</td>
<td>31.20</td>
<td>30.83</td>
<td>100.00</td>
<td>95.63</td>
<td>34.07</td>
</tr>
<tr>
<td>14</td>
<td>HarmGR11</td>
<td>22.13</td>
<td>22.40</td>
<td>24.20</td>
<td>24.73</td>
<td>23.14</td>
<td>22.87</td>
<td>23.23</td>
<td>23.60</td>
<td>32.52</td>
<td>32.52</td>
<td>31.76</td>
<td>31.37</td>
<td>95.63</td>
<td>100.00</td>
<td>34.16</td>
</tr>
<tr>
<td>15</td>
<td>HzeaGR12</td>
<td>26.27</td>
<td>26.01</td>
<td>23.58</td>
<td>24.35</td>
<td>23.39</td>
<td>23.39</td>
<td>26.53</td>
<td>26.41</td>
<td>30.52</td>
<td>30.28</td>
<td>29.93</td>
<td>30.05</td>
<td>34.07</td>
<td>34.16</td>
<td>100.00</td>
</tr>
<tr>
<td>16</td>
<td>HarmGR12</td>
<td>26.27</td>
<td>26.01</td>
<td>23.32</td>
<td>24.09</td>
<td>23.91</td>
<td>23.91</td>
<td>26.79</td>
<td>26.67</td>
<td>30.75</td>
<td>30.52</td>
<td>30.41</td>
<td>30.52</td>
<td>33.83</td>
<td>33.92</td>
<td>97.24</td>
</tr>
</tbody>
</table>
Supplementary data, *Helicoverpa zea* sugar receptor DNA and Amino acid sequences

>\text{HzeaGR4}

\begin{verbatim}
ATGGAAATTAAGCTGTGAACATTATTTTGGACAAAGAATTTGAATATGGAAGAGGAATTGCTGCA
AGAATTAAAAAGAAGAGAAGCTGAGACTCTTACACAAACTTAAACAGTGTGTTCCTCAAGGCTCGATTCTTC
GGTATATCTGGGTATGGGCTCACAATATCATATTTTGGTGAGCATATATTGAGCATGATA
\end{verbatim}

>\text{HzeaGR5}

\begin{verbatim}
ATGCAAAATGGATGGAACAACGTTATTTCGAACATCATGTTGAGATGGAACACAGTGATATATCTCTTCA
GAACATGGGAGCCTGTTGACACCAGAATGTAAGATGCTGAAAAAGAAGAGGACATGACACAC
CTTCCAAACAGCCATGACATCTGACGCTATCAGGATCAAGTGATGGCTGAAAAAGAGAAACCATGTACAAC
CTTCCAAACAGCCATGACATCTGACGCTATCAGGATCAAGTGATGGCTGAAAAAGAGAAACCATGTACAAC
\end{verbatim}

>\text{HzeaGR6}

\begin{verbatim}
ATGCAAATGTGAACTTCTGCAACATATTTCGATGATGGAATATGGAAGAGGAATTGCTGCA
\end{verbatim}
ATGGGGCAAACAAGTTTCAGAAGAAATATGTCTTTCTGGATTCCTGTTAAAAAGAACAAGGTAGACGTGGCTAAGCCTAAAGTGAAGAACATCACATCATTCCAGGATGCTCTGAGAGTTACCGTCATCATAGGACAGGTCTTCTCCTTACTCCCTTTCGTCGGAGTATTTACCAATGTGGCTAGTAATGTCAAGTTTGTGAAAACTTCATGGAAATGTGTATACTCGTTGCTGTCTCTCATTGGTCAAATGTTCATGGCAGTACTGTGTATCAATAAGCTGGCAAAGACTACTGTTTCCCTTAATGGCACATCGCCAGTAATATTCTACGTAACCACCTGCGTGACGATGATGCTGTTCTTCCAAGTGCGGGCGCTGGCCCGCGCTGGTGCAACATATCTAAAGCCGAGGACATGGATCCCAACTTCGATTGCAGCCTCACTAGGAAGTGCAATATTACTTGCGCTGTCGTCCTGATACTGGCTTTGTGTGAACACATACTATCGTTATATCAGCATTTGCGGGAGCCTCGGCCTGCTATTCAGGAATGGATACTTACGAAGGTTTCGTCACCCATTTCTATCCTTGGGTGTTCACCTATCTACCTTATTCCGTCGTATTAGGTGTTTTAACCCAGTTTCTCCACTTCAGTGCAGTTCATATGGAACTTCTCCGACTTGTTTGTGATCTGCATGAGCTACTATCTGACGTCACGACTGGAGCAGGTCAACAGGAAACTGCTGGCGGCTCAGGGCAAGTATTTACCAGAGATTTTCTGGAGGGCTACTCGTGAGGACTACTGCAAGAGCAACTCAGATAGTGCGGAAAGTAGACGAGGTGATCAGTGGAGTCGTCTTCATATCATTCGCTAACAACCTCTTCTTCATCTGTCTGCAACTCTTCAATACTTTGGAAGACGGCTTAAAAGGCACAGGAGAATGCACTCCGAAGTTGAAAAAAATCGTAGTATCAAAAACCGGGCCACTAGGGGGTCACGAGGCCGCACACTCCTCCTTCTCCCTGGTTTACTTGCTGTCTCGGTCTGTAGCAGTGTCTCTCATAGCTTCTCAAGTGAACAGTGCTTCGTCAGTACCTGCGCCTGTGCTGTATGATGTTCCTTCACCTGTGTATTGTGTTGAGGTGCAGAGGTTTTTGGACCAGGTCAACGGGACAAAGTGGCGTTGAGTGGCCTCCAGTTCTTCTCAGTAACCAGAGGATTACTTTTAACTGTGGCGGGCACATAGTGACGTATGAGCTAGTAATGTTCCAGTTCAACTCATCGACACCCACCCTCAATATACTTCTCCACTGTGTTACTCACACTATTACAAACACTAGACT

>HzeaGR7

ATGTCAGTCGAGGGTTGGAGCTTCCGGAACAATTTTCTTCCGGGATGAAAATCTACTTCTTGGCTGAAAAACACCATCATGACGACCTTCTAAACATGCGATGGATCTTGTCAATCGGAGTGCTGGTTCAGCGACATCAATACTGCCAGAATGCTTCTTTGCTCTGTGACTGACTATTATATGAGGTTAAACAGTGAATCTGAATCTGAGATTTTGGTATGAGGTTTTTAACTGTGGCGGGCACATAGTGACGTATGAGCTAGTAATGTTCCAGTTCAACTCATCGACACCCACCCTCAATATACTTCTCCACTGTGTTACTCACACTATTACAAACACTAGACT

>HzeaGR8

ATGTCAGCGAAGGAATTTTCTTCCGGGATGAAAATCTACTTCTTGGCTGAAAAACACCATCATGACGACCTTCTAAACATGCGATGGATCTTGTCAATCGGAGTGCTGGTTCAGCGACATCAATACTGCCAGAATGCTTCTTTGCTCTGTGACTGACTATTATATGAGGTTAAACAGTGAATCTGAATCTGAGATTTTGGTATGAGGTTTTTAACTGTGGCGGGCACATAGTGACGTATGAGCTAGTAATGTTCCAGTTCAACTCATCGACACCCACCCTCAATATACTTCTCCACTGTGTTACTCACACTATTACAAACACTAGACT

https://mc06.manuscriptcentral.com/genome-pubs
AGCAGCTCGTGGCCTGCTAGAGACATGAGTTGTCATCGAAGCGTAACAGCCCGCCTAGCAGGAACCATCTTCTATTCCATTCAATACTATCCCTGGTGCTGATCTCAAAACTTTACTACAATTGGAGAACTGATATAGCTGGAGTATGGGGTAAAGTGGAACGCTCTGTTGGTGTTAAAATTCCGGTGGATAAGACATTGAAATGTCGCATGTCTTTTGTCGCTGGTCTTATGACGTTCTGTTCATTCTTTGAACATGCGCTGAGCATACTAGCATCAGTAGGATTTGACTG

>HzeaGR10

ATGGGAATATGGGGTTAGATGCTAAAATTTCGAAGGAATTAGAAAGTATCAATTTCAGAGTATTGCAAGAAACTC

>HzeaGR11

ATGACATCGAAGCTATTCCTCAAGACATTCAACATGGCGACTAGACATCGGTTAGACAAGCGAGAGATATGTGGCCTGCATAGTACTGTCAGAGGAACTCTGTTCTGCAGCCGAGTGATGGGGTTACTCCCTGTCTCTGGTCTCACTTGTCCTACTAGCCGGAGGTTGAGGTTCACATTCCGGTCTCCATACACGGTCCTGTACGTGGCGAGTTTGTTCGGCCAGCTACTGATGTTCGTCATGACCTTCTGTTGGCTGATGATGAACGGCATCTCCTTAGCTAATATAACAAAGCAAGTATTCTACACATCAAGCCTGATTTCATCATTAATTCTACTACACATCGGTCGCTGCTGGCCCTCGCTGGTGGGCAGCGTGGAGACGCTGGAGCGCGAGCTGCCTCCCTTCCACAGGAACGTGGCCTCTATCAGCAACACGCTCCTCATTTTTACATCACCAGAGATCTCGTTCTTACTGTAAGTGTACCTAGT

https://mc06.manuscriptcentral.com/genome-pubs
GGAGCGCGAGCTGCCTCCATCAGCAACGGTTGTGGAACACCTCCTCTTCACTGTGTTTTATGGGTTAAAAGTAGCATGCGCTTGTGACTCTAACAATGTGGCTGAAAACTATTTTCGCTTCAATATGCC
TTGGATCTTTGATTATACTCCTTTTACTATATGGGAAGGGGGCTCTTAGTGAGTTGTTCAATATTCAGTCGACGT
TGTGTGAGTGGTGAAGCACTACCTACTACTCATGTGATCAGCATCTACCTGACGGAACATCTGCTCATACACAATG
AGCTCCTCAAGAAGGGCTGCAAGAACACGGAGCATTCTTTTCGGCTGAATTCTGGAAGATAGTG
CGCCTAGTGAGAAGGCTGAGTTAAATTGTCAGTCTGGCAGTACTATATACGTACCTTGGCGACAGACACTAGTTGGA
TATGTGACAAACTTTTTTATAGCTTGGCAGCCGCAAGGGGCTATTTTATTACCTGCTGAGATTTCA
CAGACTAAAGTGCTCTCTGCAATGAATACGAGAATCCCCTTTGCAATGGACAGCAATCTTTTTATAGCTTGAGTC
GGACGCAAAGGGATCAAAAGTAGCTCTCCTCGTACATATTTATATGTCATGCATTGCCTCAAAGAAGCCGGGAT
TACTACCAATGTGATGGAGAACTATGTGGTGCACAGGATGCCATATGTCTTCAAATATGTACCTTACAGCCTTT
TCTGGGCTTTCTTGTTTGAGTACCTGTGTCTTCAGTCTACCTTCGTTTGGAGCTTCAATGATGTCCTGATCACCT
GTTTCAGTTGGAAGCAAAAGCACATCTCGTCCACCCGTTTTATAGGCTCAATGGACAGCAAACTTTTTATAGCTTG
GAGATTGACCGGAAAGTTCTTTTATGTCACAAGGAATATGGTTCTGAAAGTAATCGGCACAATAATAACTTACGAA
ATCAGATCGTCTCTTATTTCTGCGGTTCTTCACGTGATCTTTTCTACATCTGCTGGAGCTATTAACTCTTT
GCACTGAAAATACGTCAGTCTTCAAATTTTGCAATGAACTACAGACCAAACAAGCTCTAACGTCGCCCTCATACC
TCTTTATCTATATTTCTTACCTAGTGACGTCTCCGAGCCGACAGTGCATCTCCTTGCTTCATACATGACA
CTGTGCTTGGAGACCTGTCGCTGAGGCTGTATGATGTCGACACTACATTATATGATAATAGGATGGTCTCGAT
TTCAGTTGCAAACTCATACATCAAGTAGATGACGCCGAAGGTCTTTTTATGTCACAGAAATAGTGGTCTCGT
AAAGTAAATCCGCAACAATAAAAATCTACGAAATATTTTTGCTGCACTGACATACAAACGCGGAAAACCTTTATATGAA
TGGTACAAAGTGAATATGCAATATTTTCCTCCACAGTTACG

>HzeaGR12
ATGAAAGTCCACGAACTGACAATGAGAAGATCCAACAAGCGTTGTGGTTTACATGTATGCTTACGTCACGCTA
TGCGGCTAGCCCGATGACAGAGATTCTTTTCTCCTCCTCTCAGGGTCTTGCGAAGGACACCCCTATCTGGAACAGGTA
CAAATTTTTGCTCTCTACTCTACATTTCTCCTGCCACCTTACTTGGGAACACTAGTTGCTTGCTGACTATT
CTACTCTTTCTCAACTGAAGTCACTCGTGAATTCATATACACGTCTATTTTACGTGACGTCACTAAATTTC
gCCGGTATTATTTCTGGAAGTAAAGGAAACAATGGCCGAGGGTTGATGCGAGGCGGAACTGAAAGGAGAAGGGA
GTTAACTGAACGGAACTGGCCACAAAGACTATTATTACTGTTGCTCGATGCTACAGTGGCTACTGGGACTGT
GCTTTGGTGAACACATCTCGTCCACGGAGTATTTTAACTGGGACATGCTGACTTCTCCAAAGAAGCCGGGAT
TACTACCAATGTGATGGAGAACTATGTGGTGCACAGGATGCCATATGTCTTCAAATATGTACCTTACAGCCTTT
TCTGGGCTTTCTTGTTTGAGTACCTGTGTCTTCAGTCTACCTTCGTTTGGAGCTTCAATGATGTCCTGATCACCT
GTTTCAGTTGGAAGCAAAAGCACATCTCGTCCACCCGTTTTATAGGCTCAATGGACAGCAAACTTTTTATAGCTTG
GAGATTGACCGGAAAGTTCTTTTATGTCACAAGGAATATGGTTCTGAAAGTAATCGGCACAATAATAACTTACGAA
ATCAGATCGTCTCTTATTTCTGCGGTTCTTCACGTGATCTTTTCTACATCTGCTGGAGCTATTAACTCTTT
GCACTGAAAATACGTCAGTCTTCAAATTTTGCAATGAACTACAGACCAAACAAGCTCTAACGTCGCCCTCATACC
TCTTTATCTATATTTCTTACCTAGTGACGTCTCCGAGCCGACAGTGCATCTCCTTGCTTCATACATGACA
CTGTGCTTGGAGACCTGTCGCTGAGGCTGTATGATGTCGACACTACATTATATGATAATAGGATGGTCTCGAT
TTCAGTTGCAAACTCATACATCAAGTAGATGACGCCGAAGGTCTTTTTATGTCACAGAAATAGTGGTCTCGT
AAAGTAAATCCGCAACAATAAAAATCTACGAAATATTTTTGCTGCACTGACATACAAACGCGGAAAACCTTTATATGAA
TGGTACAAAGTGAATATGCAATATTTTCCTCCACAGTTACG

https://mc06.manuscriptcentral.com/genome-pubs
MEYGLDAKISKELESINFRVLQETQLDEAKRSTKPDWVDDNQLVVKRNDVQFDTAFC
CKAMKILLVWQVTIGNLNPFVTGQLDPSMKRFTVYTWKFLFSLTMAIAQTIGITLCYKLF
FREPTISALGFVFTFTSTCETTTLFILLASKWPALSMDIVBSKLDENVSDKIIITKCRIT
CCIFGIMNEHFLSILSVARVIECSQNETDHEGEFVVKTSFPWLYDLNPYVAVAVMV
QVYNLITTANWNYIFIFVCSSTYSLSSILNQINKIELEAQKTHVPAKTNLRPDYTRA
THLVKREDVIDSIGVLVYTANDLFFICLQLYVMILSNMSKAAQLVNIKLPDPQGTFRASY
FAYLYISVLYLLVRFLTEQIVASGVNTASLLFAPILYGITPAITKEVERFQONQVNGDV
VLSGLHFHYITRDLVLTVSVFS

>HzeaGR11
MTSKLFLKTFMMATRHLDKREICGLHSTVRRGTFCSRVMGLLPVSGLTCPTSRRLRFTF
RSPYTVLYASLFGQQLMVFMTCFWMMGNSLANITNAVYTFSSLISSILLHLGRGWP
SLVGSVEETLRELPPFHRNVASINTLRELFFHPHRNVASINNVVHEHLSVFGYLKVAL
CSNNVVAENYFRFNMIFDTYPFTIKWGGALSELENQIFSTFVWLDLILIMVMSYLYTEH
LIIHNELLLLKAAEQHEFSCLEFRQYKLIRVLKLLINGQFGIYILTSGSNLWICTQLF
YSLSRTQTPHITCFKDSQTKVPFVEYENPLCPWMLGEKALNGVEHISYFTSFSPF
ILRTTLVLLAARIHNSVAPLYYLYGFRSFHIEIVERFIAQINLKVAMSGLDFYYVT
RTMILITLGTIVYELVLLQFNR

>HzeaGR12
MKVHELTMRRSNKRCGLHVCRLHAMRLARWMTGFPQLQGQAYADGTKYIKLSLYFYINF
ATLQLQISMSCFAILLFFQETVLTNSISNVIYVTLISAVIFIKLAKQPRLMARATET
EQGLETIELKQNKVIVKCVIAVAMALVEHILTCEFNLTFVMHCLKEAITNVMENY
VVRMRPYPKYPYPSLFWAFLFEYLCQSTFVWSFNDVLITCFSIYITAYFSLNQVTA
NSKDKDNMIPWSTLRVHYSKLVRLVKEIDNQIISFILLAFDLDYICLQFNSLHRNY
VSFKCFNELQTKALTSPSYLLYLYSFILVLRATMLSLFSNVMACAPFHVAVYDP
STYNDEVVRFQPQLQHTQVLGTLKFFYVTRNMLKVGITITYETVLQQYTTIPNYPYN
GTVKILNISSHSY