Published in final edited form as:

Insect Sci. 2013 October ; 20(5): 559–574. doi:10.1111/j.1744-7917.2012.01560.x.

Application of *Cydia pomonella* expressed sequence tags: Identification and expression of three general odorant binding proteins in codling moth

Stephen F. Garczynski¹, Brad S. Coates², Thomas R. Unruh¹, Scott Schaeffer³, Derick Jiwan³, Tyson Koepke³, and Amit Dhingra³

¹USDA-ARS, Yakima Agricultural Research Laboratory, Wapato, WA 98951

²USDA-ARS, Corn Insect and Crop Genetics Research Unit, Genetics Laboratory, Iowa State University, Ames, IA 50011

³Department of Horticulture and Landscape Architecture, Washington State University, Pullman, WA 99164, USA

Abstract

The codling moth, *Cydia pomonella*, is one of the most important pests of pome fruits in the world, yet the molecular genetics and the physiology of this insect remain poorly understood. A combined assembly of 8 341 expressed sequence tags was generated from Roche 454 GS-FLX sequencing of eight tissue-specific cDNA libraries. Putative chemosensory proteins (12) and odorant binding proteins (OBPs) (18) were annotated, which included three putative general OBP (GOBP), one more than typically reported for other Lepidoptera. To further characterize CpomGOBPs, we cloned cDNA copies of their transcripts and determined their expression patterns in various tissues. Cloning and sequencing of the 698 nt transcript for CpomGOBP1 resulted in the prediction of a 163 amino acid coding region, and subsequent RT-PCR indicated that the transcripts were mainly expressed in antennae and mouthparts. The 1 289 nt (160 amino acid) CpomGOBP2 and the novel 702 nt (169 amino acid) CpomGOBP3 transcripts are mainly expressed in antennae, mouthparts, and female abdomen tips. These results indicate that next generation sequencing is useful for the identification of novel transcripts of interest, and that codling moth expresses a transcript encoding for a new member of the GOBP subfamily.

Keywords

codling moth; EST; odorant binding proteins

Correspondence: Stephen F. Garczynski, USDA-ARS, Yakima Agricultural Research Laboratory, 5230 Konnowac Pass Road, Wapato, WA 98951, USA. Tel: (509) 454 6572; Fax: (509) 454 5646; steve.garczynski@ars.usda.gov.

Disclosure

This article reports the results of research only. Mention of a commercial or proprietary product does not constitute an endorsement of the product by the United States Department of Agriculture.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Introduction

The codling moth, *Cydia pomonella* (Lepidoptera: Tortricidae), is the major insect pest of apple and pear, and has a broad distribution that causes economic damage worldwide. Control of codling moth has typically been achieved with a series of chemical insecticide applications (mainly organophosphates), but this management practice has led to the emergence of insecticide resistance, growing concerns about contamination of the environment, as well as consumer and worker safety issues. The planned removal of organophosphates for use in apples in the USA has led to initiatives to implement new chemistries and to develop new codling moth control strategies (Riedl *et al.*, 2004). As an alternative to chemical insecticide treatments, orchardists have increased the use of sex pheromones and kairomones in codling moth control and monitoring programs. The use of these semiochemicals relies on the insect's ability to perceive the chemical cues that influence codling moth reproduction, host location, and larval feeding. Mating disruption tactics employ synthetic sex pheromones that interfere with the ability of males finding female moths, and is currently used to suppress codling moth populations in over 160 000 ha of apple and pear orchards worldwide (Witzgall *et al.*, 2008). Semiochemical signals and associated gene pathways involved in their perception form the basis of mating disruption and monitoring strategies that are widely used in orchard IPM programs. Increased reliance of semio-chemicals for codling moth control highlights the importance of a functional understanding of how lepidopteran chemosensory pathways work, and how chemical cues can be exploited to modify the behavior of pest insects for the control of crop damage.

Chemosensory proteins (CSPs) and odorant binding proteins (OBPs) are small soluble polypeptides thought to play key roles in perireceptor events leading to chemoreceptor signaling events (Vogt, 2005). Despite the enormous amount of structural data, there remains no clearly assigned role of either CSPs or OBPs in chemoreceptor signaling events (Pelosi *et al.*, 2006). CSPs and OBPs form two phylogenetically distinct protein families with CSPs characterized by the presence of four cysteines with a conserved spacing pattern allowing for the formation of two disulfide bridges (Angeli *et al.*, 1999), and OBPs usually have six cysteines with a conserved spacing pattern leading to the formation of three disulfide bridges (Leal *et al.*, 1999; Scaloni *et al.*, 1999; Vogt, 2005). Gene transcripts encoding CSPs have been detected in antennae, brain, legs, mouthparts, wings, and pheromone glands (Angeli *et al.*, 1999; Picimbon *et al.*, 2001; Jaquin-Joly *et al.*, 2001; Whitfield *et al.*, 2002; Ban *et al.*, 2003; Vogt, 2005; Pelosi *et al.*, 2006), and these proteins bind pheromones and other odorants (Vogt, 2005; Pelosi *et al.*, 2006). Gene transcripts encoding OBPs are mainly found in chemosensory tissues and bind pheromones and other odorants (Vogt, 2005). Functionally, OBPs have roles in behavioral responses to pheromones (Xu *et al.*, 2005; Laughlin *et al.*, 2008; Sun *et al.*, 2012) and taste perception (Matsuo *et al.*, 2007). However, for Lepidoptera, there remains no clearly assigned role of either CSPs or OBPs in chemoreceptor signaling events (Pelosi *et al.*, 2006).

Pheromone binding proteins (PBPs) and general OBPs (GOBPs) are defined as subfamilies within the clade of OBPs. As with other members of the OBP family, PBPs and GOBPs also have a tertiary structure that is coordinated by the formation of three disulfide bridges. PBPs and GOBPs are mainly expressed in antennae, with PBPs usually localized to trichoid

sensilla and GOBPs to basi-conic and trichoid sensilla (Vogt, 2005; Pelosi *et al.*, 2006 and references therein). Both PBPs and GOBPs have been shown to bind pheromones (for recent examples refer to Gong *et al.*, 2009; Leal *et al.*, 2009; Zhou *et al.*, 2009; He *et al.*, 2010; Liu *et al.*, 2010; Yin *et al.*, 2012), but GOBP2 has also been shown to bind plant volatiles (Gong *et al.*, 2009; Yin *et al.*, 2012), suggesting a possible role in host finding. Although much is known about the structure and ligand binding of PBPs and GOBPs, their roles in pheromone and host plant reception remain incomplete.

In an effort to obtain a greater understanding of the chemosensory system, we generated expressed sequence tags (ESTs) from various codling moth tissues (adult male and female antennae, mouthparts, and legs, various stages of eggs, and neonates) using next generation sequencing (NGS) technology. Once the ESTs were annotated, we focused our work to identify soluble binding proteins expressed in each of these codling moth tissues. Furthermore, we cloned transcripts encoding three GOBPs and characterized their expression profiles in various tissues dissected from males and females by RT-PCR.

Materials and methods

Insects and tissue source

Cydia pomonella eggs, neonates, and adult males and females were obtained from a colony maintained at the USDA-ARS Yakima Agricultural Research Laboratory (Wapato, WA). Eggs (30 mg each; 0–24, 24–48, 48–72, and 72–96 h) and neonate larvae (30 mg) were collected and immediately placed into 1.5 mL microfuge tubes containing 100 μ L RNAlater® (Ambion, Austin, TX). Adult males and females (0–24 h posteclosion) were chilled at 4°C, and then placed in a Petri dish on ice for dissection. Antennae, forelegs, and mouthparts were removed from 50 males and 50 females with forceps and placed directly into 1.5 mL microfuge tubes containing 100 μ L RNAlater® (Ambion). Eggs, neonates, antennae, and pooled forelegs and mouthparts in RNAlater® (Ambion) were stored at 4°C until RNA was extracted within 1 week of dissection.

RNA extraction, cDNA preparation, and transcriptome sequencing

Total RNA was extracted from eggs, neonates, antennae, pooled forelegs, and mouthparts using the RNeasy® Plus Mini Kit (Qiagen, Valencia, CA) according to manufacturer's protocol for the purification of total RNA from animal tissues. In the final step, total RNA was eluted from the RNeasy spin column (Qiagen) with 40 μ L of RNase-free water (supplied with kit). Double-stranded (ds) cDNAs were prepared using components and procedures of the SMART™ PCR cDNA Synthesis Kit (Clontech, Mountain View, CA) and SuperScript® III First-Strand Synthesis SuperMix (Invitrogen, Carlsbad, CA). Briefly, first-strand cDNA was generated from 4 μ L total RNA primed with the CDSIII/3' PCR primer and SMART IV oligonucleotide (Clontech) (which is incorporated onto the 5' end of the cDNA) in a 20 μ L reaction with components supplied with the SuperScript® III First-Strand Synthesis SuperMix (Invitrogen) for 50 min at 50°C. Second-strand cDNA synthesis was done by long distance (LD) PCR using Advantage™ 2 *Taq* Polymerase (Clontech). The LD PCR reaction contained 2 μ L of first-strand cDNA in a 100 μ L reaction set up according to the manufacturer's protocol and amplified using an initial denaturation for 90 s at 94°C,

then amplification for 20 s at 94°C, 6 min at 68°C for 26 cycles followed by a 5 min 72°C incubation. The resulting dscDNA was then purified using the High Pure PCR Product Purification Kit (Roche Applied Science, Indianapolis, IN) according to manufacturer's instructions and eluted with 100 µL of kit elution buffer. Amplification and purification procedures were repeated until 10 µg cDNA was generated. The cDNAs, 10 µg from each tissue source, were converted into sequencing libraries using individual Multiple Identifier (MIDs) tags (Roche) and pyrosequenced using a Roche 454 GS-FLX system. Raw sequence reads were separated based on individual MIDs and the tags were removed using custom scripts. Thereafter, sequence data from each library were independently assembled using the Roche GS De Novo Assembler (i.e., Newbler assembler) using default parameters (seed step: 12, seed length: 16, minimum overlap length: 40, minimum overlap identity: 90%, alignment identity score: 2, and alignment difference score: 23), and all of the nonredundant (nr) contigs were exported to a file in FASTA format by the software provided. A combined assembly generated from a pooled collection of all 454 GS-FLX data was generated using identical parameters.

Homology searches using the *C. pomonella* EST assembly

FASTA formatted sequence from 8 341 contigs 100 bp within the combined assembly of *C. pomonella* embryonic, neonate, male antennal, female antennal, male foreleg and mouthparts, and female foreleg and mouthparts ESTs were imported into the Blast2GO suite (Conesa *et al.*, 2005; Götz *et al.*, 2008). Homology searches were carried out by query of the National Center for Biotechnology Information (NCBI) nr protein database using the BLASTX algorithm (Altschul *et al.*, 1997) with cutoffs of *E* value 1×10^{-6} and Hsp 33. Two additional homology searches were performed with the 8 341 *C. pomonella* EST contigs as queries against the NCBI dbEST database using BLASTN and TBLASTX algorithms (Altschul *et al.*, 1997), and conducted by the Blast2GO suite (*E* value 1×10^{-6} ; Hsp cutoff 33). Results from homology searches were independently reported as distributions of *E* values and percent sequence similarity to database hits and distributions of species assignment of top database hits reported.

The contigs within the *C. pomonella* EST dataset that correspond to unique gene transcripts were investigated by comparison to the *Bombyx mori* gene models. The translated consensus gene set, constructed by merging all the gene sets using GLEAN, were downloaded from SilkDB (<http://www.silkdb.org/silkdb/doc/download.html>), which represents 22 987 known and predicted genes from the whole genome assembly. Downloaded data were imported into a local database using the program BioEdit (Hall, 1999), against which the *C. pomonella* ESTs were used to query the GLEAN protein database using the BLASTX algorithm, and hits with *E* values 1×10^{-25} were retained. Commonly, *C. pomonella* EST sequences appeared similar to more than one *B. mori* gene, likely resulting from the presence of paralogs or conserved protein domains present in proteins with similar functions. Hence, the BLASTX output was filtered for unique contigs matches.

Functional annotation and definition of gene families

We assigned genes to positions within cellular structures, cellular processes, and molecular functions by searching the Gene Ontology (GO) database (The Gene Ontology Consortium,

2000; <http://www.geneontology.org/>). We used the top BLASTX descriptors output from Blast2GO (*E*-value-hit filter 10^{-5} ; annotation cutoff = 55; GO weight = 5; Hsp-Hit coverage cutoff = 0), and subsequent retrieval of Enzyme Code (EC) designations directly from the GO website. Final sequence annotation steps used searches of the inclusive InterPro databases using the InterProScan (IPS) tool (<http://www.ebi.ac.uk/interpro/>) (Quevillon *et al.*, 2005; Hunter *et al.*, 2009). Biochemical pathway information was downloaded from relevant maps in the Kyoto Encyclopedia of Genes and Genomes (<http://www.genome.jp/kegg/>; Ogata *et al.*, 1999) using EC terms. We report the resulting biological process (P), molecular functions (F), and cellular component (C) at GO level 2 with statistics on distribution of sequence lengths, *E* values, percent sequence similarity, and species from which BLAST hits were derived.

The InterPro annotations of silkworm genes were downloaded from SilkDB (<http://www.silkdb.org/silkdb/doc/download.html>; file *silkworm_glean.gff.ipr.tar.gz*), and text was searched for InterPro accession numbers present within corresponding *C. pomonella* IPS results. The *B. mori* gene models predicted to carry out similar function to *C. pomonella* CSP, OBP, PBP, and GOBP gene transcripts were extracted, and defined as the paralogous gene space that putatively carry out similar functions. The relationship among *B. mori* and *C. pomonella* paralogs from selected genes was investigated following full-length cDNA cloning and sequencing from *C. pomonella* (see below).

Phylogenetic analyses of *C. pomonella* chemosensory and OBPs

Deduced amino acid sequences of putative *C. pomonella* CSPs, OBPs, PBPs, and GOBPs were initially aligned using ClustalW (Thompson *et al.*, 1994; <http://www.genome.jp/tools/clustalw/>) with default settings. A phylogenetic tree (Fig. S2, Supporting Information) was generated by selecting rooted phylogenetic tree with branch length (unweighted pair group method with arithmetic mean method) on the ClustalW results page. Boxshade 3.21 (http://www.ch.embnet.org/software/BOX_form.html) was used to create boxshade alignments.

Lepidopteran CSP-, OBP-, PBP-, and GOBP-related sequences were retrieved from the nr nucleotide database on GenBank using TBLASTN. Amino acid sequences were downloaded in FASTA format and signal sequences, predicted using the SignalP 3.0 Server (<http://www.cbs.dtu.dk/services/SignalP>; Bendtsen *et al.*, 2004), were removed. Final alignments of these diverse amino acid sequences used the ClustalW algorithm (Blossum protein weighting matrix, gap open penalty 10, extension penalty of 0.1 for pair wise alignment and 0.05 for multiple alignment; allowing hydrophilic gaps). A neighbor-joining (NJ) tree (equal rates, partial deletion), as implemented in MEGA 5 (Tamura *et al.*, 2011), was created to show general topography without node support, and a maximum parsimony (MP) tree was constructed using the program TNT version 1.1 (Goloboff *et al.*, 2008) using “new technology search” parameters with five random seed, sectorial search and ratchet, with discovery of minimum length tree 20 times; 12 246 924 475 possible tree configurations were examined.

Full-length cloning and validation of GOBP tissue expression

Rapid amplification of cDNA ends (RACE) was performed using the SMARTTM RACE cDNA amplification kit (Clontech) according to the manufacturer's protocols. For 5' RACE, first-strand cDNA was prepared as above except that total RNA was converted to cDNA using the 5' -RACE CDS Primer A and SMART IITM A oligonucleotide (Clontech). From the cDNA sequences obtained above, reverse primers (See Table S1 for list of oligonucleotides used in this study, Supporting Information) were designed for 5' RACE using Primer3 (v 0.4.0; <http://frodo.wi.mit.edu/primer3/>). Amplification was performed using the 10× Universal Primer A Mix as forward primer and sequence specific primers as follows: initial denaturation for 3 min at 94°C, then amplification for 30 s at 94°C, 2 min at 72°C for 40 cycles, followed by a final 5 min 72°C incubation. PCR products were separated on 1% agarose gels containing ethidium bromide and visualized on a UV light box. Excised bands were extracted and purified using GenElute minus EtBr spin columns (Sigma, St. Louis, MO), and the purified PCR products were cloned using the TOPO TA cloning kit for sequencing (Invitrogen) with TOP 10 *Escherichia coli* chemically competent cells. Plasmid DNA was extracted from selected colonies using the QIAprep spin mini prep kit (Qiagen). The cDNA clones were sequenced at MC Laboratories (MCLab, San Francisco, CA). To determine the identity of encoded proteins, TBLASTN on the National Center for Biotechnology Information website was used (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>).

Total RNA was extracted from males and females for the following stages and tissue types; pharate (early) pupae, late pupae heads, late pupae bodies, newly emerged adults (0–24 h) antennae, heads devoid of antennae and mouthparts, mouthparts, bodies, and female abdomen tips to include pheromone glands. The RNA was treated with Turbo DNase (Ambion) according to the manufacturer's protocol and then quantitated using the QuantiTTM RiboGreen® RNA assay kit (Invitrogen). First-strand cDNA synthesis was performed as above, using 100 ng of the DNase-treated total RNA and primed with 5' -RACE CDS Primer A. PCR amplification was performed using gene-specific primers (200 nmol/L final concentration; Table S1, Supporting Information) and cDNA template equivalent to 5 ng of input RNA with the following conditions: initial denaturation for 3 min at 94°C, then amplification for 20 s at 94°C, 20 s at 62°C, 30 s at 72°C for 35 cycles, followed by a final 5 min 72°C incubation. PCR products were separated by loading 1/4th of the total reaction onto 1.5% agarose gels containing ethidium bromide and visualized on a UV light box. To confirm identity of PCR products generated, visible bands were excised, cloned, and sequenced as described above. As a control, primers to detect CpomActin were used and it was detected in all tissues tested (Fig. 4).

Results

Transcriptome sequencing and assemblies

Sequencing of the pooled transcriptomes from eggs, neonates, male and female legs and mouthparts, and male and female antennae produced 894 230 reads totaling ~136 Mb. Assembly of these reads yielded 8 341 contigs with a minimum length of 100 bp (mean length = 478.4 ± 364.0 bp) comprising a 90.7 Mb transcriptome. Reads and bases passing

the multiplexing index filter and the assembly results for each individual tissue type are shown in Table 1. The assembled codling moth sequence data were deposited to the online database LepDB.org (Coates *et al.*, unpublished).

Homology searches using the *C. pomonella* EST assembly

Search of the NCBI nr protein database using the BLASTX algorithm and 8 341 *C. pomonella* EST contigs as queries resulted in 2 266 ESTs generating “hits” with *E* value 9.5×10^{-7} and similarities 40.05% (mean = $75.65 \pm 13.6\%$; Fig. S1a, Supporting Information). Thus, 27.2% of the inferred protein sequences from the *C. pomonella* ESTs displayed homology to known genes within current database accessions. The distribution of species to which *C. pomonella* EST top BLASTX “hits” were obtained showed that 30.4% and 12.2% were most similar to *Drosophila* and mosquito species, respectively, whereas “hits” to other species of Lepidoptera accounted for 2.0%. The greatest individual species representation among results of the BLASTX homology search was from the model insect species *Tribolium castaneum* (5.8%) and *B. mori* (4.4%; Fig. S1b, Supporting Information). Hits to sequence accessions from *C. pomonella* contributed <0.1% to the total annotated dataset (data not shown).

A total of 4 379 of the 8 341 *C. pomonella* EST sequences from the combined assembly (52.5%) generated a “hit” via a BLASTN search of NCBI dbEST, and the distribution of sequence similarity values among all “hits” to dbEST accessions is shown in Fig. 1A. From these results, 1 544 of the 4 379 top BLASTN “hits” (35.3%) were shown to be accessions from the species *Choristoneura fumiferana* (Lepidoptera: Tortricidae), followed by 907 for *Bicyclus anynana* (Lepidoptera: Nymphalidae) (20.7%; Fig. 1C). An analogous search of dbEST using the TBLASTX algorithm indicated that sequence similarity was present among database accessions for 3 604 of *C. pomonella* EST queries (43.2%; distribution of similarities among all “hits” in Fig. 1B). The species distribution among top TBLASTX “hits” showed the greatest number *B. anynana* (776; 21.5%), followed by *C. fumiferana* (648; 18.0% Fig. 1C). The BLASTN and TBLASTX results from the *C. pomonella* EST queries of the dbEST database showed a greater proportion of “hits” to species of Lepidoptera compared with the BLASTX search of the nr protein database. BLASTX results show that 1 848 of 8 341 (22.2%) *C. pomonella* combined assembly EST sequences had 1 match to the 22 987 translations from the *B. mori* GLEAN gene models.

Functional annotation and definition of gene families

A total of 7 586 GO designations were obtained for *C. pomonella* EST sequences (0.98 ± 2.49 ; range 0–51 GOs per sequence). At level 2, the distribution of GO terms among C, P, and F identified cell structure, cellular process, and binding to respectively be at the highest frequency (Figs. 2A–2C). GO terms for C show that structural proteins, including insect cuticle protein coding genes, are highly represented (InterPro accession IPR000618; Table 2). The presence of putative binding functions within predicted genes identified at GO level 2 (F) was corroborated by observing a total of 27 sequences matching an InterPro accession with OBP activity (IPR006170 and IPR006625), troponin (IPR001978; Ca⁺ binding), and chitin binding activity (IPR002557; Table 2). A total of 465 unique InterPro annotations

were obtained for *C. pomonella* EST data (Table S2, Supporting Information), and all were found in the corresponding InterPro annotation file for *B. mori*.

A total of 2 342 orthologous relationships were predicted following query of *C. pomonella* EST contigs against the GLEAN protein database using the BLASTX algorithm (Supplemental data SD1, Supporting Information). Because of a one-to-many relationship between contigs and *B. mori* gene models, likely owing to potential paralogs, the placement of *C. pomonella* gene transcripts within closely related gene families was not investigated further by phylogenetic analyses, except for those of the CSPs and OBPs (see below).

Phylogenetic analysis of *C. pomonella* chemosensory and OBPs

Functional annotations derived from the *C. pomonella* EST data were filtered by keyword leading to the identification of gene transcripts that encoded 12 CSPs, 12 OBPs, three GOBPs, and three PBP (Table S3; Fig. S2; Fig S3, Supporting Information). A preliminary phylogenetic analysis of the putative CpomCSPs, CpomOBPs, CpomGOBPs, and CpomPBPs showed that each protein family fell into distinct monophyletic groups (Fig. S2, Supporting Information). Boxshade alignments for CpomCSP, CpomOBP, CpomPBP, and CpomGOBP amino acid sequences identified that the four cysteine residues typical for CSPs are conserved among the 12 CpomCSPs, and similarly, six cysteines typical for OBPs are conserved among most of the CpomOBPs including members of the CpomPBP and CpomGOBP subfamilies (Fig. S3, Supporting Information). Functional annotation data from the tissue-specific ESTs indicate that Cpom-CSP2 transcripts were detected in all tissue types, whereas CpomCSP11 and CpomCSP12 transcripts were only detected in neonates, CpomCSP1 transcripts were present only in the egg ESTs, and CpomCSP6 and CpomCSP10 transcripts were seen only in the female and male antennal ESTs, respectively (Fig. S2, Supporting Information). Transcripts encoding putative CpomOBPs, CpomGOBPs, and CpomPBPs were found mainly in antennae. A transcript encoding CpomOBP9 was only found in neonate ESTs, transcripts encoding CpomOBP3 and CpomPBP3 were only detected in female antennae ESTs, and transcripts encoding CpomOBP1, CpomOBP11, and CpomOBP12 were also found in the samples prepared from legs and mouthparts (Fig. S2, Supporting Information).

The TBLASTN searches of the NCBI nr nucleotide database limited to arthropods using *C. Pomonella* chemosensory and OBP amino acid sequences produced 209 lepidopteran “hits,” which were aligned and a phylo-genetic hypothesis was estimated with unrooted NJ trees (Fig. 3). In total, 12 CpomCSPs aligned and grouped with 56 other lepidopteran CSPs (Fig. 3A), and the 18 CpomOBPs aligned and grouped with 140 other lepidopteran OBPs (Fig. 3B). Of the 18 putative CpomOBPs, three grouped within the PBP subfamily and three grouped within the GOBP subfamily (Fig. 3B). A single tree containing the CpomCSPs and CpomOBPs along with the other lepidopteran CSPs and OBPs was constructed using the MP algorithm implemented in TNT (Goboloff *et al.*, 2008), and resulted in the prediction of a more complicated set of relationships that divided the four major groups into smaller clades that showed high bootstrap support (Fig. S4, Supporting Information). The MP tree showed that the annotated *C. pomonella* CSPs and OBPs correspond to the nomenclature

assigned for other Lepidoptera extracted from the BLAST search of GenBank (Fig. S4, Supporting Information).

Full-length cloning and tissue expression of GOBPs

Other Lepidoptera are known to express two transcripts encoding GOBPs in antennal tissues. Our annotations predicted three putative *C. pomonella* transcripts encoding for GOBPs, indicating the potential presence of a novel paralog, which we chose for further characterization. Oligonucleotide primers designed from the transcriptome sequences were used in 3' and 5' RACE reactions (Table S1, Supporting Information) to verify the presence of and determine full-length sequences of CpomGOBP transcripts from male and female antennal-derived cDNAs. The transcript for antennal CpomGOBP1 is 698 nt, encoding for 163 amino acids (GenBank ID: JN701809; Fig. S5A, Supporting Information), that for CpomGOBP2 is 1289 nt, encoding for 160 amino acids (GenBank ID: JN701810; Fig. S5B, Supporting Information), whereas the transcript for CpomGOBP3 is 702 nt, encoding for 169 amino acids (GenBank ID: JN701811; Fig. S5C, Supporting Information). Each of the deduced proteins has a predicted signal peptide (Fig. S5, Supporting Information) and all contain conserved domains and cysteines, consistent with other GOBPs previously determined for each protein family (Fig. S3, Supporting Information).

To determine where the CpomGOBP transcripts are expressed in males and females, a tissue expression profile using RT-PCR was performed with cDNA prepared from total RNA extracted from dissected early pupae, late pupae heads and bodies, newly emerged adults (0–24 h) antennae, heads devoid of antennae and mouthparts, mouthparts, bodies, and female abdomen tips to include pheromone glands (Fig. 4). CpomGOBP1 expression was limited to pupal heads, adult antennae, and mouth parts, with PCR products more strongly visualized in adult male and female antennae (Fig. 4, top panel). CpomGOBP2 transcripts were more broadly expressed, with PCR products detected in late pupae heads, and antennae, heads, mouthparts of adults from both sexes, and female abdomen tips (Fig. 4, second panel). Expression of Cpom-GOBP3 was similar to that of CpomGOBP1 with PCR products mainly detected in antennae and mouthparts, and also faintly detected in female heads (Fig. 4, third panel).

Discussion

The use of NGS technologies for the rapid acquisition of transcriptome sequence data from a nonmodel insect species was reported by Vera *et al.* (2008), where downstream functional gene annotation, gene orthology estimation with a model species, and prediction of putative mutations was described. Although 454-based transcriptome sequencing for nonmodel organisms has since become more commonplace, few instances have been published for species of Lepidoptera (Vera *et al.*, 2008; Pauchet *et al.*, 2010; Wanner *et al.*, 2010; Margam *et al.*, 2011). NGS technology was used here to rapidly generate cDNA sequence data from eight *C. pomonella* tissues, and produced 8 341 EST contigs representing an initial transcriptome resource for codling moth. Our transcript pool of 8 341 EST contigs included 2 266 with identifiable protein coding sequences following gene annotation. A surprisingly high number of these top BLAST “hits” matched dipteran insects and only 2% were to

Lepidoptera. This may likely be due to the high representation of *Drosophila* sp. proteins within the nr database compared with the Lepidoptera. In contrast, TBLASTN searches against the Gen-BankdbEST “EST others” database showed that >50% of the top “hits” were from lepidopteran sequence accessions, which may be a consequence of the omission of model species’ data compared with the analogous searches of nr databases. Restriction of search criteria to databases with more closely related sequence accessions did increase the proportion of putative matches to *C. pomonella* contigs, but provided no gene descriptors from which GO annotations could be retrieved. A high proportion of putative genes encoded by lepidopteran ESTs typically remain uncharacterized following analysis via annotation pipelines (Coates *et al.*, 2008; Vera *et al.*, 2008). This lack of annotation information cause difficulties in the assignment to putative functions for species-and order-specific genes, and may remain a dilemma for the genomic investigations of Lepidoptera. Similarly, nearly 50% of the gene models generated from the *B. mori* genome assembly lack GO information and ~25% contain coding sequences with no homologous matches to *Drosophila melanogaster* (The International Silkworm Genome Consortium, 2008). Analogously, a high proportion of *C. pomonella* ESTs received no functional annotation information, which suggests that Lepidoptera are a rich source for novel gene discovery. Of the functional annotations that were made, a high proportion was involved in cellular structure formation, cellular process, and binding. Although these data were derived from a pooled analysis where the contribution of constituent libraries were not feasible, our data represent a resource for gene discovery in Lepidoptera similar to that described by Vera *et al.* (2008) and Pauchet *et al.* (2010).

Within the annotated portion of the *C. pomonella* ESTs, we identified gene transcripts encoding 12 putative CSPs and 18 putative OBPs (which included five members of the PBP subfamily and three members of the GOBP sub-family). These orthologs and numbers are consistent with CSPs and OBPs identified in transcriptomes of other lepidopterans (Gong *et al.*, 2007; Jordan *et al.*, 2008; Ozaki *et al.*, 2008; Zhang *et al.*, 2009; Grosse-Wilde *et al.*, 2011; Legeai *et al.*, 2011). Phylogenetic analysis of CSP and OBP proteins from *C. pomonella* showed a one-to-one orthology with genes from other Lepidoptera. Given the incomplete knowledge of PBP function in lepidopteran chemosensory pathways, discussions of the nomenclature or putative biochemical role of these binding protein families may be premature.

GOBPs are a subfamily of OBPs that are expressed in lepidopteran antennae, and have been shown to bind pheromones (Gong *et al.*, 2009; Leal *et al.*, 2009; Zhou *et al.*, 2009; He *et al.*, 2010; Liu *et al.*, 2010), and in Lepidoptera consist of two members, GOBP1 and GOBP2 (Vogt *et al.*, 1991; Krieger *et al.*, 1996; Newcomb *et al.*, 2002; Vogt *et al.*, 2002; Jin *et al.*, 2009; Leal *et al.*, 2009; Zhang *et al.*, 2009). A unique finding in this study was the identification of three *C. pomonella* transcripts that encode putative GOBPs. Because GOBPs were mainly detected in basiconic sensilla of silk moths, it was hypothesized that they have potential role in interactions with plant and general odor molecules (Steinbrecht *et al.*, 1995), and subsequently, GOBP2 from *Chilo suppressalis* (Lepidoptera: Pyralidae) was shown to bind to laurinal dehyde and benzaldehyde supporting a potential role in the perception of host plant volatiles (Gong *et al.*, 2009). Recent studies have shown that the

GOBP subfamily of proteins may bind to pheromones (Zhou *et al.*, 2009; He *et al.*, 2010; Liu *et al.*, 2010), and that the *B. mori* GOBP2 differentially bind sex pheromone components (Zhou *et al.*, 2009). In contrast, evidence from the navel orange worm, *Amyelois transitella* (Walker) (Lepidoptera: Pyralidae) suggested that GOBPs do not bind to semiochemicals that attract adult females (Liu *et al.*, 2010). Despite the evidence above, the binding of GOBPs to plant volatiles has not been well established, and the role of GOBPs in pheromone and plant volatile perception and signaling remains unresolved.

GOBPs show spatial specificity in expression, and are localized mainly in adult male and female antennae (Vogt, 2005 and references within; Gong *et al.*, 2009; Leal *et al.*, 2009; He *et al.*, 2010 among numerous others), and in larval antenna and maxilla (Vogt *et al.*, 2002). Similar to the evidence from other species of Lepidoptera, the transcripts for CpomGOBPs were detected in both male and female antennae of both adults and in late stage pupal heads with antennae attached. We also found that all CpomGOBP transcripts are expressed in mouthparts dissected from adult males and females. CpomGOBP2 transcripts were also detected in adult male and female heads (devoid of antennae and mouthparts); however, localization within the head of these transcripts was not determined. Interestingly, CpomGOBP2 and CpomGOBP3 (faintly) transcripts were also found in female abdomen tips. Depending on the ligands for CpomGOBP2 and CpomGOBP3, be they pheromones or plant volatiles, it could be easy to speculate that they may have a role in oviposition or the regulation of pheromone production or release, and these issues will be addressed in future studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Drs. Kevin Clark, Lisa Neven, and Tonja Fisher for their review of this manuscript and helpful suggestions. Laura Willett and Jennifer Stout are thanked for their technical assistance. This project is supported by grants from the Washington Tree Fruit Research Commission (awarded to SFG) and National Research Initiative Arthropod and Nematode Biology and Management Program of the USDA National Institute of Food and Agriculture, Grant no. 2008-35302-18839 (awarded to SFG).

References

- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*. 1997; 25:3389–3402. [PubMed: 9254694]
- Angeli S, Ceron F, Scaloni A, Monti M, Monteforti G, Minnocci A, Petacchi R, Pelosi P. Purification, structural characterization, cloning and immunocyto-chemical localization of chemoreception proteins from *Schistocerca gregaria*. *European Journal of Biochemistry*. 1999; 262:745–754. [PubMed: 10411636]
- Ban LP, Scaloni A, Brandazza A, Angeli S, Zhang L, Yan YH, Pelosi P. Chemosensory proteins of *Locusta migratoria*. *Insect Molecular Biology*. 2003; 12:125–134. [PubMed: 12653934]
- Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *Journal of Molecular Biology*. 2004; 340:783–795. [PubMed: 15223320]

- Coates BS, Sumerford DV, Hellmich RL, Lewis LC. Mining an *Ostrinia nubilalis* midgut expressed sequence tag (EST) library for candidate genes and single nucleotide polymorphisms (SNPs). *Insect Molecular Biology*. 2008; 17:607–621. [PubMed: 19133073]
- Conesa A, Götz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. 2005; 21:3674–3676. [PubMed: 16081474]
- Goboloff PA, Farris JS, Nixon KC. TNT, a free program for phylogenetic analysis. *Cladistics*. 2008; 24:774–786.
- Gong DP, Zhang HJ, Zhao P, Lin Y, Xia QY, Xiang ZH. Identification and expression pattern of the chemosensory protein gene family in the silkworm, *Bombyx mori*. *Insect Biochemistry and Molecular Biology*. 2007; 37:266–277. [PubMed: 17296501]
- Gong DP, Zhang HJ, Zhao P, Xia QY, Xiang ZH. The odorant binding protein gene family from the genome of silkworm, *Bombyx mori*. *BMC Genomics*. 2009; 10:332. [PubMed: 19624863]
- Götz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Mueda MJ, Robles M, Talon M, Dopazo J, Conesa A. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*. 2008; 36:3421–3435.
- Grosse-Wilde E, Kuebler LS, Bucks S, Vogel H, Wicher D, Hansson BS. Antennal transcriptome of *Manduca sexta*. *Proceedings of the National Academy of Sciences of the United States of America*. 2011; 108:7449–7454. [PubMed: 21498690]
- Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*. 1999; 41:95–98.
- He X, Tzotzos G, Woodcock C, Pickett JA, Hooper T, Field LM, Zhou J-J. Binding of the general odorant binding protein of *Bombyx mori* BmorGOBP2 to the moth sex pheromone components. *Journal of Chemical Ecology*. 2010; 36:1293–1305. [PubMed: 20981477]
- Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, Finn RD, Gough J, Haft D, Hulo N, Kahn D, Kelly E, Laugraud A, Letunic I, Lonsdale D, Lopez R, Madera M, Maslen J, McAnulla C, McDowall J, Mistry J, Mitchell A, Mulder N, Natale D, Orengo C, Quinn AF, Selengut JD, Sigrist CJA, Thimma M, Thomas PD, Valentin F, Wilson D, Wu CH, Yeats C. InterPro: the integrative protein signature database. *Nucleic Acids Research*. 2009; 37:D224–D228. [PubMed: 18974183]
- Jacquín-Joly E, Vogt RG, Francois MC, Nagnan-Le Meillour P. Functional and expression pattern analysis of chemosensory proteins expressed in antennae and pheromonal gland of *Mamestrabraccae*. *Chemical Senses*. 2001; 26:833–844. [PubMed: 11555479]
- Jin FL, Dong XL, Xu XX, Ren SX. cDNA cloning and recombinant expression of the general odorant binding protein II from *Spodoptera litura*. *Science China Series C: Life Sciences*. 2009; 52:80–87.
- Jordan MD, Stanley D, Marshall SDG, De Silva D, Crowhurst RN, Gleave AP, Greenwood DR, Newcomb RD. Expressed sequence tags and proteomics of antennae from the tortricid moth, *Epiphyas postvittana*. *Insect Molecular Biology*. 2008; 17:361–373. [PubMed: 18651918]
- Krieger J, von Nickisch-Roseneck E, Mameli M, Pelosi P, Breer H. Binding proteins from the antennae of *Bombyx mori*. *Insect Biochemistry and Molecular Biology*. 1996; 26:297–307. [PubMed: 8900598]
- Laughlin JD, Ha TS, Jones DNM, Smith DP. Activation of pheromone sensitive neurons is mediated by conformational activation of pheromone binding protein. *Cell*. 2008; 133:1255–1265. [PubMed: 18585358]
- Leal WS, Ishida Y, Pelletier J, Xu W, Rayo J, Xu X, Ames JB. Olfactory proteins mediating chemical communication in the navel orangeworm moth, *Amyelois transitella*. *PLoS ONE*. 2009; 4:e7235. [PubMed: 19789654]
- Leal WS, Nikonova L, Peng G. Disulfide structure of the pheromone binding protein from the silkworm moth, *Bombyx mori*. *FEBS Letters*. 1999; 464:85–90. [PubMed: 10611489]
- Legeai F, Malpel S, Montagne N, Monsempe C, Cousserans F, Merlin C, François MC, Maibèche-Coisné M, Gavory F, Poulain J, Jacquín-Joly E. An expressed sequence tag collection from the male antennae of the noctuid moth *Spodoptera littoralis*: a resource for olfactory and pheromone detection research. *BMC Genomics*. 2011; 12:86. [PubMed: 21276261]

- Liu Z, Vidal DM, Syed Z, Ishida Y, Leal W. Pheromone binding to general odorant-binding proteins from the navel orangeworm. *Journal of Chemical Ecology*. 2010; 36:787–794. [PubMed: 20535533]
- Margam VM, Coates BS, Bayles DO, Hellmich RL, Agunbiade T, Ba M, Sun W, Kroemer JA, Ba MN, Binso-Dabire CL, Baoua I, Ishiyaku MF, Covas FG, Srinivasan R, Armstrong J, Murdock LL, Pittendrigh BR. Transcriptome sequencing, and rapid development and application of SNP markers for the legume pod borer *Maruca vitrata* (Lepidoptera: Crambidae). *PLoS ONE*. 2011; 6:e21388. [PubMed: 21754987]
- Matsuo T, Sugaya S, Yasukawa J, Aigaki T, Fuyama Y. Odorant-binding proteins OBP57d and OBP57e affect taste perception and host-plant preference in *Drosophila sechellia*. *PLoS Biology*. 2007; 5:e118. [PubMed: 17456006]
- Newcomb RD, Sirey TM, Rassam M, Greenwood DR. Pheromone binding proteins of *Epiphya spostvittana* (Lepidoptera: Tortricidae) are encoded at a single locus. *Insect Biochemistry and Molecular Biology*. 2002; 32:1543–1554. [PubMed: 12530222]
- Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*. 1999; 27:29–34. [PubMed: 9847135]
- Ozaki K, Utoguchi A, Yamada A, Yoshikawa H. Identification and genomic structure of chemosensory proteins (CSP) and odorant binding proteins (OBP) genes expressed in foreleg tarsi of the swallowtail butterfly *Papilio xuthus*. *Insect Biochemistry and Molecular Biology*. 2008; 38:969–976. [PubMed: 18771731]
- Pauchet Y, Wilkinson P, Vogel H, Nelson DR, Reynolds SE, Heckel DG, French-Constant RH. Pyrosequencing the *Manduca sexta* larval midgut transcriptome: messages for digestion, detoxification and defence. *Insect Molecular Biology*. 2010; 19:61–75. [PubMed: 19909380]
- Pelosi P, Zhou J-J, Ban LP, Calvellido M. Soluble proteins in insect chemical communication. *Cellular and Molecular Life Sciences*. 2006; 63:1658–1676. [PubMed: 16786224]
- Picimbon J-F, Dietrich K, Krieger J, Breer H. Identity and expression pattern of chemosensory proteins in *Heliothis virescens* (Lepidoptera, Noctuidae). *Insect Biochemistry and Molecular Biology*. 2001; 31:1173–1181. [PubMed: 11583930]
- Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. InterProScan: protein domains identifier. *Nucleic Acids Research*. 2005; 33:W116–W120. [PubMed: 15980438]
- Riedl H, Walston AT, Brooks DJ, Farnsworth A, Farnsworth J, McMahon B, McCarty G. Codling moth control with old and new insecticides: the continuing search for the next Guthion. *Proceedings of Western Orchard Pest Disease Management Conference*. 2004; 78:31.
- Scaloni A, Monti M, Angeli S, Pelosi P. Structural analysis and disulfide-bridge pairing of two odorant-binding proteins from *Bombyx mori*. *Biochemical and Biophysical Research Communications*. 1999; 266:386–391. [PubMed: 10600513]
- Steinbrecht RA, Laue M, Ziegelberger G. Immunolocalization of pheromone binding protein and general odorant-binding protein in olfactory sensilla of the silk moths *Antheraea* and *Bombyx*. *Cell Tissue Research*. 1995; 282:203–217.
- Sun YF, De Biasio F, Qiao HL, Iovinella I, Yang SX, Ling Y, Riviello L, Battaglia D, Falabella P, Yang XL, Pelosi P. Two odorant-binding proteins mediate the behavioural response of aphids to the alarm pheromone (E)- β -farnesene and structural analogues. *PLoS ONE*. 2012; 7:e32759. [PubMed: 22427877]
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*. 2011; 28:2731–2739. [PubMed: 21546353]
- The Gene Ontology Consortium. Gene ontology: tool for the unification of biology. *Nature Genetics*. 2000; 25:25–29. [PubMed: 10802651]
- The International Silkworm Genome Consortium. The genome of a lepidopteran model insect, the silkworm *Bombyx mori*. *Insect Biochemistry and Molecular Biology*. 2008; 38:1036–1045. [PubMed: 19121390]
- Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*. 1994; 22:4673–4680. [PubMed: 7984417]

- Vera JC, Wheat CW, Fescemyer HW, Frilander MJ, Crawford DL, Hanski I, Marden JH. Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Molecular Ecology*. 2008; 17:1636–1647. [PubMed: 18266620]
- Vogt RG, Rogers ME, Fanco MD, Sun M. A comparative study of odorant binding protein genes: differential expression of the PBPI-GOBP2 gene cluster in *Manduca sexta* (Lepidoptera) and the organization of OBP genes in *Drosophila melanogaster* (Diptera). *Journal of Experimental Biology*. 2002; 205:719–744. [PubMed: 11914382]
- Vogt RG, Prestwich GD, Lerner MR. Odorant binding protein subfamilies associate with distinct classes of olfactory receptor neurons in insects. *Journal of Neurobiology*. 1991; 22:74–84. [PubMed: 2010751]
- Vogt, RG. Molecular basis of pheromone detection in insects.. In: Gilbert, LI.; Latro, K.; Gill, S., editors. *Comprehensive Insect Physiology, Biochemistry, Pharmacology and Molecular Biology*. Volume 3. Endocrinology. Elsevier; London, UK.: 2005. p. 753-804.
- Wanner KW, Nichols AS, Allen JE, Bunger PL, Garczynski SF, Linn CE, Robertson HM, Luetje CW. Sex pheromone receptor specificity in the European corn borer moth, *Ostrinia nubilalis*. *PLoS ONE*. 2010; 5:e8685. [PubMed: 20084285]
- Whitfield CW, Band MR, Bonaldo MF, Kumar CG, Liu L, Pardinas JR, Robertson HM, Soares MB, Robinson GE. Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee. *Genome Research*. 2002; 12:555–566. [PubMed: 11932240]
- Witzgall P, Stelinski L, Gut L, Thomson D. Codling moth management and chemical ecology. *Annual Reviews in Entomology*. 2008; 53:503–522.
- Xu P, Atkinson R, Jones DN, Smith DP. *Drosophila* OBP LUSH is required for activity of pheromone-sensitive neurons. *Neuron*. 2005; 45:193–200. [PubMed: 15664171]
- Yin J, Feng H, Sun H, Xi J, Cao J, Li K. Functional analysis of general odorant binding protein 2 from the meadow moth, *Loxostege sticticalis* L. (Lepidoptera: Pyralidae). *PLoS ONE*. 2012; 7:e33589. [PubMed: 22479417]
- Zhang ZC, Wang MQ, Lu YB, Zhang G. Molecular characterization and expression pattern of two general odorant binding proteins from the diamondback moth, *Plutella xylostella*. *Journal of Chemical Ecology*. 2009; 35:1188–1196. [PubMed: 19823915]
- Zhou JJ, Robertson G, He X, Dufour S, Hooper AM, Pickett JA, Keep NH, Field LM. Characterisation of *Bombyx mori* odorant-binding proteins reveals that a general odorant-binding protein discriminates between sex pheromone components. *Journal of Molecular Biology*. 2009; 389:529–545. [PubMed: 19371749]

Fig. 1.

The distribution of sequence similarities obtained from (A) BLASTN and (B) TBLASTX algorithm search results when *Cydia pomonella* ESTs were queried to the NCBI expressed sequence tag (EST) database (dbEST). The number of database top hits obtained for lepidopteran species using BLASTN and TBLASTX search algorithms are shown in (C).

Fig. 2.

Gene ontologies ascribed to *Cydia pomonella* EST top BLAST hit terms by search of the Gene Ontology (GO) database. Assignments for cellular complement (C), biological process (P), and molecular function (F) are shown at GO level 2 (Note: a given EST sequence may be assigned >1 GO term).

Fig. 3. Neighbor joining trees of codling moth chemosensory and binding proteins sequences (highlighted in red) with 192 other lepidopteran sequences retrieved from GenBank using BLASTP against nonredundant sequence of Lepidoptera. Panel A: alignment of 12 CpomCSPs with 56 lepidopteran chemosensory proteins. Panel B: alignment of 18 CpomOBPs with 140 lepidopteran odorant binding proteins. Tissues in which Cpom transcripts were detected are labeled as follows—eggs (E), neonates (N), male antennae (MA), female antennae (FA), male legs and mouthparts (ML), and female legs and mouthparts (FL).

Fig. 4.

CpomGOBP1, CpomGOBP2, and CpomGOBP3 expression profile in cDNA prepared from total RNA extracted from pharate pupae (Ph), male and female pupae heads (Hd) and bodies (Bod), adult male and female antennae (Ant), heads (Hd), mouthparts (MP) and bodies (Bod), and adult female abdomen tips (Ab Tip). Top panel: Detection of CpomGOBP1 using transcript specific primers in RT-PCR reactions. Second panel: Detection of CpomGOBP2 using specific primers in RTPCR reactions. Third panel: Detection of CpomGOBP3 using specific primers in RT-PCR reactions. Bottom panel: Detection of CpomActin using specific primers in RT-PCR reactions. For CpomGOBP1, CpomGOBP2, CpomGOBP3, and CpomActin, PCR products were visualized on 1.5% agarose gels by ethidium bromide staining and UV illumination. Bands produced were of the size predicted for the transcript specific primer pair.

Table 1

Summary of transcriptome sequencing.

Assembly	Combined		Neonate	Male leg and mouthparts	Female leg and mouthparts	Male antennae	Female antennae
Raw reads	894 230	138901	148 998	177384	104449	176065	146 873
Raw bases	136 032215	20152914	20 699 624	24 323 265	16 960 937	26 525 737	22198 223
Contigs (>100 bp)	8 341	1 869	1 215	1 379	974	1 803	1 638
Average contig length	478	473	472	428	444	427	415
Total contig bases	3 990 137	882 989	573 753	590 774	432 650	770 533	678 883

Frequency of 10 most encountered InterPro accessions from *C. pomonella* EST annotated data, and corresponding *B. mori* gene models with matching InterPro annotation.

Insect Sci. Author manuscript; available in PMC 2014 December 04.

InterPro	N	InterPro description	B. mori gene model (BGBMGA) with similar function
IPR001254	8	Peptidase S1/S6, chymotrypsin/Hap PA, 002281-PA.	PA, 001445-PA, 001486-PA, 001487-PA, 002362-PA, 002548-PA, 002549-PA, 002664-PA, 002770-PA, 003060-PA, 003061-PA, 003062-PA, 003063-PA, 003064-PA, 003065-PA, 003066-PA, 003067-PA, 003068-PA, 003069-PA, 003070-PA, 003071-PA, 003072-PA, 003073-PA, 003074-PA, 003075-PA, 003076-PA, 003077-PA, 003078-PA, 003079-PA, 003080-PA, 003081-PA, 003082-PA, 003083-PA, 003084-PA, 003085-PA, 003086-PA, 003087-PA, 003088-PA, 003089-PA, 003090-PA, 003091-PA, 003092-PA, 003093-PA, 003094-PA, 003095-PA, 003096-PA, 003097-PA, 003098-PA, 003099-PA, 003100-PA, 003101-PA, 003102-PA, 003103-PA, 003104-PA, 003105-PA, 003106-PA, 003107-PA, 003108-PA, 003109-PA, 003110-PA, 003111-PA, 003112-PA, 003113-PA, 003114-PA, 003115-PA, 003116-PA, 003117-PA, 003118-PA, 003119-PA, 003120-PA, 003121-PA, 003122-PA, 003123-PA, 003124-PA, 003125-PA, 003126-PA, 003127-PA, 003128-PA, 003129-PA, 003130-PA, 003131-PA, 003132-PA, 003133-PA, 003134-PA, 003135-PA, 003136-PA, 003137-PA, 003138-PA, 003139-PA, 003140-PA, 003141-PA, 003142-PA, 003143-PA, 003144-PA, 003145-PA, 003146-PA, 003147-PA, 003148-PA, 003149-PA, 003150-PA, 003151-PA, 003152-PA, 003153-PA, 003154-PA, 003155-PA, 003156-PA, 003157-PA, 003158-PA, 003159-PA, 003160-PA, 003161-PA, 003162-PA, 003163-PA, 003164-PA, 003165-PA, 003166-PA, 003167-PA, 003168-PA, 003169-PA, 003170-PA, 003171-PA, 003172-PA, 003173-PA, 003174-PA, 003175-PA, 003176-PA, 003177-PA, 003178-PA, 003179-PA, 003180-PA, 003181-PA, 003182-PA, 003183-PA, 003184-PA, 003185-PA, 003186-PA, 003187-PA, 003188-PA, 003189-PA, 003190-PA, 003191-PA, 003192-PA, 003193-PA, 003194-PA, 003195-PA, 003196-PA, 003197-PA, 003198-PA, 003199-PA, 003200-PA, 003201-PA, 003202-PA, 003203-PA, 003204-PA, 003205-PA, 003206-PA, 003207-PA, 003208-PA, 003209-PA, 003210-PA, 003211-PA, 003212-PA, 003213-PA, 003214-PA, 003215-PA, 003216-PA, 003217-PA, 003218-PA, 003219-PA, 003220-PA, 003221-PA, 003222-PA, 003223-PA, 003224-PA, 003225-PA, 003226-PA, 003227-PA, 003228-PA, 003229-PA, 003230-PA, 003231-PA, 003232-PA, 003233-PA, 003234-PA, 003235-PA, 003236-PA, 003237-PA, 003238-PA, 003239-PA, 003240-PA, 003241-PA, 003242-PA, 003243-PA, 003244-PA, 003245-PA, 003246-PA, 003247-PA, 003248-PA, 003249-PA, 003250-PA, 003251-PA, 003252-PA, 003253-PA, 003254-PA, 003255-PA, 003256-PA, 003257-PA, 003258-PA, 003259-PA, 003260-PA, 003261-PA, 003262-PA, 003263-PA, 003264-PA, 003265-PA, 003266-PA, 003267-PA, 003268-PA, 003269-PA, 003270-PA, 003271-PA, 003272-PA, 003273-PA, 003274-PA, 003275-PA, 003276-PA, 003277-PA, 003278-PA, 003279-PA, 003280-PA, 003281-PA, 003282-PA, 003283-PA, 003284-PA, 003285-PA, 003286-PA, 003287-PA, 003288-PA, 003289-PA, 003290-PA, 003291-PA, 003292-PA, 003293-PA, 003294-PA, 003295-PA, 003296-PA, 003297-PA, 003298-PA, 003299-PA, 003300-PA, 003301-PA, 003302-PA, 003303-PA, 003304-PA, 003305-PA, 003306-PA, 003307-PA, 003308-PA, 003309-PA, 003310-PA, 003311-PA, 003312-PA, 003313-PA, 003314-PA, 003315-PA, 003316-PA, 003317-PA, 003318-PA, 003319-PA, 003320-PA, 003321-PA, 003322-PA, 003323-PA, 003324-PA, 003325-PA, 003326-PA, 003327-PA, 003328-PA, 003329-PA, 003330-PA, 003331-PA, 003332-PA, 003333-PA, 003334-PA, 003335-PA, 003336-PA, 003337-PA, 003338-PA, 003339-PA, 003340-PA, 003341-PA, 003342-PA, 003343-PA, 003344-PA, 003345-PA, 003346-PA, 003347-PA, 003348-PA, 003349-PA, 003350-PA, 003351-PA, 003352-PA, 003353-PA, 003354-PA, 003355-PA, 003356-PA, 003357-PA, 003358-PA, 003359-PA, 003360-PA, 003361-PA, 003362-PA, 003363-PA, 003364-PA, 003365-PA, 003366-PA, 003367-PA, 003368-PA, 003369-PA, 003370-PA, 003371-PA, 003372-PA, 003373-PA, 003374-PA, 003375-PA, 003376-PA, 003377-PA, 003378-PA, 003379-PA, 003380-PA, 003381-PA, 003382-PA, 003383-PA, 003384-PA, 003385-PA, 003386-PA, 003387-PA, 003388-PA, 003389-PA, 003390-PA, 003391-PA, 003392-PA, 003393-PA, 003394-PA, 003395-PA, 003396-PA, 003397-PA, 003398-PA, 003399-PA, 003400-PA, 003401-PA, 003402-PA, 003403-PA, 003404-PA, 003405-PA, 003406-PA, 003407-PA, 003408-PA, 003409-PA, 003410-PA, 003411-PA, 003412-PA, 003413-PA, 003414-PA, 003415-PA, 003416-PA, 003417-PA, 003418-PA, 003419-PA, 003420-PA, 003421-PA, 003422-PA, 003423-PA, 003424-PA, 003425-PA, 003426-PA, 003427-PA, 003428-PA, 003429-PA, 003430-PA, 003431-PA, 003432-PA, 003433-PA, 003434-PA, 003435-PA, 003436-PA, 003437-PA, 003438-PA, 003439-PA, 003440-PA, 003441-PA, 003442-PA, 003443-PA, 003444-PA, 003445-PA, 00344