GigaScience

Genome of the small hive beetle (Aethina tumida, Coleoptera: Nitidulidae), a worldwide parasite of social bee colonies, provides insights into detoxification and herbivory

--Manuscript Draft--

Manuscript Number:	GIGA-D-18-00144R1
Full Title:	Genome of the small hive beetle (Aethina tumida, Coleoptera: Nitidulidae), a worldwide parasite of social bee colonies, provides insights into detoxification and herbivory
Article Type:	Research
Funding Information:	
Abstract:	Background: The small hive beetle (Aethina tumida, ATUMI) is an invasive parasite of bee colonies. ATUMI feeds on both fruits and bee nest products, facilitating its spread and increasing its impact on honey bees and other pollinators. We have sequenced and annotated the ATUMI genome, providing the first genomic resources for this species and for the Nitidulidae, a beetle family that is closely related to the extraordinarily species-rich clade of beetles known as the Phytophaga. ATUMI thus provides a contrasting view as a neighbor for one of the most successful known animal groups. Results: The ATUMI genome encodes fewer enzymes for plant digestion than the genomes of wood-feeding beetles, but nonetheless shows signs of broad metabolic plasticity. Gustatory receptors are few in number compared to other beetles, especially receptors with known sensitivity (in other beetles) to bitter substances. In contrast, several gene families implicated in detoxification of insecticides and adaptation to diverse dietary resources show increased copy numbers. The presence and diversity of homologs involved in detoxification differs substantially from the bee hosts of ATUMI. Conclusions: Our results provide new insights into the genomic basis for local adaption and invasiveness in ATUMI, and a blueprint for control strategies that target this pest without harming their honey bee hosts. A minimal set of gustatory receptors is consistent with the observation that, once a host colony is invaded, food resources are predictable. Unique detoxification pathways and pathway members can help identify which treatments might control this species even in the presence of honey bees, which are notoriously sensitive to pesticides.
Corresponding Author:	Jay D Evans, PhD USDA Agricultural Research Service Beltsville, MD UNITED STATES
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	USDA Agricultural Research Service
Corresponding Author's Secondary Institution:	
First Author:	Jay D Evans, PhD
First Author Secondary Information:	
Order of Authors:	Jay D Evans, PhD
	Duane D McKenna, PhD
	Erin D Scully
	Steven C Cook
	Benjamin Dainat
	Noble Egekwu
	Nathaniel Grubbs
	Dawn Lopez

	Marce D Lorenzen
	Steven M Reyna
	Frank D Rinkevich
	Peter Neumann
	Qiang Huang
Order of Authors Secondary Information:	
Response to Reviewers:	Reviewer #1: Positive Comments This looks like a professionally assembled and annotated genome. It fills in an important blank space within the genomic resources of beetles. Thank you! Major Comments 1. At is not "basal" to anything. It can help you infer what the LCA might have looked like, but it has had an equal amount of evolution compared with Phytophaga. Response: We clarify in this abstract passage that Aethina and the nitidulids diverged basally to what are now Phytophaga by saying that this resource provides the closest available outgroup to the highly successful Phytophaga. Specifically “ATUMI thus provides a contrasting view as a neighbor for one of the most successful known animal groups” 2. No evidence is given for metabolic plasticity, just increased copy number of some metabolic genes. Not the same thing. While we are convinced that gene-family diversity generally leads to flexibility and novel functions in these groups it is true we have not proven this in ATUMI, and we have dropped that speculation from the abstract. We have left discussion of predicted function for specific orthologs in some gene families since these are secure across the insects for which functional analyses are possible. ATUMI is not as conducive to RNAi interference functional knockdowns as other beetles (i.e., such knockdowns are not generally systemic) but we hope these targets will lead to direct tests of function via lack-of-function RNAi assays 3. Why so few shared ortho-groups? Even restricted to SC BUSCO genes, I would still expect several hundred. Dm has the lowest # of SC at ~1050, but only a 1/5 are SC across all species. Seems odd given BUSCO is SC and complete in 90% of species with the clade of interest, by definition. Just struck me as especially low. Response: As each species was added to the BUSCO, the shared single-copy genes were decreased. This reflects singular gene loss in some lineages and also weaknesses/omissions in the public gene sets of other species. Our dataset includes 11 species ranging form Hymenoptera (ants, wasps, and bees), to Coleoptera (beetles) and Diptera (flies), it is then not surprising to find a low number of shared SC BUSCO genes. The phylogenic tree we generated based on shared SC BUSCO alignments is consistent with previous published trees, which suggests the accurate assignment of SC BUSCO genes. 4. Loss and duplication of core genes from small hive beetle genome section. This seems like a lot to read into the lack of something. Only 11 species were analyzed. Either I missed what is being said or this is a sweeping analysis of very few "samples." Response: this is a follow up analysis to the duplicated and missing BUSCO genes. First, we found 337 BUSCO genes were lost from at least two beetle species. Secondly, we mapped these 337 BUSCO genes to the Pfam database. We also mapped the total 2442 BUSCO genes to Pfam database. By analyzing the distribution of function domains between the 337 genes and 2442 genes, no significantly difference was found. But methyltransferase (MT), glycosyltransferase (GT), beta-transducin repeats (WD) and zinc finger (ZF) showed high counts from the 337 BUSCO genes. Thirdly, we calculated the average duplication evens per BUSCO

genes using the equation (total number of duplication events) / (duplicated BUSCO genes). We found a significantly positive correlation between the average duplication events per gene and the number of duplicated BUSCO genes. This result suggests that BUSCO is accurately identifying gene families which shos especially labile gene/protein counts, and hence are candidates for lineage-specific novel functions.

5. GH's, Gr's, Nav's, Ace's, GST's, CP450's, etc sections. They are full of a huge amount of information that is not useful to the central message of the sections. Lots of speculation without that leading to specific hypotheses or broader meaning. It was not clear to me why every observed pattern was explained in such detail. Three examples: Line 265-274 can be deleted without any meaning being lost. Much of the information in the opening paragraphs of each sections are not revisited or used in further paragraphs. Line 346, why would that be informative? Not saying it would not be, but I see no particular reason that it would be. Line 453 paragraph, so much to read into so little evidence, only two of the analyzed species were non-beetles.

Thank you for this suggestion, which is valid, we were perhaps overly excited about some of these shifts, and time plus functional evidence will tell whether the predicted changes in function are real. We have reduced all of the gene family vignettes to what we hope are arguments most relevant to this species and its possible control with novel insecticides. Each section was reduced, on average, by one thiord, and we have in the end trimmed 25 references that were less essential for the arguments that remain. The strong focus on detoxification enzymes came in part from the reality that any steps chosen to control these beetles will be hampered by off-target affects on their sensitive hosts (honey bees), so we are posing possible weak points in the beetle. This information, we expect, will help chemists who are currently designing new controls for ATUMI.

Minor Comments

Abstract. The results have nothing to do with what is discussed with the background section.

We have changed this a bit by omitting comments that are not a focus of the current analysis

Abstract. The reader should be given some indication of gene compliment completeness before the manuscript speaks about gene copy numbers.

Thank you, we have added our BUSCO/completeness parameters as part of the abstract, we feel the captured genes in this analysis are complete and allow for the arguments for gene loss, duplication, etc.

Line 59. Awkward sentence about behaviour.
This was changed.

Line 78. Unclear why this paragraph is sandwiched in between two At life history paragraphs.
We have shifted this big-picture paragraph to follow the (much-shortened) review of ATUMI biology.

Line 107. Endnote field code errors. And a few other places throughout.
fixed

Line 107. Genera should be completely spelled out if it is the first word of a sentence.
in the end we have deleted the discussion of Kodamaea fungi.

Line 113. 343.3 million base pairs is not the number given in Table 1.
this has been corrected, the Table is correct

Line 124 How was DNA extracted? Just realized this is not the methods section. Possibly tell reader details can be found below?

This has been in the supplements and is now given in the methods section lines 410+

Line 142. Why was such as old tool, TopHat2, used? There is a whole generation of better tools; HISAT2, STAR, GSNAP (the updated version). TopHat2 consistently underperforms other tools, especially with default parameters (Baruzzo et al., 2017, Nature Methods).
Resonse: Thank you for encouraging this. we re-analyzed the expressed protein coding genes using Hisat2. The mapped mRNA increased from 99.63% to 99.73%. the mapped protein coding genes increased from 99.56% to 99.65%.

Line 150. "The size of the ATUMI genome assembly is similar to that of the red flour beetle (165.9 Mbp)." Your assemble is over double that, reported as 343 Mb.
Response: The actual assembly was only 234 MBp, and this section has been corrected. The sentence has been revised as “The size of the ATUMI genome assembly is larger than the red flour beetle (165.9 Mbp), but much smaller than the more derived Asian longhorn beetle (707.7 Mbp)”.

Line 164. 2444 needs a comma to be consistent with other number in manuscript.
fixed

Line 202. Recalcitrance. Great word.
Thank you

Line 564. I assume default parameters were used with all programs when not stated. Nice thing to say to remove doubt for all software used.
Response: the parameters have now been added.

Figure 6 legend. In the title, ATUMI is not bolded.
fixed

Reviewer #2: Evans et al., sequenced and assembled a draft genome of small hive beetle and analyzed some gene families based on this genome assembly. This is a very primary work in the filed of genome analysis. I suggested more comparative genomics analysis should be carried out. This manuscript, at its present status, is below the merit of other papers in Gigascience.

Major concerns,

1. The detail procedures of genome sequencing (illumina and PacBio) and genome assembly should be given in detail. How many individuals used for illumina sequencing and how many for PacBio sequencing.
Response: This was an inadvertent omission from the main text, more complete details of both the collections and the DNA extraction methods are now provided, and are also available along with GFF files and fasta files for the assembly and features in the background information for reviewers, thank you.

2. The methods (software and their parameters) of genome assembly should be given in detail

Response: the parameters have all been included now. Additionally, the detailed codes for the assembly have been uploaded as supplemental data for the reviewers.

3. The authors just mentioned that the genome annotation is carried out using NCBI eukaryotic annotation pipeline but without any detail information. This makes the work is hard to be followed.

A more complete citation to the pipeline as well as specific databases and annotations used the infer this gene set are now cited. The NCBI gene set reflects a balancing of assembly and transcript-based evidence with the resources available at NCBI for comparative genomics. This pipeline has proved more effective than in-house pipelines for generating insect gene sets and features.

4. P2 Line 38-40 Conclusion in the abstract. No evidence to support these

	conclusions. I do not think the author can get any in-depth conclusion based on present analysis. This statement has been removed 5. Without Treefam or CAFE analysis, please do not make any conclusions just based on the changes of gene numbers. We have not confirmed the birth and death of paralogs using CAFÉ, in part because this is the first species in it's clade just outside the phytophaga. As more genomes become available for the beetles it will be possible to confirm birth and loss of paralogs. Minor Concerns, 1. The abbreviations in this manuscript are not standard. It is hard to follow the used abbreviations, such as ATUMI, TTCAST. Please use either English name or Latin name instead. This naming scheme has been adopted for gene sets in the insects and likely other eukaryotes in the OrthoDb comparative genomics tables and as a precursor to official gene sets. It seemed awkward to some of us as well, but in the end it is the most even and consistent way to separate the proteins of different species, since a fixed 4-character name is easier to parse out than a variety of common names or species-level names. 2. GH for Glycoside hydrolyses whereas Grs for Gustarory receptors. Please use uniformed abbreviations We have capitalized both 3. For each gene families, especially for Gustatory Receptors, the authors used too many sentences (for GPCR,s they used one and a half page) to introduce the gene families. However, only several sentences were given to the data in this beetle. This should be revised before it is submitted again. These sections have been trimmed substantially. 4. The structure of this manuscript is strange. it has sections of "data description" and "Implicatoins". What is the difference between data description and "material and methods". And what is the difference between "implication" and Discussion. We have reformatted the sections 5. Most words in the section of methods (especially for gene families analysis) are repeats of results. It is unnecessary to repeat each gene family again in the methods. Please summarize the methods. Thank you, we have tried to be more concise in the methods, this was meant to make it clear where decisions were made on family boundaries but we have done our best to reduce redundant descriptions.
Additional Information:	
Question	**Response**
Are you submitting this manuscript to a special series or article collection?	No
Experimental design and statistics Full details of the experimental design and statistical methods used should be given in the Methods section, as detailed in our Minimum Standards Reporting Checklist. Information essential to interpreting the	Yes

data presented should be made available in the figure legends. Have you included all the information requested in your manuscript?	
Resources A description of all resources used, including antibodies, cell lines, animals and software tools, with enough information to allow them to be uniquely identified, should be included in the Methods section. Authors are strongly encouraged to cite Research Resource Identifiers (RRIDs) for antibodies, model organisms and tools, where possible. Have you included the information requested as detailed in our Minimum Standards Reporting Checklist?	Yes
Availability of data and materials All datasets and code on which the conclusions of the paper rely must be either included in your submission or deposited in publicly available repositories (where available and ethically appropriate), referencing such data using a unique identifier in the references and in the “Availability of Data and Materials” section of your manuscript. Have you have met the above requirement as detailed in our Minimum Standards Reporting Checklist?	Yes

Genome of the small hive beetle (*Aethina tumida*, Coleoptera: Nitidulidae), a worldwide parasite of social bee colonies, provides insights into detoxification and herbivory

Jay D. Evans[1*], Duane McKenna[2], Erin Scully[3], Steven C. Cook[1], Benjamin Dainat[4], Noble Egekwu[1], Nathaniel Grubbs[5], Dawn Lopez[1], Marcé D. Lorenzen[5], Steven M. Reyna[5], Frank D. Rinkevich[6], Peter Neumann[7], Qiang Huang[7,8*]

[1]USDA-ARS, Bee Research Laboratory, BARC-East Building 306, Beltsville, Maryland 20705, USA

[2]Department of Biological Sciences, University of Memphis, 3700 Walker Ave., Memphis, TN 38152, USA

[3]USDA-ARS, Center for Grain and Animal Health, Stored Product Insect and Engineering Research Unit, Manhattan, KS 66502, USA

[4]Agroscope, Swiss Bee Research Center, CH-3003 Bern, Switzerland

[5]Department of Entomology and Plant Pathology, North Carolina State University, 1566 Thomas Hall, Raleigh, NC 27695, USA

[6]USDA, Honey Bee Breeding, Genetics and Physiology Laboratory, 1157 Ben Hur Road, Baton Rouge, LA 70820, USA

[7]Institute of Bee Health, Vetsuisse Faculty, University of Bern, Schwarzenburgstrasse 161, CH-3097, Liebefeld, Switzerland

[8]Honey Bee Research Institute, Jiangxi Agricultural University, Zhimin Avenue 1101, 330045 Nanchang, China

* For correspondence

Email: Jay.Evans@ARS.USDA.GOV (JDE); qiang-huang@live.com (QH)

Abstract

Background: The small hive beetle (*Aethina tumida,* ATUMI) is an invasive parasite of bee colonies. ATUMI feeds on both fruits and bee nest products, facilitating its spread and increasing its impact on honey bees and other pollinators. We have sequenced and annotated the ATUMI genome, providing the first genomic resources for this species and for the Nitidulidae, a beetle family that is closely related to the extraordinarily species-rich clade of beetles known as the Phytophaga. ATUMI thus provides a contrasting view as a neighbor for one of the most successful known animal groups.

Results: The ATUMI genome encodes fewer enzymes for plant digestion than the genomes of wood-feeding beetles, but nonetheless shows signs of broad metabolic plasticity. Gustatory receptors are few in number compared to other beetles, especially receptors with known sensitivity (in other beetles) to bitter substances. In contrast, several gene families implicated in detoxification of insecticides and adaptation to diverse dietary resources show increased copy numbers. The presence and diversity of homologs involved in detoxification differs substantially from the bee hosts of ATUMI.

Conclusions: Our results provide new insights into the genomic basis for local adaption and invasiveness in ATUMI, and a blueprint for control strategies that target this pest without harming their honey bee hosts. A minimal set of gustatory receptors is consistent with the observation that, once a host colony is invaded, food resources are predictable. Unique detoxification pathways and pathway members can help identify which treatments might control this species even in the presence of honey bees, which are notoriously sensitive to pesticides.

Keywords: Coleoptera, pollination, *Apis mellifera*, invasive pest, phytophagy, glycoside hydrolase.

Introduction

The small hive beetle (= ATUMI; *Aethina tumida* Coleoptera: Nitidulidae, Murray, 1867) is a rapidly spreading invasive species originating from sub-Saharan Africa. ATUMI is now found on all continents except Antarctica [1-4]. Outside of its endemic range, it has become an economically important parasite of social bee colonies, including honey bees, bumblebees and stingless bees [2] (Figure 1). ATUMI significantly impacts beekeeping and the regulation of honey bees and hive products worldwide. ATUMI eggs are laid within colonies and developing larvae feed until they leave the colony for pupation [2]. ATUMI pupate in the soil then emerge as adults to infest social bee nests. Once inside the bee nest, adult ATUMI employ a “sit-and-wait” strategy, relying on the resources of the nest for nutrition and shelter until options for successful reproduction

arise [2]. ATUMI larvae and adults can feed on a large variety of food sources inside and outside of social bee colonies, including fruits, meat, adult bees, bee brood and bee food stores (pollen and honey) [1, 5, 6]. Beetles and their bee hosts show an elaborate set of interactions. For example, honey bees attempt to confine adult ATUMI to prisons built from plant resins [6] and beetles can also manipulate guard bees to obtain food by rubbing their antennae against the guarding bees' mandibles, inducing them to regurgitate food.

ATUMI belongs to the beetle family Nitidulidae (sap beetles; c. 4,500 species), which feed mainly on decaying vegetable matter, over-ripe fruit, or sap. The Nitidulidae belong to the superfamily Cucujoidea (sap, bark and fungus beetles), which is either the sister-group of the Phytophaga (leaf beetles, weevils, longhorned beetles and their relatives; [7], the most species rich radiation of plant-feeding animals on Earth with >125,000 described species), or forms a paraphyletic clade subtending the Phytophaga [8, 9]. In the latter case, the Phytophaga are derived from within Cucujoidea. Interestingly, the trophic habits of Nitidulidae may therefore represent a transitional stage from fungivory, saprophagy, and detritivory (the typical habit(s) of most Cucujoidea and its containing clade, series Cucujiformia) to phytophagy (feeding on plants), the typical trophic habit of Phytophaga. Comparative studies of the ATUMI genome may therefore provide new insights into the evolution and genomic basis of phytophagy in beetles.

To date, just 10 beetle genome assemblies have been released [10], of which only 7 are published, despite there being >400,000 described beetle species. These are:

Tribolium castaneum (red flour beetle, TCAST; Tenebrionoidea: Tenebrionidae: Tenebrioninae; [11], *Anoplophora glabripennis* (Asian longhorned beetle, AGLAB; Chrysomeloidea: Cerambycidae: Lamiinae; [12]), *Dendroctonus ponderosae* (mountain pine beetle, DPOND; Curculionoidea: Curculionidae: Scolytinae; [13]), *Hypothenemus hampei* (coffee berry borer beetle, HHAMP; Curculionoidea: Curculionidae: Scolytinae; [14]), *Oryctes borbonicus* (Reunion Island scarab beetle, OBORB; Scarabaeoidea: Scarabaeidae: Dynastinae; [15]), *Onthophagus taurus* (bull headed dung beetle, OTAUR; Scarabaeoidea: Scarabaeidae: Scarabaeinae; Unpublished), *Nicrophorus vespilloides* (burying beetle, NVESP; Staphylinoidea: Silphidae: Silphinae; [16]), *Agrilus planipennis* (emerald ash borer, APLAN; Buprestoidea: Buprestidae: Agrilinae; Unpublished), *Leptinotarsa decemlineata* (Colorado potato beetle, LDECE; Chrysomelidae: Chrysomelinae: Doryphorini; [17]), and *Pogonus chalceus* (salt marsh beetle, PCHAL; Carabidae: Trechinae: Pogonini; Unpublished). The ATUMI genome described here joins this group as the only representative from the superfamily Cucujoidea.

A robust reference genome assembly comprised of 243 million base pairs was used to identify and annotate 14,076 protein coding genes, over 3000 additional transcribed features and a strong complement of repetitive DNA's, tRNA's, and transposable elements. The described protein-coding genes provide strong candidates for core metabolism and development, and suggest that these beetles, like their honey bee hosts, rely on olfactory cues and less so on chemosenses related to taste. An analysis of protein groups involved in insecticide metabolism reveals a large repertoire of detoxification enzymes to mediate xenobiotic interactions. The described resources

will be useful for both chemical and non-chemical approaches for controlling this key pest of honey bees.

Results and discussion

Genome traits, genetic diversity and phylogenetic analysis

We generated a genome assembly of 243 Mbp (Million base pair) comprised of 3063 contigs (contig N_{50}=298kb; Table 1). The genome sizes of sequenced and assembled beetle species vary greatly from 160Mbp to 1.1Gbp. The size of the ATUMI genome assembly larger than that of the red flour beetle (165.9 Mbp), but much smaller than the more derived Asian longhorned beetle (707.7 Mbp). A total of 1,293,015 heterozygous single nucleotide polymorphism (SNP) positions were identified, with an average density of one SNP per 181 bp. SNP density was significantly different across contigs (T-test, $P < 0.01$). This pattern was not related to contig size. Overall, 60.2% of SNPs occurred on contigs with annotated genome features and 22.5% were within gene regions.

The NCBI eukaryotic genome annotation pipeline (https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Aethina_tumida/100/) proposed 14,076 protein-coding genes and a total of 17,436 mRNA models. When our previous RNA sequencing reads were aligned to the genome assembly alongside the predicted gene models, 99.73% of the predicted mRNA models and 99.65% of the predicted protein-coding genes were supported. It is possible that the 64 protein-coding genes undetected by RNA-Seq were not expressed, expressed too briefly, or not

captured in our pooled RNA samples. Alternatively, these might reflect partial or inaccurate gene models or pseudo-genes that are no longer functional in this beetle.

By aligning the ATUMI official protein set against 2,444 core Endopterygota Benchmarking Universal Single-Copy Orthologs (BUSCO), 97.5% of complete BUSCOs were found (Figure 2b). We further aligned the ATUMI genome assembly against Endopterygota set of BUSCOs and 92.8% of complete BUSCOs were found (Supplemental File 1). The results suggest a high level of completeness in the genome assembly, as well as the official set of gene models. By comparing single-copy orthologs among the sequenced beetles (ATUMI, TCAST, DPOND, AGLAB, ATAUR, APLAN, HHAMP, NVESP), honey bees (AMELL) and *Drosophila melanogaster* (DMELA), 181 shared ortholog groups were found. A phylogenetic tree was built by concatenating these shared 181 orthologous groups (Figure 2a). These results suggest that ATUMI is sister to TCAST and the Asian longhorned beetle (AGLAB). OrthoDB [18] orthology delineation revealed that ATUMI has 7066 conserved orthologous groups with beetles and 4554 orthologous groups shared with ten additional insect species.

Loss and duplication of BUSCO genes from the small hive beetle genome

The duplication and absence of core genes, including those represented by BUSCO, could represent important evolutionary changes in species or in lineages (Figure 3). A complete protein set of 11 insect species was used for alignment against the ATUMI BUSCO candidates. We found 337 core Endopterygota BUSCOs that were either fragmented or completely lost from at least two beetle genomes. We mapped the

common ancestor sequences of these 337 missing orthologs and full set of 2442 Endopterygota BUSCOs to the Pfam database. Among the 'lost' orthologs, 1094 protein domains were found, while among 2442 Endopterygota orthologs, 4632 protein domains were found. By comparing the count distribution of each domain between lost orthologs and overall orthologs, no significant difference was found (Pearson's Chi-square Test, $P>0.05$). Among the lost orthologs, a methyltransferase (MT), a glycosyltransferase (GT), and two proteins with beta-transducin repeats (WD) and zinc finger (ZF) domains, respectively showed the highest counts and were also absent from at least four beetle species (Figure 2c).

Glycoside hydrolases

Glycoside hydrolases (GHs) are important enzymes that aid in the digestion of plant cell walls and carbohydrates in insects [19]; however, GHs can also contribute to remodeling of the peritrophic matrix (PM) [20], lysosomal enzyme activity, glycoprotein oligosaccharide catabolism, immune response, and growth and development [21, 22]. A limited diversity of GH families was identified in the ATUMI genome when compared to other beetles. While phytophagous insects, such as AGLAB [7], DPOND [13] and HHAMP [14] harbored anywhere from 19-24 different GH families represented by 101-199 genes, only 14 GH families represented by 91 genes were identified in the ATUMI genome. Only OBORB, whose diet is unknown [15], had a lower GH family diversity and

GH copy number, with 13 different families represented by 47 different genes. No GH families unique to ATUMI were identified (Supplemental Files 2 and 3).

Using orthology searches, five orthogroups containing GHs were more prominent in the ATUMI genome compared to other beetles and two GHs lacked orthologs in other beetle genomes. The more prominent orthogroups contained genes with highest scoring BLASTP matches to GH 30 glucosylceramidase (eight copies; sphingolipid metabolism), uncharacterized GH 31 α-glucosidases (five copies), GH 16 β-1,3-glucan binding protein (five copies; exoskeleton and/or PM remodeling), GH 38 lysosomal α-mannosidase (five copies), and GH 18 chitinase (three copies). Interestingly, unigenes coding for GH 18 (20 copies), GH 31 (11 copies), and GH 38 enzymes were also among the most prominent GH families in the ATUMI genome (Figure 4). Generally, GH 38 copy numbers were high in the ATUMI genome relative to other beetles and were exceeded only by TCAST. In contrast, copy numbers of GH 18 and 31 genes were similar to those found across other beetles. Additionally, two GH genes encoded by the ATUMI genome lacked orthology to other beetle GHs, including a GH 2 family gene coding for β-mannosidase and a GH 35 family gene coding for β-galactosidase. Other beetles code for GH 2 β-mannosidases and GH 35 β-galactosidases, so it is unclear why these two genes were not assigned to orthogroups. However, the evolutionary history of genes coding for GH enzymes is complex and it may be difficult to assign orthologs in some cases.

Overall, ATUMI lacked a diverse and expansive repertoire of GHs relative to phytophagous beetles, which may reflect the ATUMI diet. Pollen generally contains high concentrations of the monosaccharides glucose and fructose [23], which are used

directly for ATP production by the glycolysis pathway (glucose) or after phosphorylation by fructokinase (fructose). Therefore, although pollen can also contain starch, sucrose, and small amounts of pectin [23], digestion of more complex carbohydrates may not be necessary, requiring a less expansive repertoire of GH enzymes relative to phytophagous beetles. Supporting this hypothesis, genes coding for enzymes capable of digesting starch were identified (α-amylase), but genes coding for invertases and polygalacturonases for sucrose and pectin digestion could not be identified. Alternatively, microbial symbionts harbored by ATUMI may facilitate the breakdown of these polysaccharides as has been observed previously in their honey bee hosts, which share a similar diet [24].

Gustatory Receptors

G-protein-coupled receptors (GPCRs) comprise a large family of integral membrane proteins found in cells of all eukaryotes [25]. GPCRs function to detect extracellular stimuli, and activate cellular signal transduction pathways that ultimately lead to physiological and behavioral responses. Gustatory receptors (GRs) belong to novel arthropod GPCR gene superfamilies, which are phylogenetically unrelated to mammalian taste receptor genes, and distinct from related insect odorant/pheromone receptor genes [26]. GRs are important components of an organism's sensory machinery; the ability of animals to distinguish between nutritious, noxious, and possibly toxic compounds is a matter of life or death. Sensory machinery has been honed over evolutionary time, and has given rise to receptors binding either sweet

(attractive) or bitter (aversive) tastants, [27, 28]. An amino-acid substitution in a ligand-binding region may affect the range at which different ligand's receptors may bind, particularly for GRs perceiving sugars [29].

Gustatory receptor genes fall into four main clades that correspond with perception of different tastants (sweet or bitter; Figure 5). Designations of the type of substance perceived by these receptors can be inferred from other taxa (e.g., *Drosophila* sp.), and the positions of uncharacterized proteins within the cladogram. A group of apparently highly conserved genes encoding proteins for perceiving sweet substances (clades *5a* and *64a-f*) is separate from other groups that show higher sequence variability; a pattern seen in other studies (e.g., [30]). Proteins of *GR5a* and *GR64a-f* can form heterodimeric complexes at receptor sites, and may or may not be necessary together for perception of different sugars [31, 32]. ATUMI appears to lack a *GR5a* gene (Table 2; Figure 5), suggesting this gene may not be necessary for perceiving sweet tastants. In this group of ATUMI *GR*s, it is interesting to note that one candidate with a very long branch-length (XP_019866072) encodes a 379 amino-acid protein derived from 3 exons, and has a very long intron. It is unclear why this gene is so distinct compared to the relatively highly conserved sequences for other related *GR* genes.

A major finding is that ATUMI has a substantially depauperate repertoire of *GR* genes compared to both AGLAB and TCAST (Figure 5). This low number of *GRs* in ATUMI is more likely the result of a lack of gene expansion in particular lineages or subfamilies of *GRs* rather than gene loss. A similarly small number of *GRs* is evident in the honey bee genome [33]. In that species, the relatively reduced *GR* gene repertoire may be a

consequence of restricted dietary breadth (specialist on pollen and nectar), and also possibly arises from the processing of collected foods by adult workers and microbes, which may reduce the load of plant secondary compounds. AMELL larvae are fed processed foods by attending nurse bees, so they may not need an expansive repertoire of GRs to discriminate among different tastants [30]. Because of the close affinity of ATUMI with honey bees, including sharing a similar diet, the evolutionary pressures limiting expansion of *GRs* in ATUMI may be similar. As an example, TCAST, a dietary generalist, shows a significant expansion in the *GR28a/b* gene complex (Table 2); genes in this complex may be important for perceiving plant secondary compounds [34].

Stemming from their importance to insect biology, *GRs* have been characterized from genomic and transcriptomic studies for a number of economically important insects, or those having an ecological and/or epidemiological significance, including TCAST [11], AGLAB [12] and now ATUMI (this study). Understanding the chemosensory abilities of insects, particularly pest insects, is important for designing possible means of control that target the insect's ability to find and/or distinguish among nutrients or to detect poisons, and/or developing baits containing insecticides formulated with highly attractive substances.

Voltage-gated sodium channel

The voltage-gated sodium channel (Na_{v1}) is responsible for generating action potentials in neurons. Sodium channel modulator insecticides such as pyrethroids and DDT act on the Na_{v1} channel by maintaining the open state of the channel via interactions with two

proposed binding sites [35, 36]. A diverse collection of mutations in Na_{v1} has been identified in many populations of pyrethroid-resistant pests and neurophysiological studies of heterologously-expressed channels have confirmed the role of these mutations in pyrethroid resistance [37].

A single transcript and protein were predicted for Na_{v1} from the ATUMI assembly However, Na_{v1} is known to possess optional and alternative exons in most insects [38-40]. Alternative exon use diversifies the physiological repertoire of the sodium channel and may affect insecticide sensitivity [41]. Further cloning experiments to determine the actual optional and alternative exon use in ATUMI Na_{v1} should be informative.

A large number of mutations in Na_{v1} have been associated with target site resistance to pyrethroids and DDT [37]. We did not identify such mutations in the predicted ATUMI Na_{v1} nor is this species known to be resistant to these insecticides. Therefore, this sequence serves as a reference for a susceptible target site for pyrethroids and DDT and a tool for developing molecular diagnostic assays to monitor changes in resistance allele frequency.

Acetylcholinesterase

Acetylcholinesterase (Ace) cleaves acetylcholine (ACh) to regulate the effect of the neurotransmitter in the synaptic cleft. Ace is the target of organophosphate (OP) and carbamate insecticides and mutations in Ace result in target-site insensitivity to these two insecticide classes [42, 43].

ATUMI is predicted to possess active forms of both Ace1 (XP_019871456.1) and Ace2 (XP_019866656.1) (Supplemental File 2). Ace mutations involved in OP resistance [43, 44] are found to be in the susceptible state in the predicted Ace proteins of ATUMI (Table 3). In the cases where an alternative amino acid was found in ATUMI (i.e, ATUMI_Ace2 position 198), that same amino acid was seen in other insects that were presumably sensitive to OPs, so it does not likely confer reduced OP sensitivity. Ace2 performs primary acetylcholine esterase activity in honey bees, while Ace1 is the primary enzyme in beetles and most other insects [45]. Therefore identifying compounds that only inhibit ATUMI_Ace1 may provide a level of ATUMI-specific control.

ATP-Binding Cassette Proteins

ATP-binding cassette (ABC) proteins are a large, diverse family of proteins found in most organisms, from bacteria to plants and vertebrates. Most ABC proteins engage in active transport of molecules across cell membranes. This family of transporters is perhaps most notable for moving toxins into or out of cells, which has resulted in the identification of several of these proteins playing a role in the resistance of cancer cells to multiple drug treatments (Multi-Drug Resistant, MDR). So it is not surprising that some of these proteins have been identified as having roles in insect susceptibility or resistance to certain insecticides (Reviewed by [46]. In spite of their importance for shaping pest control methods, these genes are under-studied in insects, with few having been fully characterized in any species. The status of ATUMI as a pest of beehives makes it important to understand what role ABC genes may play in how beekeepers control this species.

The beetle genetic model organism, TCAST, has had its full suite of ABC-family genes identified through a combination of RNA-seq and genomic analysis. In this species, 74 genes have been identified (Table 4); [47, 48]. The translation products of these genes were used to query the ATUMI genome, in which 56 ABC genes were identified (Table 4). In most respects, the makeup of ABC genes in ATUMI resemble those found in TCAST - both species have identical numbers of ABC-B, D, E, F, and H subfamily members. Indeed, the numbers of members in the D-F and H subfamilies are highly conserved, with DMELA having the same number, and clear one-to-one relationships can be seen in these subfamilies among the members from each species (Figure 6). It should be noted that members of subfamilies E and F do not function as transporters, and are highly conserved in number and sequence between

insects and humans. Moreover, RNAi targeting ABC-E and one of the ABC-F genes in TCAST resulted in complete mortality, suggesting that the critical cellular roles of these genes may also be conserved. The ABC-B subfamily also appears well conserved, and may be worth further scrutiny in ATUMI, since this subfamily has been associated with resistance to several classes of pesticides in multiple species [46].

ATUMI differed from TCAST in member counts for three ABC subfamilies (Table 4). The first was subfamily A, for which only four members could be identified in ATUMI, relative to the ten found in TCAST and DMELA, a number roughly consistent across the insects. However, it is important to note that ABC-A genes are fairly large full transporters, and as such are often complex and difficult to identify in full. So, it is likely that some of the ABC-A genes are either not present in the current genome assembly, or are too fractured to recognize. It is also interesting to note that the beetle ABC-A genes appear to segregate from those of DMELA (Figure 6), suggesting possible pesticide targets against ATUMI, which may not harm other species, including pollinators.

TCAST appears to have one more ABC-G gene than does ATUMI; specifically, ATUMI appears to lack an ortholog of the well-studied DMELA eye-pigment transporter known as Brown (Bw). However, it has been well documented that Bw orthologs has substantially diverged in TCAST [48]. It is possible that similar divergence has also prevented clear identification of a Bw ortholog in ATUMI. Otherwise, most other ABC-G genes have clear one-to-one orthologs in all three species (Figure 6).

The largest subfamily, the ABC-C genes, is known to play roles in multi-drug resistance in human disease, and some have been associated with Bt resistance in lepidopterans [46].

ATUMI has fewer ABC-Cs than TCAST, but more than DMELA. At first, this might suggest a beetle-specific expansion as well as a TCAST-specific expansion. Indeed, there is a suite of expansions that may be beetle-specific (Figure 6), although comparisons to more species would be required to confirm this. However, each species also appears to have its own expansions; TCAST and ATUMI expansions are often tandem, as can be seen by the number of genes found on the same linkage groups/scaffolds (Figure 6). Indeed, there are surprisingly few clear one-to-one orthologous relationships, suggesting rapid evolution of ABC-C genes to fill species-specific needs. To understand ATUMI responses to pesticides, these ATUMI-specific expansions may be worth further study.

Gluthatione-S-Transferase

Gluthatione-S-Transferases (GSTs) are conjugases that bind glutathione to a wide variety of substrates such as plant allelochemicals, insecticides, reactive oxygen species, and metabolic products that can provide detoxification, antioxidant, excretion, and transport functions [49], [50], [51]. Insect GSTs are widely studied due to their role in insecticide resistance [52]. Genomic analyses show that insects possess between 10 to 41 genes that encode GSTs distributed across 8 classes (i.e. Delta, Epsilon, Omega, Sigma, Theta, Zeta, Microsomal, and Unclassified) [53].

In the ATUMI genome, 49 GSTs were identified, 9 of which displayed isoforms (Figure 7; Table 5). The number of genes in the ATUMI genome is very similar to what has been identified in TCAST, especially in the Delta, Epsilon, Sigma, and Theta classes. Relative to other insects, ATUMI and TCAST have expansions in the Epsilon, Sigma, Zeta, and Microsomal GST classes,

which supports the hypothesis that these may be Coleoptera-specific class expansions [53]. The small number of genes in the Delta class for both ATUMI and TCAST suggests a class contraction or lack of expansion within the beetles.

Increases in the expression and activity Delta and Epsilon classes confer resistance to diverse classes of insecticides such organophosphates, organochlorines (DDT), and pyrethroids [50], [52]. These two GST classes tend to be the most numerous and dynamic in terms of expansions and contractions [53]. Therefore it would appear that ATUMI possesses a wide diversity of GSTs, especially in the Epsilon class, to detoxify insecticides utilized for their control.

Cytochrome P450

The Cytochrome P450 monooxygenases (CYP450s) are classified as phase I metabolic enzymes which are involved in the biosynthesis, bioactivation, and regulation of endogenous compounds such as hormones, fatty acids, and sterols as well as detoxification of xenobiotic compounds such as plant alleleochemicals and insecticides. Overexpression of CYP450s often underlies high levels of detoxification-mediated insecticide resistance in many insects [54] [55] [56]. In the 69 insect genomes that have been published, more than 7,500 P450 genes have been identified in 208 families across 4 clans (CYP2, CYP3, CYP4 and Mitochondrial) [57].

In ATUMI, we found 116 genes across the 4 CYP clans (Figure 8, Table 6). The CYP2 and mitochondrial clans contained 8 and 10 genes, respectively, and orthologs were identified in other species. The conservation in sequence and number is expected as many of the genes in these clans are involved in ecdysteroid biosynthesis [58]. In contrast to the conserved CYP2 and mitochondrial clans, there are clear expansions in CYP3 and CYP4 compared to other species.

These expansions are typified by large expansions of a single family that lacks orthologs in other species [59]. Within the CYP3 clan, the 55 genes are clustered in smaller blooms with the largest consisting of 13 genes. The 43 genes belonging to the CYP4 clan of ATUMI is among the largest seen in insects [7] with a noticeably large bloom of 20 genes. Additionally, CYPs in the CYP3 and CYP4 clans have been implicated in insecticide resistance [60] [61] [62]. Therefore, a rapid onset of insecticide resistance may be facilitated by the large number of CYPs in the CYP3 and CYP4 clans in the ATUMI genome.

Carboxyl/Choline Esterases

Carboxyl/Choline Esterases (COEs) are capable of metabolizing a wide variety of substrates and their activity is involved in a number of physiological processes such as bioactivation of juvenile hormone and regulating acetylcholine interactions at the synapse [63] [64]. Increases in the amount of esterase expression and mutations in the catalytic site of esterases confer insecticide resistance [65] [66]. Insects possess a wide variety of COEs that are broadly classified as intracellular or dietary (Clades A-C), secreted pheromone/hormone processing (Clades D-G), and neurodevelopmental (Clades H-M) [63].

The ATUMI genome contained 60 genes encoding putative COEs, with only one displaying multiple isoforms (Figure 9). The number of genes in the secreted and neurodevelopmental groups was mostly consistent with other insects (Table 7). The expansion of Clade E (Secreted β-esterase) is consistent with a similar expansion in TCAST. This expansion is not entirely characteristic of Coleoptera as DPOND and AGLAB only have 4 and 1 member of Clade E, respectively [7]. The 10 genes for neuroligins is nearly twice the number seen in other

insects [12] [67] [68]. Nevertheless, the general conservation in sequence and number suggests critical roles for these COEs across insects. In contrast to COEs in the secreted and neurodevelopmental groups, a vast majority of ATUMI COEs in the intracellular or dietary class lacked clear orthologs in TCAST, AMELL, or DMELA. This expansion of intracellular or dietary esterases is consistent with expansions observed in other insect genomes. These species-specific expansions of intracellular or dietary esterases may be due to dietary differences among these insects. Dietary esterases may also contribute to insecticide resistance [63]. Therefore, this expansive array of dietary esterases may allow ATUMI to detoxify insecticides that may be used for control.

ATUMI is an expanding invasive pest of honey bees, disrupting managed bee colonies and arguably having a strong impact on feral on naturally occurring colonies. We anticipate the resources described here will lead to novel methods to track and control this pest. The ATUMI genome also reveals numerous evolutionary distinctions relative to other sequenced arthropods. These distinctions help clarify the sensory cues used by ATUMI and the dietary habits of this beetle, and of beetles (order Coleoptera) more broadly.

Methods

DNA extraction

ATUMI adults were collected from a population maintained by the USDA-ARS Honey Bee Breeding, Genetics and Physiology Laboratory (Baton Rouge, LA) in November 2011. ATUMI larvae were collected March 8, 2014, from a continuous culture of small hive beetles maintained at the USDA-ARS Bee Research Laboratory. For adult beetles, extractions were

carried out on three whole male beetles using the Qiagen DNAEasy kit. Larval DNA was extracted from 150 second-instar larvae in 30 groups of five larvae each. Larvae were crushed using a plastic pestle in 1ml of freshly prepared CTAB buffer consisting of 100 mM TrisHCl (pH 8.0), 20 mM EDTA (pH. 8.0), 1.4 M NaCl, 2% CTAB and 0.2% β-mercaptoethanol. The suspension was incubated at 65^{o}C for 60 minutes, with gentle mixing at 0, 20, and 40 minutes. Samples were centrifuged for 2 min at 14k rpm (20817 g) in an Eppendorf microcentrifuge. 500 µl of the supernatant was moved into a new tube containing using a wide-bore pipette into a sterile tube containing 500 µl chloroform:isoamylalcohol (24:1). After gentle mixing by hand, tubes were centrifuged at 14k rpm for 15 min. Approximately 400 µl of the aqueous layer was transferred into new tubes containing 250 µl cold isopropanol, followed by gentle mixing and incubation at 4^{o}C for 30 minutes. Samples were centrifuged at 14k rpm for 30 min a 4^{o}C, and then the supernatant was poured off. Pellets were washed with 1 ml cold 75% EtOH and centrifuged again for 2 min at 14k rpm. After the supernatant was poured off, the resulting pellets were washed in 1 ml cold 100% EtOH, centrifuged for 2 min, after which the EtOH was poured off, the pellets were spun for an additional 30 seconds, and the last of the wash was removed by pipette. Pellets were air-dried for 30 minutes and the resulting DNA pellet was resuspended in 50 µl ddH20. Samples were incubated for 30 min with 2.5 µl of an RNAse cocktail at 37^{o}C, followed by gentle addition of 5 µl of 7M NaOAc and 100 µl EtOH. After 30 minutes of incubation on wet ice, the DNA samples were spun at 12k rpm for 30 min, washed once with 70% EtOH, dried and suspended in 20 µl ddH20. Extracts were pooled and assayed by gel electrophoresis to ensure DNA integrity and by Nanodrop (Thermofisher, Inc.) for quantification (180 ng/µl in 25 µl, 45 µg total DNA).

DNA sequencing

In total, 1,173,425,522 Illumina DNA reads (101 base-pairs [bp] per read with a 300 bp insert size, Hi-Seq 2500) were generated from 12 paired-end (PE) libraries generated from DNA from the three adult male beetles. An additional 1,235,055 Pacific BioSciences (PacBio) reads (average read length = 6795 bp) were generated from 40 SMRT cells (Chemistry C2, PacBio, Menlo Park, CA), using DNA derived from the pooled larval beetles. A two-step method was used to assemble the genome. First, the Sparse assembler was used to build short but accurate contigs from the Illumina reads using the settings: (LD 0 K 41 g 15 NodeCovTh 1 EdgeCovTh 0 GS 600000000) [69]. The assembled contigs were used as a backbone for further assembly. Second, the PacBio reads were error corrected by the proovread package (default settings) [70] and the error-corrected PacBio reads were used to construct long contigs by filling the gaps of the backbones using the Sparc package deployed with default settings [71]. Genes were annotated using version 7.2 of the NCBI eukaryotic annotation pipeline [72]. Illumina mRNA paired-end sequencing reads (101 bp per read, >1000x transcriptome coverage) reflecting an equimolar pool of all ATUMI life stages (described in [73] and downloaded from USDA AgDataCommons; https://tinyurl.com/ybanauxb) were used to assist gene annotation. Full annotation details for this gene set are described at https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Aethina_tumida/100/. Transcriptome sequencing reads were aligned to the constructed ATUMI genome assembly to evaluate the completeness of the gene set, using the TopHat2 package [45]. Reads were also mapped using HISAT2 (https://ccb.jhu.edu/software/hisat2/index.shtml), showing a marginal increase in

aligned reads. We further assessed the completeness of the genome assembly using BUSCO (Benchmarking Universal Single Copy Orthologs; [74]).

Phylogenetic and genetic diversity of beetles

The official protein sets of ATUMI, the red flour beetle (*Tribolium castaneum*) [11], mountain pine beetle (*Dendroctonus ponderosae*) [13], Asian longhorned beetle (*Anoplophora glabripennis*) [12], dung beetle (*Onthophagus taurus*; https://www.ncbi.nlm.nih.gov/assembly/GCA_000648695.1/), emerald ash borer (*Agrilus planipennis*; https://www.ncbi.nlm.nih.gov/assembly/GCF_000699045.1/), coffee borer beetle (*Hypothenemus hampei*) [14], burying beetle (*Nicrophorus vespilloides*) [16], scarab beetle (*Oryctes borbonicus*) [15], honey bee (*Apis mellifera*) [75], and fruit fly (*Drosophila melanogaster*) [76] were used to query the BUSCO Endopterygota ortholog set. Single copy orthologs shared by all 11 insect species were further used for phylogenetic analysis. Protein sequences of these orthologous groups (OGs) were aligned using MUSCLE using default protein settings [77]. Alignments were quality trimmed with trimAI (-w 3 –gt 0.95 –st 0.01) [78] and the orthologous groups were concatenated for use in phylogenetic analysis. A maximum likelihood (ML) tree search was implemented using the program RAxML version 8.2.9 [79] with 1000 bootstrap replicates (-N 1000 –m PROTGAMMAAUTO –f a). The final tree was viewed and edited with TreeGraph2 [80]. Microsatellite markers were identified in the ATUMI genome assembly using the Microsatellite Search and Building Database (MSDB) package and default settings [81]. The raw Illumina gDNA reads, used to assemble the ATUMI genome, were re-aligned to the assembly using BWA with default settings [82]. The aligned reads were used to

identify single nucleotide polymorphism (SNP) positions using GATK under default settings (version 3.6; [83]), and the further annotated with SNPEFF [84].

Gustatory Receptors

The repertoire of gustatory receptors has been preliminarily characterized for TCAST [85] (62 GRs) and *A. glabripennis* [86]. Additionally, online databases have listed gustatory receptors for *T. castaneum*, including UniProtKB (www.uniprot.org [87]), and BeetleBase (www.Beetlebase.org [88]). Amino acid sequences for putative and identified *GR* genes were compiled from these resources and truncated to remove redundancies. The compiled TCAST gene set contained 71 *GR* genes. To identify and enumerate gustatory receptors for AGLAB and ATUMI, amino acid sequences of TCAST gustatory receptor genes were submitted to the ATUMI RefSeq gene set and genome assembly using BLASTP and TBLASTN, respectively. Putative *GR* genes for both species were selected from hits based on an E-score ≤ to E^{-100}. Using the data set of *GR* genes compiled for *T. castaneum*, 38 and 11 putative GR proteins were identified for AGLAB and ATUMI, respectively. Sequences were aligned using MUSCLE [77]. The PhyML program (v3.1/3.0 aLRT) was used to build a phylogenetic tree using maximum likelihood method [25, 89]. The tree was further edited and visualized with the TreeDyn (v198.3) program [90]. All analyses from the sequence alignment to tree reconstruction were performed on the phylogeny.fr platform [91]. Sequences obtained in Newick format from this platform were used as input in the iTOL program to construct and visualize using an unrooted, circular phylogenetic tree [92].

ABC Transporters

Potential ATUMI ABC genes homologous to TCAST ABCs were identified using protein BLAST to search with each TCAST ABC sequence using WebApollo at https://i5k.nal.usda.gov/aethina-tumida. Protein sequences from ATUMI, TCAST, and DMELA were then compiled, and trimmed to exclude all but 51 residues around the Walker B motif of the nucleotide binding domain. This 51-amino-acid sequence was then used to build the phylogenetic tree (See Table 3 for the sequences used from ATUMI). The maximum likelihood phylogenetic tree was constructed using the program MEGA, version 7 [93], using default parameters in all categories except: LG model of amino-acid substitution with Gamma distributed substitution rates (based on Best Model determination within the MEGA program), and Partial Deletion treatment of gaps/missing data [94].

Insecticide targets and detoxification genes

The predicted proteins from the official gene set of ATUMI (taxid 116153) were queried with TCAST orthologs for gene families and pathway members related to insecticide resistance via BLASTP. Putative orthologs in ATUM were designated by >95% query coverage and E-value $<1E^{-100}$.

Additional files

File S1 MS-Word. Detailed material and methods.

File S2 MS-Excel. Orthology assignments for glycoside hydrolases (GHs) coded by ATUMI**.**

File S3 MS-Excel. Protein identifiers for orthogroup assignments.

Data and Material

Data supporting the results of this article are deposited at NCBI-Bioproject PRJNA256171.

Abbreviations

ATUMI: small hive beetle; BUSCO: Benchmarking Universal Single Copy Orthologs

Competing Interest

The authors declare no competing interest.

Funding

No external funding

Author contribution

J.D.E. and Q.H. designed the study. E.S. analyzed digestive proteins. D.M., B.D., N. G., S. C., N. E., M. L., S. R. and F. R. analyzed detoxification genes. J.D.E., D.M., B.D., N. G., S. C., N. E., M. L., S. R., P. N., F. R. and Q.H. wrote the manuscript.

Acknowledgements

We appreciate the bioinformatics support from the Competency Centre in Bioinformatics and Computational Biology, Bern, Switzerland

Tables

Table 1. Assembly statistics of the small hive beetle genome assembly

Illumina (genome coverage)	535
PacBio (genome coverage)	50
Assembly size (Mbp)	234.3
Number of contigs	3063
Largest contig (Kbp)	2683.7
Smallest contig (Kbp)	1.26
N50 (Kbp)	298.8
Number of contig > 10 Kbp	2236
Number of contig > N50	192
Number of protein coding genes	14076
Number of mRNAs	17463
Density of SNPs (bps per SNP position)	177
Density of microsatellites (loci per Kbp)	8.23

Table 2. The number of gustatory receptor (GR) genes from major groups for three coleopteran species; the small hive beetle, ATUMI, AGLAB, and TCAST and their putative coding for detecting either bitter or sweet tastants.

Species	**Gustatory receptor group**						**Tastant type**	
	2a	5a	28a/b	43a	64a-f	Total	Bitter	Sweet
AGLAB	11	1	7	1	6	26	19	7
ATUMI	3	0	2	2	4	11	5	6
TCAST	12	3	30	12	14	71	42	29

Table 3. Evaluation of resistance mutations in acetylcholine esterase and their status in ATUMI.

Torpedo Ace Position number, the resistance mutations are described in [43].

Torpedo Ace Position	ATUMI_Ace1 Position	Resistance Mutations	ATUMI_Ace1 State
119	189	G247S, G119D	G
128	198	D237E	D
201	270	A302S	A
227	296	G265A, G262A	G
290	358	F290V	F
331	399	S431F, F445W, F439C	F
Torpedo Ace Position	**ATUMI_Ace2 Position**	**Resistance Mutations**	**ATUMI_Ace2 State**
78	114	F139L, F115S	F
82	118	E81K	E
129	177	I161V/T	I
151	198	V180L	I

227	280	G265A, G262A/V	G
238	290	S291G	T
290	358	F330Y, F237Y	F
328	383	G365A, G368A	G
396	452	G488S	G

Table 4: Numbers of ABC genes in each species, by subfamily. * lower counts discussed in text.

Species	**Subfamily**								**Total**
	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	
ATUMI	4*	6	24	2	1	3	13*	3	56
TCAST	10	6	35	2	1	3	14	3	74
DMELA	10	8	14	2	1	3	15	3	56

Table 5. Comparison of the number of GSTs between ATUMI, AMELL, DMELA, and TCAST [53], [67].

GST Class	ATUMI	AMELL	DMELA	TCAST

Delta	3	1	11	3
Epsilon	19	0	14	19
Omega	1	1	5	3
Sigma	7	4	1	7
Theta	1	1	4	1
Zeta	5	1	2	1
Microsomal	6	2	1	5
Unclassified	7	0	0	2
Total	49	10	38	41

Table 6. Comparison of CYP450 genes in ATUMI, AMELL, DMELA, and TCAST.

P450 Clan	ATUMI	AMELL	DMELA	TCAST
CYP2	8	8	6	8
CYP3	55	28	36	82
CYP4	43	4	32	49
Mitochondrial	10	6	11	10

Total	116	46	85	149

Table 7. Comparison of COE from *Aethina tumida* to *Drosophila melanogaster*, *Tribolium castaneum*, and *Apis mellifera*. Nomenclature and gene counts follow McKenna 2016 [12] and Claudianos 2006. [67].

COE Subfamily	*ATUMI*	*AMELL*	*DMELA*	*TCAST*
Clades A-C (Dietary)	27	8	13	55
Clade D (Integument Esterases)	2	1	3	5
Clade E (Secreted β-esterase)	8	3	3	10
Clade F (JH Esterases)	3	1	2	1
Clade H (Glutactins)	2	0	4	2
Clade I (Unknown Function)	1	2	2	2
Clade J (Acetylcholinesterases)	2	2	1	2
Clade K (Gliotactin)	1	1	1	2
Clade L (Neuroligins)	10	5	4	5
Clade M (Neurotactins)	4	1	2	1
Total	60	24	35	85

Figure Legends

Figure 1. *Aethina tumida* (A) adult and (B) larva. Photos courtesy of Alex Wild Photography, used with permission.

Figure 2. Phylogenetic tree and estimated completeness of the genomes of 11 insect species. A) The phylogenetic tree was constructed on protein sequences of 181 single copy orthologs shared among all 11 insect species. All nodes have 100% bootstrap support. AMELL and DMELA were used as outgroups. Branch lengths are shown for each node. B) Completeness of official protein sets of each insect species were assessed by aligning to the Endopterygota sets of benchmarking universal single-copy orthologs (BUSCOs). For ATUMI, 97.5% of complete BUSCOs were found. C) The pervasiveness of gene loss during endopterygote evolution. From the domain counts of lost BUSCOs, methyltransferase (MT), glycosyltransferase (GT) and leucine rich repeats (LRR) are among the top 5% of total domains and are commonly lost from multiple species. WD and ZF red boxes indicate that the gene is lost, while white boxes indicate that the gene is maintained in each species.

Figure 3. Gene duplication events plotted against the average gene duplication event per gene. The protein sets of the 11 studied beetle species, as well as honey bee and fruit fly were searched against the Endopterygota BUSCO set using BLAST. Redundant proteins (including recent paralogs and those with known alternative splicing) were used to quantify the average number of duplication events per gene in each species.

Figure 4. Glycoside hydrolase (GH) family copy numbers identified from beetle genomes. Genes coding for glycoside hydrolases were identified using Pfam domain assignments [95] and genome assemblies and coding gene predictions were obtained from NCBI (GenBank Accession Numbers: GCA_000390285.1 AGLAB, GCA_000355655.1 DPOND, GCA_001412225.1 *N. vespilloides*, GCA_001443705.1 *O. borbonicus*, GCA_000002335.3 TCAST) with the exception of *H.s hampei*, which was downloaded from https://genome.med.nyu.edu/coffee-beetle/cbb.html. Families are color coded from green to red based on their relative abundance (total count/total number of GH genes) with red representing GH families that are highly abundant (≥25% of the total GH genes) and green representing GH families of lesser abundance (≤0.01%). Notably, the GH profiles of ATUMI and TCAST (neither of which feed on living plant material) differ strongly from the GH profiles of the phytophagous beetles, even though they all belong to the same infraorder, suggesting that diet, in part, might be driving the differences in GH family members and copy numbers. AGLAB=Asian longhorned beetle (*A. glabripennis*); HHAMP=Coffee berry borer (*H. hampei*); DPOND=Mountain pine beetle (*D. ponderosae*); NVESP=burying beetle (*N. vespilloides*); OBORB=scarab beetle (*O. borbonicus*); ATUMI=small hive beetle, and TCAST=red flour beetle (TCAST).

Figure 5. Maximum likelihood cladogram for gustatory receptor genes from three coleopteran species. The small hive beetle, *Aethina tumida* (*ATUMI*; green labels/lines), the Asian long horned beetle, *Anoplophora glabripennis* (*Agla*; red labels/lines), and the red flour beetle, *Tribolium castaneum* (*TCAST*; blue labels/lines). Individual genes are labeled with species identifier and GenBank accession number. Scale bar for branch lengths represents 0.1 amino

acid substitutions per site. Ring around cladogram indicates gene families coded for perceiving bitter (yellow) and sweet (pink) tastants.

Figure 6. Maximum likelihood phylogenetic tree of ABC proteins from ATUMI (At), TCAST (Tc), and DMELA (Dm). ATUMI genes are marked in blue, TCAST in green, and DMELA in purple. ABC subfamilies are indicated with colored lines to the right of the tree. Names for DMELA proteins were taken from Flybase (http://flybase.org/reports/FBgg0000552), and include the Flybase number for reference. TCAST names are taken from the two papers in which the genes were identified [47, 48], with the NCBI Refseq accession number provided for reference. ATUMI names were generated for this paper by combining the subfamily of the identified sequence with the scaffold on which the encoding gene may be found; if multiple ABC genes of a particular subfamily were found on the same scaffold, the sequences were given an additional letter designation based on their relative location, reading left to right on the scaffold as shown in WebApollo. For reference, the scaffold number and base coordinates for the gene have also been included.

Figure 7. Maximum Likelihood phylogenetic tree of glutathione-S-transferase (GST) proteins. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa *A. tumida* (ATUMI) in green; *A. mellifera* (AMELL) in black; *D. melanogaster* (DMELA) in blue; and *T. castaneum* (TCAST) in red identified manually using the Uniprot and Pfam databases. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. Initial tree(s) for the heuristic search were obtained

automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. All positions with less than 95% site coverage were eliminated. The tree was annotated and visualized with the iToL web tool (itol.embl.de/) [96]..

Figure 8. Maximum Likelihood phylogenetic tree of the cytochrome P450 detoxification system. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa *A. tumida* (ATUMI) in green; *A. mellifera* (AMELL) in black; *D. melanogaster* (DMELA) in blue; and *T. castaneum* (TCAST) in red identified manually using the Uniprot and Pfam databases. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. All positions with less than 95% site coverage were eliminated. P450s are clustered to CYP2, CYP3, CYP4 and mitochondrial clans. The tree was annotated and visualized with the iToL web tool (itol.embl.de/) [96].

Figure 9. Maximum likelihood phylogenetic tree of carboxylesterase (COE) genes. The maximum likelihood bootstrap consensus tree (1000 replicates) showing the relationships among COE genes from the genomes of *A. tumida* (ATUMI) in green; *A. mellifera* (AMELL) in black; *D. melanogaster* (DMELA) in blue; and *T. castaneum* (TCAST) in red, identified manually using the Uniprot and Pfam databases. Branches corresponding to partitions recovered in less

than 50% of bootstrap replicates are collapsed. Starting tree(s) for the heuristic search were obtained automatically using neighbor-joining and BioNJ algorithms applied to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with the superior log likelihood value. All positions with less than 95% site coverage were eliminated. The phylogenetically distinct clusters were named according to established nomenclature for COE genes [12]. The tree was annotated and visualized with the iToL web tool [96].

References

1. Lundie AE. The small hive beetle *Aethina tumida*, Science Bulletin 220. In: Forestry DoAa, (ed.). Pretoria, South Africa1940.
2. Neumann P, Pettis JS and Schäfer MO. Quo vadis *Aethina tumida*? Biology and control of small hive beetles. Apidologie. 2016;47 3:427-66. doi:10.1007/s13592-016-0426-x.
3. Al Toufailia H, Alves DA, Benà DDC, Bento JMS, Iwanicki NSA, Cline AR, et al. First record of small hive beetle, *Aethina tumida* Murray, in South America. Journal of Apicultural Research. 2017;56 1:76-80. doi:10.1080/00218839.2017.1284476.
4. Lee S, Hong KJ, Cho YS, Choi YS, Yoo MS and Lee S. Review of the subgenus *Aethina* Erichson s. str. (Coleoptera: Nitidulidae: Nitidulinae) in Korea, reporting recent invasion of small hive beetle, *Aethina tumida*. Journal of Asia-Pacific Entomology. 2017;20 2:553-8. doi:10.1016/j.aspen.2017.03.006.
5. Buchholz S, Schäfer MO, Spiewok S, Pettis JS, Duncan M, Ritter W, et al. Alternative food sources of *Aethina tumida* (Coleoptera: Nitidulidae). Journal of Apicultural Research. 2008;47 3:202-9. doi:10.3827/IBRA.1.47.3.08.

6. Neumann P, Pirk CWW, Hepburn HR, Solbrig AJ, Ratnieks FLW, Elzen PJ, et al. Social encapsulation of beetle parasites by Cape honeybee colonies (*Apis mellifera capensis* Esch.). Naturwissenschaften. 2001;88 5:214-6. doi:10.1007/s001140100224.

7. Haddad S and Mckenna DD. Phylogeny and evolution of the superfamily Chrysomeloidea (Coleoptera: Cucujiformia). Systematic Entomology. 2016;41 4:697-716.

8. McKenna DD, A.L. Wild, K. Kanda, C.L. Bellamy, R.G. Beutel, M.S. Caterino, C.W. Farnum, D.C. Hawks, M.A. Ivie, M.L. Jameson, R.A.B. Leschen, A.E. Marvaldi, J.V. McHugh, A.F. Newton, J.A. Robertson, M.K. Thayer, M.F. Whiting, J.F. Lawrence, A. Ślipiński, D.R. Maddison, B.D. Farrell. The beetle tree of life reveals Coleoptera survived end Permian mass extinction to diversify during the Cretaceous terrestrial revolution. Systematic Entomology. 2015:1-46. doi:10.1111/syen.12132.

9. Robertson JA, A. Ślipiński, M. Moulton, F.W. Shockley, A. Giorgi, N.P. Lord, D.D. McKenna, W. Tomaszewska, J. Forrester, K.B. Miller, M.F. Whiting, J.V. McHugh. Phylogeny and classification of the beetle superfamily Cucujoidea and the recognition of a new superfamily Coccinelloidea (Coleoptera: Cucujiformia). 2015;Systematic Entomology doi:10.1111/syen.12138.

10. McKenna DD. Beetle genomes in the 21st century: prospects, progress and priorities. . Current Opinion in Insect Science 2018;25 76-82.

11. Richards S, Gibbs RA, Weinstock GM, Brown SJ, Denell R, Beeman RW, et al. The genome of the model beetle and pest *Tribolium castaneum*. Nature. 2008;452 7190:949-55. doi:10.1038/nature06784.

12. McKenna DD, Scully ED, Pauchet Y, Hoover K, Kirsch R, Geib SM, et al. Genome of the Asian longhorned beetle (*Anoplophora glabripennis*), a globally significant invasive species, reveals key functional and evolutionary innovations at the beetle-plant interface. Genome Biology. 2016;17 227 doi:10.1186/s13059-016-1088-8.

13. Keeling Clea. Draft genome of the mountain pine beetle, *Dendroctonus ponderosae* Hopkins, a major forest pest. 2013 14:R27.

14. Vega. Draft genome of the most devastating insect pest of coffee worldwide: the coffee berry borer, *Hypothenemus hampei*. Scientific Reports. 2015;5:12525.

15. Meyer JM, Markov GV, Baskaran P, Herrmann M, Sommer RJ and Rödelsperger C. Draft Genome of the Scarab Beetle *Oryctes borbonicus* on La Réunion Island. Genome Biology and Evolution. 2016;8 7:2093-15.

16. Cunningham CB, Ji L, Axel R, Wiberg W, Shelton J, McKinney EC, et al. The Genome and Methylome of a Beetle with Complex Social Behavior, *Nicrophorus vespilloides* (Coleoptera: Silphidae) Genome Biol Evol 2015;7 12:3383-96. doi:10.1093/gbe/evv194.

17. Schoville SD, Chen YH, Andersson MN, Benoit JB, Bhandari A, Bowsher JH, et al. A model species for agricultural pest genomics: the genome of the Colorado potato beetle, *Leptinotarsa decemlineata* (Coleoptera: Chrysomelidae). bioRxiv. 2017; doi:http://dx.doi.org/10.1101/192641.

18. Emms DM and Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biology. 2015;16:157.

19. Pauchet Y, Wilkinson P and Chauhan R. Diversity of the beetle genes encoding novel plant cell wall degrading enzymes. PloS One. 2010;5:e15635.

20. Merzendorfer H and Zimoch L. Chitin metabolism in insects: structure, function and regulation of chitin synthases and chitinases. Journal of Experimental Biology. 2003;206:4393-412.

21. Kramer KJ and Muthukrishnan S. Insect chitinases: molecular biology and potential use as biopesticides. Insect Biochemistry and Molecular Biology. 1997;27:887-900.

22. Fujita Kea. A chitinase structurally related to the glycoside hydrolase family 48 is indispensable for hormonally induced diapause termination in a beetle. Biochemical and Biophysical Research Communications. 2006;345:502-7.

23. Pacini E. Types and meaning of pollen carbohydrate reserves. Sexual Plant Reproduction. 1996;362.

24. Engel P, Martinson VG and Moran NA. Functional diversity within the simple gut microbiota of the honey bee. Proceedings of the National Academy of Sciences of the United States of America. 2012;109 27:11002-7.

25. Guindon S and Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 2003;52 5:696-704.

26. Clyne PJ, Warr CG and Carlson JR. Candidate Taste Receptors in <em>Drosophila</em&gt. Science. 2000;287 5459:1830.

27. Thorne N, Chromey C Fau - Bray S, Bray S Fau - Amrein H and Amrein H. Taste perception and coding in Drosophila. Current Biology. 2004;14 12:1065-79.

28. Wang Z, Singhvi A Fau - Kong P, Kong P Fau - Scott K and Scott K. Taste representations in the Drosophila brain. Cell. 2004;117 7:981-91.

29. Isono K and Morita H. Molecular and cellular designs of insect taste receptor system. Frontiers in Cellular Neuroscience. 2010;4:20.

30. Robertson H and Wanner K. The chemoreceptor superfamily in the honey bee, *Apis mellifera*: expansion of the odorant, but no gustatory receptor family. Genome Research. 2006;16:1395-403.

31. Amrein H. An expression system for Gustatory receptors—and why it failed. Fly. 2014;8 4:232-3. doi:10.1080/19336934.2015.1039756.

32. Freeman EG, Wisotsky Z and Dahanukar A. Detection of sweet tastants by a conserved group of insect gustatory receptors. Proceedings of the National Academy of Sciences. 2014;111 4:1598-603. doi:10.1073/pnas.1311724111.

33. Honey Bee Genome Sequencing Consortium. Insights into social insects from the genome of the honeybee *Apis mellifera*. Nature 2006;443:931-49. doi:10.1038/nature05260.

34. Wanner KW and Robertson HM. The gustatory receptor family in the silkworm moth *Bombyx mori* is characterized by a large expansion of a single lineage of putative bitter receptors. Insect Molecular Biology. 2008;17 6:621-9. doi:10.1111/j.1365-2583.2008.00836.x.

35. Du Y, Nomura Y, Satar G, Hu Z, Nauen R, He SY, et al. Molecular evidence for dual pyrethroid-receptor sites on a mosquito sodium channel. PNAS. 2013;110 29:11785-90. doi:10.1073/pnas.1305118110.

36. O'Reilly AO, Khambay BPS, Williamson MS, Field LA, Wallace BA and Davies TGE. Modelling insecticide-binding sites in the voltage-gated sodium channel. Biochem J. 2006;396:255-63.

37. Rinkevich FD, Du Y and Dong K. Diversity and convergence of sodium channel mutations involved in resistance to pyrethroids Pestic Biochem Physiol. 2013;106:93-100.

38. Lee SH, Ingles PJ, Knipple DC and Soderlund DM. Developmental regulation of alternative exon usage in the house fly *Vssc1* sodium channel gene. Invert Neurosci. 2002;4:125-33.

39. Shao YM, Dong K, Tang ZH and Zhang CX. Molecuar characterization of a sodium channel gene from the silkworm *Bombyx mori*. Insect Biochem Molec Biol. 2009;39:145-51.

40. Davies T, Field L, Usherwood P and Williamson M. A comparative study of voltage-gated sodium channels in the Insecta: implications of pyrethroid resistance in Anopheline and other Neopteran species. Insect Molecular Biology. 2007;16 3:361-75.

41. Tan J. Alternative splicing of an insect sodium channel gene generates pharmacologically distinct sodium channels. J Neurosci. 2002;22:5300.

42. Fournier D and Mutero A. Modification of acetylcholinesterase as a mechanism of resistance to insecticides. Comparative Biochemistry and Physiology \108C. 1994:19-31.

43. Fournier D. Mutations of acetylcholinesterase which confer insecticide resistance in insect populations. Chemico-Biol Inteact. 2005;157:257-61.

44. Baek JH, Kim JI, Lee D-W, Chung BK, Miyata T and Lee SH. Identification and characterization of ace1-type acetylcholinesterase likely associated with organophosphate resistance in *Plutella xylostella*. Pesticide Biochemistry and Physiology. 2005;81 3:164-75.

45. Kim. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome biology. 2013.

46. Dermauw W and Van Leeuwen T. The ABC gene family in arthropods: comparative genomics and role in insecticide transport and resistance. Insect Biochemistry and Molecular Biology. 2014;45:89-110.

47. Broehan G, Kroeger T, Lorenzen M and Merzendorfer H. Functional analysis of the ATP-binding cassette (ABC) transporter gene family of Tribolium castaneum. BMC Genomics. 2013;14 6.

48. Grubbs N, Haas S, Beeman RW and Lorenzen MD. The ABCs of eye color in *Tribolium castaneum*: orthologs of the Drosophila white, scarlet, and brown Genes. Genetics. 2015;199 3:749-59.

49. Simon JY. Insect glutathione S-transferases. Zoological Studies. 1996;35 1:9-19.

50. Che-Mendoza A, Penilla RP and Rodríguez DA. Insecticide resistance and glutathione S-transferases in mosquitoes: a review. African Journal of Biotechnology,. 2009;8 8.

51. Corona M and Robinson GE. Genes of the antioxidant system of the honey bee: annotation and phylogeny. Insect Molecular Biology. 2006;15 5:687-701.

52. Enayati AA, Ranson H and Hemingway J. Insect glutathione transferases and insecticide resistance. Insect Molecular Biology. 14 1:3-8.

53. Shi H, Pei L, Gu S, Zhu S, Wang Y and al. e. Glutathione S-transferase (GST) genes in the red flour beetle, *Tribolium castaneum*, and comparative analysis with five additional insects. Genomics. 2012;100 5:327-35.

54. Daborn PJ, Yen JL, Bogwitz MR, Le Goff G, Feil E and al. e. A Single P450 Allele Associated with Insecticide Resistance in *Drosophila*. Science. 2002;297 5590:2253-6.

55. Liu N and Scott JG. Increased transcription of CYP6D1 causes cytochrome P450-mediated insecticide resistance in house fly. Insect Biochem Molec Biol. 1998;28:531-5.

56. Hardstone MC, Komagata O, Kasai S, Tomita T and Scott JG. Use of isogenic strains indicates CYP9M10 is linked to permethrin resistance in Culex pipiens quinquefasciatus. Insect Molec Biol. 2010;19:717-26.

57. Nelson DR, 2018. Cytochrome P450 diversity in the tree of life. Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 1866(1), pp.141-154. Cytochrome P450 diversity in the tree of life. iochimica et Biophysica Acta (BBA)-Proteins and Proteomics. 2018;1866 1:141-54.

58. Gilbert LI. Halloween genes encode P450 enzymes that mediate steroid hormone biosynthesis in Drosophila melanogaster. Mol Cell Endocrinol. 2004;215:1-10.

59. Feyereisen R. Arthropod CYPomes illustrate the tempo and mode in P450 evolution. Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics. 2011;1841 1:19-28.

60. Daborn P, Boundy S, Yen J and Pittendrigh B. DDT resistance in *Drosophila* correlates with Cyp6g1 over-expression and confers cross-resistance to the neonicotinoid imidacloprid. Molecular Genetics and Genomics. 2001;266 4:556-63.

61. Scott JG and Wen Z. Cytochromes P450 of insects: the tip of the iceberg. Pest management science. 2001;57 10:958-67.

62. Pridgeon JW, Zhang L and Liu N. Overexpression of CYP4G19 associated with a pyrethroid-resistant strain of the German cockroach, *Blattella germanica* (L.). Genetics. 2003;314:157-63.

63. Oakeshott JG, Claudianos C, Campbell PM, Newcomb RD and Russell RJ. Biochemical genetics and genomics of insect esterases. In: Gilbert LI, Iatrou K and Gill SS, editors. Comprehensive Molecular Insect Science. Boston: Elsevier; 2005. p. 309-81.

64. Montella IR, Schama R and Valle D. The classification of esterases: an important gene family involved in insecticide resistance-A review. Memorias do Instituto Oswaldo Cruz 2012:437-49.

65. Bass C and Field LM. Gene amplification and insecticide resistance. Pest Management Science. 2011;67:886-90.

66. Newcomb RD, Campbell PM, Ollis DL, Cheah E, Russell RJ and Oakeshott JG. A single amino acid substitution converts a carboxylesterase to an organophosphorus hydrolase and confers insecticide resistance on a blowfly. Proceedings of the National Academy of Sciences 1997;94:7464-8.

67. Claudianos C, Ranson H, Johnson RM, Biswas MA and Schuler ea. A deficit of detoxification enzymes: pesticide sensitivity and environmental response in the honeybee. Insect Molecular Biology. 2006;15 5:615-36.

68. Yu QY, Lu C, Li WL, Xiang ZH and Zhang Z. Annotation and expression of carboxylesterases in the silkworm, *Bombyx mori*. BMC genomics. 2009;10 1.

69. Ye C, Ma ZS, Cannon CH, Pop M and Yu DW. Exploiting sparseness in de novo genome assembly. BMC Bioinformatic. 2012; doi:10.1186/1471-2105-13-S6-S1.

70. Hackl TH, R.; Schultz, J.; Foerster, F. . proovread: large-scale high accuracy PacBio correction through iterative short read consensus. . 2014.

71. Ye C MZ. Sparc: a sparsity-based consensus algorithm for long erroneous sequencing reads. 2016; doi:10.7717.

72. Thibaud-Nissen F, Souvorov A, Murphy T, DiCuccio M and Kitts P. Eukaryotic Genome Annotation Pipeline. The NCBI Handbook, 2nd Edition. Bethesda, MD: National Center for Biotechnology Information (US); 2013. https://www.ncbi.nlm.nih.gov/books/NBK169439/.

73. Tarver MR, Huang Q, de Guzman L, Rinderer T, Holloway B, Reese J, et al. Transcriptomic and functional resources for the small hive beetle *Aethina tumida*, a worldwide parasite of honey bees. Genomics Data. 2016;9:97-9. doi:10.1016/j.gdata.2016.06.003.

74. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV and Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. . Bioinformatics 2015;31 19:3210-2. doi:10.1093/bioinformatics/btv351.

75. Elsik CG. Finding the missing honey bee genes: lessons learned from a genome upgrade. BMC Genomics. 2014;15 86 doi:doi.org/10.1186/1471-2164-15-86.

76. Adams. The genome sequence of *Drosophila melanogaster*. Science. 2000.

77. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32 5:1792-7. doi:10.1093/nar/gkh340.

78. Capella-Gutierrez S, Silla-Martinez JM and Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25:1972-3.

79. Stamatakis A. RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. Bioinformatics. 2014.

80. Stöver BC and Müller KF. TreeGraph 2: Combining and visualizing evidence from different phylogenetic analyses. BMC bioinformatics. 2010;11 1.

81. Du L, Li D, Zhang X and Yue B. MSDB: a user-friendly program for reporting distribution and building databases of microsatellites from genome sequences. . Journal of Heredity. 2013;104 1:154-7.

82. Li H and Durbin R. Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics. 2009;25:1754-60.

83. McKenna A HM, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data Genome Research. 2010;20:1297-303.

84. Cingolani P PA, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. 2012;6 2:80-92.

85. Abdel-latief M. A Family of Chemoreceptors in *Tribolium castaneum* (Tenebrionidae: Coleoptera). PLOS ONE. 2007;2 12:e1319. doi:10.1371/journal.pone.0001319.

86. McKenna DD, Scully ED, Pauchet Y, Hoover K, Kirsch R, Geib SM, et al. Genome of the Asian longhorned beetle (*Anoplophora glabripennis*), a globally significant invasive species, reveals key functional and evolutionary innovations at the beetle–plant interface. Genome Biology. 2016;17 1:227. doi:10.1186/s13059-016-1088-8.

87. Consortium TU. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 2017;45 D1:D158-D69. doi:10.1093/nar/gkw1099.

88. Kim HS, Murphy T, Xia J, Caragea D, Park Y, Beeman RW, et al. BeetleBase in 2010: revisions to provide comprehensive genomic information for *Tribolium castaneum*. Nucleic Acids Res. 2010;38:D437-D42. doi:10.1093/nar/gkp807.

89. Anisimova M and Gascuel O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Syst Biol. 2006;55 4:539-52. doi:10.1080/10635150600755453.

90. Chevenet F, Brun C, Banuls AL, Jacq B and Christen R. TreeDyn: towards dynamic graphics and annotations for analyses of trees. BMC Bioinformatics. 2006;7:439. doi:10.1186/1471-2105-7-439.

91. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, et al. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. 2008;36 Web Server issue:W465-9. doi:10.1093/nar/gkn180.

92. Letunic I and Bork P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics. 2007;23 1:127-8. doi:10.1093/bioinformatics/btl529.

93. Kumar S, Stecher G and Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular biology and evolution. 2016:msw054.

94. Hall BG. Building Phylogenetic Trees from Molecular Data with MEGA. Molecular Biology and Evolution. 2013;30 5:1229-35.

95. Bateman A. The Pfam protein families database. Nucleic Acids Res. 2002;30:276-80.

96. Letunic I and Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 2016;44 W1:W242-W5.

Figure 1

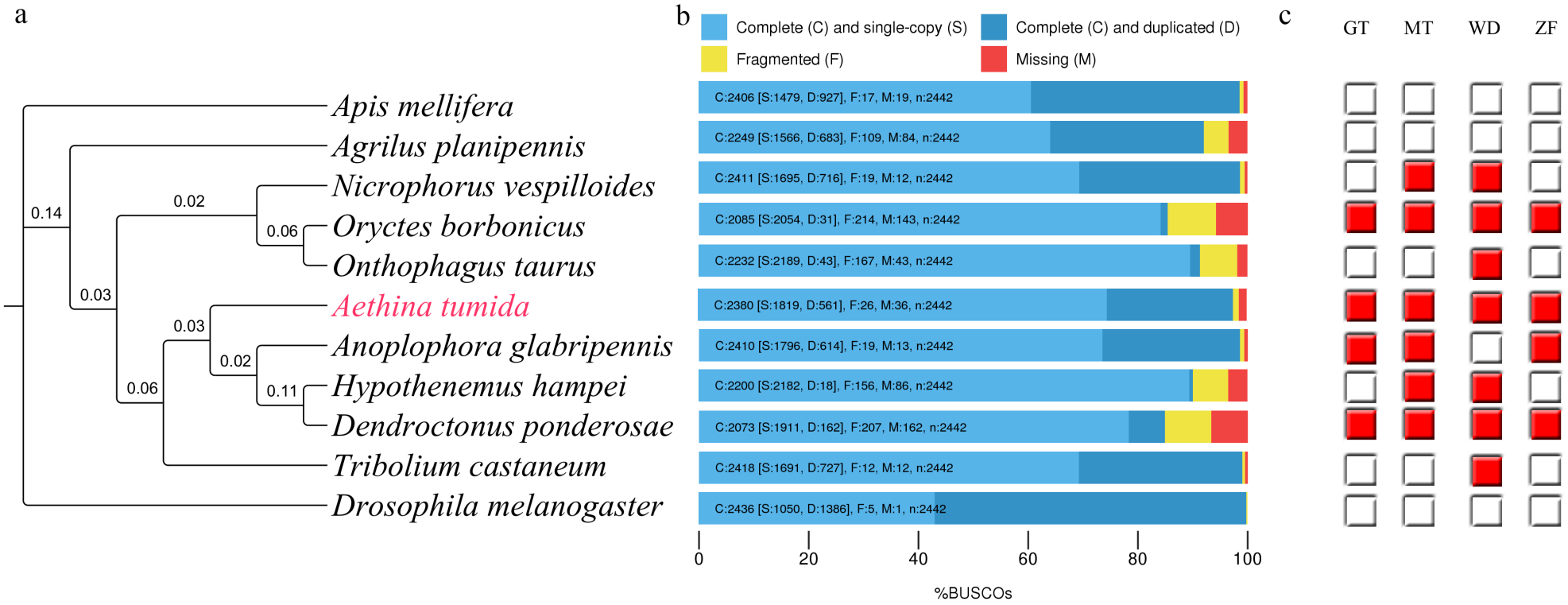

a
Apis mellifera
Agrilus planipennis
Nicrophorus vespilloides
Oryctes borbonicus
Onthophagus taurus
Aethina tumida
Anoplophora glabripennis
Hypothenemus hampei
Dendroctonus ponderosae
Tribolium castaneum
Drosophila melanogaster
0.14
0.02
0.06
0.03
0.03
0.02
0.06
0.11
b
Complete (C) and single-copy (S)
Complete (C) and duplicated (D)
Fragmented (F)
Missing (M)
C:2406 [S:1479, D:927], F:17, M:19, n:2442
C:2249 [S:1566, D:683], F:109, M:84, n:2442
C:2411 [S:1695, D:716], F:19, M:12, n:2442
C:2085 [S:2054, D:31], F:214, M:143, n:2442
C:2232 [S:2189, D:43], F:167, M:43, n:2442
C:2380 [S:1819, D:561], F:26, M:36, n:2442
C:2410 [S:1796, D:614], F:19, M:13, n:2442
C:2200 [S:2182, D:18], F:156, M:86, n:2442
C:2073 [S:1911, D:162], F:207, M:162, n:2442
C:2418 [S:1691, D:727], F:12, M:12, n:2442
C:2436 [S:1050, D:1386], F:5, M:1, n:2442
0
20
40
60
80
100
%BUSCOs
c
GT
MT
WD
ZF

Figure 2

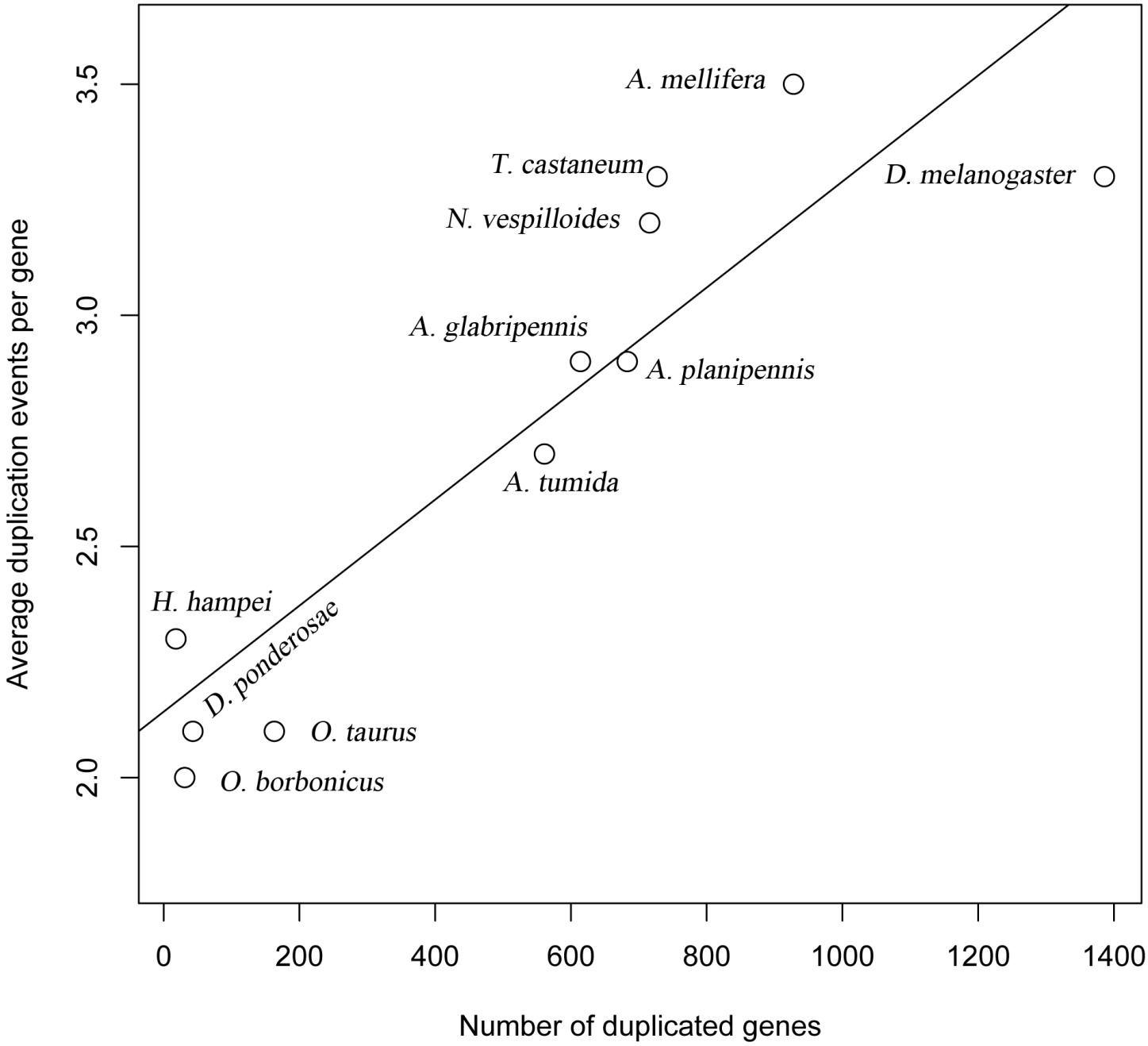
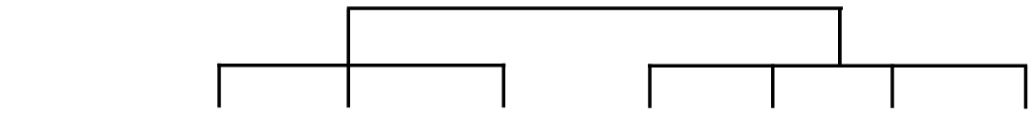

Average duplication events per gene
2.0
2.5
3.0
3.5
0
200
400
600
800
1000
1200
1400
Number of duplicated genes
A. mellifera
T. castaneum
D. melanogaster
N. vespilloides
A. glabripennis
A. planipennis
A. tumida
H. hampei
D. ponderosae
O. taurus
O. borbonicus

Figure 3



Family	ALB	CBB	DPOND	NIC	ORY	SHB	TCAS
GH1	59	21	25	8	2	8	13
GH2	5	2	3	8	1	5	6
GH5	6	2	2	0	0	0	0
GH9	1	0	0	1	1	0	1
GH10	0	2	0	0	0	0	0
GH13	1	0	0	0	0	0	0
GH15	2	2	2	4	1	3	2
GH16	4	5	11	1	3	7	3
GH18	25	19	11	18	13	18	22
GH20	8	8	10	14	8	8	10
GH22	7	0	0	0	0	0	0
GH28	17	7	23	0	0	0	0
GH30	7	1	2	3	1	8	5
GH31	17	6	10	6	7	11	8
GH32	2	0	2	0	0	0	0
GH35	10	5	10	3	0	4	4
GH37	7	0	0	0	0	0	0
GH38	10	7	7	7	5	10	11
GH39	1	0	0	1	0	1	1
GH45	2	3	9	0	0	0	0
GH47	4	4	4	5	3	6	4
GH48	1	4	8	0	0	0	0
GH63	1	1	1	1	1	0	1
GH67	0	0	0	0	0	0	1
GH79	1	1	1	2	1	1	1
GH85	1	1	1	1	0	1	1
GH99	0	0	0	1	0	0	1
Total	**199**	**101**	**142**	**84**	**47**	**91**	**95**
	Phytophagous			Other			

Figure 4

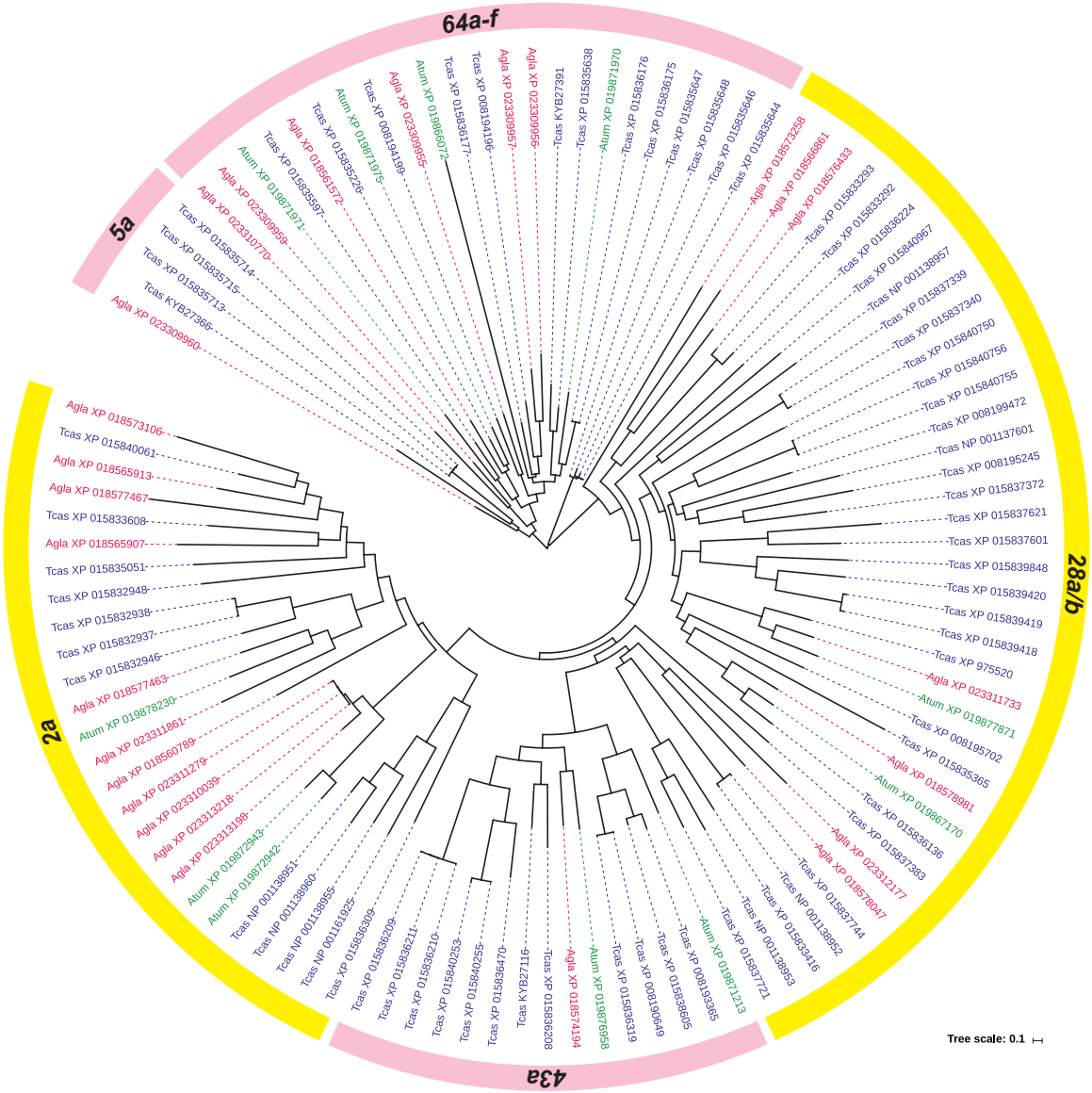

64a-f
28a/b
43a
2a
5a
Tcas XP 015836175
Tcas XP 015835647
Tcas XP 015835648
Tcas XP 015835644
Agla XP 018573258
Agla XP 018566861
Agla XP 018576433
Tcas XP 015833293
Tcas XP 015833292
Tcas XP 015836224
Tcas XP 015840967
Tcas NP 001138957
Tcas XP 015837339
Tcas XP 015837340
Tcas XP 015840750
Tcas XP 015840756
Tcas XP 015840755
Tcas XP 008199472
Tcas NP 001137601
Tcas XP 008195245
Tcas XP 015837372
Tcas XP 015837621
Tcas XP 015837601
Tcas XP 015839848
Tcas XP 015839420
Tcas XP 015839419
Tcas XP 015839418
Tcas XP 975520
Agla XP 023311733
Atum XP 019877871
Tcas XP 008195702
Tcas XP 015835365
Agla XP 018578981
Atum XP 019867170
Tcas XP 015836136
Agla XP 023312177
Agla XP 018578047
Tcas XP 015837744
Tcas NP 001138952
Tcas XP 015833416
Tcas NP 001138953
Tcas XP 015837721
Atum XP 019871213
Tcas XP 008193365
Tcas XP 015838605
Tcas XP 008190649
Tcas XP 015836319
Atum XP 019876958
Agla XP 018574194
Tcas XP 015836208
Tcas KYB27116
Tcas XP 015840255
Tcas XP 015836470
Tcas XP 015840253
Tcas XP 015836210
Tcas XP 015836211
Tcas XP 015836209
Tcas XP 015836309
Tcas NP 001161925
Tcas NP 001138955
Tcas NP 001138960
Tcas NP 001138951
Atum XP 019872942
Atum XP 019872943
Agla XP 023313198
Agla XP 023313218
Agla XP 023310039
Agla XP 023311279
Agla XP 018560789
Agla XP 023311861
Atum XP 019878230
Agla XP 018577463
Tcas XP 015832946
Tcas XP 015832937
Tcas XP 015832938
Tcas XP 015832948
Tcas XP 015835051
Agla XP 018565907
Tcas XP 015833608
Agla XP 018577467
Agla XP 018565913
Tcas XP 015840061
Agla XP 018573106
Agla XP 023309960
Tcas KYB27366
Tcas XP 015835713
Tcas XP 015835715
Tcas XP 015835714
Agla XP 023310770
Agla XP 023309959
Tcas XP 015835597
Agla XP 018561572
Atum XP 019871971
Tcas XP 015835226
Atum XP 019871975
Tcas XP 008194199
Agla XP 023309955
Atum XP 019866072
Tcas XP 015836177
Tcas XP 008194196
Agla XP 023309957
Agla XP 023309956
Tcas KYB27391
Tcas XP 015835638
Atum XP 019871970
Tcas XP 015836176
Tree scale: 0.1

Figure 5

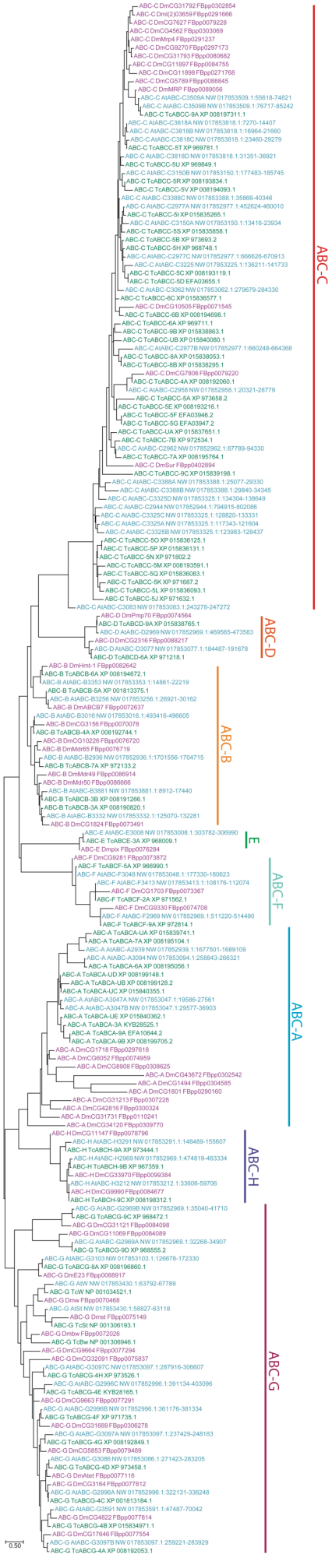

ABC-C DmCG31792 FBpp0302854
ABC-C Dml(2)03659 FBpp0291666
ABC-C DmCG7627 FBpp0079228
ABC-C DmCG4562 FBpp0303069
ABC-C DmMrp4 FBpp0291237
ABC-C DmCG9270 FBpp0297173
ABC-C DmCG31793 FBpp0080662
ABC-C DmCG11897 FBpp0084755
ABC-C DmCG11898 FBpp0271768
ABC-C DmCG5789 FBpp0088845
ABC-C DmMRP FBpp0089056
ABC-C AtABC-C3509A NW 017853509.1:55618-74821
ABC-C AtABC-C3509B NW 017853509.1:76717-85242
ABC-C TcABCC-8A XP 008197311.1
ABC-C AtABC-C3818A NW 017853818.1:7270-14407
ABC-C AtABC-C3818B NW 017853818.1:16964-21680
ABC-C AtABC-C3818C NW 017853818.1:23460-29279
ABC-C TcABCC-5T XP 969781.1
ABC-C AtABC-C3818D NW 017853818.1:31351-36921
ABC-C TcABCC-5U XP 969849.1
ABC-C AtABC-C3150B NW 017853150.1:177483-185745
ABC-C TcABCC-5R XP 008193834.1
ABC-C TcABCC-5V XP 008194093.1
ABC-C AtABC-C3388C NW 017853388.1:35866-40346
ABC-C AtABC-C2977A NW 017852977.1:452624-460010
ABC-C TcABCC-5I XP 015835265.1
ABC-C AtABC-C3150A NW 017853150.1:13416-23934
ABC-C TcABCC-5S XP 015835858.1
ABC-C TcABCC-5B XP 973693.2
ABC-C TcABCC-5H XP 968748.1
ABC-C AtABC-C2977C NW 017852977.1:666626-670913
ABC-C AtABC-C3225 NW 017853225.1:136211-141733
ABC-C TcABCC-5C XP 008193119.1
ABC-C TcABCC-5D EFA03655.1
ABC-C AtABC-C3062 NW 017853062.1:279679-284330
ABC-C TcABCC-6C XP 015836577.1
ABC-C DmCG10505 FBpp0071545
ABC-C TcABCC-6B XP 008194698.1
ABC-C TcABCC-6A XP 969711.1
ABC-C TcABCC-9B XP 015838863.1
ABC-C TcABCC-UB XP 015840080.1
ABC-C AtABC-C2977B NW 017852977.1:660248-664368
ABC-C TcABCC-8A XP 015838053.1
ABC-C TcABCC-8B XP 015838295.1
ABC-C DmCG7806 FBpp0079220
ABC-C TcABCC-4A XP 008192060.1
ABC-C AtABC-C2958 NW 017852958.1:20321-28779
ABC-C TcABCC-5A XP 973658.2
ABC-C TcABCC-5E XP 008193216.1
ABC-C TcABCC-5F EFA03946.2
ABC-C TcABCC-5G EFA03947.2
ABC-C TcABCC-UA XP 015837651.1
ABC-C TcABCC-7B XP 972534.1
ABC-C AtABC-C2962 NW 017852962.1:87789-94330
ABC-C TcABCC-7A XP 008195764.1
ABC-C DmSur FBpp0402894
ABC-C TcABCC-9C XP 015839198.1
ABC-C AtABC-C3388A NW 017853388.1:25077-29330
ABC-C AtABC-C3388B NW 017853388.1:29840-34345
ABC-C AtABC-C3325D NW 017853325.1:134304-138649
ABC-C AtABC-C2944 NW 017852944.1:794915-802086
ABC-C AtABC-C3325C NW 017853325.1:128820-133331
ABC-C AtABC-C3325A NW 017853325.1:117343-121604
ABC-C AtABC-C3325B NW 017853325.1:123983-128437
ABC-C TcABCC-5O XP 015836125.1
ABC-C TcABCC-5P XP 015836131.1
ABC-C TcABCC-5N XP 971802.2
ABC-C TcABCC-5M XP 008193591.1
ABC-C TcABCC-5Q XP 015836083.1
ABC-C TcABCC-5K XP 971687.2
ABC-C TcABCC-5L XP 015836093.1
ABC-C TcABCC-5J XP 971632.1
ABC-C AtABC-C3083 NW 017853083.1:243278-247272
ABC-C
ABC-D DmPmp70 FBpp0074564
ABC-D TcABCD-9A XP 015838765.1
ABC-D AtABC-D2969 NW 017852969.1:469565-473583
ABC-D DmCG32316 FBpp0088217
ABC-D AtABC-D3077 NW 017853077.1:184487-191678
ABC-D TcABCD-6A XP 971218.1
ABC-D
ABC-B DmHmt-1 FBpp0082642
ABC-B TcABCB-6A XP 008194672.1
ABC-B AtABC-B3353 NW 017853353.1:14861-22219
ABC-B TcABCB-5A XP 001813375.1
ABC-B AtABC-B3256 NW 017853256.1:26921-30162
ABC-B DmABCB7 FBpp0072637
ABC-B AtABC-B3016 NW 017853016.1:493416-496605
ABC-B DmCG3156 FBpp0070078
ABC-B TcABCB-4A XP 008192744.1
ABC-B DmCG10226 FBpp0076720
ABC-B DmMdr65 FBpp0076719
ABC-B AtABC-B2936 NW 017852936.1:1701556-1704715
ABC-B TcABCB-7A XP 972133.2
ABC-B DmMdr49 FBpp0086914
ABC-B DmMdr50 FBpp0086666
ABC-B AtABC-B3881 NW 017853881.1:8912-17440
ABC-B TcABCB-3B XP 008191266.1
ABC-B TcABCB-3A XP 008190820.1
ABC-B AtABC-B3332 NW 017853332.1:125070-132281
ABC-B DmCG1824 FBpp0073491
ABC-B
ABC-E AtABC-E3008 NW 017853008.1:303782-306990
ABC-E TcABCE-3A XP 968009.1
ABC-E Dmpix FBpp0076284
ABC-E
ABC-F DmCG9281 FBpp0073872
ABC-F TcABCF-5A XP 966990.1
ABC-F AtABC-F3048 NW 017853048.1:177330-180623
ABC-F AtABC-F3413 NW 017853413.1:108176-112074
ABC-F DmCG1703 FBpp0073367
ABC-F TcABCF-2A XP 971562.1
ABC-F DmCG9330 FBpp0074708
ABC-F AtABC-F2969 NW 017852969.1:511220-514490
ABC-F TcABCF-8A XP 972814.1
ABC-F
ABC-A TcABCA-UA XP 015839741.1
ABC-A TcABCA-7A XP 008195104.1
ABC-A AtABC-A2939 NW 017852939.1:1677501-1689109
ABC-A AtABC-A3094 NW 017853094.1:258843-266321
ABC-A TcABCA-6A XP 008195056.1
ABC-A TcABCA-UD XP 008199148.1
ABC-A TcABCA-UB XP 008199128.2
ABC-A TcABCA-UC XP 015840355.1
ABC-A AtABC-A3047A NW 017853047.1:19586-27561
ABC-A AtABC-A3047B NW 017853047.1:29577-36903
ABC-A TcABCA-UE XP 015840362.1
ABC-A TcABCA-3A KYB28525.1
ABC-A TcABCA-9A EFA10544.2
ABC-A TcABCA-9B XP 008199705.2
ABC-A DmCG1718 FBpp0297618
ABC-A DmCG6052 FBpp0074959
ABC-A DmCG8908 FBpp0308625
ABC-A DmCG43672 FBpp0302542
ABC-A DmCG1494 FBpp0304585
ABC-A DmCG1801 FBpp0290160
ABC-A DmCG31213 FBpp0307228
ABC-A DmCG42816 FBpp0300324
ABC-A DmCG31731 FBpp0110241
ABC-A DmCG34120 FBpp0309770
ABC-A
ABC-H DmCG11147 FBpp0076796
ABC-H AtABC-H3291 NW 017853291.1:148489-155607
ABC-H TcABCH-9A XP 973444.1
ABC-H AtABC-H2969 NW 017852969.1:474819-483334
ABC-H TcABCH-9B XP 967359.1
ABC-H DmCG33970 FBpp0099384
ABC-H AtABC-H3212 NW 017853212.1:33606-59706
ABC-H DmCG9990 FBpp0084677
ABC-H TcABCH-9C XP 008198312.1
ABC-H
ABC-G AtABC-G2969B NW 017852969.1:35040-41710
ABC-G TcABCG-9C XP 968472.1
ABC-G DmCG31121 FBpp0084098
ABC-G DmCG11069 FBpp0084089
ABC-G AtABC-G2969A NW 017852969.1:32268-34907
ABC-G TcABCG-9D XP 968555.2
ABC-G AtABC-G3103 NW 017853103.1:126678-172330
ABC-G TcABCG-8A XP 008196860.1
ABC-G DmE23 FBpp0088917
ABC-G AtW NW 017853430.1:63792-67789
ABC-G TcW NP 001034521.1
ABC-G Dmw FBpp0070468
ABC-G AtSt NW 017853430.1:58827-63118
ABC-G Dmst FBpp0075149
ABC-G TcSt NP 001306193.1
ABC-G Dmbw FBpp0072026
ABC-G TcBw NP 001306946.1
ABC-G DmCG9664 FBpp0077294
ABC-G DmCG32091 FBpp0075837
ABC-G AtABC-G3097C NW 017853097.1:287916-306607
ABC-G TcABCG-4H XP 973526.1
ABC-G AtABC-G2996C NW 017852996.1:391134-403096
ABC-G TcABCG-4E KYB28165.1
ABC-G DmCG9663 FBpp0077291
ABC-G AtABC-G2996B NW 017852996.1:361176-381334
ABC-G TcABCG-4F XP 971735.1
ABC-G DmCG31689 FBpp0306278
ABC-G AtABC-G3097A NW 017853097.1:237429-248183
ABC-G TcABCG-4G XP 008192849.1
ABC-G DmCG5853 FBpp0079489
ABC-G AtABC-G3086 NW 017853086.1:271423-283205
ABC-G TcABCG-4D XP 973458.1
ABC-G DmAtet FBpp0077116
ABC-G DmCG3164 FBpp0077812
ABC-G AtABC-G2996A NW 017852996.1:322131-336248
ABC-G TcABCG-4C XP 001813184.1
ABC-G AtABC-G3591 NW 017853591.1:47487-70042
ABC-G DmCG4822 FBpp0077814
ABC-G TcABCG-4B XP 015834971.1
ABC-G DmCG17646 FBpp0077554
ABC-G AtABC-G3097B NW 017853097.1:259221-283929
ABC-G TcABCG-4A XP 008192053.1
ABC-G
0.50

Figure 6

COE Classification

- Clade D: Integument Esterases
- Clade E: B- and Pheromone Esterases
- Clade F: Juvenile Hormone Esterases
- Clade H: Glutatactins
- Clade I: Unknown Function
- Clade J: Acetylcholinesterases
- Clade K: Gliotactins
- Clade L: Neuroligins
- Clade M: Neurtactins

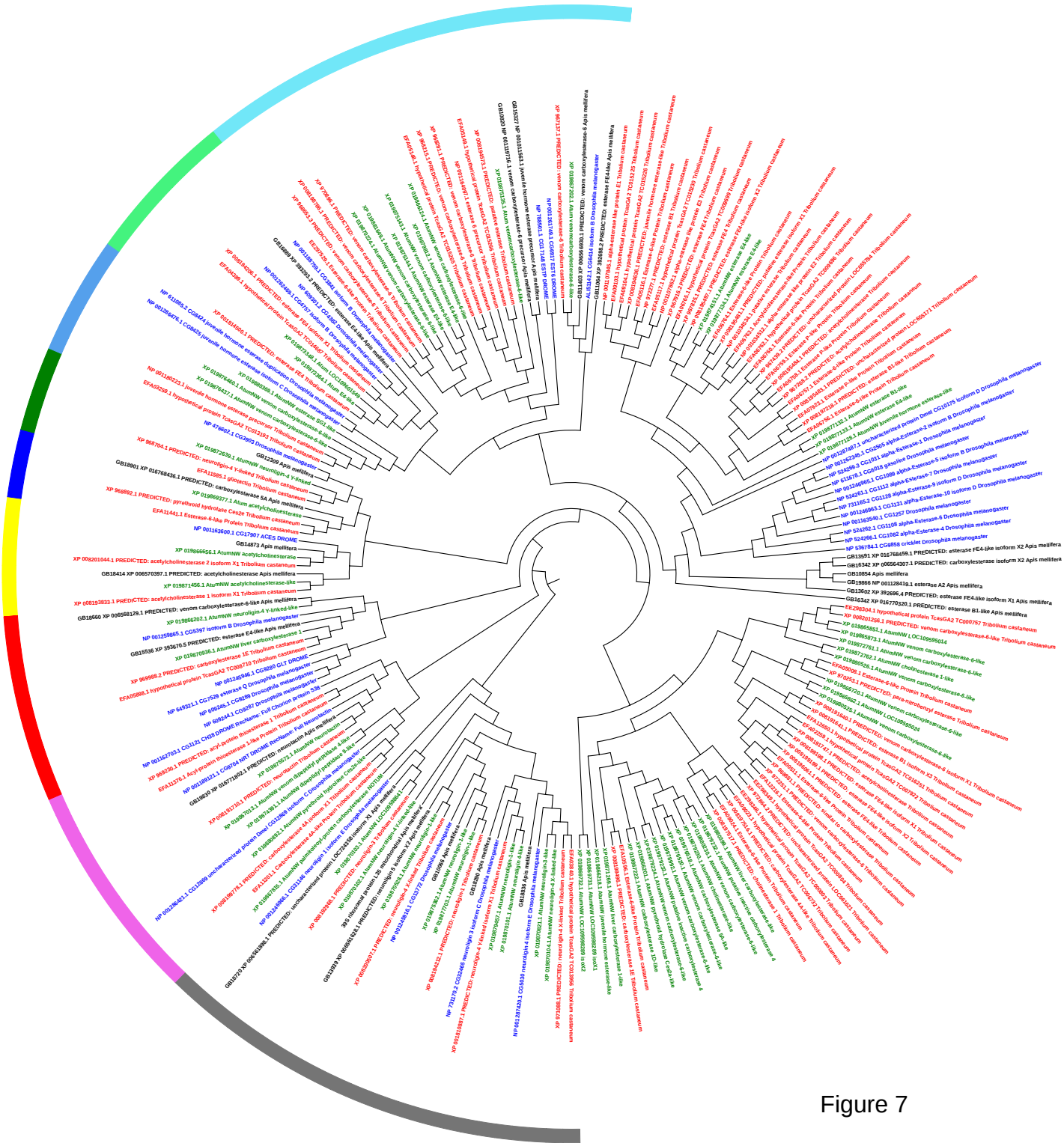


Figure 7

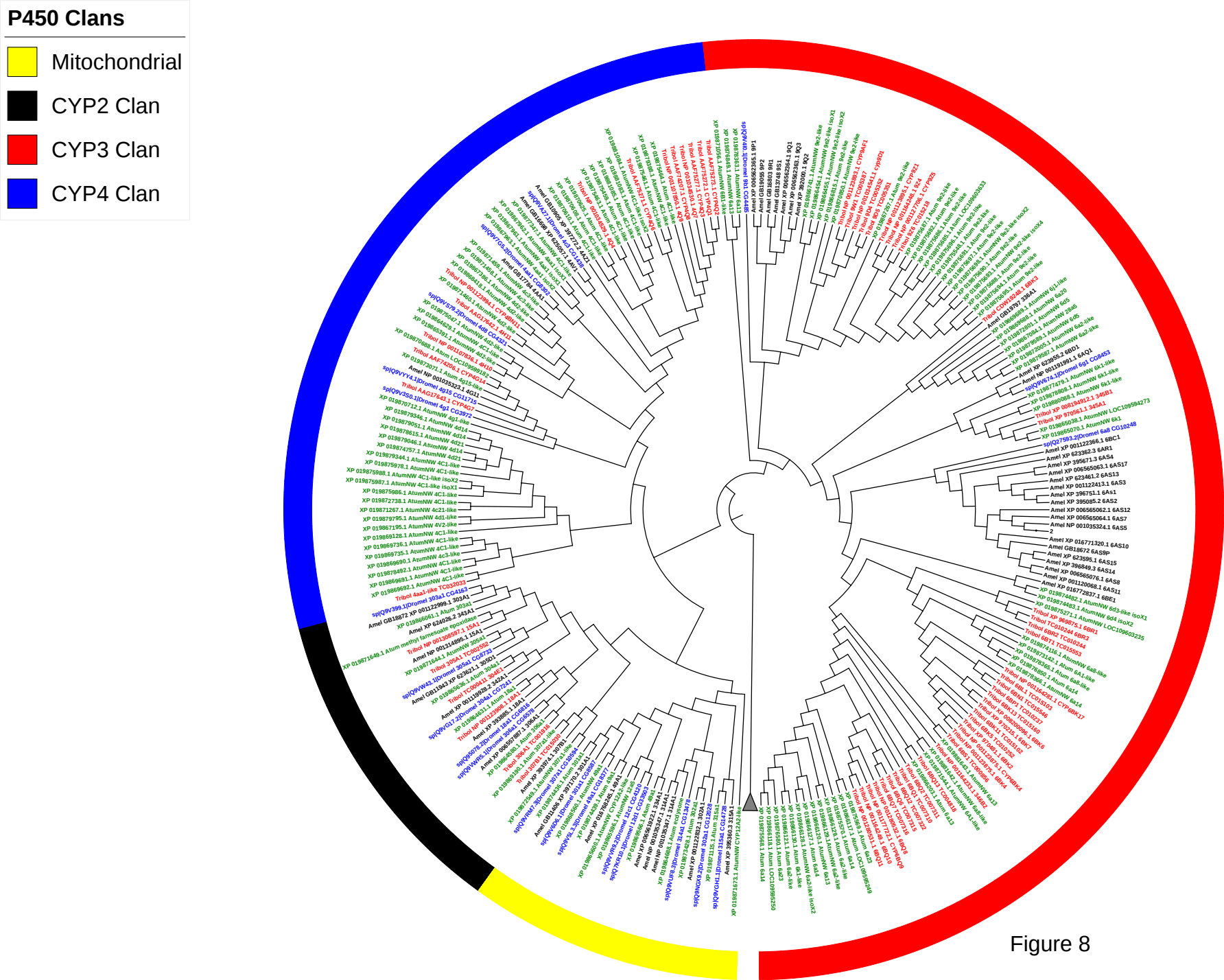


Figure 8

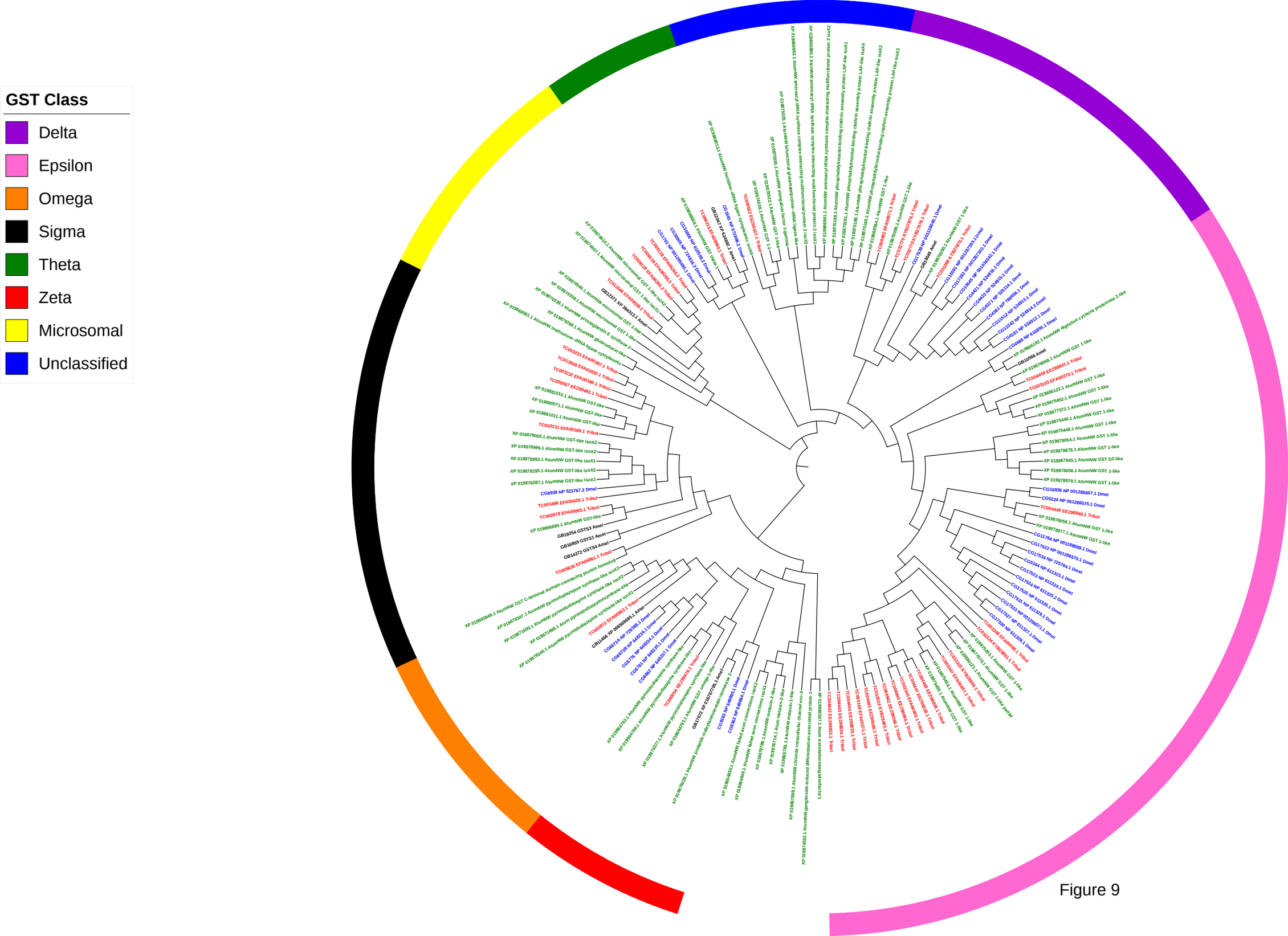

GST Class
Delta
Epsilon
Omega
Sigma
Theta
Zeta
Microsomal
Unclassified

Figure 9

Supplementary Material 3

Click here to access/download

Supplementary Material

SupplementalFile3domains.xlsx

Supplementary Material 2

United States Department of Agriculture

Research, Education, and Economics
Agricultural Research Service

Gigascience
Oxford Academic Journals

2 Sept. 2018

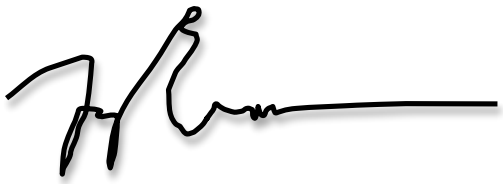
Dear Editors,

We have attached our revised manuscript "**Genome of the small hive beetle (*Aethina tumida*, Coleoptera: Nitidulidae), a worldwide parasite of social bee colonies, provides insights into detoxification and herbivory"** that we feel is significantly improved thanks to the careful comments and critiques from the reviewers. We feel we were able to address these critiques, and hope that this improved manuscript is now suitable for publication in GigaScience. Due to the unique taxonomic position of this beetle and (most importantly) its growing impact on honey bee populations, we feel this genome, gene set, and analysis will help propel a productive set of discussions and discoveries for the biology of this parasite.

We have added substantially to the methods description in the main text to clarify the parameters behind the computational routines in this paper. We have not changed the extensive metadata (described below) for analyses and resources generated during the project and hope this is helpful for the reviewers and future readers. These resources are, we believ, still housed at GigaScience from the prior submission, although they can also be downloaded via: https://drive.google.com/drive/folders/16YDMOzFp4Au4YjBFaIXrBZDvq6tdRICd.

We look forward to your reply and are open again to feedback to make this an impactful paper for *GigaScience*.

Yours sincerely,



Dr. Jay D. Evans
Research Leader USDA-ARS Bee Research Lab BARC-East Bldg. 306 Beltsville, MD 20705
301-504-5143 (ph) 240-755-6720 (cell) jay.evans@ars.usda.gov

1 - Genomic DNA and transcriptomic sequence data Evans et al, Small Hive Beetle

2 - assembly fasta

Bee Research Laboratory
Building 476, Beltsville, MD 20705

GCF_001937115.1_Atum_1.0_genomic.fna is the genome assembly file, which is available at ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/937/115/GCF_001937115.1_Atum_1.0

3 - coding gene annotations (GFF)
GCF_001937115.1_Atum_1.0_genomic.gff is the coding gene annotation file, which is available at ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/937/115/GCF_001937115.1_Atum_1.0

4 - coding gene nucleotide sequences (fasta)
GCF_001937115.1_Atum_1.0_cds_from_genomic.fna is the coding gene nucleotides sequences, which is available at ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/937/115/GCF_001937115.1_Atum_1.0

5 - coding gene translated sequences (protein fasta)
GCF_001937115.1_Atum_1.0_protein.faa is the protein sequences, which is available at ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/937/115/GCF_001937115.1_Atum_1.0

6 - repeats/transposable elements/ncRNAs /other annotations (GFF)
not applicable

7 - gene /gene family alignments (multi-fasta)
sorted_all.trim.aln is the alignment file among species of Figure 2a.

8 - phylogenetic tree files (newick)
For the species phylogenetic tree (Figure 2a), RAxML_bestTree.treeSHB_1000 and RAxML_bipartitions.treeSHB_1000 are the output files.
Figure 5:Fig5GRSeqsAlign.fas
Figure 6: Fig6_ABCWalkerAlig.fas
Figure 7: Fig7_ALN_COE.fas
Figure 8: Fig8_ALN_P450.fas
Figure 9:Fig9_ALN_GST_20180304.fas

9 - BUSCO output file(s)
All BUSCO output files are in the folder BUSCO_output

10 - gene expression levels (fpkm values tables) if performed
not applicable

11 - SNP annotations (VCF) if performed
SHB_annotated.vcf is the SNP annotation file

12 - any perl/python scripts used in analysis process
not applicable

13 - readme.txt including all file names with a brief description of each