

**The International Asian Longhorned Beetle Genome Consortium**

**III.4 Myrosinase and cyanogenic β-glycosidase-like sequences**

**Contributors**

**Leslie A. Kuhn** (Michigan State University, Departments of Biochemistry and Molecular Biology, Computer Science and Engineering, and Fisheries and Wildlife)

**Ann M. Ray** (Xavier University, Department of Biology)

**Erin Scully** (USDA-ARS, Grain, Forage, and Bioenergy Research, and University of Nebraska-Lincoln, Department of Agronomy and Horticulture, and The Pennsylvania State University, Department of Entomology and Center for Chemical Ecology)

**Introduction**

Here we analyze whether sequences putatively identified as members of the beta-glycosyl hydrolase GH1 family (referred to as beta-glycosidases for brevity) are likely to encode for myrosinase (thioglucoside hydrolase) or cyanogenic beta-glycosidase (CBG) enzymes, two subfamilies of GH1. These enzymes are involved in detoxification of host compounds and chemical communication in plants and insects. We became interested in insect myrosinases, their roles in chemical signaling, and the possibility of controlling their activity with inhibitors when we discovered a UniProt *Anoplophora glabripennis* sequence annotated by sequence homology as myrosinase-1 according to (57). A myrosinase system very similar to that in plants exists in insects such as cabbage aphids (*Brevicoryne brassicae* (L.)) and turnip aphids (*Lipaphis (Hyadaphis) erysimi* (Kaltenbach)*,* Hemiptera: Aphididae), which use the isothiocyanate products of myrosinase (ITCs) as components of an alarm pheromone in combination with *E*-β-farnesene (94, 95). Furthermore, in striped flea beetles, (*Phyllotreta striolata* (F.), Coleoptera: Chyromelidae) the ITC products of the beetle-expressed myrosinase enhance the response of beetles to male-produced aggregation pheromones (50, 96-98). Experiments are in progress to test the hypothesis that *A. glabripennis* adults and larvae produce myrosinase, and that field and laboratory-reared beetles differ in their expression of myrosinase. Additional studies are in progress to address the interesting question of whether myrosinase products act as chemical attractants for adults of *A. glabripennis*, possibly in combination with known male-produced compounds (99, 100).

Cyanogenic beta-glycosidases (CBGs) are another GH1 subfamily involved in food detoxification by insects. CBGs are known to detoxify the “cyanide bomb” that plants use to deter herbivory, with cyanide-containing compounds such as amygdalin (also known as laetrile) being synthesized in trees of the genus *Prunus* (101). Insect herbivores may metabolize cyanogenic glucosides or sequester them, also for use in defense against predators (102). A few species of Arthropoda (within the classes Diplopoda, Chilopoda, and Insecta) synthesize cyanogenic glucosides *de novo*, and some of these same species are able to sequester cyanogenic glucosides from their host plants, e.g., Insecta: Lepidoptera: Zygaenidae (forester and burnet moths) (102). Given that *Prunus* is one of the host genera for *A. glabripennis* (103), it is possible that sequences resembling myrosinase actually function as CBGs in *A. glabripennis*, since the two enzymes are closely related (95). In fact, some CBG sequences are more distantly related to each other than they are to myrosinase sequences or other GH1s (104). Yet another member of the GH1 family, lactase-phlorizin hydrolase (also known as lactase or LPH; see http://www.cazy.org/GH1.html), is found in the midguts of Lepidoptera such as the fall armyworm (*Spodoptera frugiperda* (Smith), Noctuidae) and hydrolyzes galactolipids and hemicelluloses by cleaving them at two different active sites (105).

GH1s have a classical (α/β)8 TIM barrel fold in which the two key active site glutamic acid residues are ~200 residues apart, located at the C-termini of β-strands 4 (acid/base) and 7 (nucleophile) (106). The *Sinapis alba* myrosinase, however, contains a glutamine residue in place of the classic glutamate at the end of strand 4, because instead it uses an ascorbate cofactor as the catalytic base (107). Burmeister et al. (108) suggested that myrosinase and cyanogenic beta-glucosidase act identically on substrates.

**Methods and Results**

Twenty-three potential *A. glabripennis* myrosinase amino acid sequences were identified by BLAST search (37) for the highest-scoring alignments (typical identity of ~44%) between *A. glabripennis* GH1s and *Tribolium castaneum* sequences annotated as myrosinases in the SwissProt/UniProt database (109). These sequences are labelled as follows in the *A. glabripennis* genome: AGLA004459-PA, AGLA004460-PA, AGLA004461-PA, AGLA009626-PA, AGLA009627-PA, AGLA009631-PA, AGLA009633-PA, AGLA009636-PA, AGLA009638-PA, AGLA009642-PA, AGLA009643-PA, AGLA009644-PA, AGLA014364-PA, AGLA014878-PA, AGLA016146-PA, AGLA016151-PA, AGLA016153-PA, AGLA016544-PA, AGLA016545-PA, AGLA017752-PA, AGLA018044-PA, AGLA018048-PA, and AGLA018242-PA. The *A. glabripennis* sequence lengths vary from 48 residues (likely a protein sequence fragment) to 1391 residues (probably forming a multi-domain protein). Amino acid sequence motifs have been defined for beta-glycosyl hydrolases by Henrissat and colleagues (110-114), including a set of five aligned sequence blocks available in the PRINTS database (http://www.bioinf.man.ac.uk/cgi-bin/dbbrowser/sprint/searchprintss.cgi?prints\_accn=GLHYDRLASE1). The ProSite database contains a simpler set of two sequence motifs, the first also defined by Henrissat and colleagues (http://prosite.expasy.org/cgi-bin/prosite/nicedoc.pl?PS00653) as the consensus pattern:

**F-x-[FYWM]-[GSTA]-x-[GSTA]-x-[GSTA](2)-[FYNH]-[NQ]-x-E-x- [GSTA]**

All known GH1 sequences in SwissProt at the time the ProSite motif was defined in 1995 satisfied the above pattern, and there were no false positives. The second ProSite motif for GH1s, labeled PS00572, is:

**[LIVMFSTC]-[LIVFYS]-[LIV]-[LIVMST]-E-N-G-[LIVMFAR]-[CSAGN]**

According to ProSite, SwissProt sequence matches to this motif included 105 true positives, 95 false negatives, and 59 false positives. This motif is thus less discriminatory for GH1s. A third, nonredundant GH1 motif has been defined (Fig. 3 in (115)), corresponding to the **VKYWLTINQLYSVPTR** region in *S. alba* myrosinase in which the central glutamine (Q) residue corresponds to the conserved glutamate acid/base residue found in other GH1s.

To analyze relationships between the 23 *A. glabripennis* myrosinase-like GH1 sequences and known myrosinases and CBGs, multiple sequence analysis was performed with MUSCLE (31), as implemented in the software suite Mega6 (Molecular Evolutionary Genetics Analysis; http://www.megasoftware.net). The set of 23 myrosinase-like GH1 sequences was seeded with an additional set of seven GH1 sequences of known function (Fig. S14): *B. brassicae* myrosinase 1 (sp|Q95X01), *P. striolata* myrosinase (tr|A0A059UAD7), *S. alba* myrosinase (Protein Data Bank entry 2wxd, chain M), *P. striolata* GH1 (gi|634006832), *Lepisosteus oculatus* (Winchell) lactase-like protein (lactase-phlorizin hydrolase, a CBG homolog; gi|573881201), *Trifolium repens* CBG (Protein Data Bank entry 1cbg, chain A), and a beta-glycosidase from *S. sulfataricus* (Protein Data Bank entry 1uwt, chain A). From this multiple sequence alignment, which correctly aligned the conserved GH1 motifs for the subset of 17 *A. glabripennis* sequences with complete coverage of the catalytic domain (AGLA004459, 4460, 4461, 9627, 9636, 9638, 9643, 9644, 14364, 14878, 16146, 16153, 16544, 16545, 17752, 18044, and 18242) and the 7 additional previously characterized GH1 sequences, a pairwise distance matrix between sequences was calculated in Mega6 by using the (116) model of amino acid substitution likelihood, with uniform rates across sites and pairwise deletion for handling any sequence gaps. Based on the pairwise distance matrix, agglomerative cluster analysis of these 24 sequences was performed in SciPy (117) by using the complete linkage clustering function (scipy.cluster.hierarchy.complete). Complete linkage clustering has several practical and intuitive advantages. At any given sequence similarity cutoff (using the similarity values defined in the pairwise distance matrix described above), all sequences grouped in a cluster are guaranteed to have at least that pairwise degree of similarity. Furthermore, the defined clusters are guaranteed to be the densest such grouping of sequences for a given degree of similarity. The results are deterministic and do not depend on the order in which sequences are input. Upon clustering of the 17 complete *A. glabripennis* GH1-like sequences and seven GH1 sequences of known function (results not shown), the known myrosinases split across different sequence clusters; in particular, the *S. alba* myrosinase that uses an ascorbate cofactor appears in a separate cluster from the others (as was also found by the (104) analysis of complete myrosinase sequences).

Because analyses like the above treat all parts of the amino acid sequence as equally important, clustering aimed at functional annotation may be led astray by the low degree of conservation of solvent-exposed surface residues outside the active site, due to few steric, folding, or functional constraints on their evolution. Instead, focusing on active site and functional motif sequence conservation within the context of overall protein fold conservation is a better indicator of the conservation of substrate binding and reaction mechanism. Thus, to analyze the likelihood of myrosinase function in one or more of the *A. glabripennis* GH1-like sequences, we developed a sequence fingerprint consisting of only the residues contributing to the characteristic GH1 and myrosinase motifs described above, plus additional residues that contribute to the active site based on the available *S. alba* myrosinase crystal structure (Protein Data Bank entry 2wxd). These residues are not necessarily contiguous in sequence, since active sites typically involve discontiguous residues forming a substrate interaction surface. The motifs and active site residues only are presented in N-terminal to C-terminal order in the fingerprint (Fig. S14), as extracted from the MUSCLE alignment of the complete protein sequences. This fingerprint shows several interesting features:

* AGLA009631, 9633, and 9636 have an unusual residue (Asp) in place of the highly conserved Glu at the acid/base catalyst position in the (115) motif centered on the catalytic acid/base. Unconserved but chemically similar residues are known to occur in this position in some active GH1s: Gln in the *S. alba* myrosinase that uses ascorbate as a cofactor and Asn in the *L. oculatus* CBG-like lactase. Furthermore, Asp may provide a catalytic acid/base appropriate for interacting with a larger substrate than is found in active sites with Glu, as these side chains differ by just one methylene group in length.
* AGLA009636 and 9643 are unlikely to be active GH1s based on having an Ala or Gln substitution, respectively, for the conserved Glu in ProSite motif 2 (catalytic nucleophile position).

The pairwise distance matrix for the above aligned active site sequence motifs was calculated in Mega6 (using parameters described above), followed by complete linkage cluster analysis in order to group subfamilies of these sequences according to active site and GH1 motif similarity (Fig. S15). This analysis of motifs, unlike the full sequence analysis, groups the known myrosinases (lower blue lines in Fig. S15) and suggests that one of the *A. glabripennis* sequences in particular, AGLA018242-PA, is closely similar to a known insect myrosinase (UniProt Q95X01 from *B. brassicae*).

A known CBG and CBG-related GH1, lactase-phlorizin hydrolase, also fall in a single cluster when GH1 motifs and active site residues are used to assess similarity (green lines in Fig. S15). Further analysis is needed to evaluate the red cluster of *A. glabripennis* sequences, because they are intriguingly intermediate in similarity to known myrosinases and the known cyanogenic beta-glycosidase. Husebye et al., (95) noted that *B. brassicae* myrosinase shows close sequence and structural similarity to both *T. repens* CBG (41% sequence identity) and *S. alba* myrosinase (34% identity).

The 3-dimensional compatibility of *A. glabripennis* sequences to the myrosinase atomic structure can be assessed by homology modeling. For instance, a 3-dimensional structural model (Fig. S16) was built for one of the GH1 myrosinase candidates (AGLA009643-PA, annotated by (57) as V5GKP2 in the SwissProt/UniProt database) based on its alignment with *B. brassicae* myrosinase, by using SwissModel homology modeling software (118) (http://swissmodel.expasy.org). The resulting model was judged to be of reasonable accuracy, based on the SwissModel structural evaluation of bond stereochemistry and intramolecular contacts indicating similar quality to well-resolved crystal structures. This model for the *A. glabripennis* protein structure (Fig. S16) also provides an unambiguous definition of the active site residues and their relative orientation, which is valuable for assessing the positioning of catalytic groups and the likelihood of interacting with glycosidic versus other substrates. An alternative model for this V5GKP2 sequence based on using the *T. repens* CBG structure (Protein Data Bank entry 1CBG) was of much lower stereochemical quality, consistent with the active-site and motif cluster analysis (above) indicating that the AGLA009643 GH1 motifs and active site are more closely related to myrosinases.

**Discussion**

Given that *A. glabripennis* is highly polyphagous (not limited to one food source) and not known to feed on glucosinolate-myrosinase system containing plants, it is unclear why the beetles would express one or more myrosinase-like enzymes. Myrosinase is necessary to defuse the so-called defensive “mustard oil bomb” produced by plants in the order Brassicales in order to deter herbivory (119). Oligophagous or monophagous insects that feed on these plants often employ a system for detoxification by sequestration of glucosinolates, which are then broken down and released by an insect-expressed myrosinase system such as those in flea beetles and cabbage aphids (50, 95). In their native range, the larval host plants of *A. glabripennis* are trees of the genus *Acer, Populus, Salix* and *Ulmus*, whereas in North America *A. glabripennis* larvae can develop in hosts in the native tree genera and also in *Aesculus, Albizia, Betula, Cercidiphyllum, Fraxinus, Platanus, Prunus*, and *Sorbus* (103).These genera are not known to contain the myrosinase system. However, glucosinolates could be taken up from another food source by *A. glabripennis*.

Myrosinase could also play a role in chemical signaling in *A. glabripennis* adults, as is the case for *P. striolata*. *A. glabripennis* may sequester glucosinolates or respond to the products of myrosinase hydrolysis (ITCs), and these compounds may synergize the attraction of adult beetles to the male-produced compounds, 4-(*n*-heptyloxy)butan-1-ol, 4-(*n*-heptyloxy)butanal, and (3*E*,6*E*)-α-farnesene (99, 100). This finding would be analogous to a system already known to exist in cabbage and turnip aphids, which combine ITCs with (*E*)-β-

farnesene as an alarm pheromone (94, 95). Experiments are in progress to evaluate the response of *A. glabripennis* adults to ITCs alone, and in combination with male-produced attractants (A.M.R., unpub. data).

Another intriguing possibility is that one or more *A. glabripennis* sequences is active as a CBG, as suggested by the presence of a cluster of *A. glabripennis* sequences intermediate between known CBGs and myrosinases. Toxic cyanogenic glycosides are used by plants as a defense system analogous to the myrosinase system. Cyanogenic glycoside substrates are packaged in an adjacent compartment to CBGs in plants. Herbivory breaches the compartments, initiating the breakdown of cyanogenic glycosides and release of hydrogen cyanide gas (102). This system is present in a number of plants in the Rosaceae, including trees of the genus *Prunus*, a known host of *A. glabripennis* (103). The cyanogenic glycosides prunasin and amygalin are found in tissues of *Prunus* (120), often in levels that are toxic to herbivores, including humans and domestic animals (121). Aside from amygdalin and prunasin, linamarin and lotaustralin are cyanogenic glycosides that are widespread in plants and often occur together, particularly in Fabaceae (122). Both cyanogenic glycosides and their release of hydrogen cyanide gas can affect the behavior of herbivorous insects and other animals. Specialist insect herbivores have developed CBG systems mirroring those in plants to metabolize cyanogenic glucosides or sequester them for use in predator defense, just as insects developed myrosinase systems to combat predation and aid in signaling to conspecifics (102).

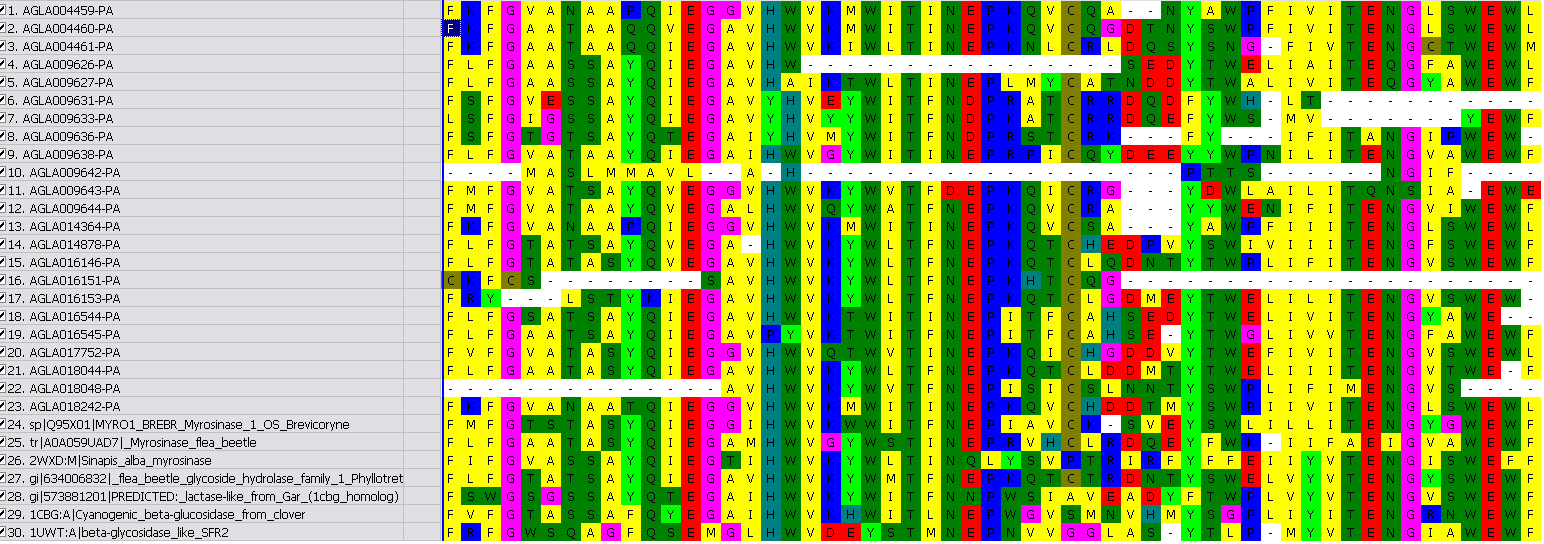
**Acknowledgments**

We sincerely thank Sebastian Raschka (Michigan State University) for his contributions to the Mega6 and SciPy cluster analysis of sequences, and also Kathryn Morris, Vanessa Lopez, Scott Gula, and Brenna Walters (Xavier University) for their help with the preliminary laboratory bioassays.

**Author contributions**

L.A.K. developed the chemical signaling hypothesis for myrosinases and cyanogenic beta-glycosidases in *A. glabripennis*, performed the sequence and structural bioinformatics analysis, and drafted this section on myrosinase and cyanogenic β-glycosidase-like sequences; A.M.R. designed and performed experiments and contributed to writing and editing the manuscript; E.S. provided myrosinase-like genomic sequence data for analysis, suggested roles for these enzymes in digestive physiology, and edited the manuscript.

**Fig. S14. GH1 and myrosinase motifs for 23 *A. glabripennis* proteins and seven known GH1s from other organisms**, as described in the Methods, colored by Mega6 according to amino acid chemistry: yellow, hydrophobic; light green, Tyr; dark green, polar; blue-green, His; pink, Gly; brown, Cys; red, Asp or Glu; dark blue, Lys or Arg. *A. glabripennis* sequences with substantial gaps (indicated by dashes on white background in the fingerprint) were omitted from the subsequent distance matrix calculation and clustering due to being incomplete sequences (unable to form the GH1 fold), though the corresponding full-length sequence, if expressed, could be a GH1. In the annotations beneath the fingerprint, “AS” indicates active site glucose or aglycone binding residues in myrosinase or CBG, as defined in Fig. 3 of (95). Additional active-site residues were identified by using PyMOL software (v. 1.5.0.5, Schrödinger, LLC, NY, NY) based on occurring within contact distance (5 Ångstroms) of the E18 inhibitor in Protein Data Bank crystal structure 2wxd (*S. alba* myrosinase).



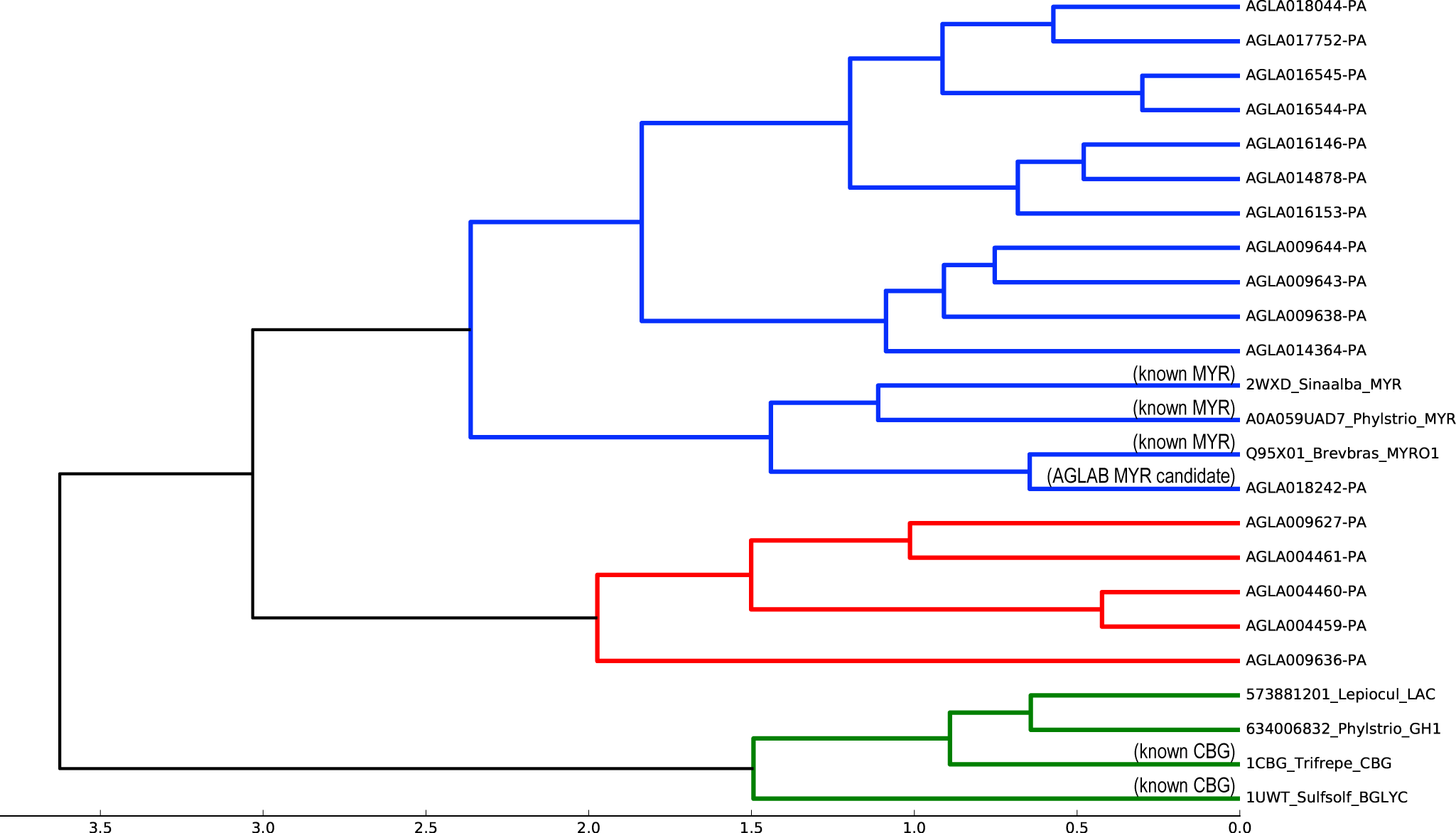
ProSite GH1 motif 1 region AS Cicek motif for GH1 including AS (discontig- ProSite GH1 motif 2; AS

conserved E, the acid/base uous residues) E is a catalytic

catalyst (Wang, 1990); preceding nucleophile

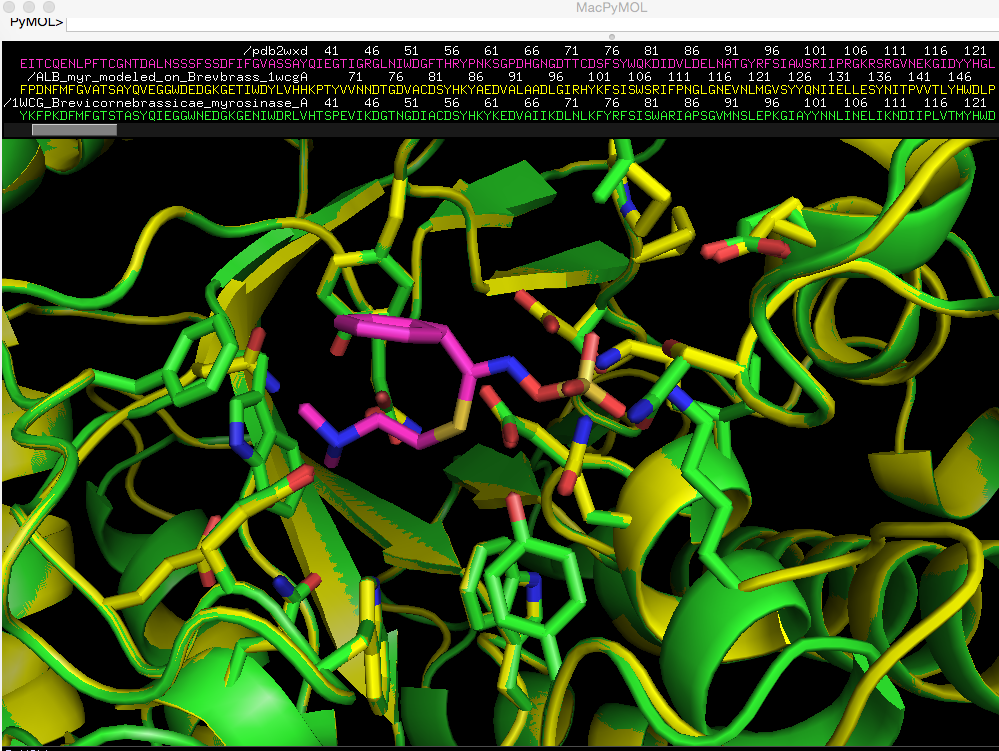
Asn H-bonds to 2-OH in GH1s ­­(Withers,1990)

**Fig. S15*.* Complete linkage clustering showing similarity in active site motif regions in AGLAB myrosinase- and cyanogenic β-glycosidase-like sequences and characterized GH1s from other organisms**,based on pairwise Jones-Taylor-Thornton amino acid substitution distance followed by complete linkage clustering. Known myrosinases (MYR) appear in a cluster with candidate MYRs from AGLAB (blue) and known cyanogenic β-glycosidases (CBG) and close relatives also form a cluster (green).



Jones-Taylor-Thornton Amino Acid Substitution Distance

**Fig. S16*.* Evaluation of a potential *A. glabripennis* myrosinase by 3-dimensional modeling** (carbon atoms and main-chain ribbons shown in yellow), using the X-ray structure of *B. brassicae* myrosinase (Protein Data Bank entry 1WCG; carbon atoms and main-chain ribbons in green) as a template for the *A. glabripennis* sequence AGLA009643-PA. The glucosinolate substrate-like E18 inhibitor from *S. alba* myrosinase is shown with carbon atoms in pink, for reference. Immediately beneath the alpha carbon of the phenyl group in E18, the GH1 conserved catalytic Glu acid/base appears in overlapping positions in the B. brassicae myrosinase and the *A. glabripennis* myrosinase-like protein. However, the Glu catalytic nucleophile conserved in GH1s including myrosinase is not conserved in this *A. glabripennis* sequence, suggesting it is not active as a myrosinase. The figure was rendered in PyMOL v. 1.5.0.5.



**Supplemental References**

1. Hare EE & Johnston JS (2011) Genome Size Determination Using Flow Cytometry of Propidium Iodide-Stained Nuclei. *Molecular Methods for Evolutionary Genetics* 772:3-12.

2. Gnerre S*, et al.* (2011) High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci U S A* 108(4):1513-1518.

3. Holt C & Yandell M (2011) MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC bioinformatics* 12:491.

4. Stanke M, Diekhans M, Baertsch R, & Haussler D (2008) Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* 24(5):637-644.

5. Korf I (2004) Gene finding in novel genomes. *BMC bioinformatics* 5.

6. Lee E*, et al.* (2013) Web Apollo: a web-based genomic annotation editing platform. *Genome biology* 14(8):R93.

7. Slater GS & Birney E (2005) Automated generation of heuristics for biological sequence comparison. *BMC bioinformatics* 6:31.

8. Poelchau M*, et al.* (2015) The i5k Workspace@NAL--enabling genomic data access, visualization and curation of arthropod genomes. *Nucleic acids research* 43(Database issue):D714-719.

9. Skinner ME, Uzilov AV, Stein LD, Mungall CJ, & Holmes IH (2009) JBrowse: a next-generation genome browser. *Genome research* 19(9):1630-1638.

10. Engsontia P*, et al.* (2008) The red flour beetle's large nose: An expanded odorant receptor gene family in Tribolium castaneum. *Insect biochemistry and molecular biology* 38(4):387-397.

11. Hill CA*, et al.* (2002) G protein-coupled receptors in Anopheles gambiae. *Science* 298(5591):176-178.

12. Kent LB, Walden KKO, & Robertson HM (2008) The Gr family of candidate gustatory and olfactory receptors in the yellow-fever mosquito Aedes aegypti. *Chem Senses* 33(1):79-93.

13. Robertson HM, Gadau J, & Wanner KW (2010) The insect chemoreceptor superfamily of the parasitoid jewel wasp Nasonia vitripennis. *Insect molecular biology* 19 Suppl 1:121-136.

14. Cantarel BL*, et al.* (2008) MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome research* 18(1):188-196.

15. Yandell M & Ence D (2012) A beginner's guide to eukaryotic genome annotation. *Nature reviews. Genetics* 13(5):329-342.

16. Keeling CI*, et al.* (2013) Draft genome of the mountain pine beetle, Dendroctonus ponderosae Hopkins, a major forest pest. *Genome biology* 14(3):R27.

17. Richards S*, et al.* (2008) The genome of the model beetle and pest Tribolium castaneum. *Nature* 452(7190):949-955.

18. Terrapon N*, et al.* (2014) Molecular traces of alternative social organization in a termite genome. *Nat Commun* 5:3636.

19. Kirkness EF*, et al.* (2010) Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proc Natl Acad Sci U S A* 107(27):12168-12173.

20. International Aphid Genomics C (2010) Genome sequence of the pea aphid Acyrthosiphon pisum. *PLoS biology* 8(2):e1000313.

21. Elsik CG*, et al.* (2014) Finding the missing honey bee genes: lessons learned from a genome upgrade. *Bmc Genomics* 15:86.

22. Weinstock GM*, et al.* (2006) Insights into social insects from the genome of the honeybee Apis mellifera. *Nature* 443(7114):931-949.

23. Werren JH*, et al.* (2010) Functional and evolutionary insights from the genomes of three parasitoid Nasonia species. *Science* 327(5963):343-348.

24. You M*, et al.* (2013) A heterozygous moth genome provides insights into herbivory and detoxification. *Nat Genet* 45(2):220-225.

25. Zhan S, Merlin C, Boore JL, & Reppert SM (2011) The monarch butterfly genome yields insights into long-distance migration. *Cell* 147(5):1171-1185.

26. Zhao C*, et al.* (2015) A massive expansion of effector genes underlies gall-formation in the wheat pest Mayetiola destructor. *Current biology : CB* 25(5):613-620.

27. Adams MD*, et al.* (2000) The genome sequence of Drosophila melanogaster. *Science* 287(5461):2185-2195.

28. Holt RA*, et al.* (2002) The genome sequence of the malaria mosquito Anopheles gambiae. *Science* 298(5591):129-149.

29. Kriventseva EV*, et al.* (2015) OrthoDB v8: update of the hierarchical catalog of orthologs and the underlying free software. *Nucleic acids research* 43(Database issue):D250-256.

30. Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22(21):2688-2690.

31. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* 32(5):1792-1797.

32. Capella-Gutierrez S, Silla-Martinez JM, & Gabaldon T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25(15):1972-1973.

33. Rice P, Longden I, & Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* 16(6):276-277.

34. Stover BC & Muller KF (2010) TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. *BMC bioinformatics* 11:7.

35. Huson DH & Scornavacca C (2012) Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Systematic biology* 61(6):1061-1067.

36. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, & Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31(19):3210-3212.

37. Altschul SF, Gish W, Miller W, Myers EW, & Lipman DJ (1990) Basic Local Alignment Search Tool. *J Mol Biol* 215(3):403-410.

38. R\_Core\_Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

39. Zheng CF & Sankoff D (2011) On the PATHGROUPS approach to rapid small phylogeny. *BMC Bioinformatics* 12.

40. Wheeler D, Redding AJ, & Werren JH (2013) Characterization of an ancient lepidopteran lateral gene transfer. *Plos One* 8(3):e59262.

41. Marvaldi AE, Duckett CN, Kjer KM, & Gillespie JJ (2009) Structural alignment of 18S and 28S rDNA sequences provides insights into phylogeny of Phytophaga (Coleoptera: Curculionoidea and Chrysomeloidea). *Zool Scr* 38(1):63-77.

42. Wei YD*, et al.* (2006) Molecular cloning, expression, and enzymatic activity of a novel endogenous cellulase from the mulberry longicorn beetle, Apriona germari. *Comp Biochem Phys B* 145(2):220-229.

43. Sugimura M, Watanabe H, Lo N, & Saito H (2003) Purification, characterization, cDNA cloning and nucleotide sequencing of a cellulase from the yellow-spotted longicorn beetle, Psacothea hilaris. *Eur J Biochem* 270(16):3455-3460.

44. Lee SJ*, et al.* (2004) cDNA cloning, expression, and enzymatic activity of a cellulase from the mulberry longicorn beetle, Apriona germari. *Comp Biochem Phys B* 139(1):107-116.

45. Chang CJ*, et al.* (2012) A novel exo-cellulase from white spotted longhorn beetle (Anoplophora malasiaca). *Insect biochemistry and molecular biology* 42(9):629-636.

46. Calderon-Cortes N, Watanabe H, Cano-Camacho H, Zavala-Paramo G, & Quesada M (2010) cDNA cloning, homology modelling and evolutionary insights into novel endogenous cellulases of the borer beetle Oncideres albomarginata chamela (Cerambycidae). *Insect molecular biology* 19(3):323-336.

47. Pauchet Y, Kirsch R, Giraud S, Vogel H, & Heckel DG (2014) Identification and characterization of plant cell wall degrading enzymes from three glycoside hydrolase families in the cerambycid beetle Apriona japonica. *Insect biochemistry and molecular biology* 49:1-13.

48. Scully ED, Hoover K, Carlson JE, Tien M, & Geib SM (2013) Midgut transcriptome profiling of Anoplophora glabripennis, a lignocellulose degrading cerambycid beetle. *Bmc Genomics* 14.

49. Cantarel BL*, et al.* (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic acids research* 37:D233-D238.

50. Beran F*, et al.* (2014) Phyllotreta striolata flea beetles use host plant defense compounds to create their own glucosinolate-myrosinase system. *P Natl Acad Sci USA* 111(20):7349-7354.

51. Pauchet Y, Kirsch R, Giraud S, Vogel H, & Heckel DG (2014) Identification and characterization of plant cell wall degrading enzymes from three glycoside hydrolase families in the cerambycid beetle Apriona japonica. *Insect Biochem Mol Biol* 49:1-13.

52. Geib SM, Tien M, & Hoover K (2010) Identification of proteins involved in lignocellulose degradation using in gel zymogram analysis combined with mass spectroscopy-based peptide analysis of gut proteins from larval Asian longhorned beetles, Anoplophora glabripennis. *Insect Sci* 17(3):253-264.

53. Keena MA (2005) Pourable artificial diet for rearing Anoplophora glabripennis (Coleoptera : Cerambycidae) and methods to optimize larval survival and synchronize development. *Ann Entomol Soc Am* 98(4):536-547.

54. Intra J, Pavesi G, & Horner DS (2008) Phylogenetic analyses suggest multiple changes of substrate specificity within the Glycosyl hydrolase 20 family. *Bmc Evol Biol* 8.

55. Scully ED*, et al.* (2014) Functional genomics and microbiome profiling of the Asian longhorned beetle (Anoplophora glabripennis) reveal insights into the digestive physiology and nutritional ecology of wood feeding beetles. *Bmc Genomics* 15.

56. Scully ED, Hoover K, Carlson J, Tien M, & Geib SM (2012) Proteomic analysis of Fusarium solani isolated from the Asian longhorned beetle, Anoplophora glabripennis. *Plos One* 7(4):e32990.

57. Scully ED, Hoover K, Carlson JE, Tien M, & Geib SM (2013) Midgut transcriptome profiling of Anoplophora glabripennis, a lignocellulose degrading cerambycid beetle. *Bmc Genomics* 14:850.

58. Jongsma MA, Bakker PL, Peters J, Bosch D, & Stiekema WJ (1995) Adaptation of Spodoptera-Exigua Larvae to Plant Proteinase-Inhibitors by Induction of Gut Proteinase Activity Insensitive to Inhibition. *P Natl Acad Sci USA* 92(17):8041-8045.

59. Patankar AG*, et al.* (2001) Complexity in specificities and expression of Helicoverpa armigera gut proteinases explains polyphagous nature of the insect pest. *Insect biochemistry and molecular biology* 31(4-5):453-464.

60. Tartar A*, et al.* (2009) Parallel metatranscriptome analyses of host and symbiont gene expression in the gut of the termite Reticulitermes flavipes. *Biotechnology for biofuels* 2:25.

61. Dereeper A*, et al.* (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic acids research* 36:W465-W469.

62. Ronquist F & Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19(12):1572-1574.

63. Tamura K*, et al.* (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol* 28(10):2731-2739.

64. Eyun SI*, et al.* (2014) Molecular Evolution of Glycoside Hydrolase Genes in the Western Corn Rootworm (Diabrotica virgifera virgifera). *Plos One* 9(4).

65. Voragen AGJ, Schols HA, & Pilnik W (1986) Determination of the degree of methylation and acetylation of pectins by h.p.l.c. *Food Hydrocolloid* 1(1):65-70.

66. Yoo SH, Fishman ML, Savary BJ, & Hotchkiss AT (2003) Monovalent salt-induced gelation of enzymatically deesterified pectin. *J Agr Food Chem* 51(25):7410-7417.

67. Highley TL (1997) Carbohydrolase Assays. *Methods in plant biochemistry and molecular biology*, ed DASHEK WV (CRC Press, Boca Raton, FL), p??

68. Lee SJ*, et al.* (2005) A novel cellulase gene from the mulberry longicorn beetle, Apriona germari: Gene structure, expression, and enzymatic activity. *Comp Biochem Phys B* 140(4):551-560.

69. Willis JD, Oppert B, Oppert C, Klingeman WE, & Jurat-Fuentes JL (2011) Identification, cloning, and expression of a GHF9 cellulase from Tribolium castaneum (Coleoptera: Tenebrionidae). *J Insect Physiol* 57(2):300-306.

70. Brett CT & Waldron KW (1996) The molecular components of the wall. *hysiology and Biochemistry of Plant Cell Walls*, eds Brett CT & Waldron KW (Springer), p??

71. O'Neill MA & York WS (2003) The composition and structure of plant primary cell walls. *The Plant Cell Wall*, ed Rose JKC (Blackwell), pp 1-54.

72. Beguin P (1983) Detection of Cellulase Activity in Polyacrylamide Gels Using Congo Red-Stained Agar Replicas. *Anal Biochem* 131(2):333-336.

73. Sezutsu H, Le Goff G, & Feyereisen R (2013) Origins of P450 diversity. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 368(1612):20120428.

74. Rewitz KF, O'Connor MB, & Gilbert LI (2007) Molecular evolution of the insect Halloween family of cytochrome P450s: phylogeny, gene organization and functional conservation. *Insect biochemistry and molecular biology* 37(8):741-753.

75. Qiu Y*, et al.* (2012) An insect-specific P450 oxidative decarbonylase for cuticular hydrocarbon biosynthesis. *Proc Natl Acad Sci U S A* 109(37):14858-14863.

76. Bock KW (2003) Vertebrate UDP-glucuronosyltransferases: functional and evolutionary aspects. *Biochemical pharmacology* 66(5):691-696.

77. Ahn SJ, Vogel H, & Heckel DG (2012) Comparative analysis of the UDP-glycosyltransferase multigene family in insects. *Insect biochemistry and molecular biology* 42(2):133-147.

78. Mackenzie PI*, et al.* (2005) Nomenclature update for the mammalian UDP glycosyltransferase (UGT) gene superfamily. *Pharmacogenet Genom* 15(10):677-685.

79. Wang XH*, et al.* (2014) The locust genome provides insight into swarm formation and long-distance flight. *Nat Commun* 5:1-9.

80. Lü FG*, et al.* (2015) Identification of carboxylesterase genes and their expression profiles in the Colorado potato beetle Leptinotarsa decemlineata treated with fipronil and cyhalothrin. *Pestic Biochem Phys* 122:86-95.

81. Wheelock CE, Shan G, & Ottea J (2005) Overview of carboxylesterases and their role in the metabolism of insecticides. *J Pestic Sci* 30(2):75-83.

82. Durand N*, et al.* (2011) Degradation of pheromone and plant volatile components by a same odorant-degrading enzyme in the cotton leafworm, Spodoptera littoralis. *Plos One* 6(12):e29147.

83. Yu QY, Lu C, Li WL, Xiang ZH, & Zhang Z (2009) Annotation and expression of carboxylesterases in the silkworm, Bombyx mori. *Bmc Genomics* 10:553.

84. Riddiford LM, Hiruma K, Zhou XF, & Nelson CA (2003) Insights into the molecular basis of the hormonal control of molting and metamorphosis from Manduca sexta and Drosophila melanogaster. *Insect biochemistry and molecular biology* 33(12):1327-1338.

85. Oakeshott J, Claudianos C, Campbell P, Newcomb R, & Russell R (2005) Biochemical genetics and genomics of insect esterases. *Comprehensive molecular insect science*, eds Gilbert L, Iatrou K, & Gill S (Elsevier, London), Vol 5, pp 309-361.

86. Devonshire AL & Moores GD (1982) A carboxylesterase with broad substrate specificity causes organophosphorus, carbamate and pyrethroid resistance in peach-potato aphids (Myzus persicae). *Pestic Biochem Phys* 18(2):235-246.

87. Newcomb RD*, et al.* (1997) A single amino acid substitution converts a carboxylesterase to an organophosphorus hydrolase and confers insecticide resistance on a blowfly. *Proc Natl Acad Sci U S A* 94(14):7464-7468.

88. Soderlund DM (1997) Molecular Mechanisms of Insecticide Resistance. *Molecular Mechanisms of Resistance to Agrochemicals,* Chemistry of Plant Protection, ed Sjut V (Springer Berlin Heidelberg, Berlin Heidelberg), Vol 13, pp 21-56.

89. Scharf ME & Boucias DG (2010) Potential of termite-based biomass pre-treatment strategies for use in bioethanol production. *Insect Sci* 17(3):166-174.

90. Despres L, David JP, & Gallet C (2007) The evolutionary ecology of insect resistance to plant chemicals. *Trends in ecology & evolution* 22(6):298-307.

91. Small GJ & Hemingway J (2000) Molecular characterization of the amplified carboxylesterase gene associated with organophosphorus insecticide resistance in the brown planthopper, Nilaparvata lugens. *Insect molecular biology* 9(6):647-653.

92. Zwickl DJ (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. dissertation (The University of Texas at Austin).

93. Abascal F, Zardoya R, & Posada D (2005) ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21(9):2104-2105.

94. Dawson GW, Griffiths DC, Pickett JA, Wadhams LJ, & Woodcock CM (1987) Plant-Derived Synergists of Alarm Pheromone from Turnip Aphid, Lipaphis (Hyadaphis) Erysimi (Homoptera, Aphididae). *J Chem Ecol* 13(7):1663-1671.

95. Husebye H*, et al.* (2005) Crystal structure at 1.1 Angstroms resolution of an insect myrosinase from Brevicoryne brassicae shows its close relationship to beta-glucosidases. *Insect biochemistry and molecular biology* 35(12):1311-1320.

96. Beran F*, et al.* (2011) Male Phyllotreta striolata (F.) Produce an Aggregation Pheromone: Identification of Male-specific compounds and Interaction with Host Plant Volatiles. *J Chem Ecol* 37(1):85-97.

97. Tóth M, Csonka E, Bartelt RJ, Cosse AA, & Zilkowski BW (2012) Similarities in pheromonal communication of flea beetles Phyllotreta cruciferae Goeze and Ph. vittula Redtenbacher (Coleoptera, Chrysomelidae). *J Appl Entomol* 136(9):688-697.

98. Tóth M*, et al.* (2005) Pheromonal activity of compounds identified from male Phyllotreta cruciferae: Field tests of racemic mixtures, pure enantiomers, and combinations with allyl isothiocyanate. *J Chem Ecol* 31(11):2705-2720.

99. Crook DJ, Lance DR, & Mastro VC (2014) Identification of a Potential Third Component of the Male-Produced Pheromone of Anoplophora glabripennis and its Effect on Behavior. *J Chem Ecol* 40(11-12):1241-1250.

100. Zhang AJ, Oliver JE, Aldrich JR, Wang BD, & Mastro VC (2002) Stimulatory beetle volatiles for the Asian longhorned beetle, Anoplophora glabripennis (Motschulsky). *Z Naturforsch C* 57(5-6):553-558.

101. Gleadow RM & Moller BL (2014) Cyanogenic glycosides: synthesis, physiology, and phenotypic plasticity. *Annual review of plant biology* 65:155-185.

102. Zagrobelny M*, et al.* (2004) Cyanogenic glucosides and plant-insect interactions. *Phytochemistry* 65(3):293-306.

103. Haack RA, Herard F, Sun JH, & Turgeon JJ (2010) Managing Invasive Populations of Asian Longhorned Beetle and Citrus Longhorned Beetle: A Worldwide Perspective. *Annu Rev Entomol* 55:521-546.

104. Rask L*, et al.* (2000) Myrosinase: gene family evolution and herbivore defense in Brassicaceae. *Plant molecular biology* 42(1):93-113.

105. Marana SR, Terra WR, & Ferreira C (2000) Purification and properties of a beta-glycosidase purified from midgut cells of Spodoptera frugiperda (Lepidoptera) larvae. *Insect biochemistry and molecular biology* 30(12):1139-1146.

106. Henrissat B*, et al.* (1995) Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proc Natl Acad Sci U S A* 92(15):7090-7094.

107. Burmeister WP, Cottaz S, Rollin P, Vasella A, & Henrissat B (2000) High resolution x-ray crystallography shows that ascorbate is a cofactor for myrosinase and substitutes for the function of the catalytic base. *J Biol Chem* 275(50):39385-39393.

108. Burmeister WP*, et al.* (1997) The crystal structures of Sinapis alba myrosinase and a covalent glycosyl-enzyme intermediate provide insights into the substrate recognition and active-site machinery of an S-glycosidase. *Structure* 5(5):663-675.

109. Magrane M & Consortium U (2011) UniProt Knowledgebase: a hub of integrated protein data. *Database : the journal of biological databases and curation* 2011:bar009.

110. El Hassouni M, Henrissat B, Chippaux M, & Barras F (1992) Nucleotide sequences of the arb genes, which control beta-glucoside utilization in Erwinia chrysanthemi: comparison with the Escherichia coli bgl operon and evidence for a new beta-glycohydrolase family including enzymes from eubacteria, archeabacteria, and humans. *Journal of bacteriology* 174(3):765-777.

111. Gonzalez-Candelas L, Ramon D, & Polaina J (1990) Sequences and homology analysis of two genes encoding beta-glucosidases from Bacillus polymyxa. *Gene* 95(1):31-38.

112. Henrissat B (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. *The Biochemical journal* 280 ( Pt 2):309-316.

113. Henrissat B & Bairoch A (1993) New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *The Biochemical journal* 293 ( Pt 3):781-788.

114. Withers SG*, et al.* (1990) Unequivocal Demonstration of the Involvement of a Glutamate Residue as a Nucleophile in the Mechanism of a Retaining Glycosidase. *J Am Chem Soc* 112(15):5887-5889.

115. Cicek M (1999) Mechanism of Substrate Specificity and Catalysis in Retaining B-Glucosidases From Maize and Sorghum. Ph D. (Department of Biology, Virginia Tech (URN etd-093099-112234) ).

116. Jones DT, Taylor WR, & Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Computer applications in the biosciences : CABIOS* 8(3):275-282.

117. Jones E, Oliphant E, Peterson P, & al. e (2001) SciPy: Open Source Scientific Tools for Python.

118. Schwede T, Kopp J, Guex N, & Peitsch MC (2003) SWISS-MODEL: An automated protein homology-modeling server. *Nucleic acids research* 31(13):3381-3385.

119. Hopkins RJ, van Dam NM, & van Loon JJ (2009) Role of glucosinolates in insect-plant relationships and multitrophic interactions. *Annu Rev Entomol* 54:57-83.

120. Hegnauer R (1964) *Chemotaxonomie der Pflanzen* (Birkhaeuser Verlag, Basel and Stuttgart).

121. USDA Forest Products Laboratory FS (1979) *Extractives in Eastern Hardwoods - A review* (Madison, Wisconsin ).

122. Bernays EA & Chapman RF (1994) *Host-Plant Selection by Phytophagous Insects,* (Chapman & Hall, New York, NY. ).

123. Sarate PJ*, et al.* (2012) Developmental and digestive flexibilities in the midgut of a polyphagous pest, the cotton bollworm, Helicoverpa armigera. *J Insect Sci* 12.

124. Cates RG (1980) Feeding Patterns of Monophagous, Oligophagous, and Polyphagous Insect Herbivores - the Effect of Resource Abundance and Plant Chemistry. *Oecologia* 46(1):22-31.

125. Ryan CA (1990) Protease Inhibitors in Plants - Genes for Improving Defenses against Insects and Pathogens. *Annu Rev Phytopathol* 28:425-449.

126. Song HK & Suh SW (1998) Kunitz-type soybean trypsin inhibitor revisited: refined structure of its complex with porcine trypsin reveals an insight into the interaction between a homologous inhibitor from Erythrina caffra and tissue-type plasminogen activator. *J Mol Biol* 275(2):347-363.

127. Kwon TH*, et al.* (2000) A masquerade-like serine proteinase homologue is necessary for phenoloxidase activity in the coleopteran insect, Holotrichia diomphalia larvae. *Eur J Biochem* 267(20):6188-6196.

128. Kanost MR*, et al.* (1990) Insect Haemolymph Proteins. *Advances in Insect Physiology*, ed Wigglesworth PDEaVB (Academic Press), Vol 22, pp 299-396.

129. Broadway RM (1995) Are Insects Resistant to Plant Proteinase-Inhibitors. *J Insect Physiol* 41(2):107-116.

130. Davison A & Blaxter M (2005) Ancient origin of glycosyl hydrolase family 9 cellulase genes. *Mol Biol Evol* 22(5):1273-1284.

131. Danchin EGJ*, et al.* (2010) Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. *P Natl Acad Sci USA* 107(41):17651-17656.

132. Rybarczyk-Mydlowska K*, et al.* (2012) Rather than by direct acquisition via lateral gene transfer, GHF5 cellulases were passed on from early Pratylenchidae to root-knot and cyst nematodes. *Bmc Evol Biol* 12.

133. Todaka N*, et al.* (2010) Phylogenetic analysis of cellulolytic enzyme genes from representative lineages of termites and a related cockroach. *Plos One* 5(1):e8636.

134. Aspeborg H, Coutinho PM, Wang Y, Brumer H, & Henrissat B (2012) Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *Bmc Evol Biol* 12.

135. Ohtoko K*, et al.* (2000) Diverse genes of cellulase homologues of glycosyl hydrolase family 45 from the symbiotic protists in the hindgut of the termite Reticulitermes speratus. *Extremophiles* 4(6):343-349.

136. Pauchet Y, Wilkinson P, Chauhan R, & Ffrench-Constant RH (2010) Diversity of Beetle Genes Encoding Novel Plant Cell Wall Degrading Enzymes. *Plos One* 5(12).

137. Xu B, Hellman U, Ersson B, & Janson JC (2000) Purification, characterization and amino-acid sequence analysis of a thermostable, low molecular mass endo-beta-1,4-glucanase from blue mussel, Mytilus edulis. *Eur J Biochem* 267(16):4970-4977.

138. Kikuchi T*, et al.* (2011) Genomic Insights into the Origin of Parasitism in the Emerging Plant Pathogen Bursaphelenchus xylophilus. *Plos Pathog* 7(9).

139. Palomares-Rius JE*, et al.* (2014) Distribution and evolution of glycoside hydrolase family 45 cellulases in nematodes and fungi. *Bmc Evol Biol* 14.

140. Steenbakkers PJ*, et al.* (2002) The major component of the cellulosomes of anaerobic fungi from the genus Piromyces is a family 48 glycoside hydrolase. *DNA sequence : the journal of DNA sequencing and mapping* 13(6):313-320.

141. Wang TY*, et al.* (2011) Functional characterization of cellulases identified from the cow rumen fungus Neocallimastix patriciarum W5 by transcriptomic and secretomic analyses. *Biotechnology for biofuels* 4:24.

142. Fujita K, Shimomura K, Yamamoto K, Yamashita T, & Suzuki K (2006) A chitinase structurally related to the glycoside hydrolase family 48 is indispensable for the hormonally induced diapause termination in a beetle. *Biochem Bioph Res Co* 345(1):502-507.

143. Kirsch R*, et al.* (2014) Horizontal gene transfer and functional diversification of plant cell wall degrading polygalacturonases: Key events in the evolution of herbivory in beetles. *Insect biochemistry and molecular biology* 52:33-50.

144. Jones P*, et al.* (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30(9):1236-1240.

145. Leal WS (2013) Odorant Reception in Insects: Roles of Receptors, Binding Proteins, and Degrading Enzymes. *Annual Review of Entomology, Vol 58* 58:373-391.

146. Sánchez-Gracia A, Vieira FG, & Rozas J (2009) Molecular evolution of the major chemosensory gene families in insects. *Heredity* 103(3):208-216.

147. Mitchell RF*, et al.* (2012) Sequencing and characterizing odorant receptors of the cerambycid beetle Megacyllene caryae. *Insect biochemistry and molecular biology* 42(7):499-505.

148. Fêret S & Maleszka R (2006) Function and evolution of a gene family encoding odorant binding-like proteins in a social insect, the honey bee (Apis mellifera). *Genome research* 16(11):1404-1413.

149. Xu PX, Zwiebel LJ, & Smith DP (2003) Identification of a distinct family of genes encoding atypical odorant-binding proteins in the malaria vector mosquito, Anopheles gambiae. *Insect molecular biology* 12(6):549-560.

150. Andersson MN*, et al.* (2013) Antennal transcriptome analysis of the chemosensory gene families in the tree killing bark beetles, Ips typographus and Dendroctonus ponderosae (Coleoptera: Curculionidae: Scolytinae). *Bmc Genomics* 14.

151. Hunt T*, et al.* (2007) A comprehensive phylogeny of beetles reveals the evolutionary origins of a superradiation. *Science* 318(5858):1913-1916.

152. Missbach C*, et al.* (2014) Evolution of insect olfactory receptors. *Elife* 3.

153. Nolte A*, et al.* (2013) In situ Tip-Recordings Found No Evidence for an Orco-Based Ionotropic Mechanism of Pheromone-Transduction in Manduca sexta. *Plos One* 8(5).

154. Sato K*, et al.* (2008) Insect olfactory receptors are heteromeric ligand-gated ion channels. *Nature* 452(7190):1002-U1009.

155. Wicher D*, et al.* (2008) Drosophila odorant receptors are both ligand-gated and cyclic-nucleotide-activated cation channels. *Nature* 452(7190):1007-U1010.

156. Kirkness EF*, et al.* (2010) Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *P Natl Acad Sci USA* 107(27):12168-12173.

157. Robertson HM, Warr CG, & Carlson JR (2003) Molecular evolution of the insect chemoreceptor gene superfamily in Drosophila melanogaster. *P Natl Acad Sci USA* 100:14537-14542.

158. Zhou X*, et al.* (2012) Phylogenetic and transcriptomic analysis of chemosensory receptors in a pair of divergent ant species reveals sex-specific signatures of odor coding. *Plos Genet* 8(8):e1002930.

159. Zhang A*, et al.* (2003) Evidence for contact sex recognition pheromone of the Asian longhorned beetle, Anoplophora glabripennis (Coleoptera: Cerambycidae). *Die Naturwissenschaften* 90(9):410-413.

160. Koh TW*, et al.* (2014) The Drosophila IR20a clade of ionotropic receptors are candidate taste and pheromone receptors. *Neuron* 83(4):850-865.

161. Miyazono K, Kamiya Y, & Morikawa M (2010) Bone morphogenetic protein receptors and signal transduction. *Journal of biochemistry* 147(1):35-51.

162. Van der Zee M, da Fonseca RN, & Roth S (2008) TGFbeta signaling in Tribolium: vertebrate-like components in a beetle. *Dev Genes Evol* 218(3-4):203-213.

163. Özüak O, Buchta T, Roth S, & Lynch JA (2014) Dorsoventral polarity of the Nasonia embryo primarily relies on a BMP gradient formed without input from Toll. *Current biology : CB* 24(20):2393-2398.

164. Richards S*, et al.* (2008) The genome of the model beetle and pest Tribolium castaneum. *Nature* 452(7190):949-955.

165. Rongo C & Lehmann R (1996) Regulated synthesis, transport and assembly of the Drosophila germ plasm. *Trends Genet* 12(3):102-109.

166. Lynch JA*, et al.* (2011) The Phylogenetic Origin of oskar Coincided with the Origin of Maternally Provisioned Germ Plasm and Pole Cells at the Base of the Holometabola. *Plos Genet* 7(4).

167. Schmitt-Engel C, Cerny AC, & Schoppmeier M (2012) A dual role for nanos and pumilio in anterior and posterior blastodermal patterning of the short-germ beetle Tribolium castaneum. *Developmental biology* 364(2):224-235.

168. Krumlauf R (1992) Evolution of the Vertebrate Hox Homeobox Genes. *Bioessays* 14(4):245-252.

169. Negre B & Ruiz A (2007) HOM-C evolution in Drosophila: is there a need for Hox gene clustering? *Trends Genet* 23(2):55-59.

170. Panfilio KA, Liu PZ, Akam M, & Kaufman TC (2006) Oncopeltus fasciatus zen is essential for serosal tissue function in katatrepsis. *Developmental biology* 292(1):226-243.

171. Panfilio KA & Akam M (2007) A comparison of Hox3 and Zen protein coding sequences in taxa that span the Hox3/zen divergence. *Dev Genes Evol* 217(4):323-329.

172. Brown SJ*, et al.* (2002) Sequence of the Tribolium castaneum Homeotic complex: The region corresponding to the Drosophila melanogaster Antennapedia complex. *Genetics* 160(3):1067-1074.

173. GómezSkarmeta JL, Diez del Corral R, delaCalleMustienes E, FerresMarco D, & Modolell J (1996) araucan and caupolican, two members of the novel iroquois complex, encode homeoproteins that control proneural and vein-forming genes. *Cell* 85(1):95-105.

174. Kehl BT, Cho KO, & Choi KW (1998) Mirror, a Drosophila homeobox gene in the iroquois complex, is required for sensory organ and alula formation. *Development* 125(7):1217-1227.

175. McNeill H, Yang CH, Brodsky M, Ungos J, & Simon MA (1997) Mirror encodes a novel PBX-class homeoprotein that functions in the definition of the dorsal-ventral border in the Drosophila eye. *Gene Dev* 11(8):1073-1082.

176. Cavodeassi F, Modolell J, & Gomez-Skarmeta JL (2001) The Iroquois family of genes: from body building to neural patterning. *Development* 128(15):2847-2855.

177. Maeso I*, et al.* (2012) An ancient genomic regulatory block conserved across bilaterians and its dismantling in tetrapods by retrogene replacement. *Genome research* 22(4):642-655.

178. Peters T, Dildrop R, Ausmeier K, & Ruther U (2000) Organization of mouse Iroquois homeobox genes in two clusters suggests a conserved regulation and function in vertebrate development. *Genome research* 10(10):1453-1462.

179. Lynch JA & Roth S (2011) The evolution of dorsal-ventral patterning mechanisms in insects. *Gene Dev* 25(2):107-118.

180. Nieto MA (2002) The snail superfamily of zinc-finger transcription factors. *Nat Rev Mol Cell Bio* 3(3):155-166.

181. Nambu JR, Lewis JO, Wharton KA, & Crews ST (1991) The Drosophila Single-Minded Gene Encodes a Helix-Loop-Helix Protein That Acts as a Master Regulator of Cns Midline Development. *Cell* 67(6):1157-1167.

182. Von Ohlen T & Doe CQ (2000) Convergence of dorsal, Dpp, and Egfr signaling pathways subdivides the Drosophila neuroectoderm into three dorsal-ventral columns. *Developmental biology* 224(2):362-372.

183. Calleja M*, et al.* (2000) Generation of medial and lateral dorsal body domains by the pannier gene of Drosophila. *Development* 127(18):3971-3980.

184. Nunes da Fonseca R*, et al.* (2008) Self-regulatory circuits in dorsoventral axis formation of the short-germ beetle Tribolium castaneum. *Developmental cell* 14(4):605-615.

185. Van der Zee M, Stockhammer O, Von Levetzow C, Da Fonseca RN, & Roth S (2006) Sog/Chordin is required for ventral-to-dorsal Dpp/BMP transport and head formation in a short germ insect. *P Natl Acad Sci USA* 103(44):16307-16312.

186. Stathopoulos A, Van Drenth M, Erives A, Markstein M, & Levine M (2002) Whole-genome analysis of dorsal-ventral patterning in the Drosophila embryo. *Cell* 111(5):687-701.

187. Lynch JA, Peel AD, Drechsler A, Averof M, & Roth S (2010) EGF Signaling and the Origin of Axial Polarity among the Insects. *Current Biology* 20(11):1042-1047.

188. Evans RM (2005) The nuclear receptor superfamily: A Rosetta Stone for physiology. *Mol Endocrinol* 19(6):1429-1438.

189. Mangelsdorf DJ*, et al.* (1995) The Nuclear Receptor Superfamily - the 2nd Decade. *Cell* 83(6):835-839.

190. Auwerx J*, et al.* (1999) A unified nomenclature system for the nuclear receptor superfamily. *Cell* 97(2):161-163.

191. Gronemeyer H, Gustafsson JA, & Laudet V (2004) Principles for modulation of the nuclear receptor superfamily. *Nat Rev Drug Discov* 3(11):950-964.

192. Moore JT, Collins JL, & Pearce KH (2006) The nuclear receptor superfamily and drug discovery. *Chemmedchem* 1(5):504-+.

193. Ribeiro RCJ, Kushner PJ, & Baxter JD (1995) The Nuclear Hormone-Receptor Gene Superfamily. *Annu Rev Med* 46:443-453.

194. Palli SR, Hormann RE, Schlattner U, & Lezzi M (2005) Ecdysteroid receptors and their applications in agriculture and medicine. *Vitam Horm* 73:59-+.

195. Enmark E & Gustafsson JA (2000) Nematode genome sequence dramatically extends the nuclear receptor superfamily. *Trends Pharmacol Sci* 21(3):85-87.

196. King-Jones K & Thummel CS (2005) Nuclear receptors - A perspective from Drosophila. *Nature Reviews Genetics* 6(4):311-323.

197. Tan A & Palli SR (2008) Identification and characterization of nuclear receptors from the red flour beetle, Tribolium castaneum. *Insect biochemistry and molecular biology* 38(4):430-439.

198. Jacobs CGC, Spaink HP, & van der Zee M (2014) The extraembryonic serosa is a frontier epithelium providing the insect egg with a full-range innate immune response. *Elife* 3.

199. LeMosy EK, Kemler D, & Hashimoto C (1998) Role of Nudel protease activation in triggering dorsoventral polarization of the Drosophila embryo. *Development* 125(20):4045-4053.

200. Sen J, Goltz JS, Stevens L, & Stein D (1998) Spatially restricted expression of pipe in the Drosophila egg chamber defines embryonic dorsal-ventral polarity. *Cell* 95(4):471-481.

201. Tomoyasu Y & Denell RE (2004) Larval RNAi in Tribolium (Coleoptera) for analyzing adult development. *Dev Genes Evol* 214(11):575-578.

202. Miller SC, Miyata K, Brown SJ, & Tomoyasu Y (2012) Dissecting Systemic RNA Interference in the Red Flour Beetle Tribolium castaneum: Parameters Affecting the Efficiency of RNAi. *Plos One* 7(10).

203. Bellés X (2010) Beyond Drosophila: RNAi in vivo and functional genomics in insects. *Annu Rev Entomol* 55:111-128.

204. Horn T, Sandmann T, & Boutros M (2010) Design and evaluation of genome-wide libraries for RNA interference screens. *Genome biology* 11(6):R61.

205. Kamath RS & Ahringer J (2003) Genome-wide RNAi screening in Caenorhabditis elegans. *Methods* 30(4):313-321.

206. Rual JF*, et al.* (2004) Toward improving Caenorhabditis elegans phenome mapping with an ORFeome-based RNAi library. *Genome research* 14(10B):2162-2168.

207. Young M, Beeman RW, & Arakane Y (2012) RNAi-based functional genomics in Tribolium castaneum and possible application for controlling insect pests. *Entomol Res* 42(1):1-10.

208. Li XX, Zhang MY, & Zhang HY (2011) RNA Interference of Four Genes in Adult Bactrocera dorsalis by Feeding Their dsRNAs. *Plos One* 6(3).

209. Tian HG*, et al.* (2009) Developmental Control of a Lepidopteran Pest Spodoptera exigua by Ingestion of Bacteria Expressing dsRNA of a Non-Midgut Gene. *Plos One* 4(7).

210. Walshe DP, Lehane SM, Lehane MJ, & Haines LR (2009) Prolonged gene knockdown in the tsetse fly Glossina by feeding double stranded RNA. *Insect molecular biology* 18(1):11-19.

211. Alves AP, Lorenzen MD, Beeman RW, Foster JE, & Siegfried BD (2010) RNA interference as a method for target-site screening in the western corn rootworm, Diabrotica virgifera virgifera. *J Insect Sci* 10.

212. Mao YB*, et al.* (2007) Silencing a cotton bollworm P450 monooxygenase gene by plant-mediated RNAi impairs larval tolerance of gossypol. *Nat Biotechnol* 25(11):1307-1313.

213. Timmons L, Court DL, & Fire A (2001) Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in Caenorhabditis elegans. *Gene* 263(1-2):103-112.

214. Miyata K*, et al.* (2014) Establishing an In Vivo Assay System to Identify Components Involved in Environmental RNA Interference in the Western Corn Rootworm. *Plos One* 9(7).

215. Willis JH (2010) Structural cuticular proteins from arthropods: annotation, nomenclature, and sequence characteristics in the genomics era. *Insect biochemistry and molecular biology* 40(3):189-204.

216. Ioannidou ZS, Theodoropoulou MC, Papandreou NC, Willis JH, & Hamodrakas SJ (2014) CutProtFam-Pred: detection and classification of putative structural cuticular proteins from sequence alone, based on profile hidden Markov models. *Insect biochemistry and molecular biology* 52:51-59.

217. Arakane Y*, et al.* (2004) Characterization of two chitin synthase genes of the red flour beetle, Tribolium castaneum, and alternate exon usage in one of the genes during development. *Insect biochemistry and molecular biology* 34(3):291-304.

218. Bird AP (1980) DNA Methylation and the Frequency of Cpg in Animal DNA. *Nucleic acids research* 8(7):1499-1504.

219. Elango N, Hunt BG, Goodisman MAD, & Yi SV (2009) DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, Apis mellifera. *P Natl Acad Sci USA* 106(27):11206-11211.

220. Glastad KM, Hunt BG, Yi SV, & Goodisman MA (2011) DNA methylation in insects: on the brink of the epigenomic era. *Insect molecular biology* 20(5):553-565.

221. Simola DF*, et al.* (2013) Social insect genomes exhibit dramatic evolution in gene composition and regulation while preserving regulatory features linked to sociality. *Genome research* 23(8):1235-1247.

222. Yi SV & Goodisman MAD (2009) Computational approaches for understanding the evolution of DNA methylation in animals. *Epigenetics* 4(8):551-556.

223. Wang Y*, et al.* (2006) Functional CpG methylation system in a social insect. *Science* 314(5799):645-647.

224. Kent CF, Minaei S, Harpur BA, & Zayed A (2012) Recombination is associated with the evolution of genome structure and worker behavior in honey bees. *P Natl Acad Sci USA* 109(44):18012-18017.

225. Rae PMM & Steele RE (1979) Absence of Cytosine Methylation at C-C-G-G and G-C-G-C Sites in the Rdna Coding Regions and Intervening Sequences of Drosophila and the Rdna of Other Higher Insects. *Nucleic acids research* 6(9):2987-2995.

226. Zemach A, McDaniel IE, Silva P, & Zilberman D (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 328(5980):916-919.

227. Hunt BG, Glastad KM, Yi SV, & Goodisman MA (2013) The function of intragenic DNA methylation: insights from insect epigenomes. *Integrative and comparative biology* 53(2):319-328.

228. Denlinger DL, Yocum GD, & Rinehart JP (2012) Hormonal control of diapause. *Insect Endocrinology*, ed Gilbert LI (Elsevier), pp 430-463.

229. Taylor F (1986) Seasonal Adaptations of Insects - Tauber,Mj, Tauber,Ca, Masaki,S. *Science* 232(4754):1152-1152.

230. Rinehart JP*, et al.* (2007) Up-regulation of heat shock proteins is essential for cold survival during insect diapause. *Proc Natl Acad Sci U S A* 104(27):11130-11137.

231. Lee RE & Denlinger DL (1985) Ontogeny of Cold-Hardiness in the Flesh Fly. *Cryobiology* 22(6):632-632.

232. Sim C & Denlinger DL (2013) Insulin signaling and the regulation of insect diapause. *Front Physiol* 4.

233. Riehle MA & Brown MR (1999) Insulin stimulates ecdysteroid production through a conserved signaling cascade in the mosquito Aedes aegypti. *Insect biochemistry and molecular biology* 29(10):855-860.

234. Nijhout HF (2003) The control of body size in insects. *Developmental biology* 261(1):1-9.

235. Brown MR*, et al.* (2008) An insulin-like peptide regulates egg maturation and metabolism in the mosquito Aedes aegypti. *P Natl Acad Sci USA* 105(15):5716-5721.

236. Antonova Y, Arik AJ, Moore W, Riehle MR, & Brown MR (2012) Insulin-like peptides: Structure, Signaling, and Function. *Insect Endocrinology*, ed Gilbert LI (Elsevier), pp 63-92.

237. Feder ME & Hofmann GE (1999) Heat-shock proteins, molecular chaperones, and the stress response: Evolutionary and ecological physiology. *Annu Rev Physiol* 61:243-282.

238. Lindquist S (1986) The heat-shock response. *Annual review of biochemistry* 55:1151-1191.

239. Denlinger DL (2002) Regulation of diapause. *Annu Rev Entomol* 47:93-122.

240. MacRae TH (2010) Gene expression, metabolic regulation and stress tolerance during diapause. *Cellular and molecular life sciences : CMLS* 67(14):2405-2424.

241. Colinet H, Lee SF, & Hoffmann A (2010) Functional Characterization of the Frost Gene in Drosophila melanogaster: Importance for Recovery from Chill Coma. *Plos One* 5(6).

242. Goto SG (2001) A novel gene that is up-regulated during recovery from cold shock in Drosophila melanogaster. *Gene* 270(1-2):259-264.

243. Liou YC, Thibault P, Walker VK, Davies PL, & Graham LA (1999) A complex family of highly heterogeneous and internally repetitive hyperactive antifreeze proteins from the beetle Tenebrio molitor. *Biochemistry* 38(35):11415-11424.

244. Zachariassen KE & Husby JA (1982) Antifreeze Effect of Thermal Hysteresis Agents Protects Highly Supercooled Insects. *Nature* 298(5877):865-867.