
This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of March 30, 2011):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/329/5995/1068.full.html>

Supporting Online Material can be found at:

<http://www.sciencemag.org/content/suppl/2010/08/25/329.5995.1068.DC1.html>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

<http://www.sciencemag.org/content/329/5995/1068.full.html#related>

This article **cites 20 articles**, 4 of which can be accessed free:

<http://www.sciencemag.org/content/329/5995/1068.full.html#ref-list-1>

This article has been **cited by** 1 article(s) on the ISI Web of Science

This article has been **cited by** 4 articles hosted by HighWire Press; see:

<http://www.sciencemag.org/content/329/5995/1068.full.html#related-urls>

This article appears in the following **subject collections**:

Genetics

<http://www.sciencemag.org/cgi/collection/genetics>

Genomic Comparison of the Ants *Camponotus floridanus* and *Harpegnathos saltator*

Roberto Bonasio,^{1*} Guojie Zhang,^{2,3*} Chaoyang Ye,^{4*} Navdeep S. Mutti,^{5*} Xiaodong Fang,^{3*} Nan Qin,^{3*} Greg Donahue,⁴ Pengcheng Yang,³ Qiye Li,³ Cai Li,³ Pei Zhang,³ Zhiyong Huang,³ Shelley L. Berger,^{4†} Danny Reinberg,^{1,6†} Jun Wang,^{3,7†} Jürgen Liebig^{5†}

The organized societies of ants include short-lived worker castes displaying specialized behavior and morphology and long-lived queens dedicated to reproduction. We sequenced and compared the genomes of two socially divergent ant species: *Camponotus floridanus* and *Harpegnathos saltator*. Both genomes contained high amounts of CpG, despite the presence of DNA methylation, which in non-Hymenoptera correlates with CpG depletion. Comparison of gene expression in different castes identified up-regulation of telomerase and sirtuin deacetylases in longer-lived *H. saltator* reproductives, caste-specific expression of microRNAs and SMYD histone methyltransferases, and differential regulation of genes implicated in neuronal function and chemical communication. Our findings provide clues on the molecular differences between castes in these two ants and establish a new experimental model to study epigenetics in aging and behavior.

As eusocial insects, ants live in populous colonies in which up to millions of individuals delegate the reproductive role to one or few queens, while nonreproductive workers carry out all tasks required for colony maintenance (1). These mutually exclusive morphologies and behaviors arise from a single genome and are typically dictated not by genetic

differences, but by environmental factors (2). The first fertilized (diploid) eggs laid by a founder queen develop into workers, but as the colony enlarges, some diploid embryos take a different developmental path to become virgin queens, which leave the nest, mate, and establish new colonies. As colonies mature, queens transition from a broad behavioral repertoire that allows them to forage, excavate nests, and rear offspring, to one restricted to egg-laying and total dependence on workers. Queens also live up to 10 times longer than workers and 500 times longer than males (3).

We compared the genomes of the ants *Camponotus floridanus* and *Harpegnathos saltator*, because of contrasts in their behavioral flexibility, caste specialization, and social organization. *C. floridanus* lives in large organized colonies, in which only the queen lays fertilized eggs; when the queen dies, so does the colony (1). Non-reproductive individuals belong to two separate castes, major and minor workers, which exhibit differences in morphology and behavior established during development purely on environmental grounds. In contrast, the *H. saltator* social

system and division of labor are more basal: dimorphism between queens and workers is limited, and when the queen dies she is replaced by workers that become functional queens, called gamergates (4).

These two ant species differ in other respects as well. *C. floridanus* are scavengers, forage diurnally and nocturnally, and lay pheromone trails that mark paths to food sources. *H. saltator* workers prey on small arthropods in a solitary and diurnal fashion. *C. floridanus* exhibits high territoriality, strong nestmate recognition, and elaborate task specialization. In contrast, *H. saltator* displays low territoriality, loses nestmate recognition in the laboratory, and has only basic task specialization.

The Illumina Genome Analyzer platform was used to sequence genomic libraries for *C. floridanus* and *H. saltator*, obtaining more than 100-fold coverage. Draft genomic assemblies reached scaffold N50 size of ~600 kb (table S1), although for *C. floridanus* most genome-wide analyses reported here were conducted on an earlier version (v3), with scaffold N50 size of 444 kb (table S1). Assembly resulted in only small gaps and large N50 size, which assured us that most genomic features, particularly gene models, were predicted with reasonable accuracy. We verified the assemblies by sequencing 9 (*C. floridanus*) and 10 (*H. saltator*) randomly selected fosmid inserts (average size, 37 kb) (table S2). Additionally, we sequenced ~5000 expressed sequence tags from each ant and mapped them to the assembled scaffolds; more than 95% matched the assemblies (table S3).

The *C. floridanus* and *H. saltator* assemblies cover more than 90% of the genomes, which we estimate at 240 and 330 Mb in size, respectively (fig. S1 and table S4). The *H. saltator* assembly contains 45% G+C, similar to *Drosophila melanogaster* (42%), and *Nasonia vitripennis* (42%), whereas the *C. floridanus* genome is A+T rich, with a G+C frequency of 34%, similar to *Apis mellifera* (33%) (5) (table S4). Organisms that use DNA methylation for gene regulation typically display a depletion of CpG dinucleotides in their genome (table S5); however, CpG dinucleotides are overrepresented in both ant genomes

¹Department of Biochemistry, New York University School of Medicine, 522 First Avenue, New York, NY 10016, USA.

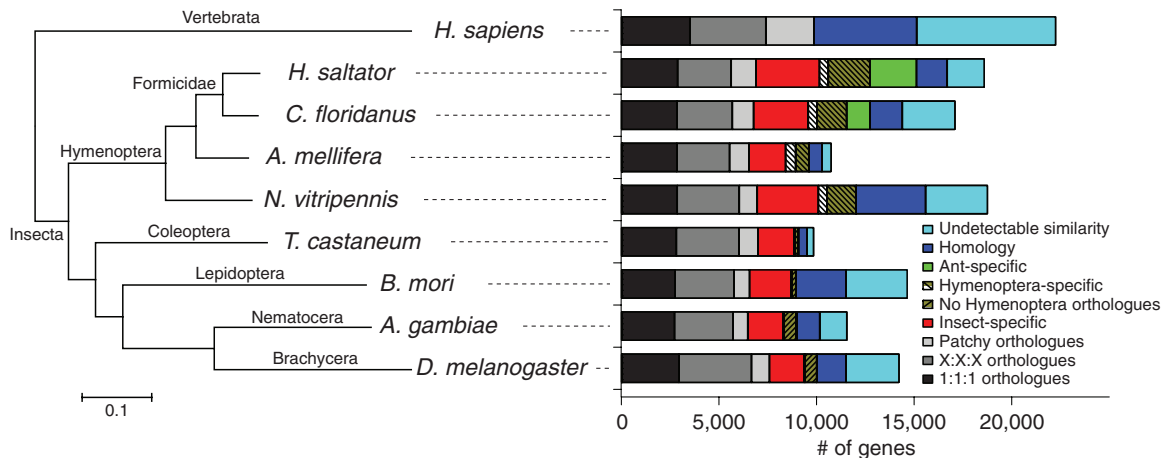
²Chinese Academy of Sciences–Max Planck Junior Research Group, State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan 650223, China. ³Beijing Genomics Institute–Shenzhen, Shenzhen 518083, China. ⁴Department of Cell and Developmental Biology, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA. ⁵School of Life Sciences, Arizona State University, Tempe, AZ 85287, USA.

⁶Howard Hughes Medical Institute, New York University Medical School, New York, NY 10016, USA. ⁷Department of Biology, University of Copenhagen, Copenhagen DK-2200, Denmark.

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: berger@mail.med.upenn.edu (S.L.B.); danny.reinberg@nyumc.org (D.R.); wangj@genomics.cn (J.W.); juergen.liebig@asu.edu (J.L.)

Fig. 1. Ant proteome. Phylogenetic tree based on maximum likelihood analyses of a concatenated alignment of single-copy proteins (left), and orthology relationships in multiple insects (right), using *Homo sapiens* as outgroup. The scale bar indicates 0.1 substitution per site. *T. castaneum*, *Tribolium castaneum*; *B. mori*, *Bombyx mori*; *A. gambiae*, *Anopheles gambiae*.



(table S4 and fig. S2), despite the presence of cytosine methylation. CpG dinucleotides are also overrepresented in *A. mellifera* (5) and in *N. vitripennis*; thus, this genomic feature may be specific to the Hymenoptera.

Repetitive elements constitute 15% of the assemblies in *C. floridanus* and 27% in *H. saltator* (table S4 and fig. S3), which is probably an underestimate, because genome regions that cannot be assembled are enriched in repeats. Both genomes contain copies of the *piggyBac* transposon [234 in total and 6 with intact open reading frames (ORFs) in *C. floridanus*, 976 in total and 23 with intact ORFs in *H. saltator*], which has been used for insect transgenesis (6, 7) and may also prove useful in ants.

Segmental duplications (SDs) account for 9.6% and 14.8% of the *C. floridanus* and *H. saltator* genomes, respectively (table S6). The 1810 (*C. floridanus*) and 2858 (*H. saltator*) gene models found in SDs are enriched in Gene Ontology (GO) terms associated with chemodetection, olfactory function, and protease/peptidase activity (table S7). *C. floridanus* SDs are also enriched for genes belonging to the cytochrome P450 family—perhaps as a result of detoxifying needs related to their generalist feeder life-style.

The predicted proteomes of *C. floridanus* and *H. saltator* share most protein families with *A. mellifera* and *N. vitripennis*, but also contain a large number of ant-specific (690) and species-specific families (3230 in *C. floridanus* and 2617 in *H. saltator*) without homologs in the other two Hymenoptera (fig. S4). When more insect ge-

nomes are included in the comparison, one-third of the ant protein-coding genes are conserved with vertebrates and two-thirds are conserved with insects, leaving 506 ant-specific protein families and 2000 to 3000 species-specific families (Fig. 1). Ant-specific genes were enriched not only in GO terms “olfactory receptor activity,” “sensory perception of smell,” “G protein-coupled receptor activity,” and “odorant binding,” but also in terms related to detoxification, including “monooxygenase activity” and “heme binding” (table S8).

We annotated 96 microRNA (miRNA) genes in *C. floridanus* and 159 in *H. saltator*. These accounted for most of the small RNA reads in adult specimens, with the exception of *H. saltator* gamergates, which showed a larger proportion of small RNAs mapping to unannotated regions of the genome (Fig. 2A and table S9). Gamergates also expressed the most diverse miRNA repertoire (table S10). The two *C. floridanus* worker castes displayed differential expression of miRNAs, with *cflo-mir-64* up-regulated in minor workers (Fig. 2B) and *cflo-mir-7* in major workers (fig. S5). This suggests that, in addition to being developmentally regulated (Fig. 2, C and D), some miRNAs might contribute to the differences among ant castes.

We examined the ant genomes and transcriptomes for signatures related to aging. Telomere shortening is a hallmark of cellular senescence in multicellular eukaryotes, and the enzyme telomerase (*TERT*), which counteracts telomere shortening, prolongs life span upon overexpression (8). *TERT* RNA levels were highest in eggs and

lower in adults in both *C. floridanus* and *H. saltator*, but they were up-regulated in *H. saltator* gamergates (Fig. 3A). This may be explained by the gamergates acquiring many physiological characteristics of queens, including longer life span (9). Aging has also been linked to the sirtuin lysine deacetylases enzymes SIRT1 and SIRT6, homologous to the *Saccharomyces cerevisiae* Sir2p implicated in replicative senescence (10). In *H. saltator* gamergates, both of these genes are expressed at higher levels compared to workers (Fig. 3B). These results suggest that the regulation of life span in gamergates may share common mechanisms with other organisms.

The caste system in ant societies in general, and the social flexibility of *H. saltator* in particular, allow us to study the role of epigenetics in behavior, aging, and development. Here, we use the term “epigenetics” as the ensemble of molecular pathways that select genomic regions (chromatin domains) for activation or repression, transmit these states through cell division, and stabilize them in differentiated cell types. These mechanisms include DNA methylation, histone posttranslational modifications, and trans-acting mechanisms involving noncoding RNAs (11).

Unlike in *Drosophila*, DNA methylation pathways in *A. mellifera* and *N. vitripennis* are similar to those of mammalian systems (5, 12). In the two ant species, we found single copies of cytosine methyltransferases *DNMT1* and *DNMT3B* genes and the noncatalytic *DNMT3L*, as well as four genes containing methyl-CpG binding domains (table S11). Consistent with these and previous observations (13), we detected the presence of 5-methylcytosine in the ants genomic DNA. *H. saltator*, which shows more ancestral characteristics, had less DNA methylation than its more derived relative, *C. floridanus* (Fig. 4A).

Histone acetyltransferases (HATs) and deacetylases (HDACs) add and remove acetyl groups on histone tails, regulate genes through chromatin structure (14), and are linked to the aging process (10). We identified 26 putative HATs in *C. floridanus* and 27 in *H. saltator*, with an expansion of GCNLI homologs (table S12). Humans have four classes of HDACs, comprising HDAC1–11 and the NAD⁺-dependent sirtuin family proteins (SIRT1–7). In contrast, only four HDACs and six sirtuins are found in *A. mellifera* (5). We identified five HDACs in *C. floridanus* and four in *H. saltator*, as well as six sirtuins in both ant species (table S11). Compared with mammals, these two ants and the honeybee lack SIRT3 (10).

Histone methylation is suspected to participate in epigenetic processes (11). We identified 27 proteins containing the conserved SET histone methyltransferase domain in *C. floridanus* and 22 in *H. saltator* (table S13). All major SET families are conserved in ants, including Polycomb/trithorax group, NSDs, SMYDs, and SETDs, and the arginine methyltransferase family. The SMYD family appears to have undergone multiple duplication events during the evolution of insects (fig. S6). We found five homologs in *C.*

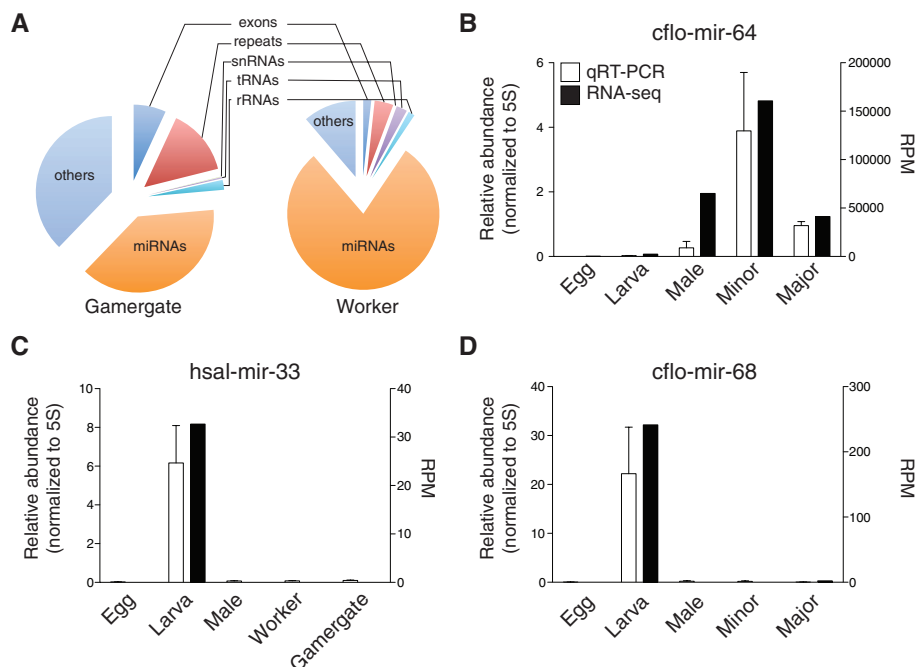


Fig. 2. Small RNA-seq and miRNA. (A) Classification of small RNA-seq reads from *H. saltator* gamergates (left) or workers (right). snRNAs, small nuclear RNAs; rRNAs, ribosomal RNAs. (B to D) Caste- and stage-specific expression of miRNAs in *C. floridanus* [(B) and (D)] and *H. saltator* (C). White bars represent quantitative reverse transcription–polymerase chain reaction (qRT-PCR) and indicate the mean $\pm$ SEM, $n = 7$ biological replicates. Black bars are scaled on the right y axis and indicate reads per million (RPM).

floridanus and six in *H. saltator* of *SMYD4*, which is involved in muscle development and breast cancer (15, 16). *SMYD* family members are differentially regulated among different ant castes and developmental stages (fig. S7); among them, a *SMYD3* homolog (Hsal_14941) is up-

regulated in gamergates (Fig. 4B), whereas a *SMYD4* homolog (Hsal_08142) is up-regulated in nonreproductive workers (Fig. 4C).

The elaborate social organization in ant colonies is attributed to a sophisticated communication system, made up of chemical signals that elicit behavioral responses via the nervous system (1). Many enriched GO terms associated with differentially expressed genes in *C. floridanus* (table S14) and *H. saltator* (table S15) castes are related to neuronal function and chemical communication. In *C. floridanus* major and minor workers, we detected differences in the expression levels of genes associated with GO terms including “postsynaptic membrane,” “ligand-gated channel activity,” “sensory perception of smell” (table S14), and “neurotransmitter binding” (Fig. 5A). This suggests that the different behaviors exhibited by distinct workers castes may in part be encoded in the brain at the transcriptional level.

Olfactory receptors (ORs) are G protein-coupled receptors (GPCRs) that function in behavioral responses and chemical communication in animals (17). We found 139 genes containing the *Drosophila* olfactory receptor domain (IPR004117) in *C. floridanus* and 105 in *H. saltator* (tables S16 and 17). Similar to the *A. mellifera* genome, the ant genomes contain fewer gustatory receptors and odorant binding proteins than *Drosophila* (tables S16 and 17). The number of identified ORs is much smaller than the number of glomeruli observed in the antennal lobe for both ant species (18, 19), apparently contradicting findings in other insects, which showed that most ORs are individually represented in the brain by a dedicated glomerulus (20). However, our homology-based approach might be insufficient to detect all ant-specific ORs.

In ants, cuticular hydrocarbons play a key role in nestmate recognition and regulation of reproduction (21). GO terms associated with the metabolism of hydrocarbons were enriched in ant-specific protein families (table S8) and in protein families expanded in ants (table S18 and fig. S8). *C. floridanus* and *H. saltator* have 19 and 14 genes, respectively, containing a beta-ketoacyl synthase domain (IPR014030), whereas *A. mellifera* has 3 and *D. melanogaster* has 4. Genes with putative roles in hydrocarbon metabolism show altered expression levels across castes and developmental stages (fig. S9). Compared to nonreproductive workers, *H. saltator* gamergates up-regulated several hydrocarbon metabolism genes, including homologs of a fatty acid synthase (*FASN*), an acetyl-coenzyme A desaturase (*SCD*), and an elongase of very long fatty acid (*ELOVL4*) (Fig. 5B). *H. saltator* gamergates advertise their reproductive status to nestmates via long cuticular hydrocarbons (22), and we found that expression of most genes with homology to human *ELOV* genes was up-regulated in gamergates (fig. S10). Because of the complexity of the metabolic pathways involved, we cannot conclude that transcriptional regulation of these genes is directly linked to communication; however, these observations suggest a molecular basis for

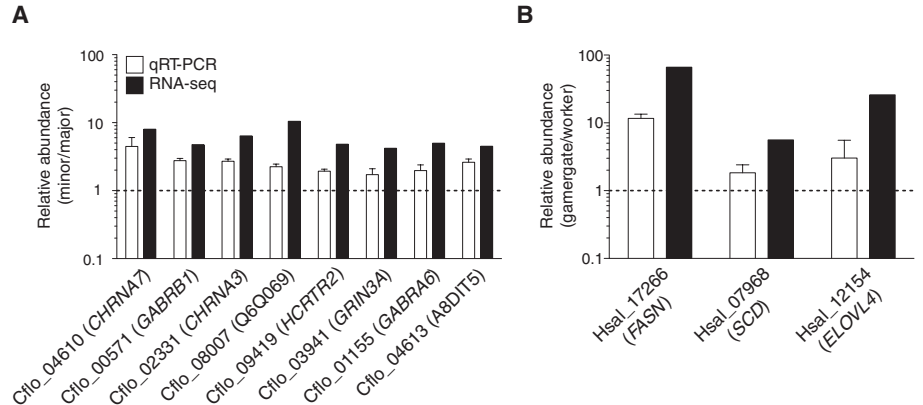
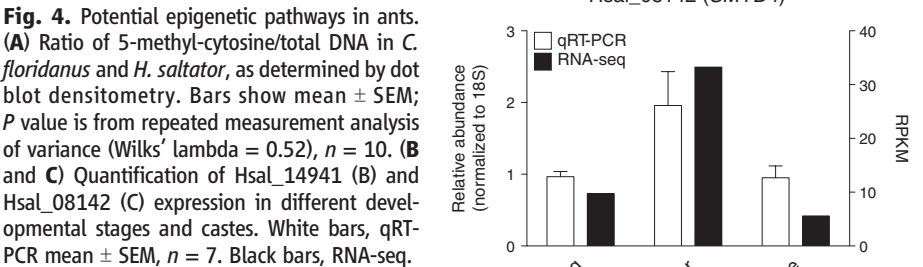
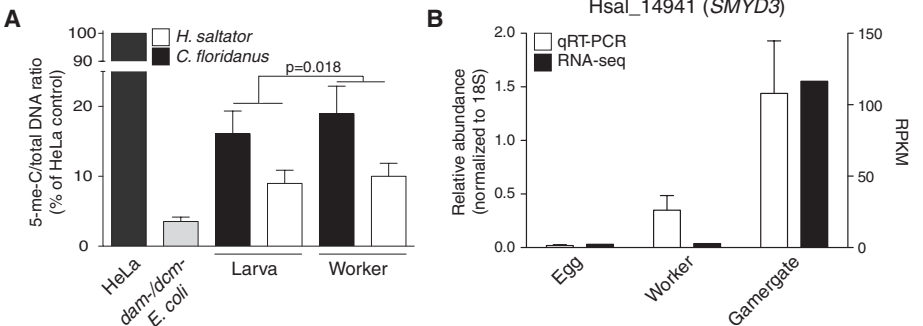
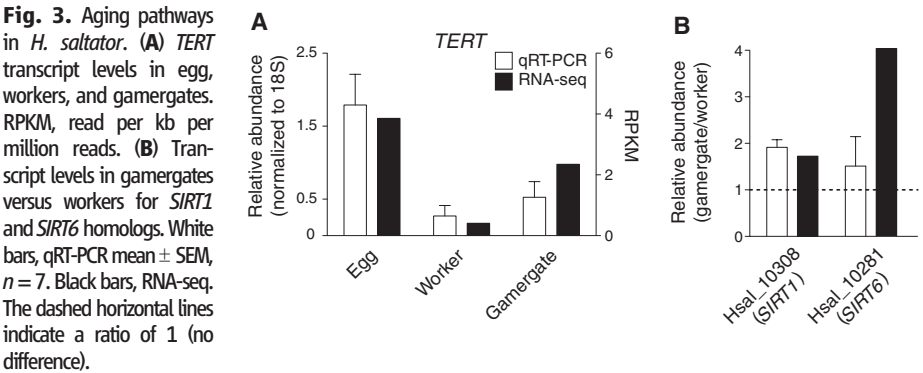


Fig. 5. Neurobiology and communication. (A) Quantification of 8 “GO:neurotransmitter binding” genes in head and thorax from major and minor workers in *C. floridanus*. White bars, qRT-PCR mean $\pm$ SEM, $n = 9$. Black bars, RNA-seq. (B) Quantification of fatty acid biosynthesis genes in *H. saltator* gamergates compared to nonreproductive workers. White bars, qRT-PCR mean $\pm$ SEM, $n = 7$. Black bars, RNA-seq.

the interplay between pheromone production, olfaction, and behavior in ants.

Our initial analysis of the genomes of two socially divergent ant species has captured major molecular features that make ants an attractive model for genomic and epigenetic studies. Ant species vary widely in the extent to which eusociality is implemented, ranging from small and less organized colonies to massive and complex societies (*1*). The diversity and flexibility of ants may provide experimental avenues to address long-standing hypotheses on the relationships among epigenetics, neurobiology, and behavior (*23*), as well as life-span regulation.

References and Notes

1. B. Hölldobler, E. O. Wilson, *The Ants* (Harvard Univ. Press, Cambridge, MA, 1990).
2. D. E. Wheeler, *Am. Nat.* **128**, 13 (1986).
3. S. Jemielity, M. Chapuisat, J. Parker, L. Keller, *Age (Omaha)* **27**, 241 (2005).
4. C. Peeters, J. Liebig, B. Hölldobler, *Insectes Soc.* **47**, 325 (2000).
5. Honeybee Genome Sequencing Consortium, *Nature* **443**, 931 (2006).
6. M. D. Lorenzen *et al.*, *Insect Mol. Biol.* **16**, 265 (2007).
7. M. Sumitani, D. S. Yamamoto, K. Oishi, J. M. Lee, M. Hatakeyama, *Insect Biochem. Mol. Biol.* **33**, 449 (2003).
8. E. Sahin, R. A. Depinho, *Nature* **464**, 520 (2010).
9. A. Hartmann, J. Heinze, *Evolution* **57**, 2424 (2003).
10. A. Vaquero, D. Reinberg, *Genes Dev.* **23**, 1849 (2009).
11. C. D. Allis, T. Jenuwein, D. Reinberg, *Epigenetics*, M. Caparros, Ed. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, ed. 1, 2007).
12. J. H. Werren *et al.*, Nasonia Genome Working Group, *Science* **327**, 343 (2010).
13. M. R. Kronforst, D. C. Gilley, J. E. Strassmann, D. C. Queller, *Curr. Biol.* **18**, R287 (2008).
14. Z. Wang *et al.*, *Cell* **138**, 1019 (2009).
15. E. C. Thompson, A. A. Travers, L. Randau, *PLoS ONE* **3**, e3008 (2008).
16. L. Hu, Y. T. Zhu, C. Qi, Y. J. Zhu, *Cancer Res.* **69**, 4067 (2009).
17. J. G. Hildebrand, G. M. Shepherd, *Annu. Rev. Neurosci.* **20**, 595 (1997).
18. C. Zube, C. J. Kleinedam, S. Kirschner, J. Neef, W. Rössler, *J. Comp. Neurol.* **506**, 425 (2008).
19. S. C. Hoyer, J. Liebig, W. Rössler, *Arthropod Struct. Dev.* **34**, 429 (2005).
20. E. Fishilevich, L. B. Vosshall, *Curr. Biol.* **15**, 1548 (2005).
21. G. J. Blomquist, A. G. Bagnères, *Insect Hydrocarbons: Biology, Biochemistry, and Chemical Ecology* (Cambridge Univ. Press, Cambridge, 2010).
22. J. Liebig, C. Peeters, N. J. Oldham, C. Markstädter, B. Hölldobler, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 4124 (2000).
23. C. Dulac, *Nature* **465**, 728 (2010).
24. We thank T. Bloss for ant colony maintenance, B. Kopenhaver for technical help, A. Alekseyenko and J. Wang for bioinformatic support, W. Wang for advice on evolutionary analysis, and P. Voigt and D. Simola for careful reading of the manuscript. R.B. is a fellow of the Helen Hay Whitney Foundation. This work was funded by Howard Hughes Medical Institute Collaborative Innovation Award #2009005 to S.L.B., D.R., and J.L. Raw sequencing data have been deposited in the National Center for Biotechnology Information as SRA020747 (*C. floridanus* genome), SRA020748 (*H. saltator* genome), and GSE22680 (RNA-seq). Assemblies and annotations have been deposited at DNA Data Bank of Japan–European Molecular Biology Laboratory–GenBank under the accession AEAB000000000 and AEAC000000000. The versions described in this paper are the first deposited versions: AEAB01000000 and AEAC01000000.

Supporting Online Material

www.sciencemag.org/cgi/content/full/329/5995/1068/DC1
Materials and Methods
SOM Text
Figs. S1 to S24
Tables S1 to S40
References

18 May 2010; accepted 20 July 2010
10.1126/science.1192428

Crystal Structure of Human Adenovirus at 3.5 Å Resolution

Vijay S. Reddy,^{1*} S. Kundhavai Natchiar,¹ Phoebe L. Stewart,² Glen R. Nemerow^{1*}

Rational development of adenovirus vectors for therapeutic gene transfer is hampered by the lack of accurate structural information. Here, we report the x-ray structure at 3.5 angstrom resolution of the 150-megadalton adenovirus capsid containing nearly 1 million amino acids. We describe interactions between the major capsid protein (hexon) and several accessory molecules that stabilize the capsid. The virus structure also reveals an altered association between the penton base and the trimeric fiber protein, perhaps reflecting an early event in cell entry. The high-resolution structure provides a substantial advance toward understanding the assembly and cell entry mechanisms of a large double-stranded DNA virus and provides new opportunities for improving adenovirus-mediated gene transfer.

Human adenoviruses (HAdV) are non-enveloped double-stranded DNA (dsDNA) viruses that are associated with acute infections (*1–3*). Although these infections are generally self-limiting, the reemergence of certain HAdV types has also been linked to potentially fatal respiratory infections in both civilian and military populations (*4*). Severe disseminated diseases also occur in patients receiving bone marrow–derived stem cells (*5, 6*). In addition to their disease associations, replication-defective or conditionally replicating HAdVs continue to be evaluated in ~25% of approved phase I to III clinical trials for vaccine and therapeutic gene

transfer (*7, 8*). However, the lack of accurate details of the virus structure limits the reengineering of HAdV vectors and prevents a better understanding of the virus life cycle. High-resolution HAdV structure determination presents a challenge because of the large size (910 Å average diameter, 150 megadalton) and complexity (pseudo-*T* = 25) of the virus. The crystal structures of the major HAdV capsid proteins, the fiber (*9*), penton base (PB) (*10*), and hexon (*11*), have been solved. The hexon and penton base crystal structures were subsequently used to derive pseudo-atomic models of the HAdV capsid at moderately high resolution (7 to 10 Å) (*12–14*) by cryoelectron microscopy (cryoEM). CryoEM structural analyses provided considerable insight into HAdV organization; however, they did not furnish detailed information on the interactions between the major and accessory (cement) proteins (IIIa, VI, VIII, and IX).

We report here the crystal structure of a recombinant HAdV-5 vector, designated Ad35F,

that is equipped with a short and flexible fiber protein derived from HAdV-35 (*15*). Details of the crystallization (*16*), diffraction data statistics (table S1), and structure determination of Ad35F at near-atomic resolution (3.5 Å) are described in (*17*).

The architecture of the HAdV capsid is shown in Fig. 1, A and B. The hexon is the most abundant protein in the capsid, with 720 subunits arranged as 240 trimers on a pseudo-*T* = 25 icosahedral lattice. Five PB monomer subunits occupy each of the icosahedral vertices and are associated with the trimeric fiber protein. Each of the 20 facets of the capsid contains 12 hexon trimers and a penton at each vertex. The icosahedral asymmetric unit consists of four hexon trimers and one PB monomer. As previously described, each hexon monomer contains two eight-stranded jelly-roll domains, V1 and V2 (*11*), whereas the PB subunit contains a single jelly-roll domain (*10*). Three sets of V1 and V2 domains give the hexon trimer a pseudo-hexagonal shape at the base, which in turn gives rise to a pseudo-*T* = 25 architecture for the HAdV capsid. Large insertions between the strands of the hexon jelly-roll domains form triangular towers on top of the hexagonal base. Representative electron density for a hexon subunit [amino acids (aa) 579 to 582] is shown in fig. S1. Some of the hypervariable region loops (aa 186 to 193 and 250 to 258) that were disordered in the isolated hexon structure are visible in the HAdV capsid crystal structure (fig. S1) because they are involved in multiple, symmetry-related interhexon contacts (figs. S2 and S3). The tertiary structures of the 12 structurally independent hexon subunits are nearly identical, having a root mean square deviation of ~1 Å upon superposition with a few differences found mainly at the amino and carboxy termini.

¹The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA. ²Vanderbilt University Medical Center, 2215 Garland Avenue, 710 Light Hall, Nashville, TN 37232, USA.

*To whom correspondence should be addressed. E-mail: gnemerow@scripps.edu (G.R.N.); reddyy@scripps.edu (V.S.R.)



Supporting Online Material for

Genomic Comparison of the Ants *Camponotus floridanus* and *Harpegnathos saltator*

Roberto Bonasio, Guojie Zhang, Chaoyang Ye, Navdeep S. Mutti, Xiaodong Fang, Nan Qin, Greg Donahue, Pengcheng Yang, Qiye Li, Cai Li, Pei Zhang, Zhiyong Huang, Shelley L. Berger,* Danny Reinberg,* Jun Wang,* Jürgen Liebig*

*To whom correspondence should be addressed. E-mail: bergers@mail.med.upenn.edu (S.L.B.); danny.reinberg@nyumc.org (D.R.); wangj@genomics.cn (J.W.); juergen.liebig@asu.edu (J.L.)

Published 27 August 2010, *Science* **329**, 1068 (2010)
DOI: 10.1126/science.1192428

This PDF file includes:

Materials and Methods
SOM Text
Figs. S1 to S24
Tables S1 to S40
References

TABLE OF CONTENTS

SUPPORTING TEXT	4
GENOME ASSEMBLY DETAILS	4
BACTERIA IN <i>H. SALTATOR</i> DIGESTIVE TRACT	4
NUCLEOTIDE DISTRIBUTION	5
POLYMORPHISM	5
REPEATS	5
TELOMERES.....	6
ANNOTATION AND EVOLUTIONARY ANALYSIS OF PROTEIN CODING GENES	6
ALTERNATIVE SPLICING	7
DEVELOPMENT GENES.....	7
<i>Embryonic development</i>	7
<i>Sex determination</i>	8
<i>Yellow and royal jelly-like proteins</i>	8
ANT BIOLOGY GENES.....	9
<i>Cuticular proteins</i>	9
<i>Digestive Proteases and Peptidases</i>	9
<i>Detoxification mechanisms</i>	9
<i>Immune defense</i>	10
<i>Hydrocarbon metabolism</i>	10
<i>Insulin pathway</i>	10
ADDITIONAL CHROMATIN BIOLOGY GENES	10
KINASES AND PHOSPHATASES	11
ADDITIONAL NEUROBIOLOGY ANNOTATIONS	11
<i>Neuropeptides</i>	11
<i>Odorant binding proteins</i>	11
RNA INTERFERENCE (RNAi)	12
MATERIALS AND METHODS.....	13
ANTS COLLECTION AND REARING.....	13
DNA ISOLATION, LIBRARY CONSTRUCTION AND SEQUENCING	13
GENOME ASSEMBLY	13
RNA ISOLATION, RNA-SEQ AND SMRNA-SEQ.....	13
REPEAT IDENTIFICATION	14
SINGLE NUCLEOTIDE POLYMORPHISM ANALYSIS	14
SEGMENTAL DUPLICATION	14
ANNOTATION OF PROTEIN CODING GENES	15
ANNOTATION OF MI RNA.....	15
ANNOTATION OF ALTERNATIVE SPLICING SITES	16
CONSTRUCTION OF INSECT GENE FAMILIES.....	16
PHYLOGENY RECONSTRUCTION FOR EIGHT INSECT SPECIES.....	19
EXPANSION AND CONTRACTION OF GENE FAMILIES	19
RNA-SEQ ANALYSIS AND GO ENRICHMENT	19

VERIFICATION OF MRNA AND MIRNA LEVELS BY QPCR	19
QUANTIFICATION OF DNA METHYLATION	20
RNAi	20
SUPPLEMENTARY FIGURES & LEGENDS	21
SUPPLEMENTARY TABLES	45
SUPPORTING REFERENCES	108

SUPPORTING TEXT

Genome assembly details

The *C. floridanus* assembly version 4 has an N50 scaffold length of 603 kb and a total length of 238 Mb (**Table S1**). The *H. saltator* assembly version 3 has N50 scaffold length of 598 kb and a total length of 297 Mb; in both assemblies the largest contig are > 400 kb and the largest scaffold > 2 Mb (**Table S1**). The total intra-scaffold gap sequence amounts to 7 Mb and 12 Mb, distributed in 16,360 (*C. floridanus*) and 20,090 (*H. saltator*) gaps with a median size of 136 bp and 234 bp in *C. floridanus* and *H. saltator* respectively. Version 4 of the *C. floridanus* genome is a refinement over version 3, a previous high quality “draft” (**Table S1**), on which most of the analyses reported in this paper were conducted. To evaluate the sequencing coverage of the assembled genome sequence, we mapped the raw sequencing data to the scaffolds using SOAP (*S1*). The peak sequencing depth were 102X for *C. floridanus* and 104X for *H. saltator* and more than 97% of both assemblies has 20-fold raw sequence coverage (**Fig. S11**). The high coverage obtained by next generation sequencing (NGS) guarantees extreme accuracy at the single nucleotide level at the expense of a more computational demanding assembly process (*S2*). Although the small gap size and the large N50 size of the scaffolds allows accurate prediction of most genomic features, we will continue to fill gaps and update genome-wide analyses and annotation until the assemblies reach “finished” genome quality (*S3*).

Bacteria in *H. saltator* digestive tract

We identified bacterial scaffold in *C. floridanus* and *H. saltator* assemblies v3 using the criterion that they should contain 2 or more adjacent ORFs with high homology to bacterial proteins in the NCBI protein database. The bacterial scaffold found in *C. floridanus* was 0.7 Mb long and covered 99% of the genome of the known *C. floridanus* symbiont, *Blochmannia floridanus* (*S4*). The 1.9 Mb bacterial scaffold found in the *H. saltator* assembly did not have a close match in the NCBI database, so we used its 16S rDNA sequence to assign a phylogenetic classification using the ribosomal database project tools (*S5*). According to its 16S rDNA sequence, this prokaryotic genome belongs to the order *Rhizobiales*, most likely of the genus *Bartonella*. Importantly, the bacterial genome could be detected in both *H. saltator* and the closely related species *H. venator*, but not in their food source (crickets), using conventional PCR (**Fig. S12A**) and qPCR (**Fig. S12B**). *Rhizobiales* bacteria are among the most prevalent ant gut symbionts, but they are usually found in herbivorous ants with a nitrogen-poor diet, while *H. saltator* feeds almost exclusively on terrestrial arthropod prey. Furthermore, the 8 other ponerine ant genera did not contain these bacteria, but *Harpegnathos* does. Only one other non-herbivorous genus, *Pheidole*, that feeds on live and dead insect prey also hosts *Rhizobiales* bacteria as symbionts (*S6*). As *H. saltator* can also feed on honey water in the laboratory, and is occasionally found in low bushes, the presence of a nitrogen-fixing bacterium in its gut suggests that when protein-rich food is scarce it may be capable of surviving on a herbivorous diet. This suggests that hosting *Rhizobiales* by predatory or scavenging ants may be a pre-adaptation to a largely herbivorous arboreal life in the tropics, where ants form dominant arthropod communities, although additional data on the occurrence of *Rhizobiales* bacteria in other ant species is needed to support this hypothesis.

Nucleotide distribution

We analyzed the distribution of G+C content in 10 kb windows. According to this parameter, *H. saltator* resembles *A. mellifera* and *N. vitripennis*, with a partially bimodal profile, whereas the distribution in *C. floridanus* is unimodal (**Fig. S13**). The distribution of genomic features according to G+C content varies between the two species (**Table S19**). In *C. floridanus*, G+C-rich domains are only mildly depleted for genes, whereas *H. saltator* genes are five times more likely to occur in A+T-rich domains. Conversely, tandem repeats are preferentially found in A+T-rich sequences, whereas they are more frequent in G+C-rich sequences in *H. saltator*.

We also measured the frequency of CpG in gene bodies, but did not observe the bimodal distribution typical of *A. mellifera* genes (**Fig. S14**), suggesting that this phenomenon is not a general feature of Hymenoptera and does not correlate with eusociality, contrary to what has been previously proposed (S7).

Polymorphism

The high coverage from Next Generation Sequencing (NGS) allows statistical approaches for the accurate detection of heterozygosity in the form of single nucleotide polymorphisms (SNPs) (S8, S9). Genomic DNA was extracted from multiple individuals belonging to the same colony. Because *C. floridanus* queens mate only once (S10), and no species from Ponerinae (the subfamily to which *H. saltator* belongs) is known to mate multiple times, the sequenced DNA contains sequences from three haploid chromosome complements: two from the diploid mother and one from a single haploid father. We mapped high quality reads to the assembly and called SNPs with a Bayesian algorithm (S9). The rate of SNPs in *C. floridanus* and *H. saltator* was 0.13% and 0.12%, respectively. Given that a triploid set of chromosomes was effectively sequenced, these numbers are comparable with SNP prevalence observed in human individuals (S8, S11). SNPs in the two ant species display a comparable skewing of transition vs. transversions, as commonly (S12, S13) (but not universally (S14)) observed. We asked whether deamination of 5-methyl-cytosine to thymidine could account for the observed prevalence of transitions, especially considering the high relative frequency of CpG dinucleotides; however, after filtering CpG dinucleotides, no changes in the frequency of C<->T and G<->A transitions were observed (**Fig. S15**), suggesting that the prevalence of transitions is a feature of the SNP distribution in these organisms, independent of cytosine methylation.

Repeats

We annotated repetitive sequences and transposable elements using a combination of homology to RepBase sequences (S15) and *de novo* prediction approaches (RepeatScout (S16), Tandem Repeats Finder (S17)). The *H. saltator* genome experienced a recent expansion of TcMar family transposons, of which we detect 49,193 copies, accounting for 15% of *H. saltator* repetitive sequence (compared to 5,168 copies in *C. floridanus*, 6,240 in *A. mellifera* and 4,466 in *N. vitripennis*). The amount of repetitive sequence in these two ant genomes is intermediate between *A. mellifera* and *N. vitripennis* (**Fig. S3**). We also analyzed how many of these transposons show signs of recent activity, as determined by the presence of intact open reading frames. The TcMar family of transposons is over-represented in active elements detected in *H. saltator*, consistent with a recent expansion (**Fig. S16**).

Telomeres

Most likely the telomeres of the insect ancestor were composed of TTAGG repeats (S18), but these were lost in certain lineages and mutated in others. For example, all Diptera, including *Drosophila* (S19) and *Anopheles gambiae* (S20) lack TTAGG repeats and the telomerase gene, while *T. castaneum* chromosomes are capped by “unconventional” TCAGG repeats and interspersed with the SART1 retrotransposon (S20). Despite these exceptions, the most common telomeric structure in insects is (TTAGG)_n, present in all Lepidoptera analyzed, including *B. mori* (which also hosts SART1 repeats in its telomeres (S21)), and all Hymenoptera analyzed (S22, S23). We identified orthologues for the *A. mellifera* and vertebrate telomerase gene (*TERT*) in both *C. floridanus* and *H. saltator* (Cf1o_14950, Hsal_18580), although incomplete at their 5' end (Fig. S17), and sought to identify telomeric repeats. Although TTAGG tandem repeats are not present in our assembled scaffolds, possibly due to the presence of subtelomeric repeats that fail to assemble, we were able to detect a large over-representation of (TTAGG)₃ in 15 k-mer from the raw Illumina reads. Specifically, this k-mer accounts for 0.033% and 0.013% of the *C. floridanus* and *H. saltator* total k-mer respectively, a frequency 10⁵ times larger than expected by chance. This strongly suggest that *C. floridanus* and *H. saltator* chromosomes are capped by TTAGG repeats, consistent with southern blot experiments on a different ant, *Manica yessensis* (S23).

Annotation and evolutionary analysis of protein coding genes

The current Official Gene Sets (OGS) v3.3 contain 17,064 (*C. floridanus*) and 18,564 (*H. saltator*) models for protein-coding genes, of which more than 60% have complete open reading frames (ORFs) and more than 80% are supported by RNA-seq or EST data (Table S20). Parameters, such as gene length, exon length, exon number, and intron length of gene models in OGS3.3 are comparable to those of other insect genomes, suggesting that our annotation pipeline does not contain systematic errors (Fig. S18). We manually curated about 400 genes of biological interest (see below) and in more than 90% the exon-intron structure was already correctly annotated by our automatic annotation pipeline; however, when minor inconsistencies were observed, we updated the annotation manually and included the corrected version in OGS3.3.

As expected, ant proteins with widespread homology to insects and human are enriched in IPR domains and GO terms related to common cellular structures, cellular housekeeping and universal molecular functions (Table S21-22), whereas insect-specific proteins are enriched in terms such as cuticle components, and digestive serine endopeptidases (Table S23). Hymenoptera-specific proteins are mostly related with the detection of chemical stimuli and their conversion to intracellular signals and transcriptional outcomes (Table S24), consistent with the specialized role of chemodetection in these insects (S24). In addition to those discussed in the main text, ant specific genes include those involved in synthesis of hydrocarbons, with 3 ant-specific acyl carrier domain-containing genes in *C. floridanus*, 4 in *H. saltator* and 4 ant-specific polyketide synthase-related genes in both species (GO: oxidoreductase activity) (Table S8). Given that complex hydrocarbon patterns are involved in signaling reproductive status among colony members (S25), some of these ant-specific proteins may participate in the biosynthesis of these and other chemical signals.

We also identified protein families that changed in size during the evolution of Hymenoptera first and ants later. The Hymenoptera genomes show a contraction relative to *D. melanogaster* in cuticle

proteins and certain protease families, but expansions of other protease families and genes for the metabolism of hydrocarbons (**Fig. S8, Table S18**). *C. floridanus* contains expanded families of juvenile hormone methyltransferases, cation transporters involved in the response to heat and humidity, short chain dehydrogenases, metalloproteases, enzymes for the synthesis of isoprenoids and sterols, and 7 homologues of human *gephyrin*, a scaffolding component of inhibitory synapses (S26). *H. saltator* contains expanded families of an organic cation transporter, polyamine biosynthesis proteins and selenocysteine methyltransferases, as well as 6 homologues of fruit fly *glait*, also implicated in CNS development (S27).

Alternative splicing

We identified sites of alternative splicing (AS) in the two ant genomes by mapping all RNA-seq reads to the assembled scaffolds with TopHat (S28). (**Fig. S19**). Overall we detected 7,583 AS events affecting 2,538 genes in *C. floridanus* and 10,185 AS events affecting 3,225 genes in *H. saltator*. Considering that only multi-exon genes represented in the RNA-seq dataset could be scored (9,550 for *C. floridanus* and 9,306 for *H. saltator*), this accounts for 27% and 35% of gene models in *C. floridanus* and *H. saltator*, respectively. As may be expected due to the lower complexity of these organisms, the percentages of genes affected by AS is much lower than the 95% estimate in humans (S29), but in line with reports in *Drosophila* (~40% (S30)); however, we anticipate that deeper sequencing of the transcriptome will reveal more AS events that are rare or restricted to certain tissues. Genes affected by AS were enriched in both species for InterPro (IPR) domains (**Table S25**) and Gene Ontology (GO) terms (**Table S26**) related to signal transduction and transcriptional regulation, such as protein kinase and small GTPase regulators, and, in *C. floridanus*, for PHD-type zinc fingers. As one third of gene models in OGS3.3 are incomplete, i.e. miss one or more exons, we repeated this analyses restricting them to genes with more than one exon, complete ORFs and RNA-seq coverage (8,858 in *C. floridanus* and 7,988 in *H. saltator*), and confirmed the enrichment of the same GO and IPR categories.

Development genes

Embryonic development

Most components of the key signaling pathways giving rise to the metazoan body plan (S31) can be identified in the genome of *C. floridanus* and *H. saltator*. In addition, the majority of genes involved embryonic induction processes such as axis formation, germ-line specification and segmentation during embryogenesis are present in both ants (**Table S27**). As observed in other non-Diptera insects, the *Drosophila* axis-specifying genes *oskar* (germ-line specification), *bicoid*, *trunk* and *torso* (anterior-posterior development), and *gurken* (dorso-ventral polarity), are also absent in both ant species. However, the downstream components of these pathways such as *hunchback*, *ocelliless*, *hedgehog* and *spatzle* are conserved (**Table S27**), and show some evidence of duplication: *ocelliless* is present in two copies in both ant genomes, consistent with its duplication in *A. mellifera*, whereas *hunchback* and *hedgehog* also appear to be duplicated in these ants, but not in honeybee. InterProScan identified 86 homeobox-containing proteins in *C. floridanus* and 80 in *H. saltator* (**Table S28**). The HOX cluster genes are all contained in one single locus measuring ~0.75 Mb in *C. floridanus* and 0.8 Mb in *H. saltator* and syntenic to both *A. mellifera* and *N. vitripennis*. As in *Drosophila* all HOX cluster genes are expressed from the same strand.

Sex determination

Sex determination in Hymenoptera correlates with ploidy: haploid individuals originating from unfertilized eggs develop into males, whereas fertilized eggs usually develop into females (S32). The molecular mechanisms that translate ploidy information into a developmental trajectory vary among Hymenoptera species and remain for the most part unknown. In *A. mellifera*, heterozygosity (which implies diploidy) of the *csd* locus dictates female development, whereas hemi- or homozygosity result in male development (complementary sex determination, CSD) (S33); however, the duplication of the *feminizer* gene that gave rise to *csd* occurred during recent evolution in the bee lineage (S34) and *N. vitripennis* has evolved a different mechanism for sex determination (S35). We annotated *C. floridanus* and *H. saltator* orthologues for known components of the sex determination machinery in insects and found that, like *N. vitripennis*, *C. floridanus* and *H. saltator* lack unambiguous orthologues of *csd* (**Table S29**). In contrast with *N. vitripennis*, however, the *feminizer* locus is duplicated in both *C. floridanus* and *H. saltator*, and both copies are closer to the *A. mellifera* *feminizer* than to *csd* (**Fig. S20**). In *A. mellifera* *feminizer* is downstream of *csd*, which directs its alternative splicing into a non-functional variant in males (S34). Whether a similar mechanism, though independent of *csd* is functioning in ants remains unknown, given that at the sequencing depth reached by the RNA-seq dataset we could not detect any evidence of sex-specific alternative splicing of *feminizer*, *doublesex* or *transformer*. It is possible that the newly originated copies of *feminizer* in ants might provide a mechanism analogous to CSD in honeybee.

Yellow and royal jelly-like proteins

Yellow and royal-jelly-like proteins form a family specific to arthropods and a few other bacteria and fungi and control expression of genes affecting cuticular pigmentation, development, sexual maturation and behavior (S36). The number of genes encoding yellow-like proteins varies from 8 in *D. melanogaster* to 20 in *A. mellifera* and 26 in *N. vitripennis*, whereas *C. floridanus* and *H. saltator* contain only 10 yellow-like proteins; however, the genomic organization of these genes into a single cluster is conserved (**Fig. S21**). In *A. mellifera* and *N. vitripennis*, repeated duplications of an ancestral member of the *yellow* cluster have given rise to the major royal jelly protein (MRJP) sub-family, which regulate physiology, development, and behavior (S37). There are 9 MRJPs in *A. mellifera* and 10 in *N. vitripennis*. Surprisingly, although the structural organization of this cluster is conserved in *C. floridanus* and *H. saltator* (**Fig. S21**), which diverged from *A. mellifera* more recently than *N. vitripennis*, a single copy of MRJP is found in this locus in the ants' genomes, and *H. saltator* possesses 2 MRJP paralogues elsewhere. This implies that either MRJPs have been selectively lost in ants during evolution, or that the duplications occurred independently in the bee and wasp lineages. It has been suggested that duplication and functional diversification of the MRJPs correlates with the emergence of social behavior in *A. mellifera* (S37), however the observation that the solitary *N. vitripennis* possesses 10 genes encoding MRJPs, whereas the highly eusocial *C. floridanus* only possesses one, argues against this hypothesis. More Hymenoptera genomes will have to be sequenced before we fully understand the evolutionary history and functional significance of the *yellow* locus.

Ant biology genes

Cuticular proteins.

Insect cuticles are layered structures consisting primarily of chitin and its associated proteins and lipids (S38). They are produced as extracellular secretions from the epidermis and the most abundant family of cuticular proteins has an extended Rebers and Riddiford (R&R) consensus motif (pfam00379) that binds to chitin (S39). Similar to *A. mellifera* (28 genes), both *C. floridanus* and *H. saltator* show a reduced number of cuticular genes compared to *N. vitripennis* (63) and *D. melanogaster* (93), with 35 putative genes in *C. floridanus* and 33 *H. saltator* (**Table S30**). In addition, both ants contain second class (13 for *C. floridanus*, 30 for *H. saltator*) and third class of chitin-binding proteins (6 for *C. floridanus*, 7 for *H. saltator*) with a cysteine rich chitin binding domain and chitin deacetylase domain respectively and including multiple copies of the peritrophic membrane proteins associated with pfam01607 domain (**Table S30**). The reduced repertoire of cuticular proteins in the two ant genomes, compared to solitary insects like *Drosophila* and *Anopheles*, may be explained by the ants' social lifestyle, which provides them with a protective nest environment, especially during embryonic, larval, and pupal development (S22).

Digestive Proteases and Peptidases

A distinguishing feature of ants is their ability to feed from a wide variety of food sources (S40). This is likely to impose a selective pressure that facilitates duplication and diversification of digestive enzymes such as proteases and peptidases (S41). Indeed, we find 66 gene models containing aspartic protease domains (IPR001461) in *C. floridanus*, compared to 1 in *H. saltator*, 1 in *A. mellifera*, and 2 in *D. melanogaster* (**Table S31**). The abundance of these acidic endopeptidases may be explained by the extremely low pH in the *C. floridanus* gaster due to the presence of large amounts of formic acid in the poison gland (S42). Metallopeptidases are also expanded in *C. floridanus*, particularly those of the M1 (IPR014782) and M10 (IPR001818) family (**Table S31**), involved in the final step of protein digestion and extracellular matrix processing, respectively (S43). The genome of *H. saltator* shows an expansion of genes containing S1/S6 chymotrypsin peptidase domains (IPR001254) with 112 gene models, compared with 73 in *C. floridanus* and 62 in *A. mellifera*, but a contraction when compared to *D. melanogaster* (256). These proteins are secreted into the gut lumen and usually serve as digestive enzymes (S44, S45). In addition to their roles in digestion, most of these genes function in regulating other processes like insect metamorphosis and innate immune responses including regulating expression of antimicrobial peptides, possibly explaining their large number and functional diversification in ants and other insects (S44).

Detoxification mechanisms.

Ants are exposed to a wide array of xenobiotic compounds as they live and forage for food primarily on the ground, where they are exposed to a variety of natural toxins. This is particularly the case for *C. floridanus*, which is a generalist feeder and a scavenger. Three enzyme families, the cytochrome P450 monooxygenases (P450), glutathione transferases (GST) and carboxy/cholinesterases (CCE) catalyze a wide range of detoxification reactions. With 128 genes (**Table S32**) the *C. floridanus* genome comprises the second largest expansion of the P450 family in insects, only surpassed by *A.*

aegypti (160), whereas the *H. saltator* genome only contains 95 members. In contrast, the GST family shows no expansion in Hymenoptera (**Table S32**). Finally, the *C. floridanus* genome comprises 75 members of the carboxy/cholinesterases, a number similar to that in *D. melanogaster* (77 genes), whereas the *H. saltator* genome contains 51, closer to the number in *A. mellifera* (49) (**Table S32**). The expansion of the detoxification tool kit in ants, particularly *C. floridanus*, might have allowed them to adapt to different ecological niches.

Immune defense

Ants live in large communal societies that rival human societies in terms of sophistication and size. Communal living and a crowded nest environment with immobile brood and stored food resources make ant colonies an attractive target to disease-causing microorganisms. We identified and annotated homologues to 228 *A. mellifera* genes implicated in immune pathways (*S46*) and found that ants, as honeybees have a reduced repertoire of immunity-related genes when compared to *Drosophila* (**Table S33**). We probed further signaling pathways related to immune defense and observed a similar decrease of immune genes in the eusocial Hymenoptera compared to *Drosophila* (**Table S33**). Genes functioning in the TOLL, IMD, JAK/STAT, and JNK pathways were present in both genomes (**Table S34**). Interestingly, both ant genomes contain orthologues for the BCL-2 family member, *BCLIXL*, an anti-apoptotic gene implicated in JAK kinase pathway, which is not found in *Drosophila*, *A. mellifera* or *N. vitripennis*. The reduced immune capabilities suggested by contractions in immune gene families could be partially compensated by behavioral traits associated with eusocial organisms like such as grooming of nestmates and nest hygiene. Additionally, the ability to secrete broad spectrum antimicrobial agents, may aid in disease control in ant colonies (*S47*)

Hydrocarbon metabolism.

Several genes with putative roles in hydrocarbon metabolism other than those discussed in the main text displayed differential regulation in different ant castes and developmental stages, including ant-specific acyl-carriers (**Fig. S9A-B**), ant-specific polyketide synthases (**Fig. S9C-D**) and a group of 6 genes homologue to the *Drosophila* long-chain fatty acyl-CoA synthetase CG6178 (*S48*) (**Fig. S9E**).

Insulin pathway.

The insulin/insulin-like growth factor-like signaling (IIS) pathway is also related to aging as downregulation prolongs lifespan in *Drosophila* and *C. elegans* (*S49, S50*). The complete IIS and TOR pathways are conserved in ants (**Table S35**). A single insulin-like peptide was identified in each ant, 4 insulin/insulin-like growth factor receptors in *C. floridanus*, and 5 in *H. saltator*, and some components of the IIS pathway were upregulated in *H. saltator* gamergates compared to workers (**Fig. S22**), consistent with some findings in honeybees (*S51*).

Additional chromatin biology genes

In addition to the histone modifiers (acetyltransferases, deacetylases and methyl-transferases) described in the main text, we also annotated manually other categories of genes known to participate in chromatin regulation. 20 proteins in *C. floridanus* and 21 in *H. saltator* contain bromodomains —acetyl-lysine binding modules (**Table S11**).

For histone demethylases, we identified 13 (*C. floridanus*) and 12 (*H. saltator*) proteins containing JmjC-domains and two homologues of LSD1 in each ant species (**Table S36**). There are also numerous proteins containing predicted methyl-lysine binding modules, such as PHD fingers (S52) (38 in *C. floridanus*, 33 in *H. saltator*), chromodomains (S52) (14 in *C. floridanus* and 20 in *H. saltator*), and MBT repeats (S53) (4 in both ant species) (**Table S11**).

Kinases and phosphatases

Kinases and phosphatases are involved in many aspects of cellular biology. By directed homology searches, we identified 207 kinases and 74 phosphatases in the *C. floridanus* genome, and 211 kinases and 72 phosphatases in the *H. saltator* genome. Domain prediction from Interpro Scan yielded more genes containing putative kinase/phosphatase domains (**Fig. S23**). Other Hymenoptera (*A. mellifera* and *N. vitripennis*), *D. melanogaster*, and *H. sapiens* were also included in the comparison. In most cases the numbers of genes containing a given IPR domain is comparable among Hymenoptera, whereas more fluctuations are found in the more divergent *D. melanogaster* and *H. sapiens* genomes. A tyrosine kinase gene family, encoding for homologue to the human *ROS1* gene, is dramatically expanded in *C. floridanus*, with several specific duplication events compared to the single homologue found in *H. saltator*. *ROS1* is a protooncogene member of the *sevenless* subfamily of receptor tyrosine kinase (S54): its specific expansion in *C. floridanus* is intriguing.

Additional neurobiology annotations

Neuropeptides

Neuropeptides are important messenger molecules that regulate a range of physiological processes, including regulation of growth, development, behavior, learning and lifespan (S55). To date we have identified 20 neuropeptides in the genome of *C. floridanus* and 19 in the genome of *H. saltator* (**Table S37**), but we expect many more to be unveiled after directed experiments, as their small size and lack of easily identifiable domains complicates bioinformatic annotation. Both *C. floridanus* and *H. saltator* have only one insulin-like peptides, one allatostatin and no allatotropin orthologues. The latter is also missing from the *A. mellifera* genome, which is surprising because the two form a counteracting pair inhibiting and stimulating, respectively, the production of juvenile hormone, a key mediator of caste determination and longevity in honeybees (S56). Unlike honeybees, *C. floridanus* and *H. saltator* possess orthologues for oxytocin and vasopressin; to date these genes have only been found in *T. castaneum* and *N. vitripennis* and is absent in most other insects including *Drosophila*. Genes for *tachykinin*, *myosuppressin-MS* and *FMRF* were only found in *C. floridanus* but not in *H. saltator*, whereas *neuropeptide-F* and *corazonin* were found in *H. saltator* but not in *C. floridanus*, though it remains possible that the missing neuropeptides genes are present but cannot be identified by bioinformatic approaches alone.

Odorant binding proteins

Odorant binding proteins (OBPs) participate in olfaction by shuttling odorants through the sensillar lymph to membrane-bound odorant receptors on the dendritic membranes of olfactory neurons (S57). Similar to honeybees, ants show a reduction in OBPs compared to *Drosophila*, with 22 genes in *C. floridanus* and 27 in *H. saltator* (**Table S38-39**).

RNA interference (RNAi)

RNAi approaches allow genetic manipulation of organisms for which conventional forward genetic approaches have not (or cannot) be established. Genes encoding for core components of the RNAi pathway, such as Dicer and Drosha, RISC components, such as the Argonaute family of proteins, as well as dsRNA binding proteins R2D2, Pasha and Loquacious are present in both ant genomes, with high degree of conservation with other insect species (**Table S40**). Both genomes also harbor a putative orthologue for *SID-1*, which encodes for a multi-transmembrane protein implicated in dsRNA uptake and systemic RNAi in *C. elegans* (S58), but is not present in *D. melanogaster*. However, the molecular details of SID-1 function remain poorly understood. The presence of a *SID-1* orthologue encouraged us to pursue RNAi approaches in ants as systemic spreading and transgenerational transmission of an RNAi phenotype is a very desirable feature in a model organism and may alleviate the difficulties of establishing conventional genetic approaches in eusocial insects (S59). Consistent with the conservation of RNAi pathway genes, injection of siRNAs caused the downregulation of *PKG* (a cGMP-dependent protein kinase involved in foraging behavior (S60)) in *C. floridanus* (**Fig. S24**).

MATERIALS AND METHODS

Ants collection and rearing

The *C. floridanus* colony was founded in the laboratory from a queen captured in the wild in Sugarloaf Key, Florida in 2002. *C. floridanus* ants were fed twice a week with water, 30% sugar water solution, artificial diet and boiled pieces of beetle larvae (*Zophobas morio*). The *H. saltator* colony was originally collected as a gamergate colony in Karnataka, India in 1999 and bred in the laboratory for 10 years. In July 2009, the *H. saltator* colony was split in half equally dividing the gamergates, workers and the brood between the two daughter colonies. *H. saltator* ants are easily maintained under laboratory conditions. As they are active predators and hunt for food, they were provided with water and 8-10 crickets 2-3 times a week. Both species were housed in plastic boxes containing a dental plaster nest with a modified nest chamber covered with a glass plate and were maintained at room temperature (25°C) under 12 hours alternating cycles of light and darkness (L:D 12:12).

DNA isolation, library construction and sequencing

Genomic DNA was isolated from pooled males and workers belonging to the same colony using standard molecular biology techniques. For small insert libraries, 5 µg of DNA from each species were sheared to fragments of 200-500 bp, end-repaired, A-tailed and ligated to Illumina paired-end adapters (Illumina, San Diego, CA). The ligated fragments were size selected at 200, 350 and 500 bp on agarose gel and amplified by LM-PCR to yield the corresponding short insert libraries. For long insert size mate-pair library construction, 20-40 µg of genomic DNA were sheared to the desired insert size using nebulization for 2 kb or HydroShear (Covaris, Woburn, MA) for 5kb and 10kb. Next, the DNA fragments were end-repaired using biotinylated nucleotide analogues (Illumina), size selected at 2, 5 and 10 kb and circularized by intramolecular ligation. Circular DNA molecules were sheared with Adaptive Focused Acoustic (Covaris) to an average size of 500 bp. Biotinylated fragments were purified on magnetic beads (Invitrogen, Carlsbad, CA), end-repaired, A-tailed and ligated to Illumina paired-end adapters, size-selected again and purified by LM-PCR. All libraries were sequenced on the Illumina Genome Analyzer platform.

Genome assembly

As we recently sequenced and assembled *de novo* a mammalian genome using exclusively high coverage Illumina Genome Analyzer sequences (S61), we took a similar approach to sequence the genomes of *C. floridanus* and *H. saltator*. We generated and sequenced libraries with insert size ranging from 200 to 10,000 bp and obtained 30.3 Gb and 49.3 Gb of purity-filtered reads for *C. floridanus* and *H. saltator*, respectively, resulting in more than 100-fold sequence coverage for both species. We assembled the reads into contigs and scaffolds with a stepwise strategy, as previously described (S2, S61).

RNA isolation, RNA-seq and smRNA-seq

We harvested 400 eggs, 50 larvae from all instars, 60 minor workers and 30 major workers from two *C. floridanus* colonies and 50 males from a third colony. For *H. saltator*, we harvested 300 eggs, 40 larvae from all instars, 20 non-reproductive workers, 10 gamergates and 50 males from 3 different

colonies. Adult individuals of various age and performing different tasks in the nest were collected to minimize confounding effects due to age or behavior. RNA was purified using TRIzol (Invitrogen). For RNA-seq, poly-A⁺ RNA was isolated with oligo-dT-coupled beads from 20 μ g total RNA of each sample. First strand cDNA synthesis was performed with random hexamers and Superscript II reverse transcriptase (Invitrogen). The second strand was synthesized with *E. coli* DNA PolI (Invitrogen). Double stranded cDNA was purified with Qiaquick PCR purification kit (Qiagen, Germantown, MD), and sheared with a nebulizer (Invitrogen) to 100-500 bp fragments. After end repair and addition of a 3' dA overhang the cDNA was ligated to Illumina PE adapter oligo mix, and size selected to 200 $\pm$ 20 bp fragments by gel purification. After 15 cycles of PCR amplification the libraries were sequenced using a 1G Illumina Genome Analyzer and the paired-end sequencing module. For smRNA-seq, we gel-purified 18-30 nts RNAs from the samples utilized for RNA-seq. Illumina 5' and 3' RNA adapter were sequentially ligated to the RNA fragments and the ligated products were size-selected on denaturing polyacrylamide gels. The adapter-linked RNA was reverse transcribed with small RNA RT primers and amplified with 15 cycles of PCR using small RNA PCR primer 1 and 2 (Illumina). The libraries were sequenced with the Illumina Genome Analyzer.

Repeat identification

We identified known transposable elements (TEs) in the two ant genomes using RepeatMasker v3.2.6 (S62) against the Repbase TE library v2009-09-01 (S63). This step identified 8.6 Mb known TEs in *C. floridanus* and 30.1Mb in *H. saltator*. Next, we constructed an *ab initio* repeat library for ants using RepeatScout (S16) with default parameters, and the consensus sequence for each repeat family was used as a custom library in RepeatMasker to identify additional high and medium copy repeats (>10 copies). With this method, we identified an additional 20 Mb repetitive sequence in *C. floridanus* and 46.3 Mb in *H. saltator*. Next, we identified non-interspersed repeat sequences by RepeatMasker with the “-noint” option, including Simple_repeat, Satellite, and Low_complexity repeats. We also predicted tandem repeats using Tandem Repeat Finder (S17), with parameters set to “Match=2, Mismatch=7, Delta=7, PM=80, PI=10, Minscore=50, and MaxPeriod=12”. TEs were classified according to Wicker *et al.* (S64).

Single nucleotide polymorphism analysis

High quality reads (average quality larger than 30) from small insert size libraries were aligned onto the assembled scaffolds using SOAP (S1), allowing a maximum of 2 mismatches. We calculated the probability of each possible genotype at every base from the alignment of short reads onto the reference genome. The allelic sequence with the highest probability was used as reference sequence and any other high probability alleles were called as heterozygous SNPs. This analysis was performed with SOAPsnp (S9).

Segmental duplication

We used the whole-genome assembly comparison method to identify segmental duplications (SDs) (S65). The self-alignment for each genome was implemented by LASTZ with parameters T=2, Y=9400 (http://www.bx.psu.edu/miller_lab). We defined as SD two sequences larger than 1 kb with identity higher than 80% but lower than 98%, to exclude improperly assembled allelic variants due to the ‘draft’ status of the genome.

Annotation of protein coding genes

We combined homology information, *de novo* prediction and expression profiling to annotate the genome of *C. floridanus* and *H. saltator* for protein coding genes. Protein sequences from the *A. mellifera* (S22), *H. sapiens* (S66) and *D. melanogaster* (S67) genome projects were used for a first round of homology-based gene predictions on the unmasked ant assemblies. The pipeline included the following steps: 1) homology search using TBLASTN with $E < 10^{-5}$; 2) selection of the most homologous protein for each genome locus showing multiple matches; 3) exclusion of regions with homology to less than 30% of the query protein; 4) protein sequence to genomic sequence alignment with GeneWise v2.0 (S68) to generate gene model structures (exon-intron boundaries). We then masked repeats and performed homology searches using protein sequences from *T. castaneum* (S69), *N. vitripennis* (S70), *B. mori* (S71), and *A. pisum* (S72), using the above pipeline. In addition, we used two *de novo* gene prediction softwares, Augustus (S73) and SNAP (S74), after training them on protein sequences of insects closely related to ants. All gene models were combined with RNA-seq and EST data using GLEAN (S75), a tool that creates consensus gene lists by integrating evidence from homology, *de novo* prediction, and RNA-seq/EST data. Next, we analyzed the gene models rejected by GLEAN's strict quality controls and included in the final annotation gene models with strong support from homology, or *de novo* predicted gene models with clear evidence of transcription in the RNA-seq or EST datasets. We manually curated the gene structure of more than 400 genes of interest, by performing TBLASTN searches using orthologous protein sequences from other organisms, and reconciling them with the assemblies using GeneWise (S68). To assign preliminary GO terms to predicted gene models we performed InterProScan (S76) on the two Official Gene Sets using the blastprodom, hmmpfam, hmmsmart, and profilescan algorithms. To avoid bias in comparing these gene sets with those from other organisms annotated differently, we repeated these analysis on all genomes used for comparisons in this study. In addition to Cflo and Hsal gene IDs, each protein coding gene model was assigned a preliminary homologous in human and *Drosophila* by performing a BLASTP search of all Swiss-Prot entries for these organisms. Unless indicated otherwise, the gene names in parentheses throughout this study specify the closest human homologue.

Annotation of miRNA

Illumina reads were aligned against the ant genomes using SOAP (S1), allowing for one mismatch. We filtered reads mapping to annotated exons of protein coding genes and to other non-coding RNA genes. To analyze the RNA secondary structures surrounding putative miRNAs, we extracted 100 nucleotides of genomic sequence flanking each side of mapped reads, predicted the secondary structure using RNAfold (S77) and analyzed by MIREAP (<https://sourceforge.net/projects/mireap/>), an in-house computational tool designed to identify genuine miRNAs from small-RNA seq; it considers miRNA biogenesis, sequencing data and structural features to improve miRNA identification. Stem-loop hairpins were considered only when: mature miRNAs are present in one arm of the hairpin precursors; the secondary structures of the hairpins are stable (free energy lower than -20 kcal/mol); and hairpins are located in intergenic regions or introns. All remaining candidates were analyzed with MiPred, a random forest-based method for classification of genuine pre-miRNAs and pseudo-pre-miRNAs using a hybrid feature (including local contiguous structure sequence composition, minimum of free energy of the secondary structure and P-value of randomization test)

(S78). We computed the relative read density for each putative miRNA and only considered those supported by 5 reads per million in at least one sample. To determine homology relationships between miRNAs from different species we used miRAlign (S79).

Annotation of alternative splicing sites

Short paired-end reads from the RNA-seq datasets were aligned onto the genome assemblies using TopHat (S28), a software package that identifies splicing junction sites. TopHat first aligns the raw reads onto the reference allowing maximal 2bp mismatches, and reports the “initially unmapped reads” (IUM reads) that do not map to the genome. Next, TopHat finds reads that span junctions using a seed-and-extend strategy from the IUM reads. Based on the junction sites identified by TopHat and our predicted gene structures, we classified the alternative splicing events into seven categories: exon-skipping (ES), intron-retention (IR), mutually exclusive exon (MXE), alternative 5' splice site (A5SS), alternative 3' splice site (A3SS), alternative first exon (AFE), and alternative last exon (ALE). These putative splice variants were determined according to the position of the observed junction sites compared to the annotated gene models. To prevent cryptic unannotated exons to be erroneously included in the IR class, we only counted as *bona fide* IR events those observed in gene models for which the affected intron region had read coverage over 90%.

Construction of insect gene families

We used the Treefam methodology (S80) to define a gene family as a group of genes descending from a single gene in the last common ancestor. We constructed a pipeline to cluster individual genes into gene families and perform phylogeny analysis: 1) Data preparation. The protein-coding genes for 8 insect species (*H. saltator*, *C. floridanus*, *A. mellifera*, *N. vitripennis*, *Tribolium castaneum*, *Bombyx mori*, *A. gambiae* and *D. melanogaster*) and human were used in this analysis. We kept only the longest transcript isoform for each gene, and only considered proteins larger than 30 aa. 2) Pairwise relation assignment (graph building). We performed BLASTP on all protein sequences against the database containing protein dataset of all the species with E-value cut-off of 10^{-5} , and conjoined fragmental alignments for each gene pairs using Solar. We assigned a connection (edge) between two nodes (genes) if more than 1/3 of the region was aligned in both genes. A H-score ranging from 0 to 100 was used to weigh the similarity (edge). For two genes G1 and G2, the H-score was defined as $\text{score}(G1G2) / \max(\text{score}(G1G1), \text{score}(G2G2))$, (score = BLAST raw score). 3) Gene family construction. We used the average distance for the hierarchical clustering algorithm, requiring the minimum edge weight (H-score) to be larger than 5, the minimum edge density (total number of edges / theoretical number of edges) to be larger than 1/3. The clustering for a gene family was terminated when the presence of one or more outgroup genes was detected. 4) Phylogeny and orthology analysis. We performed multiple alignments of protein sequences for each gene family using MUSCLE (S81), and converting the protein alignments to CDS alignments using the following Perl script.

```
#!/usr/bin/perl
=head1 Name
pepMfa_to_cdsMfa.pl  --  convert protein alignment to cds alignment
=head1 Description
```

```

=head1 Version
  Author: Fan Wei, fanw@genomics.org.cn
  Version: 1.0,  Date: 2006-12-6
  Note:
=head1 Usage
  --verbose    output running progress information to screen
  --help       output help information to screen
=head1 Exmple
=cut
use strict;
use Getopt::Long;
use FindBin qw($Bin $Script);
use File::Basename qw(basename dirname);
use Data::Dumper;
use File::Path; ## function " mkpath" and "rmtree" deal with directory
##get options from command line into variables and set default values
my ($Verbose,$Help);
GetOptions(
    "verbose"=>\$Verbose,
    "help"=>\$Help
);
die `pod2text $0` if (@ARGV == 0 || $Help);
my $aa_align_file = shift;
my $cds_file = shift;
my %aa;

##read the protein alignment information
open IN, $aa_align_file || die "fail open $aa_align_file\n";
$/=">"; <IN>; $/="\n";
while (<IN>) {
    my $name = $1 if(/^(S+)/);
    $/=">";
    my $seq = <IN>;
    chomp $seq;
    $seq =~ s/\s//g;
    $/="\n";
    $aa{$name} = $seq;
}
close IN;

##convert protein alignment to cds alignment
my $out;

```

```

open IN, $cds_file || die "fail open $cds_file\n";
$/=">"; <IN>; $/="\n";
while (<IN>) {
    $out .= ">".$_;
    my $name = $1 if(/^(S+)/);
    $/=">";
    my $seq = <IN>;
    chomp $seq;
    $seq =~ s/\s//g;
    $seq =~ s/---//g;
    $seq = uc($seq);
    $/="\n";
    my $cds;
    my $prot = $aa{$name};
    my $len_prot = length($prot);
    my $j = 0;
    for (my $i=0; $i<$len_prot; $i++) {
        my $aa = substr($prot,$i,1);
        if ($aa ne '-') {
            $cds .= substr($seq,$j,3);
            $j += 3;
        }else{
            $cds .= '---';
        }
    }
    Display_seq($cds);
    $out .= $cds;
}
close IN;
print $out;
#display a sequence in specified number on each line
#usage: disp_seq($string,$num_line);
#         disp_seq($string);
#####
sub Display_seq{
    my $seq_p=shift;
    my $num_line=(@_) ? shift : 50; ##set the number of charcters in each
line
    my $disp;
    $$seq_p =~ s/\s//g;
    for (my $i=0; $i<length($$seq_p); $i+=$num_line) {
        $disp .= substr($$seq_p,$i,$num_line)."\n";
    }
}

```

```

    }
    $$seq_p = ($disp) ? $disp : "\n";
}
#####

```

We built phylogenetic trees using TreeBeST, which takes advantage of both codon-based and aa-based algorithms (nj-dn, nj-ds, nj-mm, phym1-aa and phym1-nt) and adjusting them to the topology of the species tree, to form a more accurate consensus tree. We inferred all the orthology (from speciation) and paralogy (from duplication) gene relations from the gene phylogeny tree.

Phylogeny reconstruction for eight insect species

1,032 single-copy gene families were identified using the methodology above. Gene families with poor alignments or excessive gaps were removed. The remaining single-copy gene families were used to reconstruct the phylogeny of the 8 species. 4-fold degenerate sites were extracted from each family and concatenated into a supergene for each species. We used MODELTEST (S82) to select the best substitution model (GTR+gamma+I) and MRBAYES (S83) to reconstruct the phylogenetic tree. The chain length was set to 50,000,000 (1 sample/1000 generations) and the first 1,000 samples were burned in.

Expansion and contraction of gene families

We inferred the rate and direction of change in gene family size for ants and honeybee using a stochastic birth-death model with fruit fly as outgroup (S84). We first estimated the average rate of gene turnover across the ants, λ , which is a rate at which the size of gene family is expected to expand or contract overtime due to the gain or loss. Then the phylogenetic tree topology and branch lengths were taken into account to infer the significance of change in gene family size in each branch.

RNA-seq analysis and GO enrichment

Expression values were evaluated from RNA-seq data using RPKM normalization according to the method of Mortazavi *et al.* (S85). Briefly, the read-count for a given predicted gene was divided by the number of millions of reads and the number of kilobases in the exon model; this is meant to correct gross inconsistencies in the efficiency of different sequencing reactions. Differential expression for *C. floridanus* major and minor workers and *H. saltator* gamergates and workers was determined by setting a fold-change cutoff of at least 4. To assess GO category enrichment across all differentially expressed genes, orthologs to predicted genes were discovered by querying the UniProt and KEGG databases and the best-aligned ortholog in each case was used as a proxy input to GoMiner (S86). GoMiner results were parsed with a p-value cutoff of 0.05, an FDR cutoff of 0.1, and a constraint that the category contains at least 5 differentially regulated genes. GO terms related to transposon activity were excluded as well as GO level 2 terms. When a more specific term “child” was enriched along with the term “parent” while referring to the same exact genes, the term “parent” was discarded.

Verification of mRNA and miRNA levels by qPCR

Samples for qPCR verification were prepared by collecting 7 separate pools of eggs, larvae, males, workers (minors and majors for *C. floridanus*) and gamergates (for *H. saltator*), obtained from

different colonies. For Figure 5A we tested independently head and thorax from 3 individual major workers or 2 pools of minor workers. For both species larvae from all instars were used and the individuals all belonged to various colonies, different from that analyzed by RNA-seq. Although we carefully selected for individuals covering a range of morphologies, and presumably ages, we did not have the possibility to accurately control for age in these experiments. The gaster from *C. floridanus* minor and major workers were removed to prevent RNA degradation, due to the presence of large amounts of formic acid. For *H. saltator* worker and gamergates, the digestive tract and poison gland were removed prior to RNA purification. Ovaries were further removed from *H. saltator* gamergates to eliminate RNA from developing eggs in the reproductive tract. RNA was extracted using TRIzol Reagent (Invitrogen, Carlsbad, CA). For mRNA verification, 1 μ g total RNA from each caste was reversed transcribed using Multiscript and random priming (ABI, Carlsbad, CA). The abundance of specific cDNAs was quantified using SYBR ReadyMix (Sigma, St. Louis, MO) and an ABI 7900HT real-time PCR system. Each sample was tested in triplicates and normalized to 18S RNA. For miRNA quantification, 1 μ g total RNA was poly-A tailed and reverse transcribed using the NCode VILO miRNA cDNA synthesis kit (Invitrogen, Carlsbad, CA), then quantified with SYBR Readymix and normalized to 5S RNA.

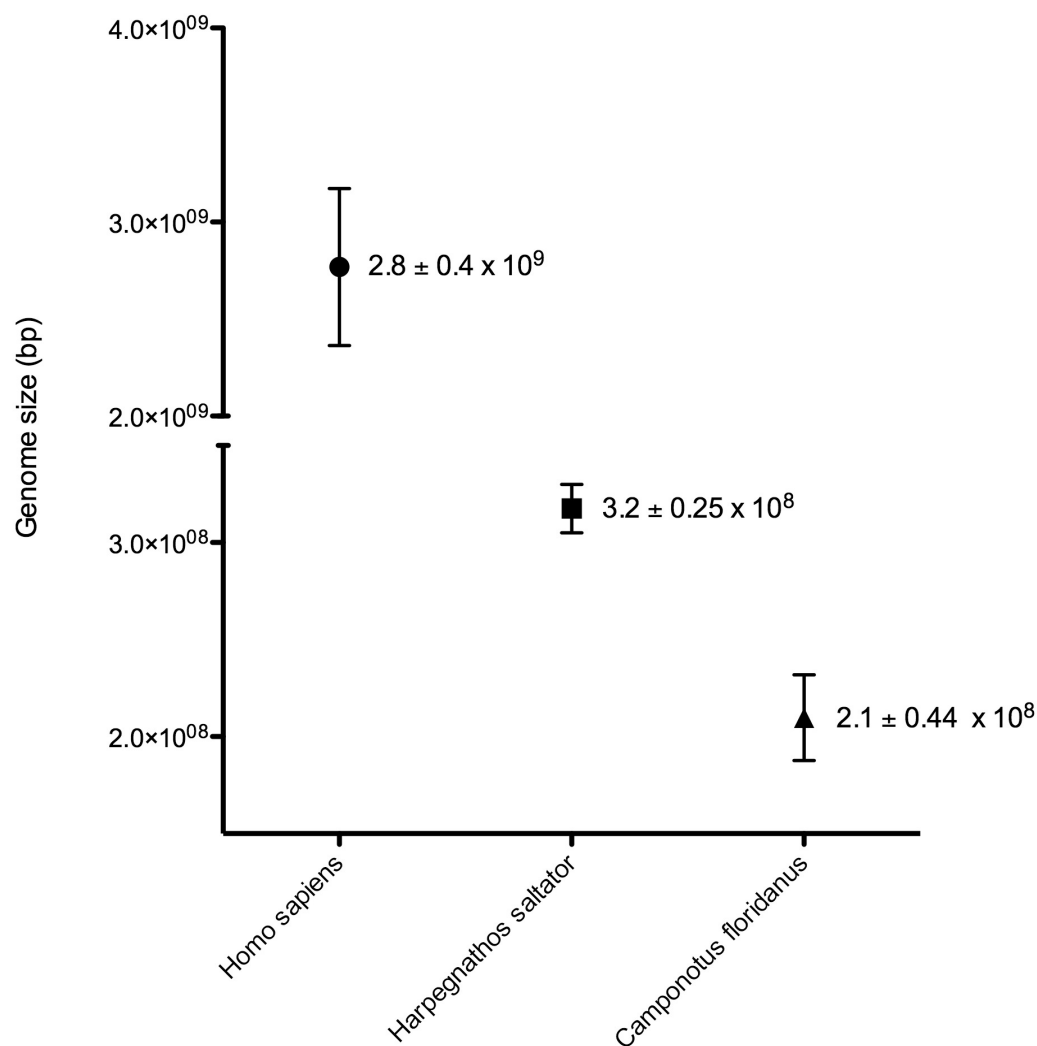
Quantification of DNA methylation

Genomic DNA was extracted from larvae or adult workers of *C. floridanus* and *H. saltator* and from HeLa cells as a positive control as well as from a strain of *E. coli* defective for all forms of DNA methylation (*dam-/dcm-*) as a negative control. 300 ng of genomic DNA and serial 1:3 dilutions were spotted on nitrocellulose and blotted with anti-5-methyl-cytosine antibodies (MAb-5MECYT, Diagenode, NJ) the signal was quantified, then antibodies were stripped and the membrane was re-probed with anti-total DNA antibodies (MAB030, Millipore, MA), to verify equal loading. Films were scanned and the intensity of each signal measured by densitometry using ImageJ (S87). Values from the middle of the linear range were used to determine the 5-methyl-cytosine/total DNA ratio, which was finally normalized to the same ratio for HeLa cells, which was considered to be constant across experiments.

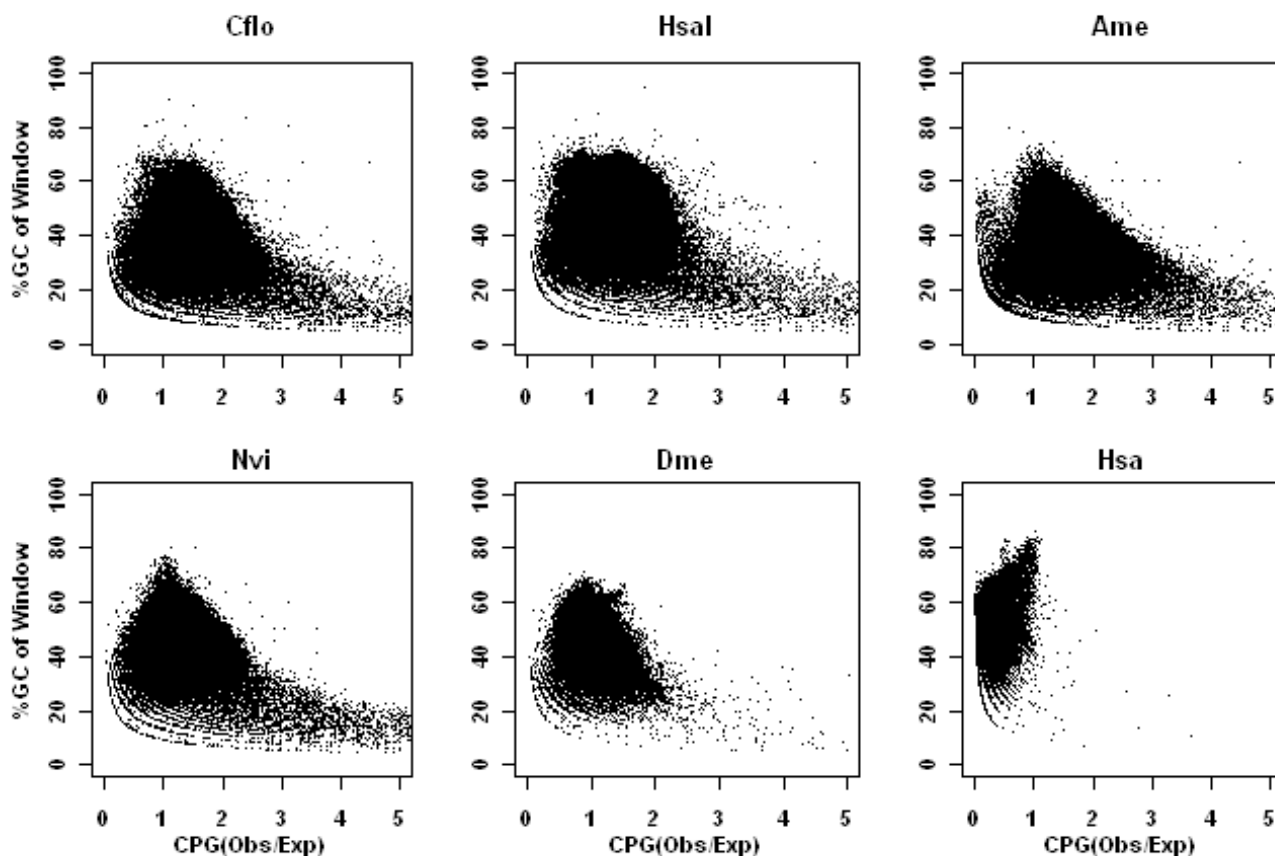
RNAi

Newly emerged *C. floridanus* adult workers were injected with approx 100 nl of an equimolar mixture (40 μ M) of two siRNA: stealth_67 Sense, GAG CUC GCU AUU CUU UAC AAC UGC A; antisense, UGC AGU UGU AAA GAA UAG CGA GCU C; stealth_425 Sense, GCG AUU UCU UCG GCG AGA AAG CUU U; antisense, AAA GCU UUC UCG CCG AAG AAA UCG C. The injections of siRNAs were performed with an Oxford micromanipulator (Singer Instruments, UK), a picospritzer II (General valve corporation, Fairfield, NJ), and an ordinary stereomicroscope. The injection pipettes were made from borosilicate glass capillary tubes (o.d. 1 mm, i.d.: 0.75 mm) (World Precision Instruments, Sarasota, FL), and pulled by a Brown micropipette puller (Model P-2000, Sutter Instrument Company, Novato, CA). The injection time was 0.12 s, the injection pressure was 85 kPa, and the balance pressure was 10 kPa. Injected adults were placed in a small nest separate from the originating colony and harvested directly into liquid nitrogen for RNA extraction.

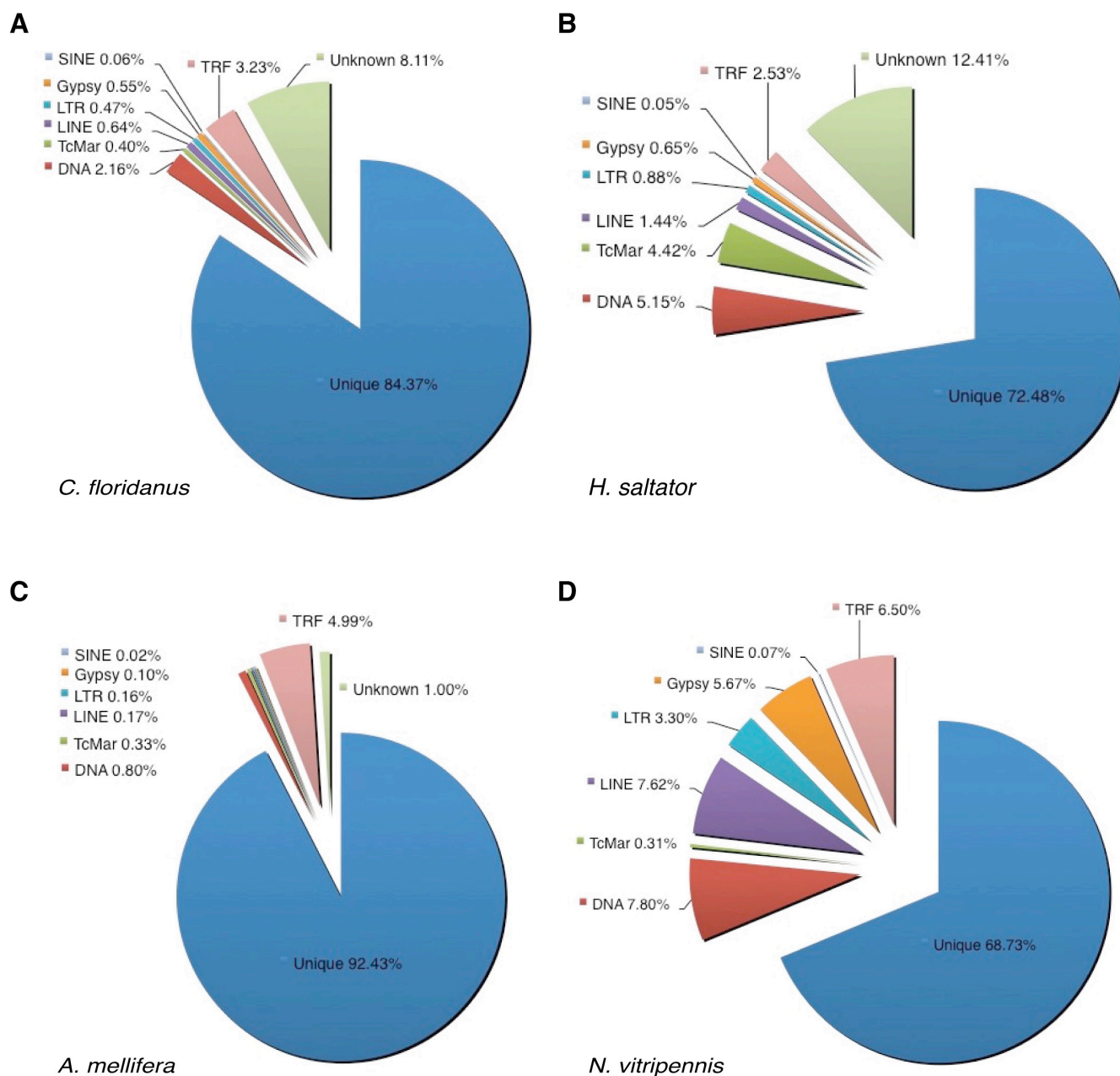
SUPPLEMENTARY FIGURES & LEGENDS



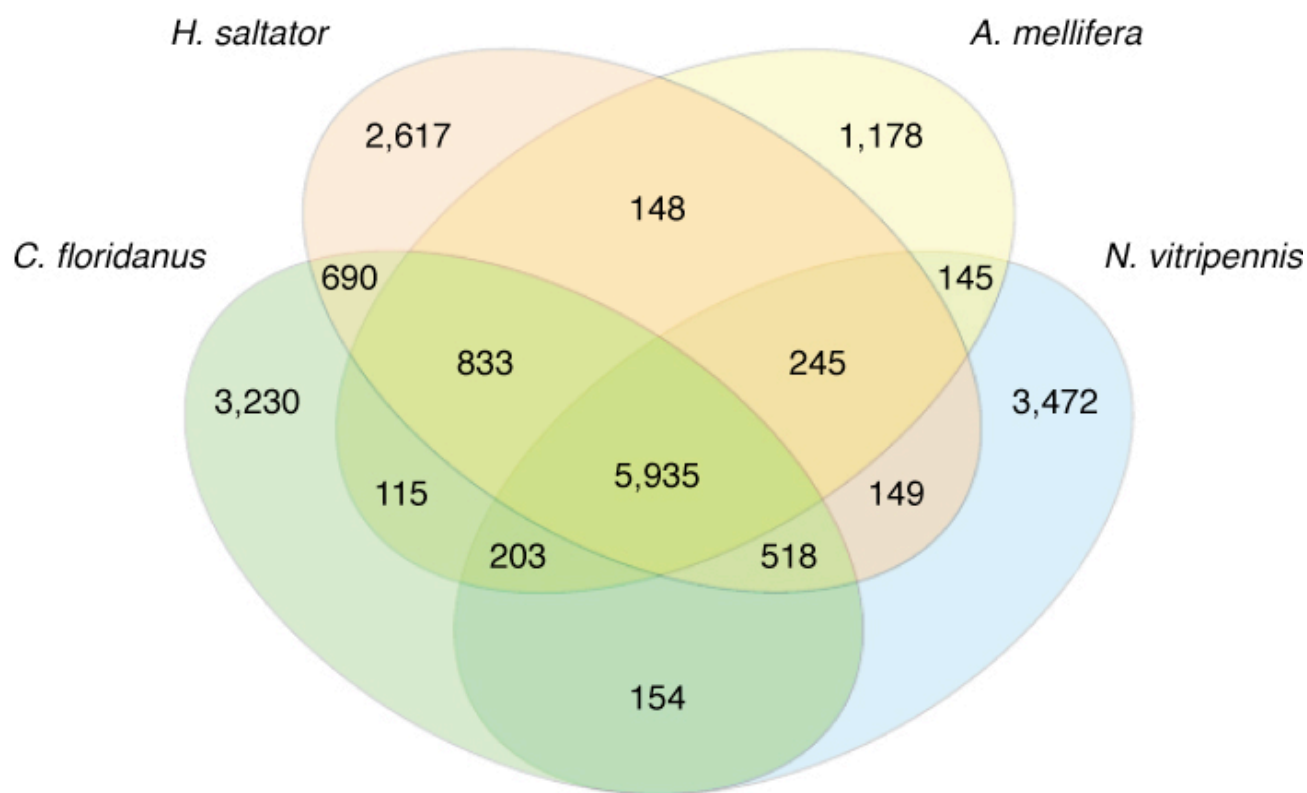
Supplementary figure S1. Size estimate for the genomes of *C. floridanus* (black triangle), *H. saltator* (black square) and *H. sapiens* (black circle) by qPCR (S88). Values shown are averaged from 2 independent qPCR reactions, each on 2 separate genomic loci and performed on 4 independent genomic DNA preparations.



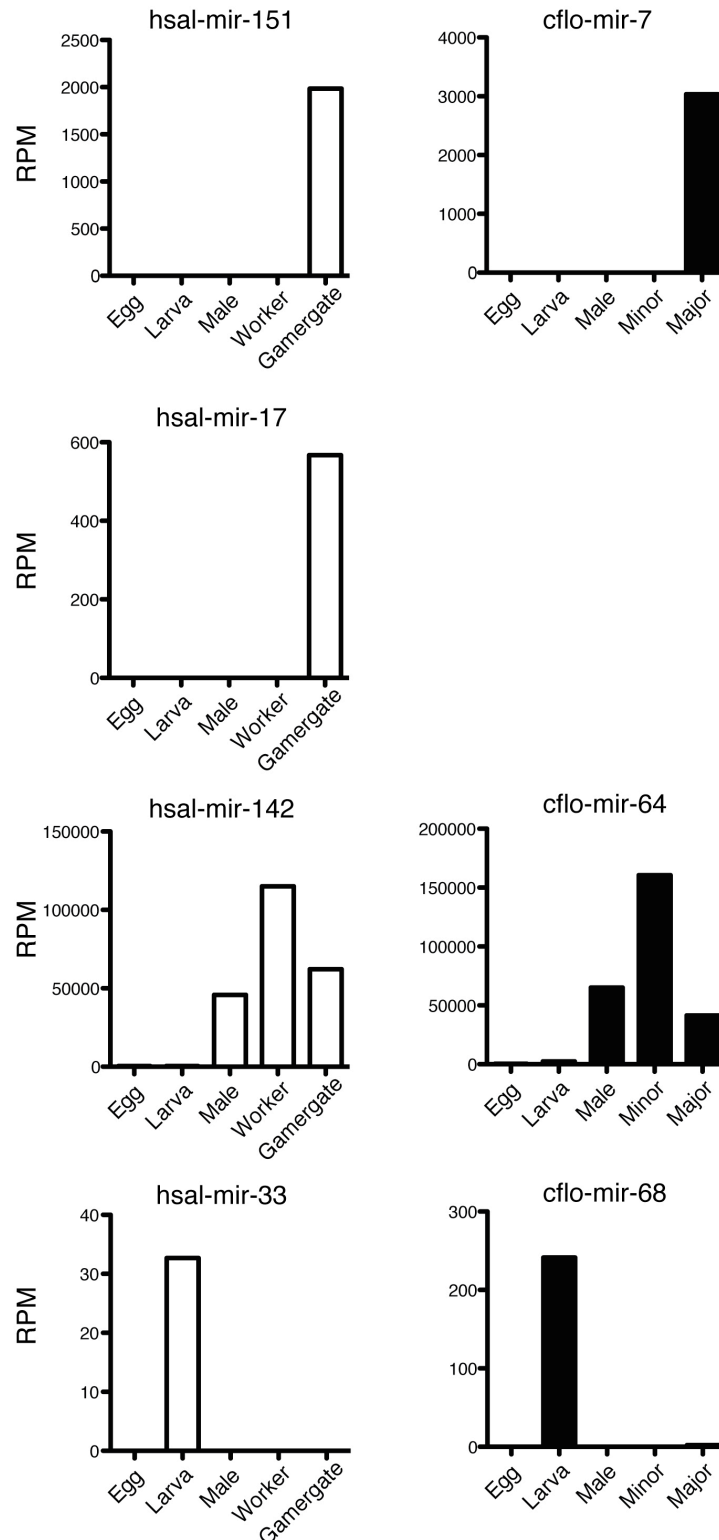
Supplementary figure S2. CpG observed/expected (Obs/Exp) distribution in insect and human genomes compared to G+C distribution. For this analysis a 500 bp sliding window was utilized. Data is plotted as in (S89). Note that in *H. sapiens* genome there is a suppression of CpG sequences, due to evolutionary fixation of meC->T deaminations (S89), whereas all Hymenoptera genomes have abundant CpG dinucleotides, especially in low G+C domains, despite possessing functional DNA methyltransferases.



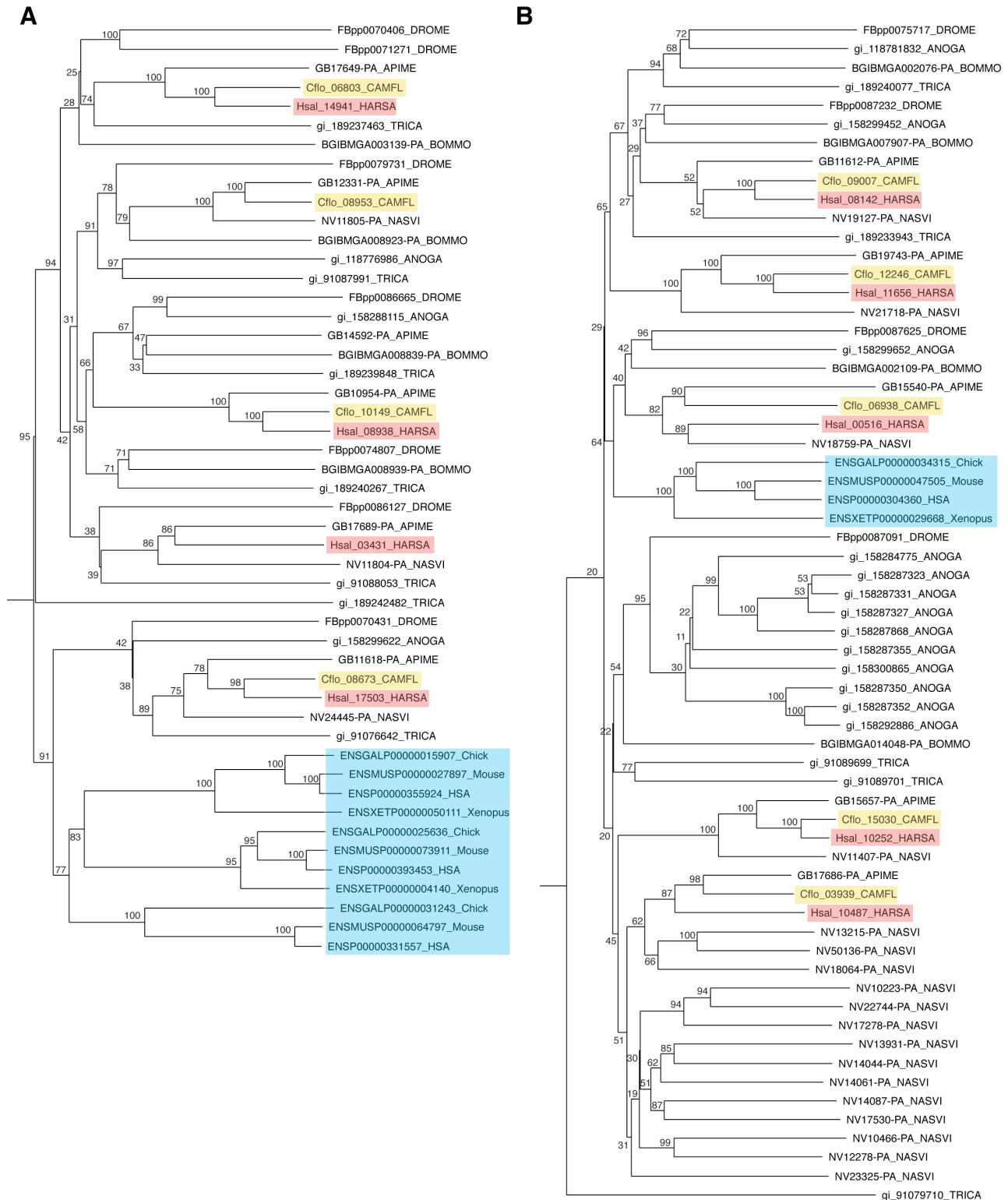
Supplementary figure S3. Repetitive sequences. The composition of the four sequenced Hymenoptera genomes in unique sequences and major classes of repetitive sequences is displayed as pie charts. TRF: sequences flagged by Tandem Repeat Finder; Unknown: sequences flagged by RepeatScout but not matching any known item in RepBase.



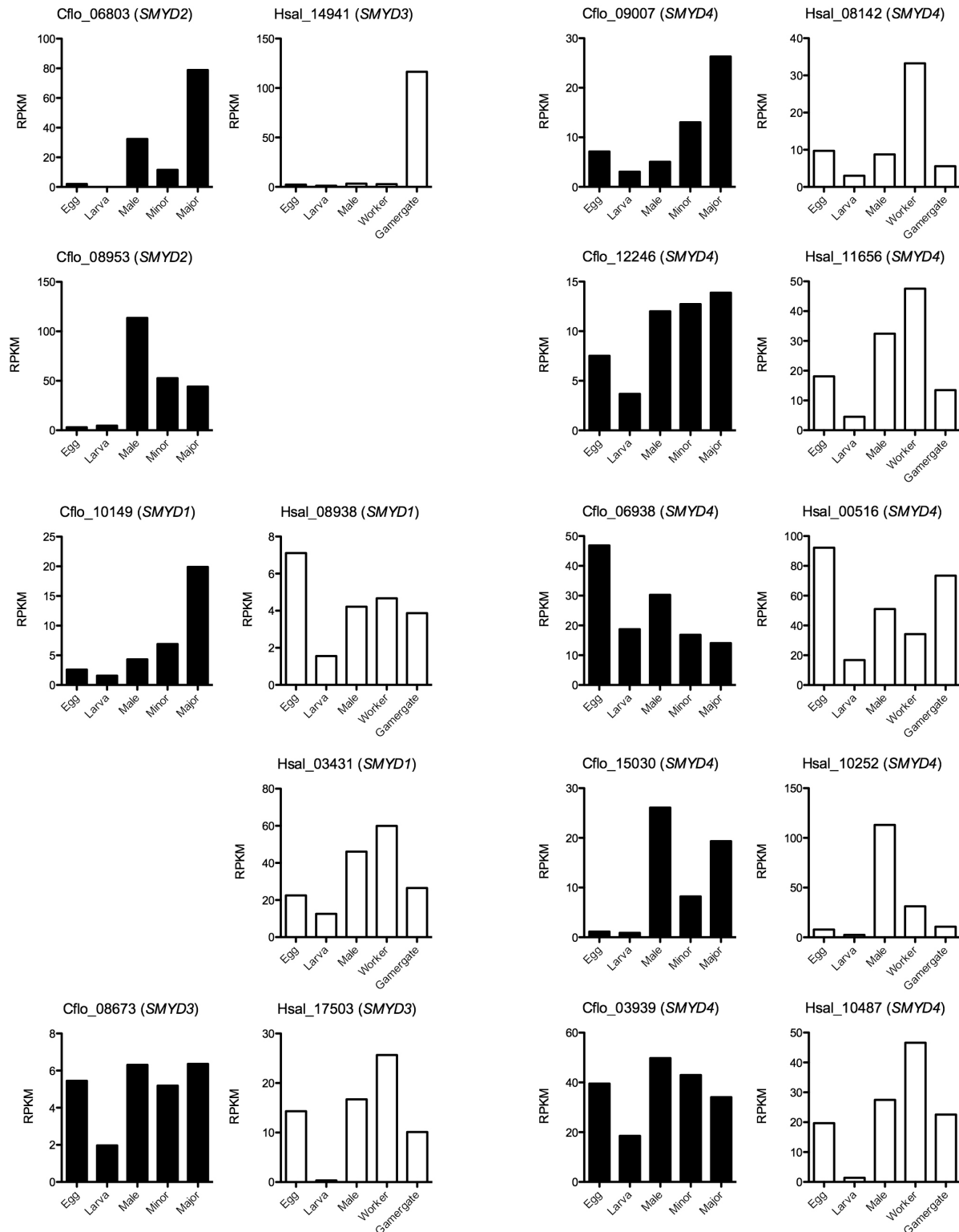
Supplementary figure S4. Conservation of protein families in four Hymenoptera. Orthology relationships were determined in all sequenced Hymenoptera and are shown in a non-weighted Euler-Venn diagram. The numbers indicate protein families in each subset.



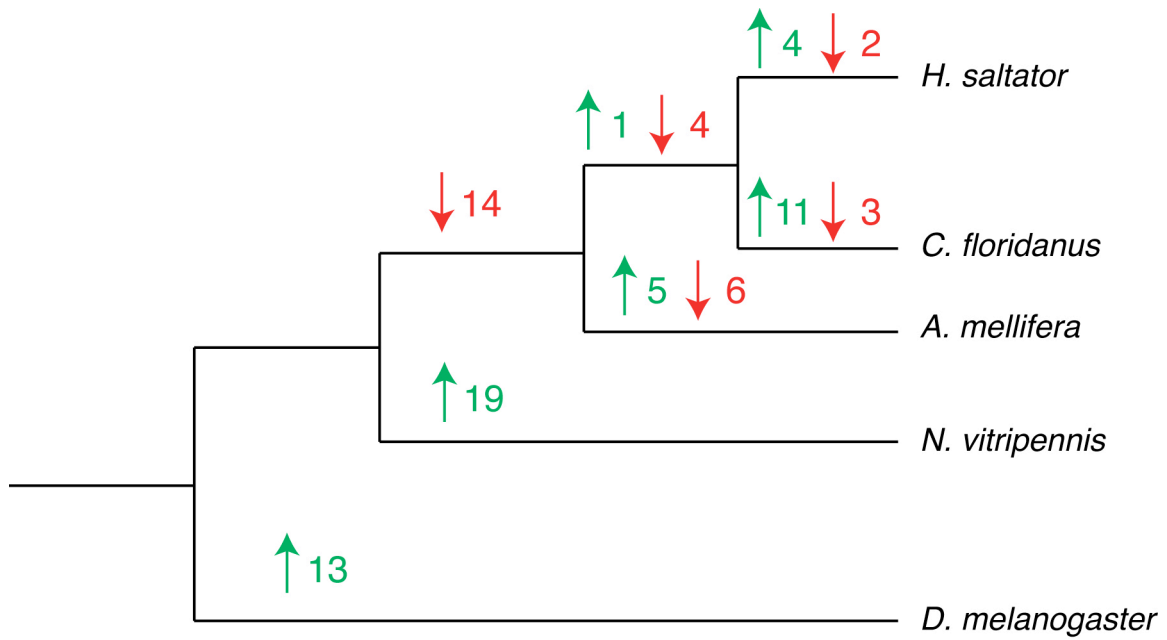
Supplementary figure S5. Differential expression of conserved miRNA in *C. floridanus* and *H. saltator*. The expression levels of the indicated miRNA (see SOM text for details) were determined by small-RNA seq and are shown in the bar plots are Read Per Million reads (RPM).



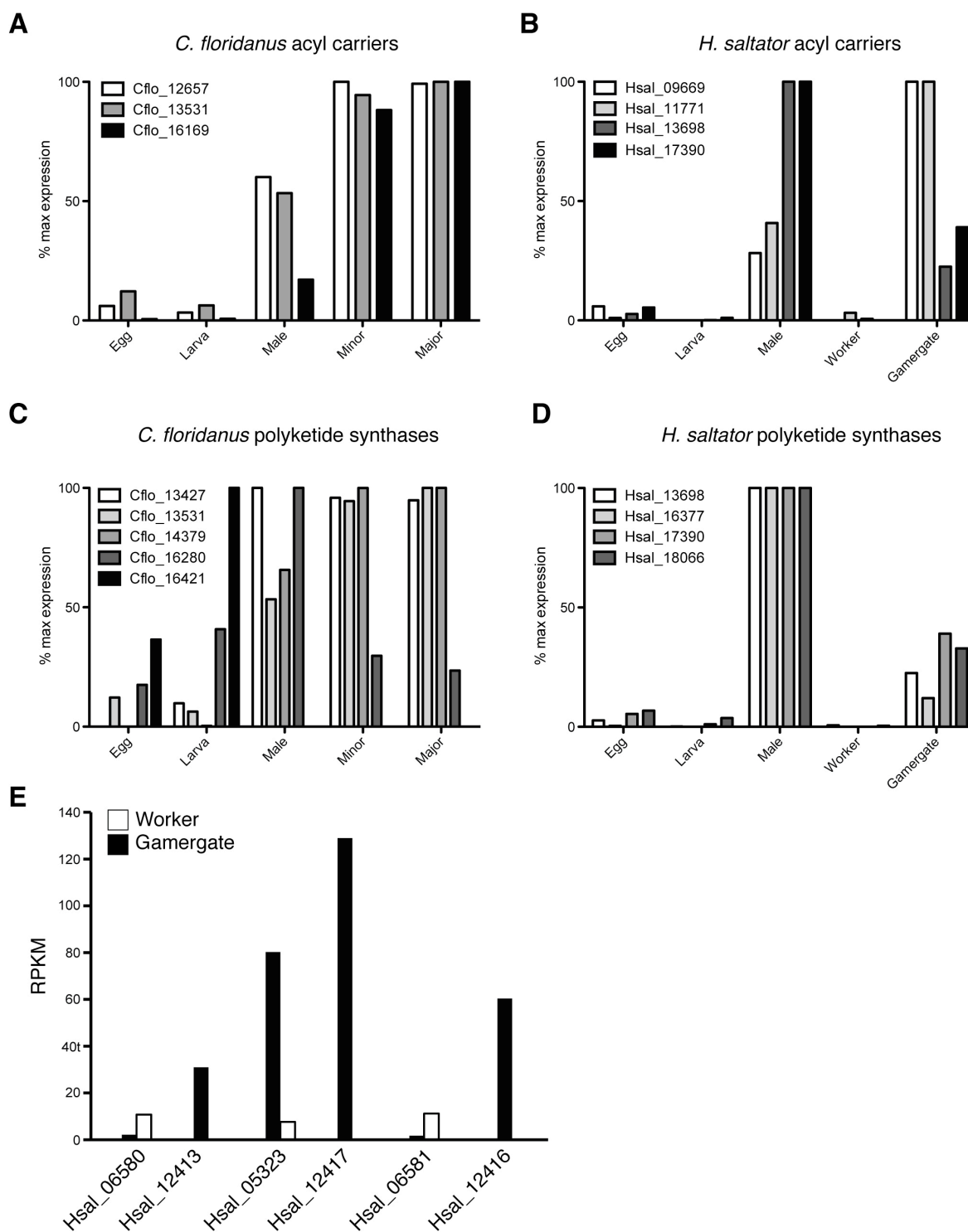
Supplementary figure S6. Expansion of the *SMYD* family of lysine methyl-transferases in insects. Phylogenetic tree for the *SMYD1-3* (A) and *SMYD4* (B) families in select vertebrate and insect genomes, constructed with the neighbor-joining method in MUSCLE (S81). Bootstrap values are from 1,000 trials. *C. floridanus* genes are highlighted in yellow, *H. saltator* in red, and vertebrates homologues are highlighted in blue.



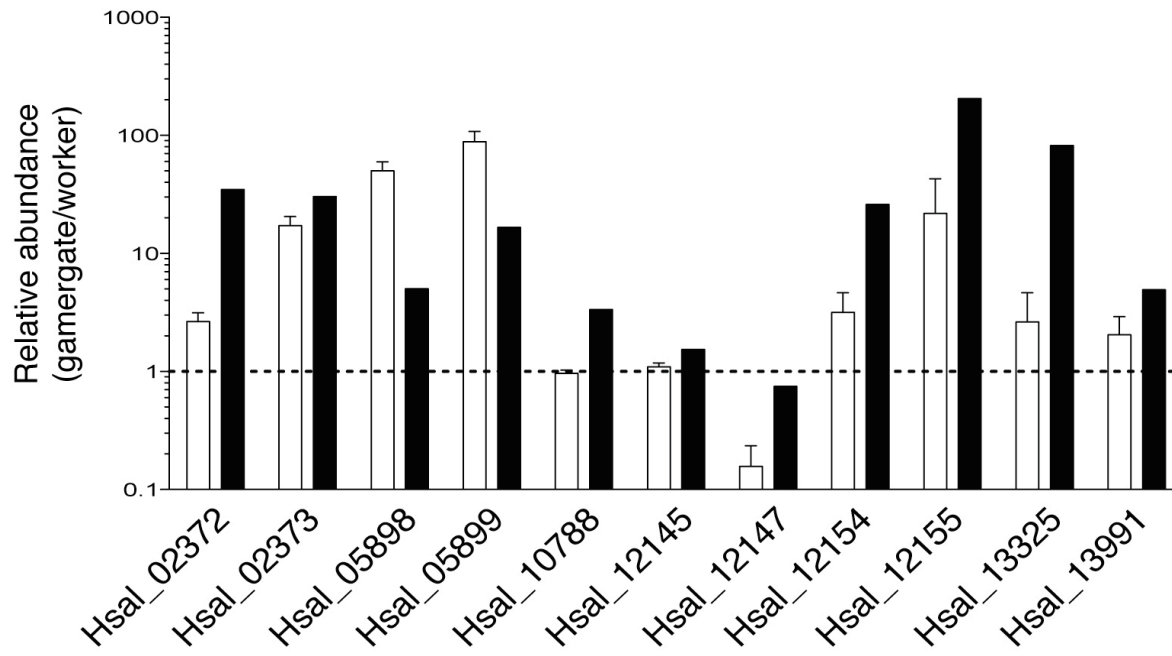
Supplementary figure S7. Differential expression of selected SMYD family genes in *C. floridanus* (black bars) and *H. saltator* (white bars) analyzed by RNA-seq. Orthologous genes, as determined by the tree shown in **Fig. S6**, are depicted side by side. Gene names in parentheses indicate the closest human homologue as determined by BLASTP.



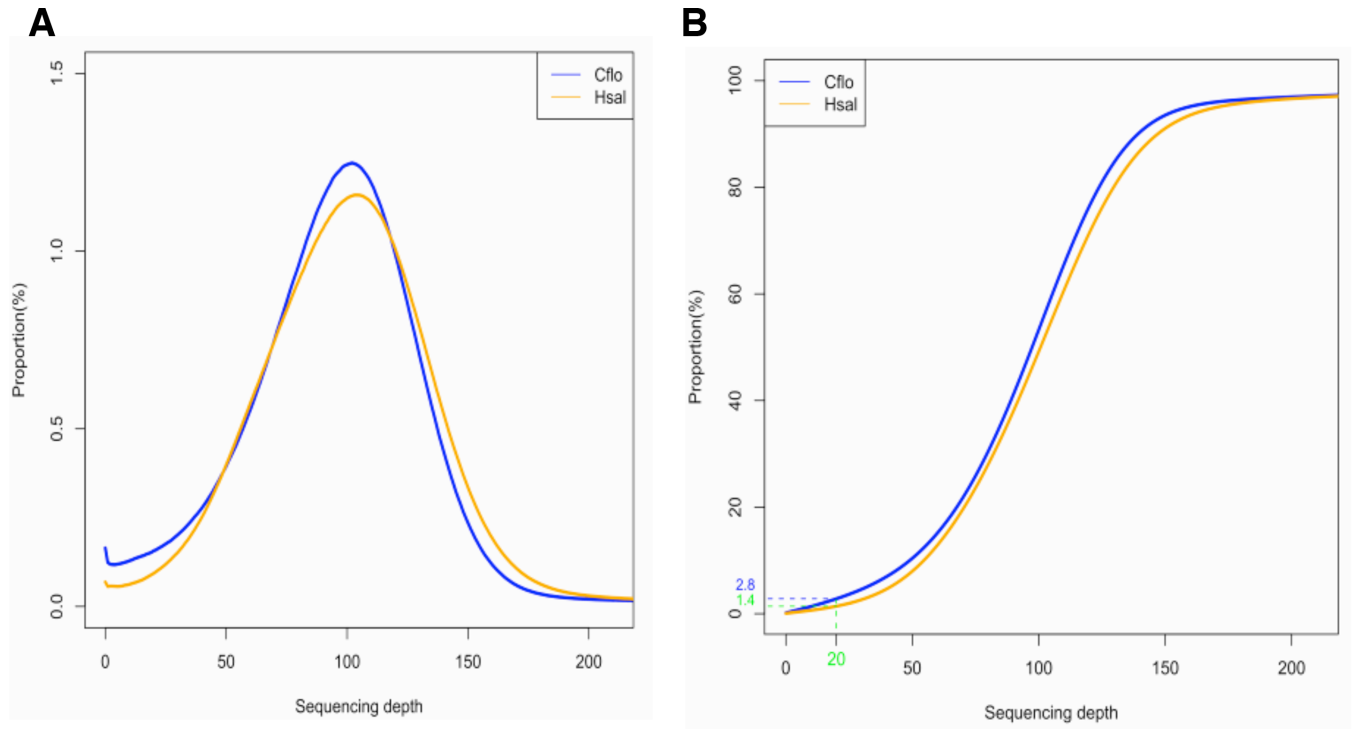
Supplementary figure S8. Size variation in orthologous protein families between 4 Hymenoptera and *D. melanogaster*. The number of protein families detected as significantly expanded (green) or contracted (red) is indicated for each branch. The identity of the protein families is reported in **table S18**.



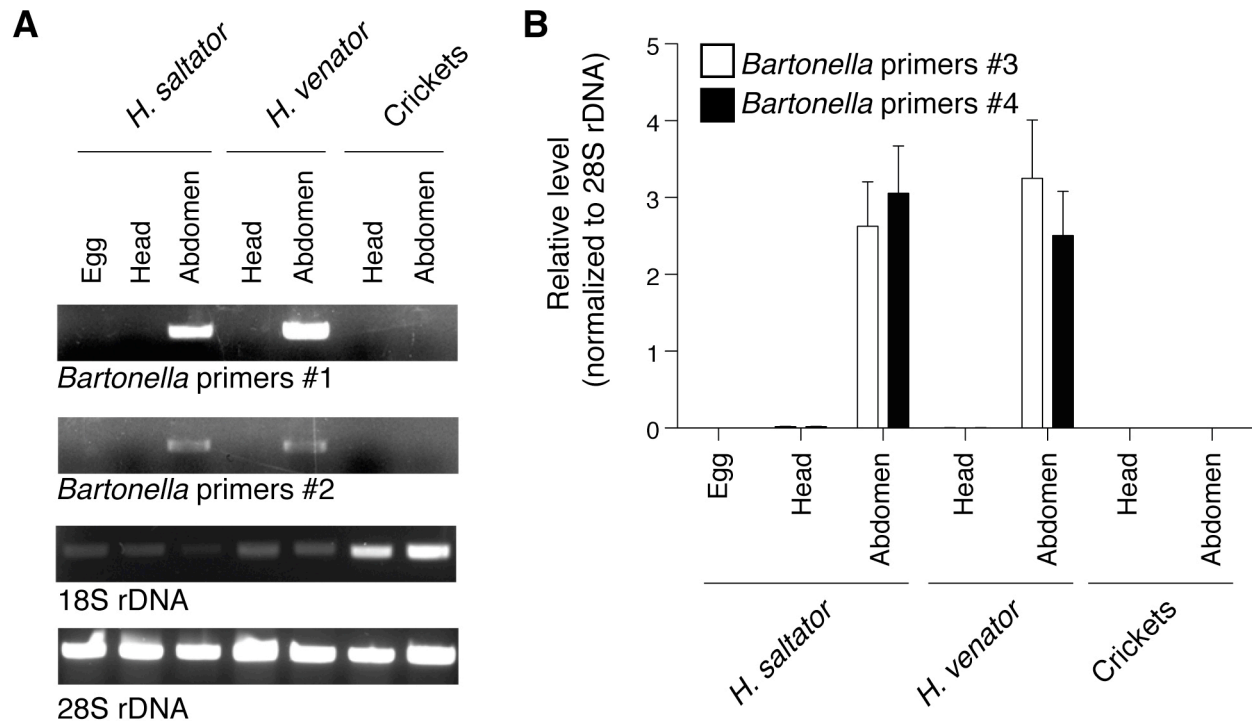
Supplementary figure S9. Differential expression of ant-specific genes containing acyl carrier protein (IPR009081) (A,B) polyketide synthase, enoylreductase domain (IPR020843) (C,D) and homologue to *Drosophila* CG6178, a long-chain fatty acyl-CoA synthetase.



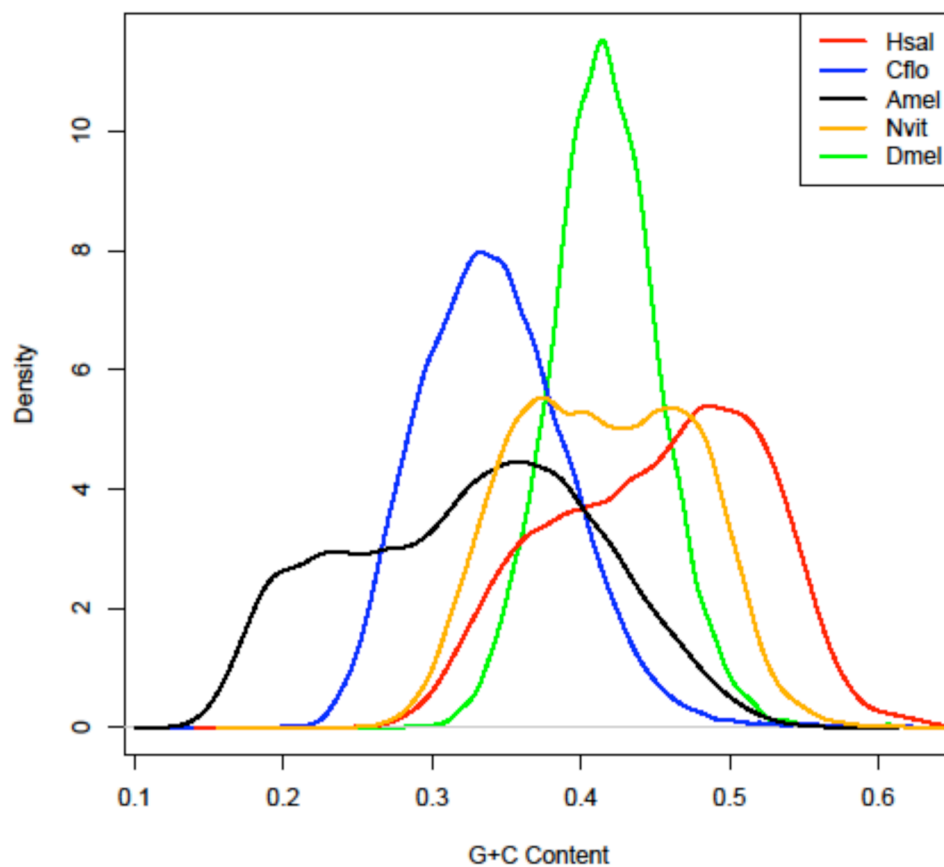
Supplementary figure S10. Upregulation of human ELOV homologs in *H. saltator* gamergates. Transcript level ratio in *H. saltator* gamergates vs. workers for the indicated genes is shown. White bars, qRT-PCR mean + SEM, n=7. Black bars RNA-seq.



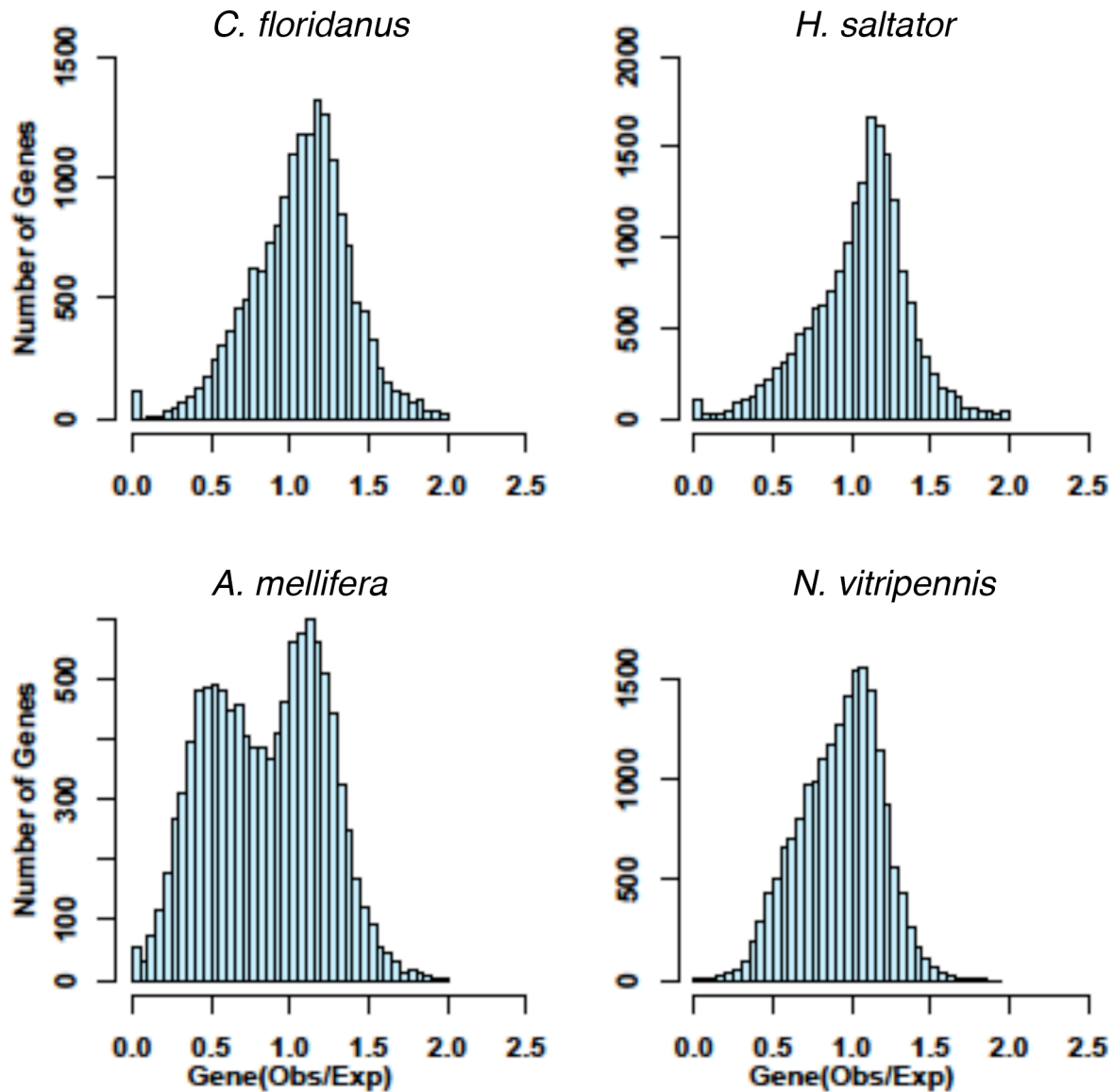
Supplementary figure S11. Sequence coverage plots. Raw sequencing reads were mapped to assembled scaffolds with SOAP (*SI*). Local (A) and cumulative (B) sequence coverage is shown. Regions with less than 20-fold coverage account for 2.8% of total assembled sequence for *C. floridanus* (Cflo) and 1.4% of total assembled sequence for *H. saltator* (Hsal).



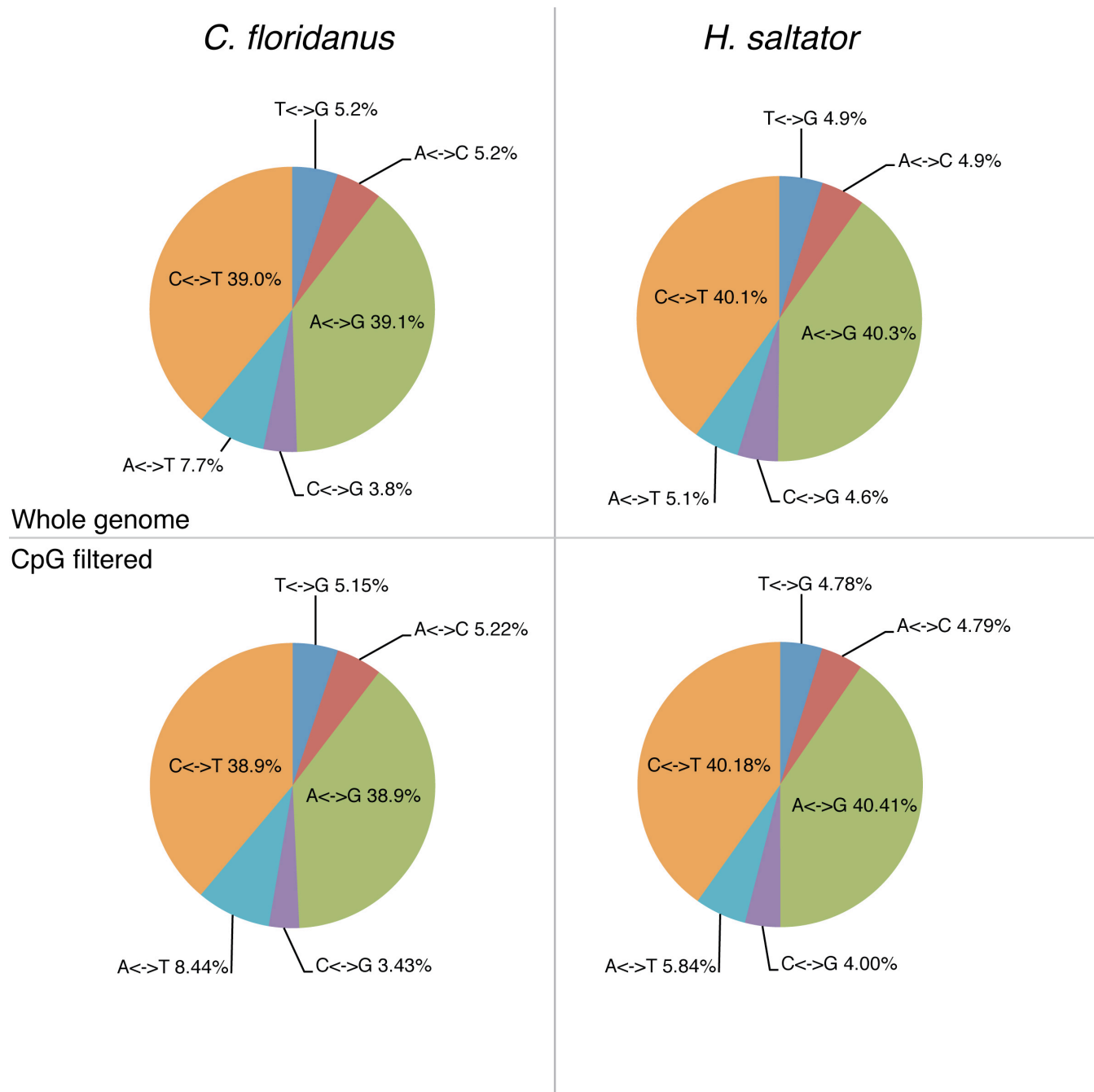
Supplementary figure S12. Detection of bacteria in *Harpegnathos* digestive tract. DNA samples extracted from *H. saltator* and *H. venator* egg, head, and abdomen sections or corresponding cricket sections were subjected to PCR analysis with primers for the putative *Bartonella* symbiont or the insects' 18S and 28S rDNA. Products were resolved on agarose gel (A) or quantified by qPCR (B).



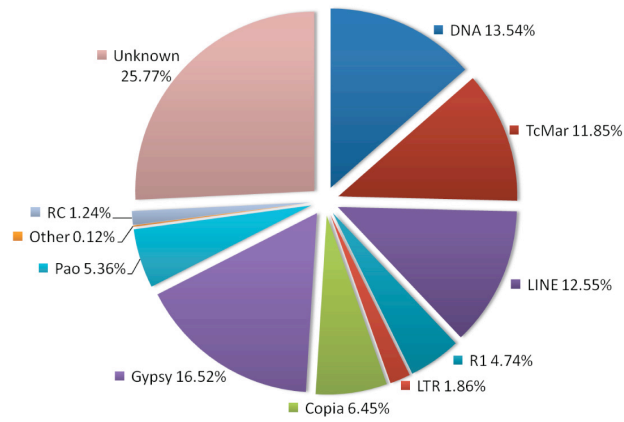
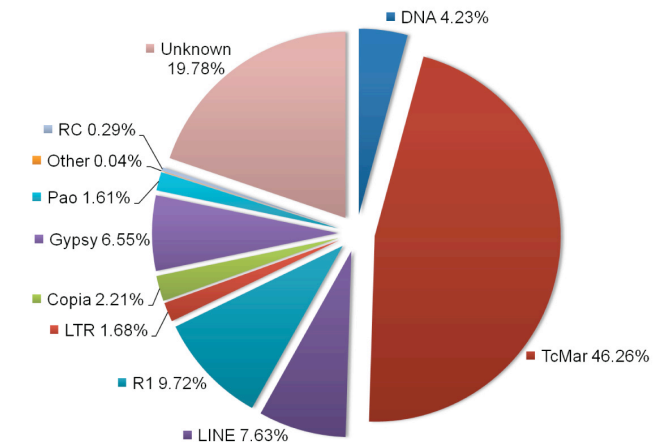
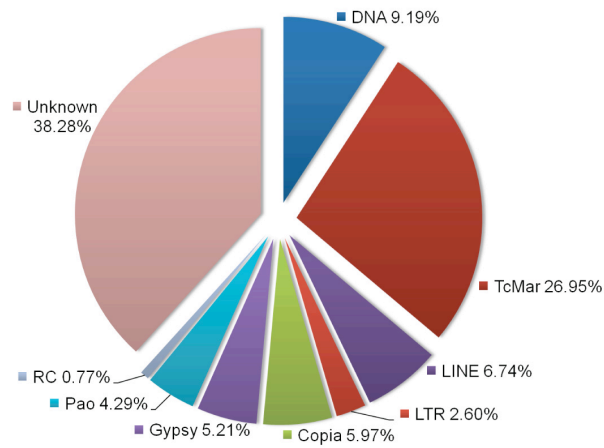
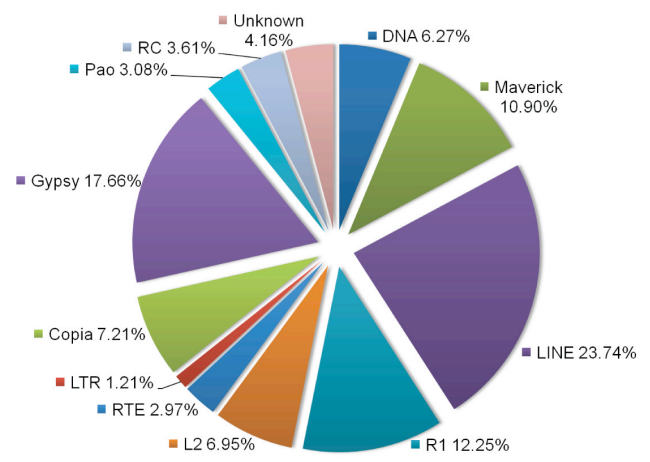
Supplementary figure S13. Comparison of G+C distribution in 5 insect genomes. A 10 kb sliding windows was passed through the genomes indicated in the legend and the fraction of G+C plotted.



Supplementary figure S14. Prevalence of CpG dinucleotides in the body of annotated genes. The observed/expected (O/E) ratio of CpG dinucleotides, based on G+C content, was calculated for each annotated gene in the four Hymenopteran genomes. Note that only *A. mellifera* displays a bimodal distribution, with a clearly distinct class of low-CpG genes.

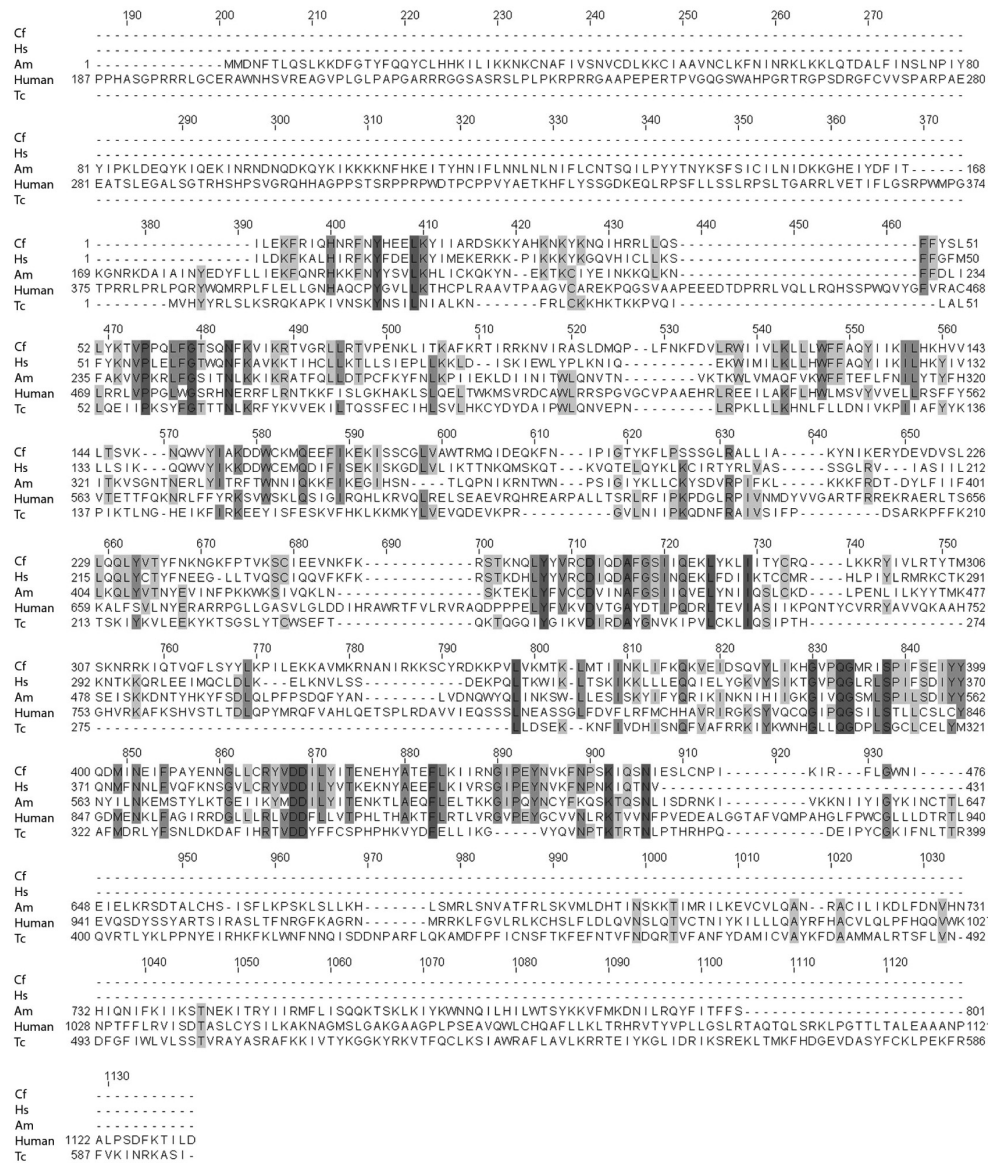


Supplementary figure S15. Polymorphism in the ant genomes. The pie charts represent the SNP distribution in the whole genome (upper panels) or after excluding CpG dinucleotides (lower panel).

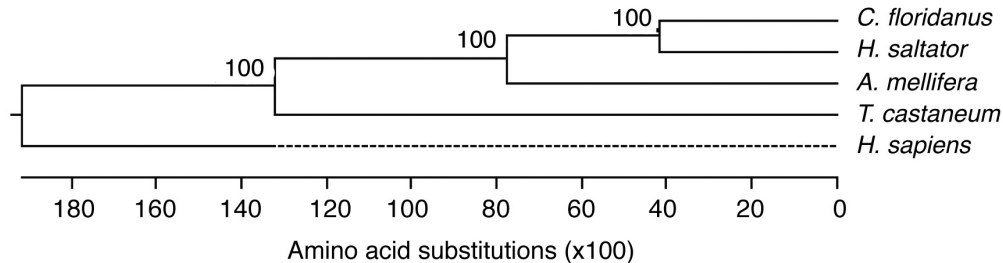
A*C. floridanus***B***H. saltator***C***A. mellifera***D***N. vitripennis*

Supplementary figure S16. Classification of active transposons in four Hymenoptera according to the RepBase database.

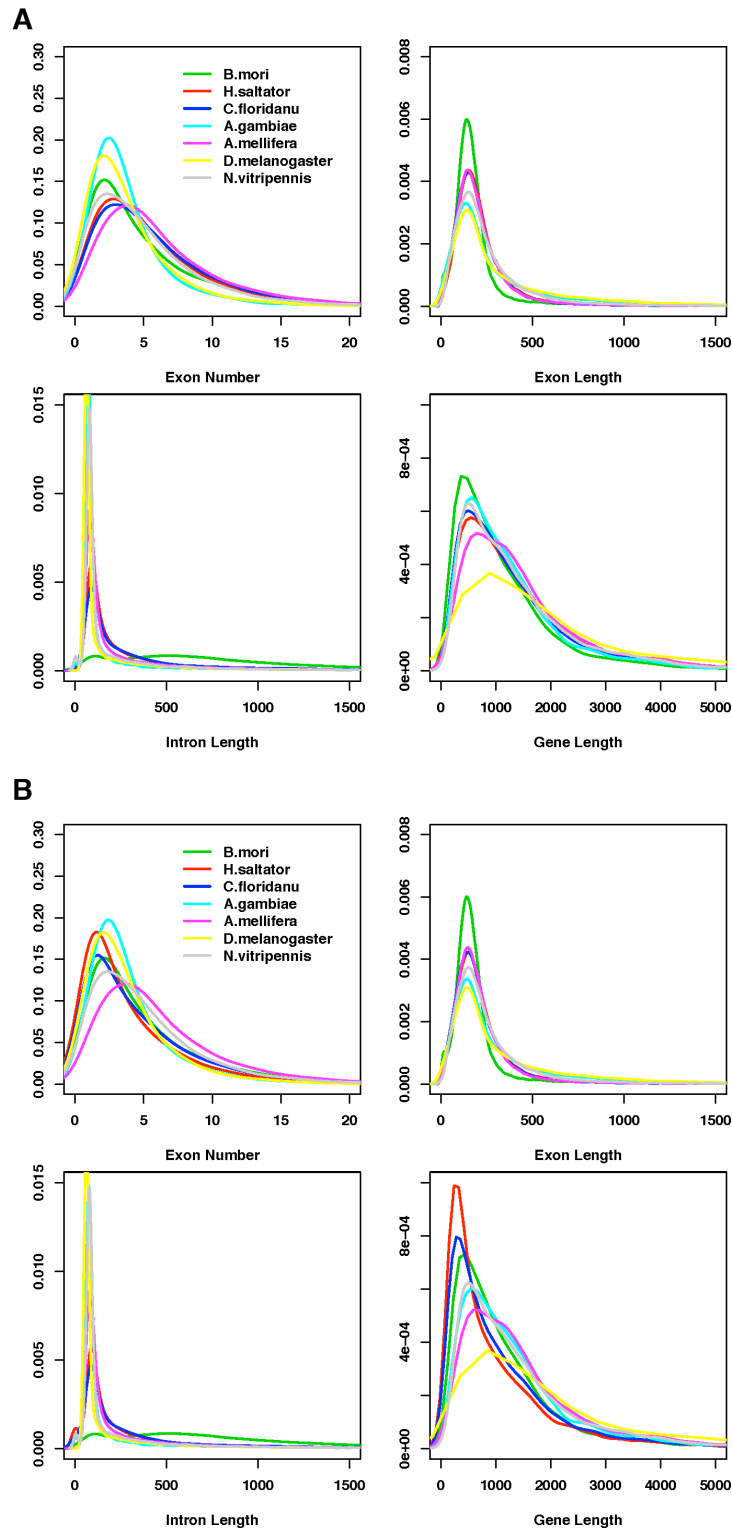
A



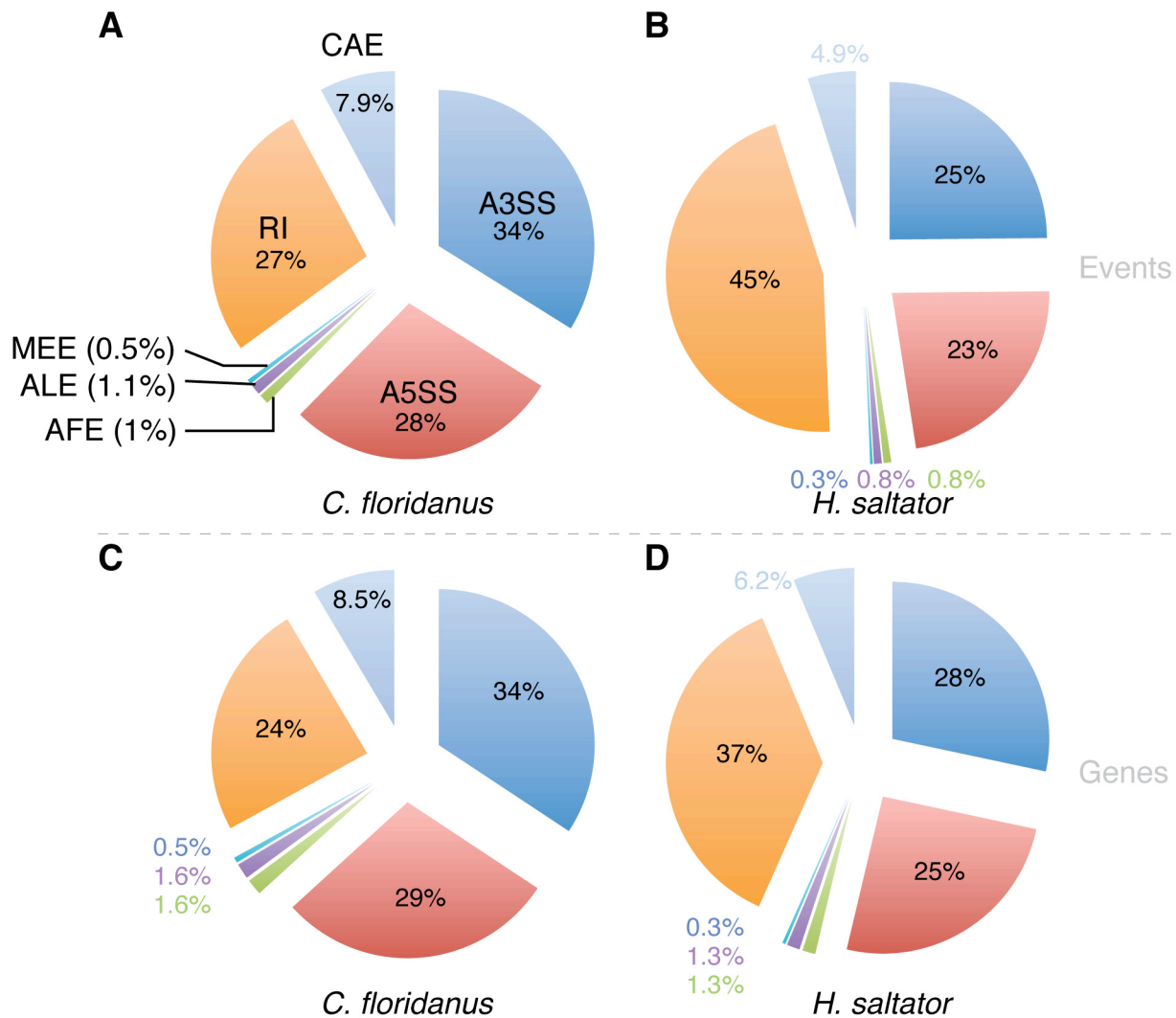
B



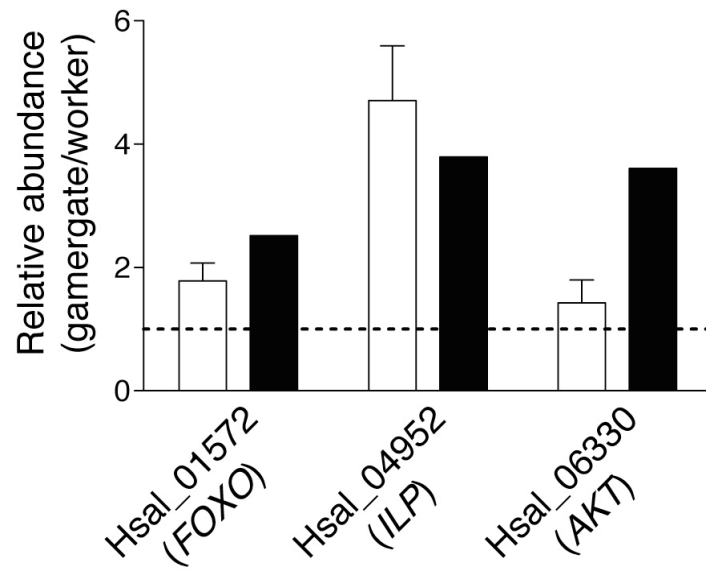
Supplementary figure S17. TERT gene in ants. (A) Alignment of putative TERT proteins from *C. floridanus* (Cf), *H. saltator* (Hs), *A. mellifera* (Am), *T. castaneum* (Tc) and *H. sapiens* (Human) generated by ClustalW with default parameters (Gonnet series, gap penalty=10, gap length penalty=0.2). Conserved amino acids are highlighted. (B) Phylogenetic tree generated by ClustalW with default parameters, and bootstrap values (1,000 trials).



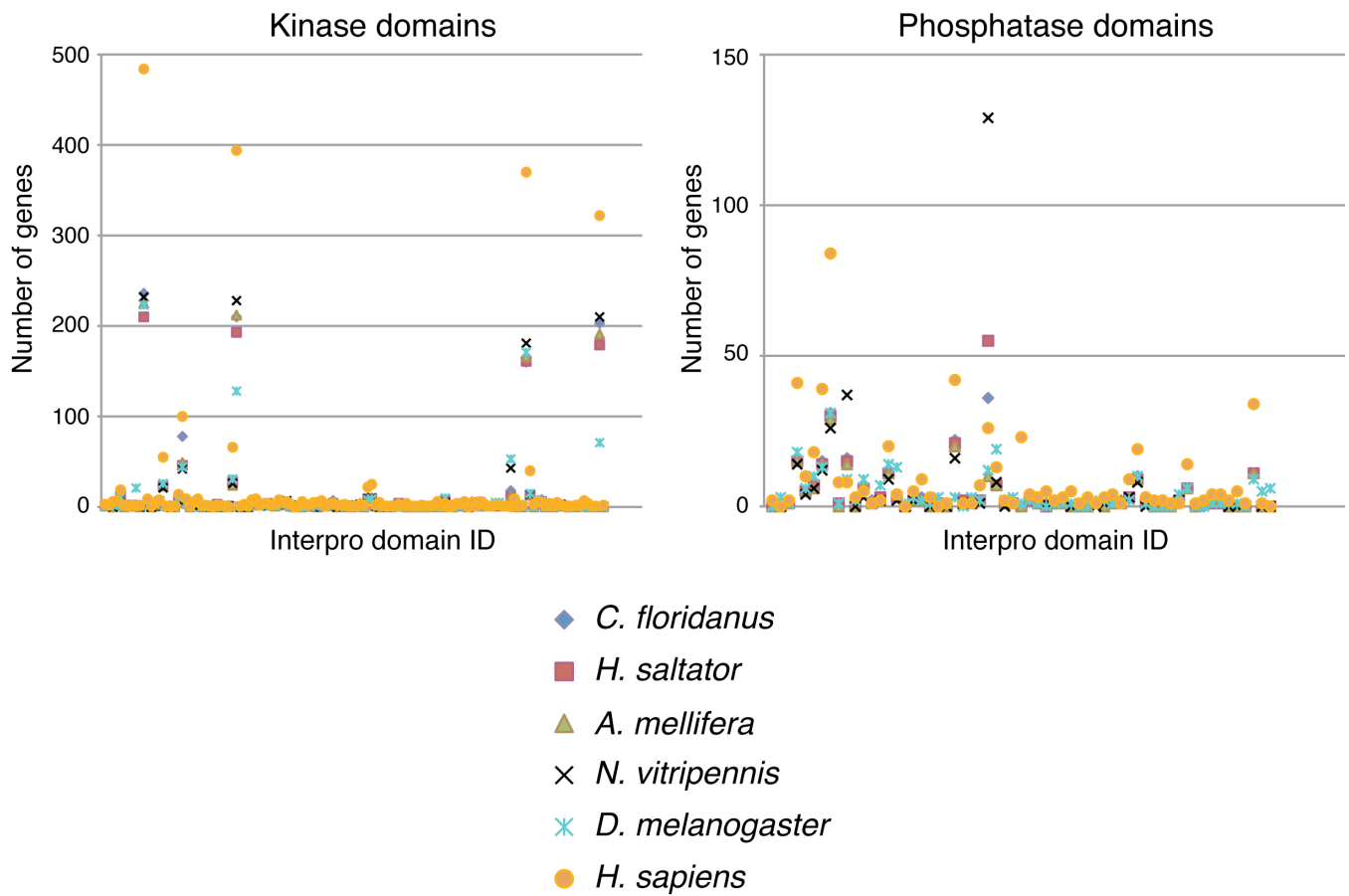
Supplementary figure S18. Quality of the gene models in annotation v3.3. The distribution for basic gene structure parameters in OGS3.3 is plotted and compared to the annotated genes in other insect genomes using (A) only gene models with complete ORFs, or (B) all annotated gene models.



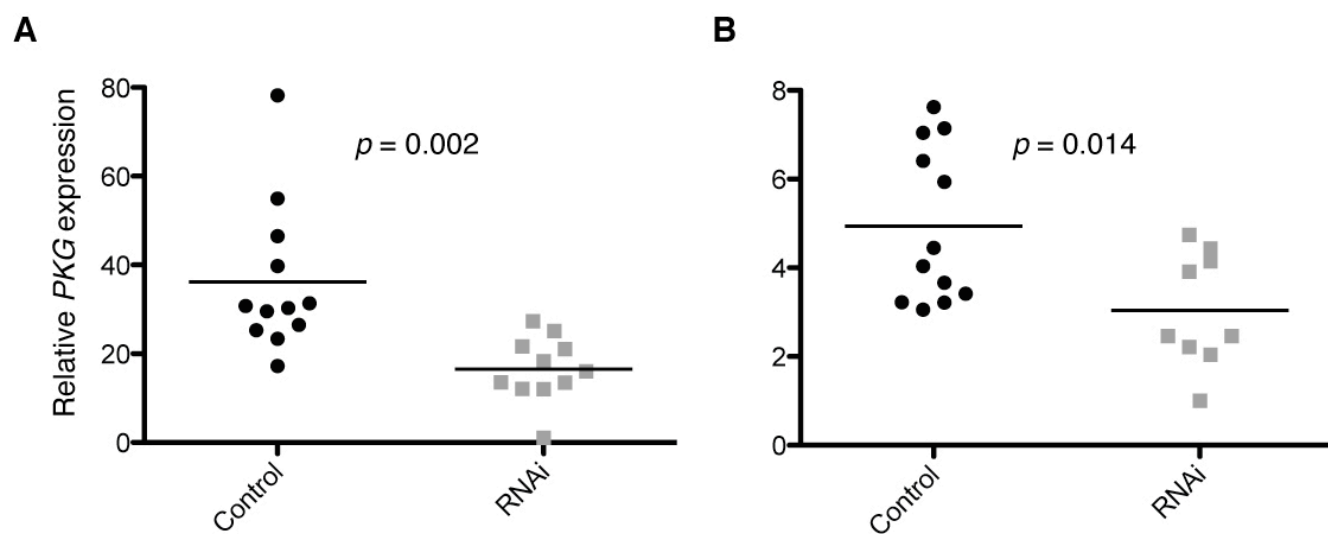
Supplementary figure S19. Genome-wide alternative splicing. Seven categories of alternative splicing events were evaluated in the genomes of *C. floridanus* (A, C) and *H. saltator* (B, D): cassette alternative exons (CAE), alternative 5' splice sites (A5SS), alternative 3' splice sites (A3SS), retained introns (RI), mutually exclusive alternative exons (MEE), alternative first exon (AFE), and alternative last exon (ALE). (A, B) Relative frequency of each category among all AS events. (C, D) Relative frequency of genes affected by one or more events in each category.



Supplementary figure S22. IIS pathway upregulation in *H. saltator* gamergates. Transcript level ratio in *H. saltator* gamergates vs. workers for *FOXO*, *ILP*, and *AKT* orthologues. White bars, qRT-PCR mean + SEM, n=7. Black bars RNA-seq. Gene names in parentheses indicate the closest human homologue as determined by BLASTP.



Supplementary figure S23. IPR kinase/phosphatase domain representation in gene models from different species. Individual IPR domains associated with kinases (left) or phosphatases (right) were sorted by IPR ID along the x axis and the number of genes in each species containing the domain is plotted on the y axis. Values for 130 kinase and 61 phosphatase domain are shown.



Supplementary figure S24. Knock-down of *PKG* expression in *C. floridanus* workers by RNAi. (A) RNAi in minor workers. (B) RNAi in major workers. Each square represent a single injected individual.

SUPPLEMENTARY TABLES

Supplementary table S1. Assembly statistics for the *C. floridanus* and *H. saltator* genomes.

<i>C. floridanus</i> (v4)				
	contig size (bp)	contig #	scaffold size (bp)	scaffold #
Longest	432,819	1	2,766,119	1
N50	24,134	2,523	602,923	109
N90	4,782	10,612	46,351	614
Total	231,207,942	41,879	238,222,169	25,494
<i>H. saltator</i> (v3)				
	contig size (bp)	contig #	scaffold size (bp)	scaffold #
Longest	477,250	1	2,276,656	1
N50	38,027	2,055	598,192	145
N90	5,790	9,333	36,568	885
Total	284,867,512	41,437	296,783,122	21,347
<i>C. floridanus</i> (v3)				
	contig size (bp)	contig #	scaffold size (bp)	scaffold #
Longest	708,316	1	2,671,896	1
N50	19,315	2,842	444,386	149
N90	3,310	13,824	32,987	816
Total	227,253,388	47,418	235,584,879	24,031

Supplementary table S2. Fosmid validation of the genomic assemblies.

<i>C. floridanus</i> (V4)				
fosmid ID	length	coverage (%)	inconsistencies	scaffold size
dantcaxa	33,353	98.3	4	315,556
dantcbxa	41,458	99.4	6	1,146,372
dantccxa	37,471	98.1	3	1,193,351
dantcdxa	34,485	98.2	6	1,177,991
dantcexa	35,892	100	2	1,360,401
dantcfxa	41,712	95.9	4	1,360,401
dantchxa	38,623	99.8	7	434,544
dantcjxa	32,233	98.5	3	1,111,256
dantckxa	35,565	99.4	4	433,063
Mean	36,755	98.6	4.3	948,104
<i>H. saltator</i> (V3)				
fosmid ID	length	coverage	inconsistencies	scaffold size
danthaxa	40828	98.6	6	290,101
danthcxa	33654	99.7	4	1,978,266
danthdxa	44574	98.5	4	573,047
danthexa	33015	98.7	3	1,163,245
danthfxa	38031	97.8	5	699,624
danthgxa	31208	98.3	6	761,569
danthhxa	38136	98.4	12	771,335
danthjxa	45695	99.5	4	472,718
danthkxa	34599	99.4	9	984,739
danthlxa	36357	98.1	2	2,276,656
Mean	37610	98.6	6	997,130

Supplementary table S3. EST validation of the genomic assemblies.

<i>C. floridanus</i> (V4)	Total used	Total matched		50% length cutoff		90% length cutoff	
Length	EST #	#	match %	#	match %	#	match %
total	4,354	4,332	99.5	4,311	99	4,190	96.2
≥ 200 nts	4,254	4,239	99.6	4,218	99.2	4,108	96.6
≥ 500 nts	3,072	3,062	99.7	3,044	99.1	2,979	97
<i>H. saltator</i> (V3)							
total	4,345	4,335	99.8	4,334	99.7	4,137	95.2
≥ 200 nts	4,284	4,276	99.8	4,275	99.8	4,089	95.4
≥ 500 nts	2,933	2,927	99.8	2,927	99.8	2,825	96.3

Supplementary table S4. Genomic features of *C. floridanus* and *H. saltator* compared to other Hymenoptera and *D. melanogaster*.

	<i>C. floridanus</i>	<i>H. saltator</i>	<i>A. mellifera</i>	<i>N. vitripennis</i>	<i>D. melanogaster</i>
Genome size (Mb)	238.2	296.7	228.5	295.1	168.7
G+C (%)	34	45	33	42	42
CpG O/E	1.58	1.49	1.65	1.35	0.94
# gene models	17064	18564	10738	18729	14108
Average CDS (bp)	1496.59	1540.17	1650.6	1480.8419	1616.16
Average intron (bp)	793.5	689.64	1266.9	1111.05	922.43
Gene density (kb/gene)	14.2	16.5	17.8	21	12.2
Repeats (%)	15.05	26.86	6.86	24.31	27.38

Supplementary table S5. G+C and dinucleotide composition of two ant genomes compared to *A. mellifera*, *N. vitripennis*, CM and *H. sapiens*.

	G+C	Dinucleotide Occurrence (Observed/Expected)								
	(%)	ApA	ApC	ApG	ApT	CpA	CpC	CpG	CpT	GpA
<i>C. floridanus</i>	34	0.8	0.83	0.82	1.07	0.92	0.78	1.58	0.82	1.07
<i>H. saltator</i>	45	0.87	0.85	0.87	1.04	0.85	0.7	1.49	0.87	1.13
<i>A. mellifera</i>	33	0.83	0.77	0.8	1.03	0.83	0.91	1.65	0.8	1.15
<i>N. vitripennis</i>	42	0.83	0.85	0.94	0.99	0.92	0.71	1.35	0.94	1.04
<i>D. melanogaster</i>	42	0.87	0.87	0.89	0.97	1.11	0.87	0.94	0.89	0.92
<i>H. sapiens</i>	41	0.81	0.83	1.16	0.88	1.2	0.99	0.34	1.16	0.98

Supplementary table S6. Size and gene content of segmental duplications in 5 insects.

Species	Genome (Mbp)	# SD	SD length (Mbp)	genome %	Genes in SD	Paralog gene pairs	Pairs shared in ants
<i>H. saltator</i>	296.8	256838	44.1	14.8	2858	705	79
<i>C. floridanus</i> (v3)	235.6	67407	22.6	9.6	1810	458	79
<i>A. mellifera</i>	228.6	22380	13	5.7	225	53	-
<i>N. vitripennis</i>	295.1	353548	45.6	15.5	2409	551	-
<i>D. melanogaster</i>	168.7	10413	6.3	3.7			

Supplementary table S7. GO enrichment analysis for gene models found in segmental duplications.

<i>C. floridanus</i>					
GO ID	GO Term	Class	Level	P value	Adj. P value
GO:0004497	monooxygenase activity	MF	4	1.6E-46	6.9E-44
GO:0009055	electron carrier activity	MF	2	1.5E-45	3.2E-43
GO:0020037	heme binding	MF	4	1.1E-44	1.1E-42
GO:0004984	olfactory receptor activity	MF	7	3.7E-41	2.7E-39
GO:0007608	sensory perception of smell	BP	8	3.7E-41	2.7E-39
GO:0005549	odorant binding	MF	3	9.0E-39	3.9E-37
GO:0070011	peptidase activity, acting on L-amino acid peptides	MF	5	1.8E-32	5.0E-31
GO:0016491	oxidoreductase activity	MF	3	2.5E-30	6.3E-29
GO:0004190	aspartic-type endopeptidase activity	MF	7	2.0E-25	4.4E-24
GO:0004888	transmembrane receptor activity	MF	5	6.8E-23	1.4E-21
GO:0004175	endopeptidase activity	MF	6	2.9E-18	5.2E-17
GO:0008237	metallopeptidase activity	MF	6	6.5E-14	9.6E-13
GO:0006508	proteolysis	BP	6	2.9E-13	3.9E-12
GO:0003824	catalytic activity	MF	2	3.7E-13	4.8E-12
GO:0043285	biopolymer catabolic process	BP	5	3.6E-12	4.4E-11
GO:0016891	endoribonuclease activity, producing 5'-phosphomonoesters	MF	8	6.3E-12	7.3E-11
GO:0004518	nuclease activity	MF	5	1.9E-11	2.1E-10
GO:0009056	catabolic process	BP	3	5.1E-10	5.2E-09
GO:0004713	protein tyrosine kinase activity	MF	7	1.1E-09	1.0E-08
<i>H. saltator</i>					
GO ID	GO Term	Class	Level	Pvalue	Adj. P value
GO:0046914	transition metal ion binding	MF	6	1.1E-66	3.6E-64
GO:0008270	zinc ion binding	MF	7	4.6E-63	7.6E-61
GO:0046872	metal ion binding	MF	5	1.2E-53	1.3E-51
GO:0003676	nucleic acid binding	MF	3	1.1E-21	5.2E-20
GO:0008236	serine-type peptidase activity	MF	5	3.2E-15	1.1E-13
GO:0004252	serine-type endopeptidase activity	MF	6	6.5E-15	2.0E-13
GO:0004984	olfactory receptor activity	MF	7	8.2E-13	1.9E-11
GO:0007608	sensory perception of smell	BP	8	8.2E-13	1.9E-11
GO:0005549	odorant binding	MF	3	1.9E-12	4.0E-11
GO:0004497	monooxygenase activity	MF	4	8.4E-10	1.2E-08
GO:0005488	binding	MF	2	9.9E-10	1.3E-08
GO:0003735	structural constituent of ribosome	MF	3	1.5E-09	1.8E-08
GO:0005840	ribosome	CC	4	1.5E-09	1.8E-08
GO:0030529	ribonucleoprotein complex	CC	3	2.2E-08	2.4E-07
GO:0020037	heme binding	MF	4	1.7E-07	1.6E-06
GO:0009055	electron carrier activity	MF	2	8.4E-07	6.2E-06

GO:0004523	ribonuclease H activity	MF	9	2.5E-06	1.6E-05
GO:0005506	iron ion binding	MF	7	5.2E-06	3.1E-05
GO:0017038	protein import	BP	5	9.0E-06	4.6E-05
GO:0070011	peptidase activity, acting on L-amino acid peptides	MF	5	2.1E-04	8.0E-04

Supplementary table S8. GO enrichment in protein coding genes found exclusively in *C.floridanus*, *H. saltator*, or both, but not in all other genomes analyzed.

<i>Ant specific</i>				
GO ID	GO Term	Class	Level	Adj. P value
GO:0004984	olfactory receptor activity	MF	7	2.29E-06
GO:0007608	sensory perception of smell	BP	8	2.29E-06
GO:0004930	G-protein coupled receptor activity	MF	6	2.29E-06
GO:0005549	odorant binding	MF	3	2.29E-06
GO:0004497	monooxygenase activity	MF	4	2.97E-04
GO:0020037	heme binding	MF	4	9.32E-04
GO:0046914	transition metal ion binding	MF	6	2.50E-03
GO:0008237	metallopeptidase activity	MF	6	3.19E-03
GO:0005515	protein binding	MF	3	1.43E-02
GO:0005102	receptor binding	MF	4	2.03E-02
GO:0016491	oxidoreductase activity	MF	3	2.29E-02
GO:0000036	acyl carrier activity	MF	4	2.96E-02
GO:0004221	ubiquitin thiolesterase activity	MF	6	4.58E-02
Found only in <i>C.floridanus</i>				
GO:0070011	peptidase activity, acting on L-amino acid peptides	MF	5	9.22E-30
GO:0008237	metallopeptidase activity	MF	6	3.11E-23
GO:0004175	endopeptidase activity	MF	6	1.67E-15
GO:0004190	aspartic-type endopeptidase activity	MF	7	7.72E-14
GO:0007606	sensory perception of chemical stimulus	BP	7	1.00E-08
GO:0005549	odorant binding	MF	3	1.19E-08
GO:0004984	olfactory receptor activity	MF	7	1.77E-08
GO:0007608	sensory perception of smell	BP	8	1.77E-08
GO:0006508	proteolysis	BP	6	4.82E-08
GO:0032501	multicellular organismal process	BP	2	1.35E-07
GO:0004930	G-protein coupled receptor activity	MF	6	2.39E-07
GO:0005578	proteinaceous extracellular matrix	CC	4	1.41E-06
GO:0005576	extracellular region	CC	2	2.44E-06
GO:0004713	protein tyrosine kinase activity	MF	7	3.32E-05
GO:0046983	protein dimerization activity	MF	4	2.23E-04
GO:0016787	hydrolase activity	MF	3	5.01E-04
GO:0008270	zinc ion binding	MF	7	1.08E-03
GO:0045263	proton-transporting ATP synthase complex, coupling factor F(o)	CC	4	1.25E-03
GO:0004518	nuclease activity	MF	5	1.52E-03
GO:0046914	transition metal ion binding	MF	6	6.11E-03
Found only in <i>H. saltator</i>				
GO:0017038	protein import	BP	5	9.50E-07
GO:0046914	transition metal ion binding	MF	6	3.72E-04
GO:0008270	zinc ion binding	MF	7	1.65E-04
GO:0016788	hydrolase activity, acting on ester bonds	MF	4	1.66E-03
GO:0008237	metallopeptidase activity	MF	6	1.28E-02

Supplementary table S9. Small RNA-seq in *C. floridanus* and *H. saltator*.

<i>C. floridanus</i>										
	Egg		Larva		Male		Major		Minor	
	Reads (x10 ³)	%	Reads (x10 ³)	%	Reads (x10 ³)	%	Reads (x10 ³)	%	Reads (x10 ³)	%
exon	572.2	6.9	601.5	5.2	340.9	3.5	179.9	2.1	314.3	2.5
repeat	1214.4	14.7	806.1	7	295	3	193.7	2.3	228.1	1.8
snRNA	23.4	0.3	1.7	0	6.7	0.1	0.5	0	0.9	0
tRNA	756.7	9.2	718.7	6.2	103.5	1.1	176.6	2.1	138.1	1.1
rRNA	159.5	1.9	53	0.5	113.8	1.2	54.8	0.6	115.1	0.9
miRNA	1027.8	12.4	2758.6	23.9	7213.9	73.6	5135.9	60.6	9734.5	76.6
others	4515.6	54.6	6626.2	57.3	1728.4	17.6	2735	32.3	2184.5	17.2
mapped	8269.8		11565.9		9802.1		8476.3		12715.5	
total	9739.6		14071.8		11571.5		11506.7		14348	
<i>H. saltator</i>										
	Egg		Larva		Male		Gamergate		Worker	
	Reads (x10 ³)	%	Reads (x10 ³)	%	Reads (x10 ³)	%	Reads (x10 ³)	%	Reads (x10 ³)	%
exon	890.8	8.4	466.4	6.2	383.1	3.9	692.8	6.9	142.7	1.7
repeat	3326.9	31.4	2299.5	30.6	615.9	6.2	1435.3	14.2	336.8	4
snRNA	1.2	0	2.5	0	7.3	0.1	7.1	0.1	2.8	0
tRNA	24.2	0.2	16.7	0.2	35.5	0.4	34	0.3	192.9	2.3
rRNA	30.4	0.3	257.3	3.4	258.9	2.6	216	2.1	119.6	1.4
miRNA	576.7	5.4	1693.8	22.5	6627.3	67.1	3891.1	38.6	6755	79.3
others	5740.9	54.2	2790.1	37.1	1947.7	19.7	3814.5	37.8	964	11.3
mapped	10591.1		7526.3		9875.6		10090.7		8513.7	
total	15005.5		10038.8		11803.2		13104.1		11299.4	

Supplementary table S10. Number of miRNA detected in different castes and developmental stages.
RPM: reads per million.

<i>C. floridanus</i>			
sample	RPM >= 5	RPM >= 10	RPM >= 100
egg	47	43	20
larva	64	60	31
male	64	54	25
major	59	48	21
minor	62	54	28
all	96	85	42
<i>H. saltator</i>			
sample	RPM >= 5	RPM >= 10	RPM >= 100
egg	82	51	26
larva	65	51	28
male	63	54	32
gamergate	102	67	35
worker	57	49	28
all	159	99	47

Supplementary table S11. Number of genes containing IPR domains related to chromatin and epigenetics.

Domain name	Interpro ID	CF	HS	AM	NV	DM	HSAP
Bromo-	IPR001487	20	21	22	21	19	43
Chromo-	IPR000953	14	20	12	30	20	32
HDAC	IPR000286	5	4	5	4	5	11
MBD	IPR001739	4	4	4	2	5	11
MBT	IPR004092	4	4	4	3	3	9
SET	IPR001214	27	22	30	49	31	46
SirT	IPR003000	6	6	6	6	5	7
C-5 Dnmt	IPR001525	3	3	6	5	1	4
SWIRM	IPR007526	3	3	4	1	3	6
JmjC	IPR003347	12	10	9	14	9	25
PHD finger	IPR001965	38	33	45	42	42	91

Supplementary table S12. Best homologous Swissprot proteins for histone acetyltransferases and deacetylases.

HATs				
<i>C. floridanus</i>	<i>H. saltator</i>	<i>Human</i>	<i>Drosophila</i>	Overall
Cflo_04916	—	ARD1A	MAK3	ARD1A_MOUSE
Cflo_08735	Hsal_06647	CA149	—	CA149_HUMAN
Cflo_01850	Hsal_04540	CBP	FSH	CBP_HUMAN
Cflo_11072	Hsal_00949	CSR2B	—	CSR2B_HUMAN
Cflo_09038	Hsal_06069	ELP3	ELP3	ELP3_CHICK
Cflo_13909	Hsal_01205	GCNL2	NU301	GCNL2_HUMAN
Cflo_13910	Hsal_01206	GCNL2	NU301	GCNL2_HUMAN
—	Hsal_01207	GCNL2	NU301	GCNL2_HUMAN
Cflo_01287	Hsal_02721	GNA1	GNA1	GNA1_DROME
Cflo_05900	Hsal_01257	HAT1	—	HAT1_RAT
Cflo_00776	Hsal_06740	MAK3	MAK3	MAK3_HUMAN
Cflo_00689	Hsal_00447	MYST1	MOF	MYST1_RAT
Cflo_08305	Hsal_07667	MYST2	TIP60	MYST2_HUMAN
Cflo_07969	Hsal_03629	MYST4	TIP60	MYST4_HUMAN
Cflo_01919	Hsal_12495	NAT11	—	NAT11_MOUSE
Cflo_04900	Hsal_01843	NAT13	SAN	SAN_DROME
Cflo_07951	Hsal_06880	NAT5	—	NAT5_XENTR
Cflo_07865	Hsal_14265	NAT6	—	NAT6_HUMAN
Cflo_06528	Hsal_02548	NAT9	NAT9	NAT9_NEMVE
Cflo_07743	Hsal_13399	NATU	SAN	NATU_MOUSE
Cflo_12696	Hsal_00085	TAD3L	—	TAD3L_HUMAN
Cflo_05966	Hsal_00380	TIP60	TIP60	TIP60_PONAB
Cflo_07448	Hsal_01504	—	—	—
Cflo_05055	Hsal_14525	—	—	—
Cflo_05450	Hsal_03546	—	—	—
SIRTs				
Cflo_12644	Hsal_10308	SIRT1	—	SIRT1_HUMAN
Cflo_11736	Hsal_06906	SIRT2	—	SIRT2_MOUSE
Cflo_01594	Hsal_03143	SIRT4	—	SIRT4_MOUSE
Cflo_07163	Hsal_03602	SIRT5	—	SIRT5_BOVIN
Cflo_08981	Hsal_10281	SIRT6	—	SIRT6_HUMAN
Cflo_13003	Hsal_02222	SIRT7	—	SIRT7_BOVIN
HDACs				
Cflo_10463	—	HDAC1	HDAC1	HDAC1_DROME
Cflo_10465	Hsal_06723	HDAC2	HDAC1	HDAC1_DROME
Cflo_10682	Hsal_06872	HDAC3	HDAC1	HDAC3_XENTR
Cflo_07521	Hsal_10875	HDAC4	HDAC1	HDAC4_RAT
Cflo_07569	Hsal_11500	HDAC6	HDAC1	HDAC6_MOUSE

Supplementary table S13. Best homologous Swissprot proteins for SET domain proteins in the ant genomes.

<i>C. floridanus</i>			
Gene ID	Human	<i>Drosophila</i>	overall
Cflo_10932	ASH1L	ASH1	ASH1L_MOUSE
Cflo_12188	EHMT1	SUV39	EHMT1_HUMAN
Cflo_10335	EZH2	EZ	EZH2_XENTR
Cflo_05575	MLL3	TRR	TRR_DROME
Cflo_01543	MLL4	TRX	TRX_DROVI
Cflo_12296	MLL5	ASH1	MLL5_HUMAN
Cflo_08954	—	MSTAA	MSTAA_DROME
Cflo_01290	NSD1	MES4	NSD1_MOUSE
Cflo_09151	PRD10	SUHW	PRD10_HUMAN
Cflo_04377	PRDM1	GLAS	PRDM1_HUMAN
Cflo_12528	SET1A	TRX	SET1B_XENLA
Cflo_09617	SETB1	SETB1	SETB1_DROPS
Cflo_10268	SETD2	C1716	C1716_DROME
Cflo_09801	SETD3	NA	SETD3_HUMAN
Cflo_06609	SETD4	NA	SETD4_MOUSE
Cflo_04573	SETD8	SETD8	SETD8_DROME
Cflo_12501	SETMR	SUV39	SETMR_MOUSE
Cflo_10149	SMYD1	MSTAB	MSTAB_DROME
Cflo_06803	SMYD2	MSTAB	MSTAB_DROME
Cflo_03939	SMYD4	NA	SMYD4_MOUSE
Cflo_06938	SMYD4	MSTAB	SMYD4_MOUSE
Cflo_09007	SMYD4	NA	SMYD4_MOUSE
Cflo_12246	SMYD4	NA	SMYD4_MOUSE
Cflo_15030	SMYD4	NA	SMYD4_HUMAN
Cflo_05356	SMYD5	NA	SMYD5_XENLA
Cflo_05373	SUV41	SUV42	SUV42_DROME
Cflo_00891	TTL12	NA	TTL12_MOUSE
<i>H. saltator</i>			
Gene ID	Human	<i>Drosophila</i>	overall
Hsal_10054	ASH1L	ASH1	ASH1_DROME
Hsal_12111	EHMT1	SUV39	EHMT2_MOUSE
Hsal_05330	EZH2	EZ	EZH2_XENTR
Hsal_NA	MLL4	TRX	TRX_DROVI
Hsal_08754	MLL5	NA	MLL5_HUMAN
Hsal_10488	NA	NA	YL678_MIMIV
Hsal_10061	PRDM4	KRUH	PRDM4_HUMAN
Hsal_01244	SET1A	TRX	SET1B_XENLA
Hsal_08596	SETB1	SETB1	SETB1_DROME
Hsal_09716	SETD2	C1716	C1716_DROME
Hsal_06753	SETD3	NA	SETD3_HUMAN

Hsal_05593	SETD8	SETD8	SETD8_DROPS
Hsal_14066	SETMR	SUV39	SETMR_HUMAN
Hsal_14941	SMYD3	MSTAB	MSTAB_DROME
Hsal_00516	SMYD4	NA	SMYD4_HUMAN
Hsal_08142	SMYD4	NA	SMYD4_MOUSE
Hsal_10252	SMYD4	NA	SMYD4_HUMAN
Hsal_10487	SMYD4	NA	SMYD4_MOUSE
Hsal_10489	SMYD4	NA	YL678_MIMIV
Hsal_11656	SMYD4	NA	SMYD4_HUMAN
Hsal_14332	SMYD5	NA	SMYD5_XENLA
Hsal_07893	SUV41	SUV42	SUV42_DROME

Supplementary table S14. Top 25 GO terms enriched in genes displaying 4-fold changes in expression levels between *C. floridanus* major and minor workers.

<i>C. floridanus</i> - major vs. minor workers			
GO ID	Term	# Genes	P-Value
GO:0004871	signal transducer activity	43	0
GO:0004872	receptor activity	37	0
GO:0070011	peptidase activity, acting on L-amino acid peptides	34	0
GO:0005975	carbohydrate metabolic process	34	0
GO:0008233	peptidase activity	34	0
GO:0006066	alcohol metabolic process	29	0
GO:0004888	transmembrane receptor activity	25	0
GO:0005976	polysaccharide metabolic process	21	0
GO:0006022	aminoglycan metabolic process	18	0
GO:0005216	ion channel activity	18	0
GO:0044456	synapse part	17	0
GO:0022836	gated channel activity	16	0
GO:0030247	polysaccharide binding	16	0
GO:0006030	chitin metabolic process	15	0
GO:0006044	N-acetylglucosamine metabolic process	15	0
GO:0008061	chitin binding	14	0
GO:0045211	postsynaptic membrane	14	0
GO:0022834	ligand-gated channel activity	13	0
GO:0015276	ligand-gated ion channel activity	13	0
GO:0005230	extracellular ligand-gated ion channel activity	10	0
GO:0004984	olfactory receptor activity	8	0
GO:0005231	excitatory extracellular ligand-gated ion channel activity	8	0
GO:0004190	aspartic-type endopeptidase activity	8	0
GO:0007608	sensory perception of smell	8	0
GO:0005549	odorant binding	7	0

Supplementary table S15. Top 25 GO terms enriched in genes displaying 4-fold changes in expression levels between *H. saltator* gamergates and non-reproductive workers.

<i>H. saltator</i> - gamergates vs. low-ranking workers			
GO ID	Term	# genes	P-Value
GO:0007600	sensory perception	40	0
GO:0004930	G-protein coupled receptor activity	37	0
GO:0007586	digestion	16	0
GO:0004888	transmembrane receptor activity	51	1.00E-04
GO:0050890	cognition	46	1.00E-04
GO:0007186	G-protein coupled receptor protein signaling pathway	45	1.00E-04
GO:0042219	cellular amino acid derivative catabolic process	11	1.00E-04
GO:0009881	photoreceptor activity	8	1.00E-04
GO:0016703	oxidoreductase activity, (...)	6	1.00E-04
GO:0004252	serine-type endopeptidase activity	24	2.00E-04
GO:0007389	pattern specification process	39	3.00E-04
GO:0018298	protein-chromophore linkage	8	3.00E-04
GO:0050908	detection of light stimulus involved in visual perception	10	4.00E-04
GO:0007602	phototransduction	10	4.00E-04
GO:0047077	Photinus-luciferin 4-monooxygenase (ATP-hydrolyzing) activity	5	4.00E-04
GO:0008218	bioluminescence	6	5.00E-04
GO:0005694	chromosome	54	6.00E-04
GO:0017171	serine hydrolase activity	27	6.00E-04
GO:0008236	serine-type peptidase activity	27	6.00E-04
GO:0006576	biogenic amine metabolic process	17	6.00E-04
GO:0009880	embryonic pattern specification	15	6.00E-04
GO:0050906	detection of stimulus involved in sensory perception	12	6.00E-04
GO:0004872	receptor activity	76	7.00E-04
GO:0009584	detection of visible light	10	7.00E-04
GO:0007608	sensory perception of smell	10	7.00E-04

Supplementary table S16. GPCRs in *C. floridanus*. The protein ID used for automatic annotation is shown with a suffix indicating the species. Gene symbols for manual annotation are from *Anopheles gambiae* with the exception of olfactory and gustatory receptors, which are from *A. mellifera*.

Ant Gene ID	Automatic annotation	Manual annotation
Muscarinic Acetylcholine Family		
Cflo_05841	XP_395760.3_APIME	GPRmac1
Cflo_08083	XP_395477.3_APIME	GPRmac2
Dopamine Family		
Cflo_01706	NP_001014983.1_APIME	GPRdop3
Cflo_09026	NP_001011595.1_APIME	GPRdop1
Cflo_00047	NP_001011567.1_APIME	GPRdop2
Cflo_01703	NP_001014983.1_APIME	GPRdop2
Histamine Family		
Cflo_02283	XP_001602335.1_NASVI	GPRhis
Melatonin Family		
Cflo_07483	XP_392683.1_APIME	GPRmtn
Octopamine/Tyramine Family		
Cflo_07626	NP_001011594.1_APIME	GPRtyr
Cflo_20006	NP_001035262.1_HUMAN	GPRtyr
Cflo_08380	NP_001011565.1_APIME	GPRoar1
Serotonin Family		
Cflo_05076	XP_001122425.1_APIME	GPR5ht7
Cflo_06547	XP_393915.3_APIME	GPR5ht1b
Cflo_10556	XP_624897.2_APIME	GPR5htorph2
Cflo_02065	XP_001122856.1_APIME	GPR5htorph2
Cflo_10559	XP_394798.1_APIME	GPR5ht2a
Glycoprotein Hormone Family		
Cflo_01416	XP_001122003.1_APIME	GPRrk
Cflo_04882	XP_001121943.1_APIME	GPRrk
Cflo_03285	XP_393713.2_APIME	GPRrk
Cflo_05914	XP_001121890.1_APIME	GPRrk
Cflo_09171	XP_394301.2_APIME	GPRrk
Cflo_11842	XP_001120061.1_APIME	GPRrk
Cflo_11064	XP_395206.3_APIME	GPRrk
Galanin/Allatostatin Family		
Cflo_04788	XP_396660.2_APIME	GPRals2

Cflo_10650	XP_397024.1_APIME	GPRals1
Gastrin/Bombesin Family		
Cflo_01829	XP_395101.3_APIME	GPRgrp1
Cflo_09222	XP_396992.3_APIME	GPRgrp2
Gastrin/Cholecystokinin Family		
Cflo_00833	XP_396660.2_APIME	GPRcck1
Gonadotrophin Releasing Hormone Family		
Cflo_11527	NP_001035354.1_APIME	GPRgnr1
Cflo_03230	XP_392570.3_APIME	GPRgnr2
Releasing Hormone Family		
Cflo_04787	XP_395081.2_APIME	GPRnpr2
Cflo_20005	NP_057652.1_HUMAN	GPRnpr1
Neuropeptide Y Family		
Cflo_04910	XP_001123033.1_APIME	GPRnpy3
Opioid Family		
Cflo_15140	XP_392683.1_APIME	GPRopr
Cflo_12702	XP_001120910.1_APIME	GPRopr
Cflo_07482	XP_001606666.1_NASVI	GPRopr
Somatostatin Family		
Cflo_05419	XP_396335.1_APIME	GPRsms
Vasopressin Family		
Cflo_12132	NP_001035354.1_APIME	GPRvpr2
Cflo_03493	XP_001122652.1_APIME	GPRvpr1
Purine/Adenosine Family		
Cflo_11996	NP_001011639.1_APIME	GPRop6
Cflo_12364	XP_394893.2_APIME	GPRads
Cflo_00324	NP_001011606.1_APIME	GPRop9
Cflo_05948	NP_001035057.1_APIME	GPRop12
Cflo_00708	NP_001011605.1_APIME	GPRop8
Cflo_11995	NP_001011639.1_APIME	GPRop6
Orphan/Putative Class A Family		
Cflo_11078	XP_396970.3_APIME	GPRorpha11
Cflo_02522	XP_001122248.1_APIME	GPRorpha18
Cflo_06056	XP_394231.2_APIME	GPRorpha4a

Cflo_04226	XP_396348.3_APIME	GPRorpha2
Cflo_05037	XP_397077.3_APIME	GPRorpha2
Cflo_04830	XP_001120030.1_APIME	GPRorpha5
Cflo_03977	XP_396660.2_APIME	GPRorpha7
Cflo_04864	XP_396491.1_APIME	GPRorpha21
Cflo_05036	XP_397139.2_APIME	GPRorpha2
Cflo_09309	XP_001120499.1_APIME	GPRorpha1
Cflo_07430	XP_001122248.1_APIME	GPRorpha18
Cflo_04890	XP_001119993.1_APIME	GPRorpha11
Calcitonin/Diuretic Hormone Family		
Cflo_01634	XP_396046.3_APIME	GPRcal1
Cflo_04806	XP_395896.3_APIME	GPRcal3
Diuretic Insect Hormone/Kinin/CRF Family		
Cflo_06182	XP_001122670.1_APIME	GPRdih2
Cflo_20003	XP_395896.3_APIME	GPRdih1
Growth Hormone Releasing Hormone Family		
Cflo_00959	XP_623966.1_APIME	GPRghp3
Cflo_10113	XP_001122475.1_APIME	GPRghp3
Cflo_01841	XP_001122135.1_APIME	GPRghp2
Latrophilin Family		
Cflo_08649	XP_001120657.1_APIME	GPRmth6
Cflo_11203	XP_001120757.1_APIME	GPRmth3
Cflo_06317	XP_001120657.1_APIME	GPRmth4
Cflo_11471	XP_001120657.1_APIME	GPRmth4
Cflo_06109	XP_624729.1_APIME	GPRmth2
Cflo_08650	XP_001120657.1_APIME	GPRmth5
Cflo_10735	XP_001120757.1_APIME	GPRmth3
Cflo_13535	XP_001120657.1_APIME	GPRmth4
Cflo_01901	XP_001120657.1_APIME	GPRmth4
Cflo_07346	XP_001120679.1_APIME	GPRmth4
Cflo_16048	XP_001120657.1_APIME	GPRmth4
Orphan/Putative Class B Family		
Cflo_07988	XP_394034.3_APIME	GPRorphb1
Cflo_07231	XP_624181.2_APIME	GPRorphb1
Cflo_04053	XP_394619.1_APIME	GPRorphb1
Cflo_07113	XP_393717.2_APIME	GPRorphb1
Cflo_04606	XP_396683.3_APIME	GPRorphb1
Cflo_11198	XP_001120678.1_APIME	GPRorphb1
Cflo_07584	XP_001122057.1_APIME	GPRorphb1

Cflo_08117	XP_393712.2_APIIME	GPRorphb1
Cflo_02851	XP_394576.2_APIIME	GPRorphb3
Cflo_10991	XP_001121284.1_APIIME	GPRorphb1
Cflo_01542	NP_001013379.1_APIIME	GPRorphb1
Cflo_03572	XP_001121669.1_APIIME	GPRorphb1
Cflo_10082	XP_001121376.1_APIIME	GPRorphb1
Cflo_09994	XP_393627.2_APIIME	GPRorphb1
Cflo_01928	XP_392166.3_APIIME	GPRorphb1
Cflo_00206	XP_624030.1_APIIME	GPRorphb2
Cflo_03264	XP_001120937.1_APIIME	GPRorphb1
Cflo_11199	XP_001120543.1_APIIME	GPRorphb1
Cflo_07226	XP_393717.2_APIIME	GPRorphb1
Cflo_08282	XP_396017.2_APIIME	GPRorphb1
Cflo_02456	XP_001122003.1_APIIME	GPRorphb1
Cflo_03562	XP_001121610.1_APIIME	GPRorphb1
Cflo_00269	XP_395331.3_APIIME	GPRorphb1
Cflo_08124	XP_393713.2_APIIME	GPRorphb1
Cflo_01824	XP_624030.1_APIIME	GPRorphb2
Cflo_01194	XP_393750.3_APIIME	GPRorphb1
Cflo_02111	XP_396158.1_APIIME	GPRorphb1
Metabotropic Glutamate Family		
Cflo_02049	XP_395227.3_APIIME	GPRmgl5
Cflo_10368	XP_397009.3_APIIME	GPRmgl
Cflo_11764	XP_392670.2_APIIME	GPRmgl4
Cflo_07177	XP_392015.2_APIIME	GPRmgl4
Cflo_02862	NP_001011623.1_APIIME	GPRmgl5
Cflo_07403	XP_397009.3_APIIME	GPRmgl2
Cflo_11626	NP_001011624.1_APIIME	GPRmgl5
GABA-B Family		
Cflo_11334	XP_393623.3_APIIME	GPRgbb2
Cflo_02011	XP_392294.3_APIIME	GPRgbb1
Cflo_03027	XP_397038.3_APIIME	GPRgbb2
Class D Atypical Fried/Smoothened Family		
Cflo_00954	XP_396277.3_APIIME	GPRstn
Cflo_04518	XP_396474.3_APIIME	GPRqfc3
Cflo_09065	XP_623678.1_APIIME	GPRqfc21
Cflo_01564	XP_396152.3_APIIME	GPRqfc1
Cflo_10677	XP_001123118.1_APIIME	GPRqfc18
Cflo_20002	XP_319897.4_ANOGA	GPRqfc18
Cflo_03610	XP_397259.3_APIIME	GPRf2
Cflo_08959	XP_393848.3_APIIME	GPRstn

Cflo_01899	XP_623435.1_APIME	GPRqfc7
Cflo_02349	XP_392736.3_APIME	GPRstn
Cflo_11470	XP_396742.2_APIME	GPRqfc4
Cflo_01086	XP_001122300.1_APIME	GPRqfc18
Cflo_03973	XP_396659.3_APIME	GPRstn
Cflo_11315	XP_001121997.1_APIME	GPRorphd1
Cflo_08471	XP_623523.1_APIME	GPRf2
Cflo_00480	XP_392416.2_APIME	GPRorphd2
Cflo_01357	XP_396248.3_APIME	GPRstn
Cflo_03151	XP_396829.1_APIME	GPRqfc18
Cflo_06158	XP_624236.2_APIME	GPRstn
Cflo_01456	XP_393220.3_APIME	GPRstn
Cflo_05107	XP_624183.1_APIME	GPRqfc15
Cflo_00585	XP_624183.1_APIME	GPRqfc15
Cflo_01382	XP_392300.3_APIME	GPRstn
Cflo_06415	XP_394096.2_APIME	GPRqfc16
Cflo_02052	XP_392713.1_APIME	GPRqfc8
Cflo_09324	XP_001120706.1_APIME	GPRqfc11
Cflo_11831	XP_623523.1_APIME	GPRf4
Cflo_07227	XP_001122151.1_APIME	GPRqfc18
Cflo_00115	XP_396231.2_APIME	GPRstn
Cflo_07507	XP_624967.1_APIME	GPRstn
Cflo_09442	XP_001121114.1_APIME	GPRf1a
Cflo_03788	XP_001121711.1_APIME	GPRqfc18
Cflo_02248	XP_392299.2_APIME	GPRqfc1
Cflo_09408	XP_393100.1_APIME	GPRqfc17
Cflo_13529	XP_392436.3_APIME	GPRqfc17
Cflo_05554	XP_394063.3_APIME	GPRstn
Cflo_06408	XP_392436.3_APIME	GPRqfc17
Cflo_09233	XP_001121399.1_APIME	GPRqfc19
Cflo_02179	XP_396476.2_APIME	GPRqfc18
Cflo_03515	XP_391941.2_APIME	GPRstn
Cflo_08973	XP_395373.2_APIME	GPRsmo
Cflo_10206	XP_396734.3_APIME	GPRstn
Cflo_10750	XP_624809.2_APIME	GPRqfc5
Cflo_00431	XP_394632.2_APIME	GPRqfc18
Cflo_01420	XP_394496.3_APIME	GPRstn
Cflo_03719	XP_393497.3_APIME	GPRstn
Cflo_10968	XP_001121984.1_APIME	GPRbos
Cflo_06871	XP_392099.3_APIME	GPRstn
Cflo_01948	XP_396579.1_APIME	GPRqfc14
Cflo_00129	XP_624091.1_APIME	GPRqfc2
Cflo_07479	XP_392099.3_APIME	GPRstn
Cflo_01387	XP_392300.3_APIME	GPRstn

Cflo_07933	XP_001119987.1_APIME	GPRstn
Cflo_08897	XP_624568.1_APIME	GPRqfc18
Cflo_07504	XP_394631.3_APIME	GPRstn
Cflo_06778	XP_001602501.1_NASVI	GPRqfc12
Cflo_04977	XP_623523.1_APIME	GPRf1a
Cflo_01221	XP_392888.2_APIME	GPRqfc5
Cflo_04967	XP_001121416.1_APIME	GPRstn
Neurokinin/Tachykinin Family		
Cflo_20004	FBpp0074640_DROME	GPRtak1
Cflo_20001	FBpp0074640_DROME	GPRtak1
Class E -Chemosensory receptors		
Odorant receptors		
Cflo_01723	XP_001121358.1_APIME	Odorant receptor 30a CG13106-PA
Cflo_01724	XP_001121358.1_APIME	Odorant receptor 30a CG13106-PA
Cflo_01725	XP_001121358.1_APIME	Odorant receptor 30a CG13106-PA
Cflo_01726	XP_393094.3_APIME	Odorant receptor 67a
Cflo_01727	XP_393094.3_APIME	Odorant receptor 67a
Cflo_02935	XP_001121741.1_APIME	Odorant receptor 85b
Cflo_02937	XP_001121741.1_APIME	Odorant receptor 85b
Cflo_02938	XP_001121864.1_APIME	Odorant receptor 2a
Cflo_02939	XP_001121864.1_APIME	Odorant receptor 2a
Cflo_02940	XP_001121864.1_APIME	Odorant receptor 2a
Cflo_02941	XP_001121864.1_APIME	Odorant receptor 2a
Cflo_02942	XP_001121864.1_APIME	Odorant receptor 2a
Cflo_02943	XP_001121864.1_APIME	Odorant receptor 2a
Cflo_02944	XP_001121741.1_APIME	Odorant receptor 85b
Cflo_02945	XP_001121659.1_APIME	Odorant receptor 85b
Cflo_02946	XP_001121659.1_APIME	Odorant receptor 85b
Cflo_02947	XP_001121864.1_APIME	Odorant receptor 2a
Cflo_02948	XP_001121741.1_APIME	Odorant receptor 85b
Cflo_02949	XP_001121741.1_APIME	Odorant receptor 85b
Cflo_02950	XP_001121741.1_APIME	Odorant receptor 85b
Cflo_02951	XP_001121741.1_APIME	Odorant receptor 85b
Cflo_02952	XP_001121741.1_APIME	Odorant receptor 85b
Cflo_02953	XP_001121741.1_APIME	Odorant receptor 85b
Cflo_02954	XP_001121741.1_APIME	Odorant receptor 85b
Cflo_02955	XP_001121864.1_APIME	Odorant receptor 2a
Cflo_02956	XP_001121741.1_APIME	Odorant receptor 85b
Cflo_02957	XP_001121741.1_APIME	Odorant receptor 85b
Cflo_02958	XP_001121741.1_APIME	Odorant receptor 85b
Cflo_02959	XP_001121741.1_APIME	Odorant receptor 85b
Cflo_02960	XP_001121741.1_APIME	Odorant receptor 85b

Cflo_02961	XP_001121864.1_APIME	Odorant receptor 2a
Cflo_02962	XP_001121741.1_APIME	Odorant receptor 85b
Cflo_02963	XP_001121741.1_APIME	Odorant receptor 85b
Cflo_02964	XP_001121741.1_APIME	Odorant receptor 85b
Cflo_02965	XP_001121741.1_APIME	Odorant receptor 85b
Cflo_02966	XP_001121741.1_APIME	Odorant receptor 85b
Cflo_02967	XP_001121080.1_APIME	Odorant receptor 83b
Cflo_02968	XP_001121080.1_APIME	Odorant receptor 83b
Cflo_02991	XP_001123098.1_APIME	Odorant receptor 85b,
Cflo_03281	XP_001122971.1_APIME	Odorant response abnormal 4
Cflo_03364	XP_001122936.1_APIME	Odorant receptor 2a
Cflo_03365	XP_001123098.1_APIME	Odorant receptor 85b,
Cflo_03367	XP_001123098.1_APIME	Odorant receptor 85b,
Cflo_03370	XP_001123098.1_APIME	Odorant receptor 85b,
Cflo_03372	XP_001123098.1_APIME	Odorant receptor 85b,
Cflo_03845	XP_001122191.1_APIME	Odorant receptor 30a
Cflo_03848	XP_001120660.1_APIME	Odorant receptor 9a
Cflo_03849	XP_001120660.1_APIME	Odorant receptor 9a
Cflo_03862	XP_001120948.1_APIME	Odorant receptor 30a
Cflo_03873	XP_001123098.1_APIME	Odorant receptor 85b,
Cflo_03879	XP_001120948.1_APIME	Odorant receptor 30a
Cflo_03978	XP_001120859.1_APIME	Odorant receptor 13a
Cflo_04072	XP_001121145.1_APIME	Odorant receptor 83b
Cflo_04466	XP_623850.2_APIME	Odorant receptor 13a
Cflo_05121	XP_623850.2_APIME	Odorant receptor 13a
Cflo_05301	XP_001122649.1_APIME	Odorant receptor 13a
Cflo_05303	XP_001122649.1_APIME	Odorant receptor 13a
Cflo_05954	XP_001121080.1_APIME	Odorant receptor 83b
Cflo_06363	XP_623850.2_APIME	Odorant receptor 13a
Cflo_06383	XP_001122936.1_APIME	Odorant receptor 2a
Cflo_07195	XP_001120948.1_APIME	Odorant receptor 30a
Cflo_08391	XP_623850.2_APIME	Odorant receptor 13a
Cflo_08397	XP_623850.2_APIME	Odorant receptor 13a
Cflo_09139	XP_001120948.1_APIME	Odorant receptor 30a
Cflo_09722	XP_001122191.1_APIME	Odorant receptor 30a
Cflo_10526	XP_623850.2_APIME	Odorant receptor 13a
Cflo_11177	XP_001120660.1_APIME	Odorant receptor 9a
Cflo_11244	XP_001122552.1_APIME	Odorant receptor 43a
Cflo_11245	XP_001122552.1_APIME	Odorant receptor 43a
Cflo_12208	XP_001121358.1_APIME	Odorant receptor 30a
Cflo_12339	XP_001120660.1_APIME	Odorant receptor 9a
Cflo_13083	XP_001120660.1_APIME	Odorant receptor 9a
Cflo_13154	XP_001120660.1_APIME	Odorant receptor 9a
Cflo_13960	XP_001122552.1_APIME	Odorant receptor 43a

Cflo_14670	XP_001121358.1_APIME	Odorant receptor 30a
Cflo_14715	XP_001121358.1_APIME	Odorant receptor 30a
Cflo_15236	XP_001120948.1_APIME	Odorant receptor 30a
Cflo_15269	XP_001120660.1_APIME	Odorant receptor 9a
Cflo_15570	XP_001119972.1_APIME	Odorant receptor 13a
Cflo_15582	XP_001121864.1_APIME	Odorant receptor 2a
Cflo_15941	XP_001121864.1_APIME	Odorant receptor 2a
Cflo_16341	XP_001121741.1_APIME	Odorant receptor 85b
Gustatory receptors		
Cflo_10865	XP_001121326.1_APIME	Gustatory receptor 43a
Cflo_03028	XP_001121326.1_APIME	Gustatory receptor 43a
Cflo_13378	XP_001121326.1_APIME	Gustatory receptor 43a
Cflo_13379	XP_001121326.1_APIME	Gustatory receptor 43a
Cflo_06080	XP_001121165.1_APIME	Gustatory receptor 68a
Cflo_14505	XP_001121326.1_APIME	Gustatory receptor 43a
Cflo_00157	XP_001121326.1_APIME	Gustatory receptor 43a
Cflo_02072	XP_001121326.1_APIME	Gustatory receptor 43a
Cflo_15115	XP_001121326.1_APIME	Gustatory receptor 43a
Cflo_07924	XP_001123138.1_APIME	Gustatory receptor 64f
Cflo_07925	XP_397125.3_APIME	Gustatory receptor 64f

Supplementary table S17. GPCRs in *H. saltator*. The protein ID used for automatic annotation is shown with a suffix indicating the species. Gene symbols for manual annotation are from *Anopheles gambiae* with the exception of olfactory and gustatory receptors, which are from *A. mellifera*.

Ant Gene ID	Automatic annotation	Manual annotation
Muscarinic Acetylcholine Family		
Hsal_10554	XP_395477.3_APIME	GPRmac2
Hsal_06712	XP_395760.3_APIME	GPRmac1
Dopamine Family		
Hsal_03180	NP_001014983.1_APIME	GPRdop3
Hsal_03176	NP_001014983.1_APIME	GPRdop2
Hsal_11748	NP_001011567.1_APIME	GPRdop2
Hsal_00349	NP_001011595.1_APIME	GPRdop1
Histamine Family		
Hsal_10951	XP_001602335.1_NASVI	GPRhis
Melatonin Family		
Hsal_01614	XP_001120910.1_APIME	GPRmtn
Hsal_08878	XP_392683.1_APIME	GPRmtn
Octopamine/Tyramine Family		
Hsal_03256	NP_001011565.1_APIME	GPRoar1
Hsal_03880	NP_001011594.1_APIME	GPRtyr
Hsal_04014	XP_001122075.1_APIME	GPRtyr
Hsal_20003	NP_001035262.1_HUMAN	GPRoar1
Hsal_20005	FBpp0099565_DROME	GPRoar1
Serotonin Family		
Hsal_12799	XP_624897.2_APIME	GPR5htorph2
Hsal_11808	XP_394798.1_APIME	GPR5ht2a
Hsal_08960	XP_393915.3_APIME	GPR5ht1a
Hsal_05441	XP_001122425.1_APIME	GPR5ht7
Glycoprotein Hormone Family		
Hsal_00261	XP_001121943.1_APIME	GPRrk
Hsal_00958	XP_395206.3_APIME	GPRrk
Hsal_02766	XP_394301.2_APIME	GPRrk
Hsal_09731	XP_001120543.1_APIME	GPRrk
Galanin/Allatostatin Family		
Hsal_06235	XP_397024.1_APIME	GPRals1

Gastrin/Bombesin Family		
Hsal_07362	XP_396992.3_APIME	GPRgrp1
Gastrin/Cholecystokinin Family		
Hsal_05756	XP_396660.2_APIME	GPRcck1
Gonadotrophin Releasing Hormone Family		
Hsal_01712	XP_392570.3_APIME	GPRgnr2
Releasing Hormone Family		
Hsal_01910	XP_396025.2_APIME	GPRnpr2
Hsal_20004	FBpp0083455_DROME	GPRnpr1
Neurokinin/Tachykinin Family		
Hsal_10556	XP_395081.2_APIME	GPRtak1
Neuropeptide Y Family		
Hsal_01860	XP_001123033.1_APIME	GPRnpy3
Opioid Family		
Hsal_08879	XP_392683.1_APIME	GPRopr
Somatostatin Family		
Hsal_12025	XP_396335.1_APIME	GPRsms
Vasopressin Family		
Hsal_00453	NP_001035354.1_APIME	GPRvpr2
Hsal_05876	XP_001122652.1_APIME	GPRvpr1
Purine/Adenosine Family		
Hsal_12731	NP_001011639.1_APIME	GPRop6
Hsal_12635	NP_001035057.1_APIME	GPRop12
Hsal_12730	NP_001011639.1_APIME	GPRop6
Hsal_04811	NP_001011606.1_APIME	GPRop9
Hsal_09407	NP_001011605.1_APIME	GPRop8
Hsal_10115	XP_394893.2_APIME	GPRads
Orphan/Putative Class A Family		
Hsal_04013	NP_001011594.1_APIME	GPRorpha19
Hsal_04680	XP_001120335.1_APIME	GPRorpha9
Hsal_03462	XP_396660.2_APIME	GPRorpha7
Hsal_04196	XP_394231.2_APIME	GPRorpha4a
Hsal_10269	XP_396970.3_APIME	GPRorpha11
Hsal_07910	XP_001120030.1_APIME	GPRorpha5

Hsal_12213	XP_397139.2_APIME	GPRorpha2
Hsal_04093	XP_001122248.1_APIME	GPRorpha18
Hsal_01833	XP_001119993.1_APIME	GPRorpha11
Hsal_11968	XP_397077.3_APIME	GPRorpha3
Hsal_02180	XP_001606982.1_NASVI	GPRorpha10
Hsal_10065	NP_001011595.1_APIME	GPRorpha10
Hsal_09241	XP_396348.3_APIME	GPRorpha2
Hsal_01393	XP_396491.1_APIME	GPRorpha21
Hsal_11966	XP_397077.3_APIME	GPRorpha2
Hsal_10785	XP_001122248.1_APIME	GPRorpha17
Hsal_12506	XP_001120499.1_APIME	GPRorpha1
Calcitonin/Diuretic Hormone Family		
Hsal_07681	XP_395896.3_APIME	GPRcal3
Hsal_03155	XP_396046.3_APIME	GPRcal1
Hsal_17080	XP_396046.3_APIME	GPRcal1
Growth Hormone Releasing Hormone Family		
Hsal_04505	XP_623966.1_APIME	GPRghp3
Hsal_02759	XP_001122135.1_APIME	GPRghp2
Hsal_02758	XP_395140.2_APIME	GPRghp2
Hsal_20001	XP_307871.1_ANOGA	GPRmth3
Latrophilin Family		
Hsal_14116	XP_001120657.1_APIME	GPRmth6
Hsal_13160	XP_001120657.1_APIME	GPRmth4
Hsal_14115	XP_001120657.1_APIME	GPRmth5
Hsal_12442	XP_001120679.1_APIME	GPRmth4
Hsal_06863	XP_624729.1_APIME	GPRmth4
Orphan/Putative Class B Family		
Hsal_01922	XP_393383.2_APIME	GPRorphb1
Hsal_08073	XP_001121610.1_APIME	GPRorphb1
Hsal_00868	XP_394576.2_APIME	GPRorphb3
Hsal_10694	XP_393750.3_APIME	GPRorphb1
Hsal_10223	XP_001122057.1_APIME	GPRorphb1
Hsal_11122	XP_001120937.1_APIME	GPRorphb1
Hsal_12313	XP_396017.2_APIME	GPRorphb1
Hsal_09729	XP_001120678.1_APIME	GPRorphb1
Hsal_03019	XP_001121376.1_APIME	GPRorphb1
Hsal_03656	XP_394034.3_APIME	GPRorphb1
Hsal_02132	XP_624030.1_APIME	GPRorphb2
Hsal_08913	NP_001013379.1_APIME	GPRorphb1
Hsal_08869	XP_396158.1_APIME	GPRorphb1

Hsal_04045	XP_393713.2_APIME	GPRorphb1
Hsal_10101	XP_393717.2_APIME	GPRorphb1
Hsal_12465	XP_393383.2_APIME	GPRorphb1
Hsal_07108	XP_393713.2_APIME	GPRorphb1
Hsal_00060	XP_001121284.1_APIME	GPRorphb1
Hsal_04037	XP_393712.2_APIME	GPRorphb1
Hsal_00066	XP_001120061.1_APIME	GPRorphb1
Hsal_06127	XP_001122003.1_APIME	GPRorphb1
Hsal_06933	XP_393713.2_APIME	GPRorphb1
Hsal_08159	XP_001121669.1_APIME	GPRorphb1
Hsal_10295	XP_001121284.1_APIME	GPRorphb1
Hsal_05528	XP_396683.3_APIME	GPRorphb1
Hsal_16070	XP_001122300.1_APIME	GPRorphb1
Hsal_00439	XP_395331.3_APIME	GPRorphb1
Hsal_07113	XP_624181.2_APIME	GPRorphb1
Metabotropic Glutamate Family		
Hsal_00163	XP_395227.3_APIME	GPRmgl5
Hsal_03589	XP_392015.2_APIME	GPRmgl4
Hsal_04729	XP_392670.2_APIME	GPRmgl4
Hsal_07455	NP_001011624.1_APIME	GPRmgl5
Hsal_02401	NP_001011623.1_APIME	GPRmgl5
Hsal_08608	XP_396520.2_APIME	GPRmgl4
Hsal_06468	XP_397009.3_APIME	GPRmgl2
Hsal_02402	NP_001011623.1_APIME	GPRmgl5
Hsal_10291	XP_397009.3_APIME	GPRmgl
GABA-B Family		
Hsal_02738	XP_397038.3_APIME	GPRgbb2
Hsal_11703	XP_393623.3_APIME	GPRgbb2
Hsal_05375	XP_392294.3_APIME	GPRgbb1
Class D Atypical Frieded/Smoothened Family		
Hsal_00127	XP_001121399.1_APIME	GPRqfc19
Hsal_20006	XP_001121943.1_APIME	GPRqfc18
Hsal_03950	XP_393497.3_APIME	GPRstn
Hsal_01494	XP_394063.3_APIME	GPRstn
Hsal_00769	XP_392099.3_APIME	GPRstn
Hsal_12339	XP_396231.2_APIME	GPRstn
Hsal_05205	XP_001121416.1_APIME	GPRstn
Hsal_07431	XP_001121711.1_APIME	GPRqfc18
Hsal_12330	XP_393831.3_APIME	GPRstn
Hsal_07109	XP_001122151.1_APIME	GPRqfc18
Hsal_09744	XP_396829.1_APIME	GPRqfc18

Hsal_07704	XP_396579.1_APIME	GPRqfc14
Hsal_04114	XP_396152.3_APIME	GPRqfc1
Hsal_09596	XP_392416.2_APIME	GPRorpd2
Hsal_06677	FBpp0078249_DROME	GPRqfc12
Hsal_13902	XP_623678.1_APIME	GPRqfc21
Hsal_13859	XP_396476.2_APIME	GPRqfc18
Hsal_01264	XP_624809.2_APIME	GPRqfc5
Hsal_02216	XP_624568.1_APIME	GPRqfc18
Hsal_05704	XP_623523.1_APIME	GPRf4
Hsal_07321	XP_392300.3_APIME	GPRstn
Hsal_00766	XP_392099.3_APIME	GPRstn
Hsal_13725	XP_396118.3_APIME	GPRstn
Hsal_01617	XP_394631.3_APIME	GPRstn
Hsal_00990	XP_393220.3_APIME	GPRstn
Hsal_04512	XP_396277.3_APIME	GPRstn
Hsal_06587	XP_623435.1_APIME	GPRqfc7
Hsal_13595	XP_001120706.1_APIME	GPRqfc11
Hsal_01636	XP_394632.2_APIME	GPRqfc18
Hsal_01041	XP_396734.3_APIME	GPRstn
Hsal_06939	XP_393848.3_APIME	GPRstn
Hsal_10575	XP_396742.2_APIME	GPRqfc4
Hsal_04414	XP_624183.1_APIME	GPRqfc15
Hsal_08062	XP_392436.3_APIME	GPRqfc17
Hsal_03465	XP_396659.3_APIME	GPRstn
Hsal_02922	XP_397259.3_APIME	GPRf2
Hsal_12096	XP_392436.3_APIME	GPRqfc17
Hsal_13543	XP_392436.3_APIME	GPRqfc17
Hsal_12703	XP_392736.3_APIME	GPRstn
Hsal_03975	XP_624091.1_APIME	GPRqfc2
Hsal_04905	XP_001121984.1_APIME	GPRbos
Hsal_03213	XP_391941.2_APIME	GPRstn
Hsal_10590	XP_623523.1_APIME	GPRf1a
Hsal_05231	XP_395373.2_APIME	GPRsmo
Hsal_00248	XP_396476.2_APIME	GPRqfc18
Hsal_01623	XP_624967.1_APIME	GPRstn
Hsal_10074	XP_624236.2_APIME	GPRstn
Hsal_09332	XP_001121890.1_APIME	GPRqfc18
Hsal_00166	XP_392713.1_APIME	GPRqfc8
Hsal_12474	XP_624183.1_APIME	GPRqfc15
Hsal_07316	XP_392300.3_APIME	GPRstn
Hsal_08483	XP_396474.3_APIME	GPRqfc3
Hsal_12375	XP_393848.3_APIME	GPRstn
Hsal_05665	XP_394096.2_APIME	GPRqfc16
Hsal_01941	XP_392888.2_APIME	GPRqfc5

Hsal_06869	XP_001123118.1_APIME	GPRqfc18
Hsal_01127	XP_396248.3_APIME	GPRstn
Hsal_00619	XP_397259.3_APIME	GPRf2
Hsal_01406	XP_001121114.1_APIME	GPRf1a
Hsal_05003	XP_392299.2_APIME	GPRqfc1
Leukokinin Family		
Hsal_20002	XP_396660.2_APIME	GPRlkk
Neurokinin/Tachykinin Family		
Hsal_20007	FBpp0074640_DROME	GPRtak1
Class E -Chemosensory receptors		
Odorant receptors		
Hsal_00588	XP_001121145.1_APIME	Odorant receptor 83b
Hsal_02698	XP_001120660.1_APIME	Odorant receptor 9a
Hsal_02699	XP_001120660.1_APIME	Odorant receptor 9a
Hsal_02701	XP_001120660.1_APIME	Odorant receptor 9a
Hsal_02730	XP_001120948.1_APIME	Odorant receptor 30a
Hsal_02732	XP_001120948.1_APIME	Odorant receptor 30a
Hsal_02733	XP_001120660.1_APIME	Odorant receptor 9a
Hsal_03460	XP_001120702.1_APIME	Odorant receptor 13a
Hsal_03461	XP_001120702.1_APIME	Odorant receptor 13a
Hsal_03844	XP_001121864.1_APIME	Odorant receptor 2a
Hsal_03845	XP_001121741.1_APIME	Odorant receptor 85b
Hsal_03846	XP_001121864.1_APIME	Odorant receptor 2a
Hsal_03847	XP_001121741.1_APIME	Odorant receptor 85b
Hsal_03848	XP_001121741.1_APIME	Odorant receptor 85b
Hsal_03849	XP_001121741.1_APIME	Odorant receptor 85b
Hsal_03850	XP_001121741.1_APIME	Odorant receptor 85b
Hsal_03851	XP_001121864.1_APIME	Odorant receptor 2a
Hsal_03984	XP_393094.3_APIME	Odorant receptor 67a
Hsal_04518	XP_001119972.1_APIME	Odorant receptor 13a
Hsal_04519	XP_001119972.1_APIME	Odorant receptor 13a
Hsal_04520	XP_001119972.1_APIME	Odorant receptor 13a
Hsal_04521	XP_001119972.1_APIME	Odorant receptor 13a
Hsal_04522	XP_001119972.1_APIME	Odorant receptor 13a
Hsal_04523	XP_001119972.1_APIME	Odorant receptor 13a
Hsal_04527	XP_001121864.1_APIME	Odorant receptor 2a
Hsal_04782	XP_623850.2_APIME	Odorant receptor 13a
Hsal_04783	XP_623850.2_APIME	Odorant receptor 13a
Hsal_05300	XP_393094.3_APIME	Odorant receptor 67a
Hsal_05539	XP_001121864.1_APIME	Odorant receptor 2a

Hsal_05540	XP_001121864.1_APIME	Odorant receptor 2a
Hsal_05541	XP_001121864.1_APIME	Odorant receptor 2a
Hsal_05542	XP_001121864.1_APIME	Odorant receptor 2a
Hsal_05543	XP_001121864.1_APIME	Odorant receptor 2a
Hsal_05544	XP_001121864.1_APIME	Odorant receptor 2a
Hsal_05545	XP_001121741.1_APIME	Odorant receptor 85b
Hsal_05546	XP_001121864.1_APIME	Odorant receptor 2a
Hsal_05547	XP_001121864.1_APIME	Odorant receptor 2a
Hsal_05548	XP_001121741.1_APIME	Odorant receptor 85b
Hsal_05549	XP_001121741.1_APIME	Odorant receptor 85b
Hsal_05550	XP_001121659.1_APIME	Odorant receptor 85b
Hsal_05551	XP_001121659.1_APIME	Odorant receptor 85b
Hsal_05552	XP_001121741.1_APIME	Odorant receptor 85b
Hsal_05553	XP_001121741.1_APIME	Odorant receptor 85b
Hsal_05554	XP_001121741.1_APIME	Odorant receptor 85b
Hsal_05555	XP_001121659.1_APIME	Odorant receptor 85b
Hsal_05556	XP_001121741.1_APIME	Odorant receptor 85b
Hsal_05557	XP_001121080.1_APIME	Odorant receptor 83b
Hsal_05558	XP_001121080.1_APIME	Odorant receptor 83b
Hsal_06096	XP_001121741.1_APIME	Odorant receptor 85b
Hsal_06098	XP_001120660.1_APIME	Odorant receptor 9a
Hsal_06688	XP_623850.2_APIME	Odorant receptor 13a
Hsal_06690	XP_623850.2_APIME	Odorant receptor 13a
Hsal_06691	XP_623850.2_APIME	Odorant receptor 13a
Hsal_06692	XP_623850.2_APIME	Odorant receptor 13a
Hsal_06693	XP_623850.2_APIME	Odorant receptor 13a
Hsal_06816	XP_001123098.1_APIME	Odorant receptor 85b
Hsal_06818	XP_001122936.1_APIME	Odorant receptor 2a
Hsal_06819	XP_001123098.1_APIME	Odorant receptor 85b
Hsal_07035	XP_001120660.1_APIME	Odorant receptor 9a
Hsal_07219	XP_001121741.1_APIME	Odorant receptor 85b
Hsal_07296	XP_001122649.1_APIME	Odorant receptor 13a
Hsal_07622	XP_001123098.1_APIME	Odorant receptor 85b,
Hsal_07623	XP_001123098.1_APIME	Odorant receptor 85b,
Hsal_07952	XP_001122552.1_APIME	Odorant receptor 43a
Hsal_07953	XP_001122552.1_APIME	Odorant receptor 43a
Hsal_07954	XP_001122552.1_APIME	Odorant receptor 43a
Hsal_07955	XP_001122552.1_APIME	Odorant receptor 43a
Hsal_08922	XP_001123098.1_APIME	Odorant receptor 85b
Hsal_08924	XP_001120660.1_APIME	Odorant receptor 9a
Hsal_08925	XP_001122936.1_APIME	Odorant receptor 2a
Hsal_08926	XP_001123098.1_APIME	Odorant receptor 85b
Hsal_08927	XP_001123098.1_APIME	Odorant receptor 85b
Hsal_08928	XP_001123098.1_APIME	Odorant receptor 85b

Hsal_08930	XP_001123098.1_APIME	Odorant receptor 85b
Hsal_09259	XP_001120702.1_APIME	Odorant receptor 13a
Hsal_09318	XP_623850.2_APIME	Odorant receptor 13a
Hsal_09505	XP_001122971.1_APIME	Odorant response abnormal 4
Hsal_10531	XP_623850.2_APIME	Odorant receptor 13a
Hsal_10532	XP_623850.2_APIME	Odorant receptor 13a
Hsal_11045	XP_623850.2_APIME	Odorant receptor 13a
Hsal_11047	XP_623850.2_APIME	Odorant receptor 13a
Hsal_11857	XP_623850.2_APIME	Odorant receptor 13a
Hsal_11971	XP_001120948.1_APIME	Odorant receptor 30a
Hsal_13478	XP_623850.2_APIME	Odorant receptor 13a
Hsal_13774	XP_001120660.1_APIME	Odorant receptor 9a
Hsal_14251	XP_001121358.1_APIME	Odorant receptor 30a
Hsal_15170	XP_001122552.1_APIME	Odorant receptor 43a
Hsal_15840	XP_001121659.1_APIME	Odorant receptor 85b
Hsal_15988	XP_001123098.1_APIME	Odorant receptor 85b
Hsal_15990	XP_001122936.1_APIME	Odorant receptor 2a
Hsal_15991	XP_001123098.1_APIME	Odorant receptor 85b
Hsal_15992	XP_001123098.1_APIME	Odorant receptor 85b
Hsal_15993	XP_001123098.1_APIME	Odorant receptor 85b
Hsal_16050	XP_001123098.1_APIME	Odorant receptor 85b
Hsal_17824	XP_001120948.1_APIME	Odorant receptor 30a
Hsal_18122	XP_001123098.1_APIME	Odorant receptor 85b
Hsal_18146	XP_001122552.1_APIME	Odorant receptor 43a
Hsal_18150	XP_001123098.1_APIME	Odorant receptor 85b
# Gustatory receptors		
Hsal_02737	XP_001121326.1_APIME	Gustatory receptor 43a
Hsal_10419	XP_001121326.1_APIME	Gustatory receptor 43a
Hsal_03101	XP_001121165.1_APIME	Gustatory receptor 68a
Hsal_11870	XP_001121326.1_APIME	Gustatory receptor 43a
Hsal_10478	XP_397125.3_APIME	Gustatory receptor 64f
Hsal_10479	XP_001123138.1_APIME	Gustatory receptor 64f

Supplementary table S18. Family size variation in 4 Hymenoptera and *D. melanogaster*. Number of genes in each genome belonging to orthologous families showing significant size variation between organisms are shown. CF: *C. floridanus*; HS: *H. saltator*; AM: *A. mellifera*; NV: *N. vitripennis*; DM: *D. melanogaster*. Putative functions were obtained by annotations as available in order of priority from 1) human homologues; 2) *D. melanogaster* homologues; 3) InterProScan and relative GO terms.

<i>D. melanogaster</i> branch					
Putative function	CF	HS	AM	NV	DM
Serine protease	2	3	3	7	17
Cuticle	15	11	14	17	44
GST	5	5	2	8	29
Metallopeptidase	1	1	1	1	12
Lysozyme-like	2	1	2	1	12
UDP glucuronosyltransferase	11	16	10	20	34
Odorant binding	7	5	6	5	19
Lipase	11	7	7	13	26
NA (CG9168)	4	4	3	3	16
Histone	3	2	2	15	24
Serine protease inhibitor	5	4	4	6	22
Protein glycosylation	5	5	4	5	21
Galactosyl-transferase	1	1	1	1	11
<i>N. vitripennis</i> branch					
Serine protease (kallikrein)	2	4	2	13	6
NA (CG31664)	1	1	1	14	2
CDK9	1	1	1	12	1
short chain dehydrogenase	6	7	5	19	9
15-hydroxyprostaglandin dehydrogenase	6	10	6	17	6
choline dehydrogenase	19	17	20	45	15
poly(A) polymerase	2	1	1	5	1
Lysosomal phosphatase	11	10	9	27	5
NA (CG11158)	1	3	1	6	3
Multicopper oxidase	6	8	6	12	5
Membrane metallopeptidase	4	7	5	18	5
Cholin kinase	5	2	3	20	4
Cholin kinase	2	2	2	13	2
Cysteine-rich secretory protein	3	1	2	10	1
Lipase	3	4	3	10	1
NA (CG33998)	1	1	1	13	1
Kynurenine-oxoglutarate transaminase	1	1	1	9	1
Metalloendopeptidase	2	2	1	6	1
ribonuclease T2	1	1	1	7	1
Vespoidea branch					
Chymotripsin	4	4	2	9	44
Serine protease	2	3	3	7	17

Histone H2B	4	5	3	18	23
Histone H1	1	1	1	11	23
Cuticle (larva)	6	9	9	21	24
Carboxypeptidase	4	5	3	12	17
GST	5	5	2	8	29
UDP glucuronosyltransferase	11	16	10	20	34
Odorant binding	1	1	1	6	3
NA (CG3106)	1	2	1	9	11
Acyltransferase	5	4	4	20	21
Histone H3	3	2	2	15	24
S-phase kinase-associated	3	1	1	8	7
Phosphatidylethanolamine binding	2	2	2	7	8
<i>A. mellifera</i> branch					
Fatty acid acil-CoA reductase	13	12	8	21	17
Tachykinin receptor	2	1	7	3	5
Cytochrome P450 4V2	22	11	5	30	31
Su(Hw), Jim, Meics	12	1	15	10	15
Retinaldehyde-binding	15	10	7	13	17
Acyl-CoA synthetase	8	16	5	13	15
Glutamate transporter	2	1	6	3	1
E3 ligase (Delta receptor)	2	1	5	1	1
Esterase/lipase	36	7	4	26	23
Odorant binding	11	11	6	9	4
Geranyltransferase	1	2	7	1	1
Formicidae branch					
Juvenile hormone methyltransferase	8	6	1	1	1
Aldo-keto reductase	3	3	5	11	8
Heat shock response	3	5	9	8	10
Yellow	6	10	20	20	11
Pantetheine hydrolase	1	1	3	6	4
<i>C. floridanus</i> branch					
Juvenile hormone methyltransferase	8	6	1	1	1
Response to heat and humidity	12	3	3	3	2
Cuticle and larval cuticle	6	9	9	21	24
Short chain dehydrogenase	14	2	1	1	1
NA	7	3	3	3	8
Alpha-beta hydrolase	1	3	2	4	7
Heat shock response	3	5	9	8	10
Mitochondrial ribosome	4	1	1	5	1
Metalloprotease	6	3	1	2	2
Carbohydrate metabolism	11	6	5	8	11
Gephyrin/cinnamon (synaptic function?)	7	1	1	1	1
Nucleolar RNP	5	1	1	1	1
Isoprenoid and sterol synthesis	9	1	1	1	1

Eye pigment/guanine metabolism	7	1	1	1	1
<i>H. saltator</i> branch					
Carboxylesterase	23	12	19	36	27
Su(Hw), Jim, Meics	12	1	15	10	15
Organic cation transport	2	7	4	3	5
CNS development (glaikit)	1	6	1	2	1
Polyamine biosynthesis	3	5	1	4	2
Selenocysteine methyltransferase	2	7	2	1	2

Supplementary table S19. Distribution of genomic features according to G+C content in four Hymenoptera. O/E: observed versus expected ratio; TR: tandem repeats.

<i>C. floridanus</i>						
GC%	Length	gene # (O/E)	exon (%)	LINE (%)	SINE (%)	TR (%)
38-60	45,163,308	0.80	6.98	1.55	0.017	2.70
34-38	58,020,258	0.78	8.11	0.58	0.021	3.43
30-34	63,228,163	1.10	10.25	0.61	0.020	3.90
26-30	44,095,177	1.27	10.52	0.51	0.028	4.15
10-26	20,980,898	1.16	7.51	0.56	0.018	4.69
<i>H. saltator</i>						
GC%	Length	gene # (O/E)	exon (%)	LINE (%)	SINE (%)	TR (%)
51-67	59,504,313	0.35	2.77	3.18	0.091	4.94
46-51	75,196,246	0.52	4.56	1.73	0.085	4.30
40-46	67,759,303	1.01	8.39	1.56	0.057	3.64
34-40	51,317,092	1.77	11.61	1.49	0.039	3.34
15-34	34,621,385	1.99	9.67	1.02	0.030	3.33
<i>A. mellifera</i>						
GC%	Length	gene # (O/E)	exon (%)	LINE (%)	SINE (%)	TR (%)
40-60	45,214,692	0.30	4.33	0.30	0.073	3.73
34-40	58,402,032	0.36	4.11	0.20	0.026	6.01
29-34	45,165,628	0.72	6.53	0.22	0.016	6.34
22-29	48,007,778	2.07	14.96	0.25	0.013	4.79
5-22	30,108,729	2.00	9.76	0.21	0.007	6.54
<i>N. vitripennis</i>						
GC%	Length	gene # (O/E)	exon (%)	LINE (%)	SINE (%)	TR (%)
46-65	53,289,430	0.46	7.04	1.56	0.092	3.01
42-46	47,848,515	0.80	9.92	2.69	0.063	5.73
38-42	47,201,779	1.29	14.27	5.25	0.034	7.96
34-38	44,682,486	1.40	14.38	6.21	0.017	11.19
20-34	43,710,024	1.15	10.68	6.37	0.013	7.20

Supplementary table S20. Statistics for the automated annotation of protein coding genes in *C. floridanus* and *H. saltator*.

		<i>C. floridanus</i>	<i>H. saltator</i>
Total	Gene models	17,064	18,564
	Expressed (%)	81	84
GLEAN	Gene models	13,386	13,341
	Expressed (%)	82	84
Rejected by GLEAN			
Homology	Gene models	2,005	3,033
	Expressed (%)	62	70
<i>De novo</i>	Gene models	1,673	2,190
	Expressed (%)	100	100

Supplementary table S21. Top 40 GO terms enriched in 1:1 orthologues conserved between ants, other insects, and humans.

GO ID	GO Term	Class	Level	Adj. P value
GO:0044237	cellular metabolic process	BP	3	9.89E-39
GO:0044424	intracellular part	CC	3	2.56E-36
GO:0009987	cellular process	BP	2	1.40E-33
GO:0044260	cellular macromolecule metabolic process	BP	4	1.21E-31
GO:0034960	cellular biopolymer metabolic process	BP	5	1.21E-31
GO:0005737	cytoplasm	CC	4	1.41E-28
GO:0044267	cellular protein metabolic process	BP	5	2.93E-23
GO:0005622	intracellular	CC	3	8.56E-23
GO:0006412	translation	BP	5	3.32E-22
GO:0032991	macromolecular complex	CC	2	4.11E-18
GO:0032555	purine ribonucleotide binding	MF	5	1.90E-17
GO:0000166	nucleotide binding	MF	3	1.02E-16
GO:0043229	intracellular organelle	CC	3	1.58E-16
GO:0009058	biosynthetic process	BP	3	1.64E-15
GO:0017076	purine nucleotide binding	MF	4	1.80E-15
GO:0044444	cytoplasmic part	CC	4	2.54E-15
GO:0034660	ncRNA metabolic process	BP	6	2.73E-15
GO:0044249	cellular biosynthetic process	BP	4	1.92E-14
GO:0010467	gene expression	BP	4	3.85E-14
GO:0008152	metabolic process	BP	2	7.13E-14
GO:0003723	RNA binding	MF	4	7.94E-14
GO:0016874	ligase activity	MF	3	1.32E-13
GO:0043283	biopolymer metabolic process	BP	4	1.91E-13
GO:0005524	ATP binding	MF	7	2.13E-13
GO:0030529	ribonucleoprotein complex	CC	3	2.51E-13
GO:0009059	macromolecule biosynthetic process	BP	4	2.81E-13
GO:0034645	cellular macromolecule biosynthetic process	BP	5	3.50E-13
GO:0044428	nuclear part	CC	4	1.49E-12
GO:0001882	nucleoside binding	MF	3	3.14E-12
GO:0030554	adenyl nucleotide binding	MF	5	3.81E-12
GO:0043284	biopolymer biosynthetic process	BP	5	4.79E-12
GO:0034961	cellular biopolymer biosynthetic process	BP	6	7.52E-12
GO:0006399	tRNA metabolic process	BP	7	4.21E-11
GO:0006396	RNA processing	BP	5	9.16E-11
GO:0016741	transferase activity, transferring one-carbon groups	MF	4	9.42E-11
GO:0044238	primary metabolic process	BP	3	9.70E-11
GO:0006418	tRNA aminoacylation for protein translation	BP	6	2.26E-10
GO:0006807	nitrogen compound metabolic process	BP	3	2.28E-10
GO:0008168	methyltransferase activity	MF	5	2.39E-10

Supplementary table S22. Top 40 GO terms enriched in many:many homologues conserved between ants, other insects, and humans (adjusted P-values calculated on the *C. floridanus* genome).

GO ID	GO Term	Class	Level	Adj. P value
GO:0003824	catalytic activity	MF	2	5.29E-36
GO:0016491	oxidoreductase activity	MF	3	2.51E-23
GO:0005215	transporter activity	MF	2	2.09E-17
GO:0016021	integral to membrane	CC	5	1.00E-16
GO:0006810	transport	BP	3	1.00E-16
GO:0055085	transmembrane transport	BP	3	2.76E-16
GO:0051179	localization	BP	2	4.12E-16
GO:0020037	heme binding	MF	4	8.04E-16
GO:0031224	intrinsic to membrane	CC	4	6.32E-15
GO:0005506	iron ion binding	MF	7	2.24E-14
GO:0009055	electron carrier activity	MF	2	7.71E-13
GO:0044425	membrane part	CC	3	5.08E-11
GO:0016787	hydrolase activity	MF	3	2.38E-10
GO:0004497	monooxygenase activity	MF	4	4.56E-10
GO:0016020	membrane	CC	3	6.42E-09
GO:0015294	solute:cation symporter activity	MF	7	9.83E-08
GO:0015291	secondary active transmembrane transporter activity	MF	5	2.24E-07
GO:0015075	ion transmembrane transporter activity	MF	5	2.24E-07
GO:0022857	transmembrane transporter activity	MF	3	4.87E-07
GO:0015370	solute:sodium symporter activity	MF	8	7.64E-07
GO:0016614	oxidoreductase activity, acting on CH-OH group of donors	MF	4	1.59E-06
GO:0009056	catabolic process	BP	3	2.01E-06
GO:0008238	exopeptidase activity	MF	6	2.21E-06
GO:0022891	substrate-specific transmembrane transporter activity	MF	4	2.99E-06
GO:0005887	integral to plasma membrane	CC	6	3.59E-06
GO:0022836	gated channel activity	MF	6	4.57E-06
GO:0022804	active transmembrane transporter activity	MF	4	9.14E-06
GO:0003924	GTPase activity	MF	8	1.08E-05
GO:0009057	macromolecule catabolic process	BP	4	1.15E-05
GO:0048037	cofactor binding	MF	3	1.37E-05
GO:0003993	acid phosphatase activity	MF	7	1.39E-05
GO:0043285	biopolymer catabolic process	BP	5	2.19E-05
GO:0034622	cellular macromolecular complex assembly	BP	4	2.57E-05
GO:0015276	ligand-gated ion channel activity	MF	7	3.25E-05
GO:0005975	carbohydrate metabolic process	BP	4	3.25E-05
GO:0004190	aspartic-type endopeptidase activity	MF	7	3.97E-05
GO:0016209	antioxidant activity	MF	2	5.46E-05
GO:0005230	extracellular ligand-gated ion channel activity	MF	8	6.81E-05
GO:0004553	hydrolase activity, hydrolyzing O-glycosyl compounds	MF	5	7.29E-05

Supplementary table S23. Top 40 GO terms enriched in protein conserved between insects but not humans (adjusted P-values calculated on the *C. floridanus* genome).

GO ID	GO Term	Class	Level	Adj. P value
GO:0042302	structural constituent of cuticle	MF	3	1.93E-30
GO:0004252	serine-type endopeptidase activity	MF	6	4.48E-12
GO:0003700	transcription factor activity	MF	3	2.12E-09
GO:0008061	chitin binding	MF	5	1.03E-08
GO:0030247	polysaccharide binding	MF	4	1.60E-08
GO:0043565	sequence-specific DNA binding	MF	5	4.49E-08
GO:0030246	carbohydrate binding	MF	3	1.11E-07
GO:0006030	chitin metabolic process	BP	6	1.97E-07
GO:0006355	regulation of transcription, DNA-dependent	BP	6	4.80E-07
GO:0045449	regulation of transcription	BP	6	1.64E-06
GO:0006351	transcription, DNA-dependent	BP	6	3.69E-06
GO:0005344	oxygen transporter activity	MF	4	3.69E-06
GO:0004872	receptor activity	MF	4	3.88E-06
GO:0004888	transmembrane receptor activity	MF	5	8.46E-06
GO:0031323	regulation of cellular metabolic process	BP	4	1.41E-05
GO:0030528	transcription regulator activity	MF	2	1.81E-05
GO:0003677	DNA binding	MF	4	1.81E-05
GO:0005576	extracellular region	CC	2	2.46E-05
GO:0005488	binding	MF	2	2.46E-05
GO:0046872	metal ion binding	MF	5	3.16E-05
GO:0006350	transcription	BP	5	3.76E-05
GO:0043169	cation binding	MF	4	6.15E-05
GO:0005921	gap junction	CC	7	1.05E-04
GO:0046914	transition metal ion binding	MF	6	1.11E-04
GO:0004871	signal transducer activity	MF	3	1.59E-04
GO:0007156	homophilic cell adhesion	BP	5	1.59E-04
GO:0008270	zinc ion binding	MF	7	1.91E-04
GO:0005102	receptor binding	MF	4	2.02E-04
GO:0004713	protein tyrosine kinase activity	MF	7	8.01E-04
GO:0005515	protein binding	MF	3	1.37E-03
GO:0005549	odorant binding	MF	3	1.43E-03
GO:0050794	regulation of cellular process	BP	3	2.48E-03
GO:0005158	insulin receptor binding	MF	5	3.20E-03
GO:0004930	G-protein coupled receptor activity	MF	6	6.99E-03
GO:0045087	innate immune response	BP	4	7.45E-03
GO:0006955	immune response	BP	3	7.55E-03
GO:0004857	enzyme inhibitor activity	MF	3	1.18E-02
GO:0030414	peptidase inhibitor activity	MF	4	1.20E-02
GO:0065007	biological regulation	BP	2	1.46E-02
GO:0005044	scavenger receptor activity	MF	6	1.55E-02

Supplementary table S24. Top GO terms enriched in protein conserved between Hymenoptera.

GO ID	GO Term	Class	Level	Adj. P value
GO:0007606	sensory perception of chemical stimulus	BP	7	1.52E-67
GO:0004984	olfactory receptor activity	MF	7	1.98E-61
GO:0007608	sensory perception of smell	BP	8	1.98E-61
GO:0005549	odorant binding	MF	3	5.79E-58
GO:0032501	multicellular organismal process	BP	2	4.31E-53
GO:0004930	G-protein coupled receptor activity	MF	6	1.85E-52
GO:0008270	zinc ion binding	MF	7	6.87E-43
GO:0004888	transmembrane receptor activity	MF	5	2.27E-35
GO:0004872	receptor activity	MF	4	9.86E-31
GO:0046914	transition metal ion binding	MF	6	1.24E-28
GO:0004871	signal transducer activity	MF	3	6.72E-23
GO:0005488	binding	MF	2	1.14E-21
GO:0046872	metal ion binding	MF	5	7.19E-20
GO:0003676	nucleic acid binding	MF	3	1.30E-18
GO:0004523	ribonuclease H activity	MF	9	2.93E-15
GO:0022900	electron transport chain	BP	4	3.74E-08
GO:0006091	generation of precursor metabolites and energy	BP	4	8.22E-08
GO:0045333	cellular respiration	BP	6	1.44E-07
GO:0042773	ATP synthesis coupled electron transport	BP	6	4.01E-07
GO:0008137	NADH dehydrogenase (ubiquinone) activity	MF	7	2.12E-05
GO:0004518	nuclease activity	MF	5	3.86E-05
GO:0006119	oxidative phosphorylation	BP	5	4.10E-05
GO:0050909	sensory perception of taste	BP	8	1.08E-03
GO:0030247	polysaccharide binding	MF	4	7.98E-03
GO:0007548	sex differentiation	BP	4	9.35E-03
GO:0008061	chitin binding	MF	5	1.26E-02
GO:0022890	inorganic cation transmembrane transporter activity	MF	7	1.45E-02
GO:0042775	mitochondrial ATP synthesis coupled electron transport	BP	7	1.57E-02
GO:0030246	carbohydrate binding	MF	3	1.62E-02
GO:0015078	hydrogen ion transmembrane transporter activity	MF	9	1.87E-02
GO:0045263	proton-transporting ATP synthase complex, coupling factor F(o)	CC	4	1.94E-02
GO:0005179	hormone activity	MF	5	2.29E-02
GO:0006030	chitin metabolic process	BP	6	2.34E-02
GO:0006123	mitochondrial electron transport, cytochrome c to oxygen	BP	6	2.61E-02
GO:0016020	membrane	CC	3	3.23E-02
GO:0004129	cytochrome-c oxidase activity	MF	5	4.18E-02

Supplementary table S25. IPR domain enrichment analysis for alternatively spliced genes in *C. floridanus* and *H. saltator*.

<i>C. floridanus</i>		
IPR ID	Domain name	Adj. P value
IPR002290	Serine/threonine-protein kinase domain	2.9E-06
IPR001452	Src homology-3 domain	2.9E-06
IPR020635	Tyrosine-protein kinase, subgroup, catalytic domain	1.0E-05
IPR017442	Serine/threonine-protein kinase-like domain	2.7E-05
IPR000719	Protein kinase, catalytic domain	5.1E-05
IPR000198	RhoGAP	1.6E-04
IPR008144	Guanylate kinase	3.7E-04
IPR008145	Guanylate kinase/L-type calcium channel region	3.7E-04
IPR001715	Calponin-like actin-binding	6.8E-04
IPR001849	Pleckstrin homology	1.1E-03
IPR001965	Zinc finger, PHD-type	1.4E-03
IPR003961	Fibronectin, type III	1.4E-03
IPR001202	WW/Rsp5/WWP	3.4E-03
<i>H. saltator</i>		
IPR ID	Domain name	Adj. P value
IPR006210	EGF-like	1.4E-03
IPR000742	EGF-like, type 3	3.8E-02
IPR003961	Fibronectin, type III	3.8E-02
IPR001849	Pleckstrin homology	6.0E-02
IPR001452	Src homology-3 domain	7.1E-02
IPR001881	EGF-like calcium-binding	7.1E-02
IPR001660	Sterile alpha motif SAM	7.6E-02
IPR020635	Tyrosine-protein kinase, subgroup, catalytic domain	7.6E-02

Supplementary table S26. GO enrichment analysis for alternatively spliced genes in *C. floridanus* and *H. saltator*.

<i>C. floridanus</i>				
GO ID	GO Term	Class	Level	Adj. P value
GO:0016773	phosphotransferase activity, alcohol group as acceptor	MF	5	2.5E-08
GO:0016301	kinase activity	MF	5	1.1E-07
GO:0004674	protein serine/threonine kinase activity	MF	7	5.2E-06
GO:0016772	transferase activity, transferring phosphorus-containing groups	MF	4	1.2E-05
GO:0001882	nucleoside binding	MF	3	2.5E-05
GO:0005524	ATP binding	MF	7	2.5E-05
GO:0030554	adenyl nucleotide binding	MF	5	2.5E-05
GO:0004672	protein kinase activity	MF	6	2.9E-05
GO:0006468	protein amino acid phosphorylation	BP	7	3.4E-05
GO:0043687	post-translational protein modification	BP	7	3.5E-05
GO:0051056	regulation of small GTPase mediated signal transduction	BP	6	6.5E-05
GO:0009966	regulation of signal transduction	BP	5	2.3E-04
GO:0030695	GTPase regulator activity	MF	4	2.3E-04
GO:0006464	protein modification process	BP	6	4.6E-04
GO:0005083	small GTPase regulator activity	MF	5	5.0E-04
GO:0046578	regulation of Ras protein signal transduction	BP	7	5.8E-04
<i>H. saltator</i>				
GO ID	GO Term	Class	Level	Adj. P value
GO:0003824	catalytic activity	MF	2	2.8E-08
GO:0008152	metabolic process	BP	2	6.8E-06
GO:0001882	nucleoside binding	MF	3	2.5E-04
GO:0030554	adenyl nucleotide binding	MF	5	2.5E-04
GO:0051056	regulation of small GTPase mediated signal transduction	BP	6	2.9E-04
GO:0016301	kinase activity	MF	5	3.8E-04
GO:0005083	small GTPase regulator activity	MF	5	4.4E-04
GO:0007242	intracellular signaling cascade	BP	5	6.4E-04
GO:0016740	transferase activity	MF	3	6.6E-04
GO:0030234	enzyme regulator activity	MF	2	6.9E-04
GO:0016773	phosphotransferase activity, alcohol group as acceptor	MF	5	7.7E-04
GO:0000166	nucleotide binding	MF	3	7.7E-04
GO:0060589	nucleoside-triphosphatase regulator activity	MF	3	7.7E-04
GO:0030695	GTPase regulator activity	MF	4	8.3E-04
GO:0050790	regulation of catalytic activity	BP	4	9.4E-04
GO:0005524	ATP binding	MF	7	9.5E-04
GO:0046578	regulation of Ras protein signal transduction	BP	7	1.0E-03

Supplementary table S27. *C. floridanus* and HS homologues of *D. melanogaster* genes involved in embryonic development.

<i>D. melanogaster</i>	<i>C. floridanus</i>	<i>H. saltator</i>
<i>trunks</i>	-	-
<i>torso</i>	-	-
<i>gurken</i>	-	-
<i>bicoid</i>	-	-
<i>oskar</i>	-	-
<i>orthodenticle</i>	Cflo_05730, Cflo_05731	Hsal_07656, Hsal_07660
<i>hunchback</i>	Cflo_04393, Cflo_02791	Hsal_07655, Hsal_10154
<i>kruppel</i>	Cflo_09059	Hsal_20013
<i>knirps</i>	Cflo_03607	Hsal_02933
<i>caudal</i>	Cflo_06253	Hsal_13323
<i>giant</i>	Cflo_13414	Hsal_00371
<i>nanos</i>	Cflo_02681	Hsal_04193
<i>tailless</i>	Cflo_06716	Hsal_06558
<i>huckebein</i>	—	—
<i>vasa</i>	Cflo_00068	Hsal_20014
<i>ultrabithorax</i>	Cflo_00848	Hsal_01804
<i>tfg-beta</i>	Cflo_10644	Hsal_06239
<i>hedgehog</i>	Cflo_03964, Cflo_03966	Hsal_01588, Hsal_01589
<i>dorsal</i>	Cflo_08314	Hsal_01917
<i>spatzle</i>	Cflo_09498	Hsal_04331
<i>wingless</i>	Cflo_00154	Hsal_05398

Supplementary table S28. Best homologous Swissprot entries for proteins containing homeobox domains (IPR001356).

<i>C. floridanus</i>	<i>H. saltator</i>	<i>Human</i>	<i>Drosophila</i>	Overall
Cflo_09380	Hsal_01309	ALX1	AL	ALX1_HUMAN
Cflo_00163	Hsal_05386	ARX	AL	AL_DROME
Cflo_03631	Hsal_11135	ARX	AL	ARX_HUMAN
Cflo_05070	Hsal_15570	ARX	AL	ARX_DANRE
Cflo_10143	Hsal_03582	ARX	AL	ARX_HUMAN
Cflo_02573	Hsal_10764	BARH1	BARH1	BARH1_HUMAN
Cflo_01350	Hsal_01133	BARH2	BARH2	HM19_CAEEL
Cflo_08609	Hsal_12287	BARH2	BARH1	BARH1_DROME
Cflo_13663	Hsal_12039	BARH2	BARH2	BARH2_MOUSE
Cflo_06955	Hsal_10254	BSH	BSH	BSH_DROME
Cflo_06253	Hsal_13323	CDX4	CAD	HMD1_CHICK
Cflo_11771	Hsal_04719	CSN1	CSN1	CSN1_DROME
Cflo_11187	Hsal_04356	CT026	—	CT026_HUMAN
Cflo_05239	Hsal_01674	CUX1	CUT	CUT_DROME
Cflo_02583	Hsal_10771	DLX6	DLL	DLL_DROME
Cflo_04366	Hsal_07250	EMX2	EMS	EMS_DROME
Cflo_10805	Hsal_11276	EVX2	EVE	EVX2_HUMAN
Cflo_13121	Hsal_00195	GBX2	UNPG	GBX2_HUMAN
Cflo_08687	Hsal_07068	GSC	GSC	GSC_DROME
Cflo_02933	Hsal_03841	GSX1	ZEN1	GSX1_HUMAN
Cflo_10458	Hsal_11910	HLX	HMH2	HLX_BOVIN
Cflo_06767	Hsal_20010	HME2	HMEN	HME60_APIME
Cflo_06768	Hsal_07413	HME2	HMEN	HMEN_DROVI
Cflo_07826	—	HNF6	ONEC	ONEC_DROME
Cflo_10727	—	HXA1	LAB	LAB_DROME
Cflo_00859	—	HXA10	ABDB	MAB5_CAEEL
Cflo_10303	Hsal_01793	HXA4	DFD	DFD_DROME
Cflo_00861	Hsal_01796	HXA5	SCR	SCR_DROME
Cflo_00841	Hsal_01810	HXA9	ABDB	ABDB_DROME
Cflo_10730	Hsal_01789	HXB2	HMPB	HMPB_DROME
Cflo_00848	Hsal_01804	HXB6	UBX	UBX_DROSI
Cflo_00844	—	HXB7	ABDA	ABDA_ANOGA
Cflo_00855	Hsal_01799	HXB7	ANTP	ANTP_DROSU
—	Hsal_01798	HXB7	ANTP	HLOX2_HELRO
Cflo_10299	—	HXD3	HMPB	LIN39_CAEEL
Cflo_05547	Hsal_07615	IRX4	ARA	ARA_DROME
Cflo_05544	Hsal_07613	IRX6	CAUP	CAUP_DROME
Cflo_03812	Hsal_09201	ISL1	AWH	ISL1_HUMAN
Cflo_06331	Hsal_20011	LBX1	BARH1	LBX1_XENTR
Cflo_02985	Hsal_05578	LHX1	GSBN	LHX1_HUMAN

Cflo_05886	Hsal_04277	LHX2	APTE	APTE_DROME
Cflo_06863-	Hsal_00750	LHX3	AWH	LHX3_XENLA
Cflo_13399	—	LHX61	AWH	AWH_DROME
—	Hsal_01220	LHX61	AWH	AWH_DROME
Cflo_13398	Hsal_01221	LHX8	AWH	AWH_DROME
Cflo_05883	Hsal_04279	LHX9	APTE	LHX9_HUMAN
Cflo_07902	Hsal_05363	LMX1B	AWH	LMX1B_MOUSE
Cflo_12378	Hsal_06629	MEOX2	DFD	MEOX2_XENLA
Cflo_11297	Hsal_07816	MNX1	ROUGH	ROUGH_DROME
Cflo_12078	Hsal_06516	MNX1	ANTP	MNX1_MOUSE
—	Hsal_06440	MSX1	HMSH	HOX71_XENLA
Cflo_06334	Hsal_06438	MSX2	HMSH	HMSH_DROME
Cflo_06335	Hsal_06439	MSX2	HMSH	HM17_APIME
Cflo_10523	Hsal_00498	NKX21	VND	NKX21_HUMAN
Cflo_02610	Hsal_09219	NKX22	VND	NX22A_DANRE
Cflo_06333	Hsal_06437	NKX26	VND	NKX26_HUMAN
Cflo_06366	Hsal_06171	NKX61	HMX	NKX61_HUMAN
Cflo_03636	Hsal_11140	OTP	OTP	OTP_LYTVA
Cflo_05731	Hsal_07656	OTX1	HMOC	OTX1_HUMAN
Cflo_05730	—	OTX2	HMOC	OTX5B_XENLA
—	Hsal_07660	OTX2	HMOC	OTX2_HUMAN
Cflo_06472	Hsal_08388	PAX3	GSB	GSB_DROME
Cflo_07008	Hsal_04609	PAX3	PRD	PRD_DROME
Cflo_02505	Hsal_03768	PAX5	PAX6	RX2_ORYLA
Cflo_00481	Hsal_09597	PAX6	PAX6	PAX6_XENLA
Cflo_06474	Hsal_08389	PAX7	GSBN	GSBN_DROME
Cflo_02624	Hsal_09224	PBX2	EXD	EXD_DROPS
Cflo_03443	Hsal_06849	PHX2A	RX	PROP1_CEBAP
Cflo_04174	Hsal_11877	PHX2B	AL	PHX2B_HUMAN
Cflo_01523	—	PITX3	PITX1	PITX3_HUMAN
Cflo_06953	Hsal_10256	PKNX2	HTH	HTH_DROME
Cflo_10687	Hsal_07370	PO2F1	PDM1	PO2F1_PIG
Cflo_06086	Hsal_03104	PO3F3	CF1A	CF1A_DROME
Cflo_05860	Hsal_10079	PO4F1	IPOU	IPOU_DROME
Cflo_02868	Hsal_02406	PO6F2	CF1A	PO6F2_HUMAN
Cflo_02794	Hsal_05271	PRRX2	GSBN	GSBN_DROME
Cflo_03633	—	RX	RX	RX2_CHICK
Cflo_07225	Hsal_07106	SATB1	—	SATB1_HUMAN
Cflo_05350	Hsal_12739	SHOX2	AL	SHOX2_HUMAN
Cflo_03402	Hsal_07526	SIX1	SO	SIX1_HUMAN
Cflo_00678	Hsal_01755	SIX3	OPTIX	SIX3_CHICK
Cflo_08297	Hsal_06384	SIX4	SO	SIX4_HUMAN
Cflo_06525	Hsal_02547	TGIF1	HTH	TGIF1_HUMAN
Cflo_08940	Hsal_05447	TLX3	HMX	TLX3_HUMAN

Cflo_05286	Hsal_09127	UNC4	UNC4	UNC4_DROME
Cflo_09678	Hsal_04407	VAX1	EMS	NOT2_XENLA
—	Hsal_11904	VSX1	GSC	VSX1_DANRE
Cflo_10715	—	ZEB2	ZFH1	ZFH1_DROME
Cflo_00361	Hsal_04077	ZFHX3	ZFH2	ZFHX3_HUMAN
—	Hsal_04718	—	—	—
Cflo_10606	Hsal_09622	—	—	—

Supplementary table S29. Annotation of *C. floridanus* and *H. saltator* homologues of genes implicated in sex determination pathways in other insects.

Gene	<i>A. mellifera</i>	<i>N. vitripennis</i>	<i>C. floridanus</i>	<i>H. saltator</i>
<i>complementary sex determiner</i>	EU101390	—	—	—
<i>feminizer</i>	NP_001128300	NP_001128299	Cflo_08057 Cflo_08058	Hsal_09737 Hsal_14962
<i>doublesex</i>	XP_001122464.1	ACJ65501	Cflo_02164	Hsal_04837
<i>transformer-2</i>	XP_001121070	XP_001601106	Cflo_13288	Hsal_11431
<i>fruitless</i>	XP_392552.3	NP_001157607	Cflo_05182	Hsal_10000
<i>dissatisfaction</i>	XP_624265.2	XP_001599315	Cflo_04265	Hsal_12124
<i>extra-marcochaetae</i>	NP_001135436	NP_001135435	Cflo_04121	Hsal_05337
<i>groucho</i>	XP_392425.3 XP_392421.2 XP_394649.2	XP_001604872.1 XP_001601479.1	Cflo_02304 Cflo_02308 Cflo_09970	Hsal_02320 Hsal_02297
<i>deadpan</i>	XP_001120814	XP_001601677	Cflo_02876	Hsal_02408
<i>female lethal(2)d</i>	XP_625211.1	XP_001605285	Cflo_01978	Hsal_05262
<i>virilizer</i>	XP_395082.3	XP_001606740	Cflo_07571	Hsal_08589
<i>daughterless</i>	XP_394032.3	XP_001604081	Cflo_06641	Hsal_01424
<i>intersex</i>	XP_395989	XP_001601927	Cflo_11260	Hsal_11631
<i>runt</i>	XP_001121886.1	XP_001601776.1	Cflo_03910	Hsal_12321
<i>hopscotch</i>	XP_001121783.1	XP_001602854	Cflo_06362	Hsal_06174
<i>sex lethal</i>	XP_394166.3	XP_001603257	Cflo_07621	Hsal_03886

Supplementary table S30. Number of genes containing domains related to cuticle biology.

Domain	Domain ID	CF	HS	AM	NV	DM
Chitinase	KOG4701	13	30	22	23	38
GH18 type II chitinases	cd00598	11	11	11	14	17
Chitinase class I	pfam00182	0	0	0	1	0
Insect cuticle protein	pfam00379	35	34	40	63	93
Chitin binding Peritrophin-A domain	pfam01607	29	35	26	38	51
Chitooligosaccharide deacetylase	pfam01522	2	2	2	2	3
Polysaccharide deacetylase	IPR002509	4	5	3	4	5

Supplementary table S31. Number of protease/peptidase in four Hymenoptera and *Drosophila* genomes according to InterProScan.

Domain name	IPR ID	CF	HS	AM	NV	DM
Peptidase S8/S53, subtilisin/kexin/sedolisin	IPR000209	6	6	6	5	5
Peptidase T1A, proteasome beta-subunit	IPR000243	0	0	0	0	7
Peptidase T2, asparaginase 2	IPR000246	4	2	6	4	5
Peptidase M41	IPR000642	3	3	3	3	3
Peptidase C1A, papain C-terminal	IPR000668	7	6	6	8	11
Peptidase C15, pyroglutamyl peptidase I	IPR000816	1	0	1	1	0
Peptidase M17, leucyl aminopeptidase, C-terminal	IPR000819	2	2	3	1	10
Peptidase M14, carboxypeptidase A	IPR000834	10	11	11	21	24
Peptidase M22, glycoprotease	IPR000905	3	2	2	2	2
Peptidase M24, structural domain	IPR000994	8	8	7	9	9
Peptidase C13, legumain	IPR001096	1	1	1	1	1
Peptidase A22A, presenilin	IPR001108	1	1	1	1	1
Peptidase S60, transferrin lactoferrin	IPR001156	4	5	3	4	3
Peptidase S1/S6, chymotrypsin/Hap	IPR001254	73	112	62	197	256
Peptidase C2, calpain	IPR001300	5	5	4	4	4
Peptidase C14, ICE, catalytic subunit p20	IPR001309	5	5	6	9	7
Peptidase S9, prolyl oligopeptidase, catalytic domain	IPR001375	7	7	7	7	9
Peptidase C19, ubiquitin carboxyl-terminal hydrolase 2	IPR001394	29	25	24	29	24
Peptidase A1	IPR001461	66	1	1	2	13
Peptidase M12A, astacin	IPR001506	5	3	3	9	13
Peptidase M2, peptidyl-dipeptidase A	IPR001548	6	5	4	3	6
Peptidase S10, serine carboxypeptidase	IPR001563	4	2	2	12	5
Peptidase M3A/M3B, thimet/oligopeptidase F	IPR001567	2	2	2	2	2
Peptidase M8, leishmanolysin	IPR001577	1	1	1	1	1
Peptidase C12, ubiquitin carboxyl-terminal hydrolase 1	IPR001578	3	3	3	3	4
Peptidase M12B, ADAM/reprolysin	IPR001590	12	12	15	21	9
Peptidase M24, methionine aminopeptidase	IPR001714	0	0	0	0	3
Peptidase M10A/M12B, matrixin/adamalysin	IPR001818	13	2	2	4	2
Peptidase S14, ClpP	IPR001907	1	1	1	1	1
Peptidase M48, Ste24p	IPR001915	2	2	2	2	3
Peptidase S1C, HrtA/DegP2/Q/S	IPR001940	0	0	0	0	1
Peptidase S16, Lon protease, C-terminal	IPR001984	0	0	0	0	3
Peptidase C14, caspase non-catalytic subunit p10	IPR002138	5	2	4	8	7
Peptidase S9B, dipeptidylpeptidase IV N-terminal	IPR002469	5	5	6	8	7
Peptidase S9A, prolyl oligopeptidase	IPR002470	0	0	0	0	1
Peptidase S54, rhomboid	IPR002610	5	6	7	6	8
Peptidase M12B, propeptide	IPR002870	7	4	9	8	4
Peptidase M20	IPR002933	3	4	2	4	7
Peptidase S16, lon N-terminal	IPR003111	3	3	4	3	2
Peptidase C48, SUMO/Sentrin/Ubl1	IPR003653	3	3	4	8	11
Peptidase S9A, oligopeptidase, N-terminal beta-propeller	IPR004106	1	1	1	1	2

Peptidase C1B, bleomycin hydrolase	IPR004134	1	1	1	1	2
Peptidase C54	IPR005078	3	2	2	1	2
Peptidase C50, separase	IPR005314	1	1	0	1	1
Peptidase M49, dipeptidyl-peptidase III	IPR005317	1	1	1	1	1
Peptidase S51, dipeptidase E	IPR005320	0	0	0	0	1
Peptidase, metallopeptidases	IPR006026	11	5	5	11	15
Peptidase C19, ubiquitin-specific peptidase, DUSP domain	IPR006615	2	2	2	1	2
Peptidase A22, presenilin signal peptide	IPR006639	3	3	3	3	3
Peptidase S59, nucleoporin	IPR007230	1	1	1	1	1
Peptidase M28	IPR007484	4	2	5	3	16
Peptidase M16, C-terminal	IPR007863	8	9	8	8	7
Peptidase M24B, X-Pro dipeptidase/aminopeptidase P N-terminal	IPR007865	2	2	2	2	2
Peptidase M19, renal dipeptidase	IPR008257	6	5	6	5	4
Peptidase S16, lon C-terminal	IPR008269	1	1	1	1	1
Peptidase M17, leucyl aminopeptidase, N-terminal	IPR008283	1	1	0	0	8
Peptidase M13	IPR008753	8	12	10	25	19
Peptidase S28	IPR008758	3	4	3	3	6
Peptidase M50	IPR008915	1	1	1	1	1
Peptidase M41, FtsH extracellular	IPR011546	3	3	3	3	3
Peptidase C14, caspase catalytic	IPR011600	6	5	6	9	7
Peptidase M20, dimerisation	IPR011650	3	4	2	3	7
Peptidase C26	IPR011697	1	1	1	1	2
Peptidase M16, N-terminal	IPR011765	6	6	5	7	6
Peptidase M12B, GON-ADAMTSs	IPR012314	1	1	1	1	1
Peptidase C78, ubiquitin fold modifier-specific peptidase 1/ 2	IPR012462	2	3	3	2	2
Peptidase C1A, propeptide	IPR012599	0	0	0	0	1
Propeptide, peptidase A1	IPR012848	1	1	1	1	2
Peptidase M12B, ADAM-TS	IPR013273	0	0	0	0	5
Peptidase M16C associated	IPR013578	1	1	1	1	1
Peptidase M1, membrane alanine aminopeptidase, N-terminal	IPR014782	62	28	10	16	24
Peptidase M1, leukotriene A4 hydrolase, aminopeptidase C-terminal	IPR015211	1	1	0	1	1
Peptidase C26, gamma-glutamyl hydrolase	IPR015527	1	1	0	1	2
Peptidase C14, caspase precursor p45, core	IPR015917	4	3	5	8	7
Peptidase M22, glycoprotease, subgroup	IPR017861	0	0	0	0	2
Protease inhibitor I8, cysteine-rich trypsin inhibitor-like subgroup	IPR018453	2	4	10	8	6
Peptidase M13, neprilysin, C-terminal	IPR018497	9	12	9	24	25
Peptidase aspartic, eukaryotic predicted	IPR019103	1	1	0	1	0
Peptidase M76, ATP23	IPR019165	1	1	1	1	1
Peptidase C65, otubain	IPR019400	1	1	1	3	1
Peptidase S24/S26A/S26B, conserved region	IPR019759	3	3	4	1	3

Supplementary table S32. Number of genes with domains belonging to detoxification enzyme families in *C. floridanus* and *H. saltator*.

Domain	Domain ID	CF	HS	AM	NV	DM
Cytochrome P450						
Cytochrome P450	IPR001128	132	93	60	96	87
Cytochrome P450, E-class, group I	IPR002401	0	0	0	0	63
Cytochrome P450, C-terminal	IPR017973	0	0	0	0	70
Glutathione S-transferases						
Glutathione S-transferase, N-terminal	IPR004045	14	14	12	20	39
Glutathione S-transferase, C-terminal	IPR004046	14	16	10	21	39
Glutathione S-transferase/chloride channel, C-terminal	IPR017933	15	16	11	24	41
Carboxy/cholinesterases						
Carboxylesterase, type B	IPR002018	43	27	29	48	35
Alpha/beta hydrolase fold-1	IPR000073	31	23	19	43	41
Divalent ion tolerance protein, CutA1	IPR004323	1	1	1	1	1

Supplementary table S33. Number of genes in immune defense pathways in *C. floridanus* and *H. saltator*.

Pathway	CF	HS	AM	DM
Immune defense	231	240	228	361
TOLL/IMD/JAK-STAT/JNK	76	69	68	98

Supplementary table S34. *C. floridanus* and *H. saltator* genes involved in immune defense.

Gene symbol	Kegg ID	<i>C. floridanus</i>	<i>H. saltator</i>
JAK-STAT pathway			
SHP2	K07293	Cflo_08744	Hsal_12992
GRB	K04364	Cflo_08580, Cflo_01423	Hsal_02341, Hsal_06942
SOS	K03099	Cflo_11707, Cflo_15117, Cflo_11706, Cflo_15357, Cflo_14630, Cflo_13527, Cflo_15116, Cflo_15190	Hsal_12097
PI3K	K00922 K02649	Cflo_06959, Cflo_03425	Hsal_03492, Hsal_08213
AKT	K04456	Cflo_11866, Cflo_08257	Hsal_06330
PIAS	K04706	Cflo_01683, Cflo_02091	1 unannotated homolog
STAT	K11224 K11221	Cflo_10787	Hsal_04889, Hsal_15821
STAM	K04705	Cflo_02172, Cflo_06578, Cflo_00318	Hsal_08616, Hsal_16595, Hsal_08640, Hsal_11412
JAK	K04447	Cflo_06362	Hsal_06174
Cb1	K04707	Cflo_03820, Cflo_13374	Hsal_02469, Hsal_02471, Hsal_00319
CBP	K04498	Cflo_01850	Hsal_04540, Hsal_04541
SOCS	K04696 K04700 K04695	Cflo_00764, Cflo_01181, Cflo_12558	Hsal_16949, Hsal_08343, Hsal_01868
CycD	K10151	Cflo_04644, Cflo_04643	Hsal_11722
BclXL	K04570	Cflo_03246	Hsal_11788
Spred	K04703	Cflo_07910	—
Sprouty	K04704	Cflo_08046	Hsal_09547
Toll pathway			
GNBP	K01199	Cflo_03103	Hsal_06973
	K01238	Cflo_13129	Hsal_10999
PGRP	K01426	Cflo_08146	Hsal_10514, Hsal_08530
	K01446	Cflo_03308, Cflo_03358, Cflo_08145	Hsal_10515, Hsal_10515
Toll-like	K06850	Cflo_07113, Cflo_08124	Hsal_04037
	K06838	—	Hsal_04045, Hsal_10101
	K08129	—	Hsal_07108
	K08764	—	Hsal_06933
	K06260	—	Hsal_08869
	K04309	Cflo_07226, Cflo_02111	—
	K06839	Cflo_08117	—
	K01292	Cflo_03285	—
	K10160	Cflo_02110	—
MyD88	K04729	Cflo_04273	Hsal_12131

Spz	—	Cflo_07408	Hsal_03005
Tube	—	Cflo_00704	Hsal_05266
Pelle	K08286	Cflo_01454	Hsal_02978
Catus	K09259	Cflo_07939, Cflo_07940	—
	K09553	—	Hsal_12521
	K10380	—	Hsal_00601
Dif/Dorsal	—	Cflo_08314	Hsal_01917
IMD pathway			
IMD	K02861	Cflo_02373	Hsal_10173
FADD	—	Cflo_10976	Hsal_09499
Dredd	—	Cflo_14662	Hsal_15238
Tab	K04404	Cflo_07135	Hsal_01491, Hsal_01490
TAK	K04443	Cflo_09489	—
	K08544	—	Hsal_08730
Sickie	K01516	Cflo_04973	Hsal_05211
Dnr-1	K10637	Cflo_01429	Hsal_04270
Ird5	K07209	Cflo_01566	Hsal_12609
Relish	K09255	Cflo_04731	Hsal_11822
Defensin	—	Cflo_01169	Hsal_06861

Supplementary table S35. *C. floridanus* and *H. saltator* genes involved in IIS signaling.

Gene	<i>Drosophila</i>	<i>C. floridanus</i>	<i>H. saltator</i>
3'-Phosphoinositide-dependent kinase-1- PDK1	dPDK1	Cflo_05676	Hsal_00114
eIF-4E binding protein (4EBP)	Thor	Cflo_00553	Hsal_06475
forkhead box, subgroup O	dFOXO	Cflo_05519	Hsal_01572
IGF_insulin_like	—	Cflo_08910	Hsal_03795
IGF_insulin_like	—	Cflo_14922	—
ILP	dILP	Cflo_08906	Hsal_04952
Insulin receptor substrate-IRS	CHICO	Cflo_02187	Hsal_05249
Insulin receptor-INR	dINR	Cflo_09206	Hsal_11112
p110 Phosphoinositide 3-kinase-P110 PI3K	Dp110	Cflo_06959	Hsal_03492
p60 PI3K-P60 PI3K	Dp60	Cflo_03425	Hsal_08213
Phosphatase and tensin homolog-PTEN	PTEN	Cflo_02813	Hsal_05284
Protein kinase B-AKT	PKB	Cflo_08257	Hsal_06330
Ras homolog enriched in brain	dRheb	Cflo_08291	Hsal_03092
S6 Kinase-S6K	S6K	Cflo_01998	Hsal_10301
similar to insulin receptor	—	—	Hsal_01375
similar to Insulin receptor precursor (IR)	—	Cflo_03027	Hsal_02738
similar to insulin-like growth factor 1 receptor isoform 1	—	Cflo_05946	Hsal_09512
similar to insulin-like growth factor 2 receptor	—	Cflo_02109	Hsal_08870
slimfast	slif	Cflo_01271	Hsal_03367
slimfast	slif	—	Hsal_03368
Target of Rapamycin-TOR	TOR	Cflo_04476	Hsal_03855
triple sex comb	dTSC1/2	Cflo_06283	Hsal_08170
Insulin/insulin-like growth factor/relaxin family	—	Cflo_20000	—

Supplementary table S36. Best homologous Swissprot entries for potential histone demethylases in *C. floridanus* and *H. saltator*.

<i>C. floridanus</i>	<i>H. saltator</i>	Human	<i>Drosophila</i>	overall
Cflo_05434	Hsal_00881	ADFP	LSD1	LSD1_DROME
Cflo_00913	Hsal_02205	HSPBAP1	NA	HBAP1_BOVIN
Cflo_12071	Hsal_11314	JAD1A	JAD1	JAD1A_HUMAN
Cflo_12870	Hsal_01820	JARD2	JAD1	JARD2_MOUSE
Cflo_02982	Hsal_05571	JHD1A	JHD1	JHD1_DROME
Cflo_05916	Hsal_09334	JHD1D	JHD1	JHD1D_MOUSE
Cflo_02692	Hsal_04177	JHD2B	NA	JHD2B_HUMAN
Cflo_04695	Hsal_05659	JHD3B	JHD3B	JHD3B_DROME
Cflo_02427	Hsal_08503	JMJD4	JMJD4	JMJD4_CHICK
Cflo_10200	Hsal_01064	JMJD5	NA	JMJD5_HUMAN
Cflo_11068	Hsal_00954	JMJD6	PTDSR	PTDSR_DROME
Cflo_10320	Hsal_14245	JMJD7	NA	JMJD7_MOUSE
Cflo_15319	Hsal_04050	LSD1	LSDA	LSD1_HUMAN
Cflo_12116	Hsal_07799	UTX	NA	UTX_HUMAN
Cflo_04528	—	UTX	NA	UTX_HUMAN

Supplementary table S37. Homology-based annotation of neuropeptide in *C. floridanus* and *H. saltator*.

Peptide	<i>A. gambiae</i>	<i>D. melanogaster</i>	<i>C. floridanus</i>	<i>H. saltator</i>
Adipokinetic hormone-AKH	—	AKH CG1171	Cflo_05554	Hsal_01494
Allatostatin-AST	agCP7920	AST CG13633	Cflo_00949	Hsal_03787
Allatostatin 2-AST2	agCP8174	Ast2 CG14919	—	—
Allatotropin 1-AT	agCP5165	—	—	—
Allatotropin 2-AT2	agCP9503	—	—	—
Bursicon-BSN	—	CG13419	Cflo_06681	Hsal_00733
capa/CAP-2b-CAPA	—	Capa CG15520	—	—
Corazonin-CRZ	agCP10813	Crz CG3302	—	CG3302(tblastn)
Crustacean cardioactive peptide-CCAP	agCP8078	Ccap CG4910	—	—
Diuretic hormone/Corticotropin releasing factor-DH	—	Dh CG8348	Cflo_09705	Hsal_11526
Diuretic hormone 31/ Calcitonin-like peptide-DH31	agCP12293	Dh31 CG13094	Cflo_02395	Hsal_07334
Ecdysis-triggering hormone-ETH	—	Eth CG18105	CG18105(tblastn)	CG18105(tblastn)
Eclosion hormone-EH	—	Eh CG5400	Cflo_01676	Hsal_09164
FMRFa peptides-FMRF	—	Fmrf CG2346	Cflo_09872	—
IFa peptide-IF	—	Ifa CG4681	Cflo_13002	CG4681(tblastn)
Ion transport peptide/ Crustacean hyperglycemic hormone-ITP	agCP15037	CG13586	Cflo_13840	Hsal_07939
		Leucokinin CG13480	—	—
Leucokinins-KIN	—			
Myoinhibitory-like peptide/ Allatostatin B-MIP	agCP9104	MIP CG6456	—	—
Myosuppressin-MS	agCP12084	Dms CG6440	Cflo_06655	Hsal_01899
Myotropin-MY	—	CG15520	—	—
Neuropeptide F-NPF	agCP4465	Npf CG10342	—	—
		NPFF NM_003717(hsa)	—	—
Neuropeptide FF-NPFF	—			
Neuropeptide-like precursor 1-NPP1	agCP10689	CG3441	—	—
Ovary ecdysteroidogenic hormone/Neuroparsin-OEH	agCP13249	—	—	—
Pyrokinins-PK	agCP14788	Hug CG6371	—	—
Pigment dispersing hormone-PDH	agCP6031	Pdf CG6496	Cflo_04480	Hsal_04702
Proctolin-PT	—	Proc-R CG6986	Cflo_07430	Hsal_04093
Prothoracicotropic hormone-PTTH	—	CG13687	—	—
Short Neuropeptide F/ LRLRFa-SNPF	—	SNPF CG13968	—	CG13968(tblastn)
Sulfakinin-SK	—	Dsk CG18090	Cflo_02222	Hsal_09481
Tachykinin-TK	agCP8706	Tk CG14734	Cflo_09451	—
Insulin-like peptide-ILP1	—	ILP1 CG14173	Cflo_08906	Hsal_04952

Insulin-like peptide-ILP2	agCP11852	IRP CG8167	—	—
Insulin-like peptide-ILP3	agCP11855	CG14167	Cflo_08906	Hsal_04952
Insulin-like peptide-ILP4	agCP11853	CG6736	—	—
Insulin-like peptide-ILP5	—	CG33273	—	—

Supplementary table S38. OBPs genes in *C. floridanus*. The protein ID used for automatic annotation is shown with a suffix indicating the species. Gene symbols for manual annotation are from *A. mellifera*.

Ant Gene ID	Automatic annotation	Manual annotation
Cflo_00046	NP_001035315.1_APIME	PBP_GOBP
Cflo_03166	NP_001011640.1_APIME	JHBP
Cflo_03710	XP_395658.3_APIME	CG3246-PA
Cflo_06181	XP_623967.2_APIME	CG5867-PA
Cflo_07918	XP_393105.2_APIME	CG5867-PA isoform 1
Cflo_07919	XP_396221.1_APIME	CG1124-PA isoform 3
Cflo_07920	XP_001121672.1_APIME	CG15497-PA
Cflo_07921	XP_396220.2_APIME	CG2016-PB
Cflo_07922	XP_001122696.1_APIME	CG14661-PA
Cflo_08198	XP_393105.2_APIME	CG5867-PA isoform 1
Cflo_08199	XP_393105.2_APIME	CG5867-PA isoform 1
Cflo_09210	NP_001011590.1_APIME	PBP_GOBP
Cflo_09290	NP_001035314.1_APIME	PhBP
Cflo_09956	NP_001035294.1_APIME	PBP_GOBP
Cflo_09957	NP_001035316.1_APIME	PBP_GOBP
Cflo_11760	XP_001122741.1_APIME	CG14661-PA
Cflo_11762	XP_001122724.1_APIME	CG14661-PA
Cflo_11763	XP_625151.1_APIME	CG14661-PA
Cflo_11956	XP_391872.1_APIME	hypothetical protein
Cflo_11957	XP_391872.1_APIME	hypothetical protein
Cflo_12533	XP_001122696.1_APIME	CG14661-PA
Cflo_12676	NA	NA

Supplementary table S39. OBP's genes in *H. saltator*. The protein ID used for automatic annotation is shown with a suffix indicating the species. Gene symbols for manual annotation are from *A. mellifera*.

Ant Gene ID	Automatic annotation	Manual annotation
Hsal_01105	XP_623967.2_APIME	CG5867-PA
Hsal_02800	XP_001122724.1_APIME	CG14661-PA
Hsal_02801	XP_001122706.1_APIME	CG1124-PA
Hsal_02802	XP_001122706.1_APIME	CG1124-PA
Hsal_02803	XP_001122696.1_APIME	CG14661-PA
Hsal_02804	XP_396220.2_APIME	CG2016-PB
Hsal_02806	XP_001121672.1_APIME	CG15497-PA
Hsal_02807	XP_396221.1_APIME	CG1124-PA isoform 3
Hsal_02808	XP_393105.2_APIME	CG5867-PA isoform 1
Hsal_03053	NP_001035294.1_APIME	PBP_GOBP
Hsal_03056	NP_001035316.1_APIME	PBP_GOBP
Hsal_03960	XP_395658.3_APIME	CG3246-PA
Hsal_10712	NP_001011588.1_APIME	PBP_GOBP
Hsal_10713	NA	NA
Hsal_11110	NP_001011590.1_APIME	PBP_GOBP
Hsal_11326	NA	NA
Hsal_11328	NA	NA
Hsal_11742	NP_001035314.1_APIME	PBP_GOBP
Hsal_11750	NP_001035315.1_APIME	PBP_GOBP
Hsal_12063	XP_001122741.1_APIME	CG14661-PA
Hsal_12064	XP_001122734.1_APIME	CG10407-PA
Hsal_12182	XP_391872.1_APIME	hypothetical protein
Hsal_12867	XP_001122741.1_APIME	CG14661-PA
Hsal_12868	XP_001122734.1_APIME	CG10407-PA
Hsal_15428	XP_396221.1_APIME	CG1124-PA isoform 3
Hsal_15973	XP_396221.1_APIME	CG1124-PA isoform 3
Hsal_15987	XP_001122696.1_APIME	CG14661-PA

Supplementary table S40. *C. floridanus* and *H. saltator* genes with homologues involved in RNAi pathways.

Gene	<i>A. mellifera</i>	<i>N. vitripennis</i>	<i>C. floridanus</i>	<i>H. saltator</i>
<i>AGO1</i>	XP_624444	XP_001601049.1	Cflo_03440	Hsal_06852
<i>AGO2</i>	XP_395048	XP_001607164.1, XP_001607156.1	Cflo_07272	Hsal_12302
<i>AGO3</i>	XP_001120996.1	XP_001603582.1	Cflo_02183	Hsal_05250
<i>Armitage</i>	XP_001121242.1	XP_001605981.1	Cflo_07747	Hsal_02941
<i>AUBERGINE</i>	NP_001159378	XP_001602384.1, XP_001607362.1, XP_001605719.1	Cflo_05817, Cflo_09745	Hsal_05340, Hsal_12917
<i>Belle</i>	XP_391829.3	XP_001605842.1	Cflo_07793	Cflo_07793
<i>Dcp-1</i>	XP_391963.3	XP_001601191.1	Cflo_11812	Hsal_07564
<i>Dcp-2</i>	XP_395977.1	XP_001607167.1	Cflo_09272	Hsal_04359
<i>DICER-1</i>	NP_001116485	XP_001605287.1	Cflo_10232	Hsal_06745
<i>Dicer-2</i>	XP_001122487	XP_001602524.1	Cflo_01234	Hsal_03382
<i>DROSHA</i>	XP_394444.2	—	Cflo_01214	Hsal_08494
<i>eri-1</i>	XP_001120094.11	XP_001603726.1	Cflo_08846	Hsal_07261
<i>Fmr1</i>	XP_394058	XP_001604888.1	Cflo_11188	Hsal_04354
<i>Gawky</i>	XP_395115.3	hmm225054	Cflo_06844	Hsal_07123
<i>loqs</i>	XP_624606	XP_001601132.1	Cflo_08406	Hsal_06701
<i>Pacman</i>	XP_393481.2	XP_001603129.1, XP_001606765.1	Cflo_08846, Cflo_10520	Hsal_07261, Hsal_00496
<i>pasha</i>	XP_397444	hmm328774	Cflo_09075	Hsal_01471
<i>PIWI</i>	XP_001120996.1	XP_001603532.1	Cflo_02183	Hsal_05250
<i>SID-1</i>	XP_395167	XP_001605484.1	Cflo_06958	Hsal_14070
<i>SID-2</i>	—	—	—	—
<i>Spindle-E</i>	XP_001122827	XP_001600067.1	Cflo_08589, Cflo_09821	Hsal_10182, Hsal_02002
<i>Tudor-SN</i>	XP_624638.2	NP_001153329.1	Cflo_03667	Hsal_04204
<i>vig</i>	XP_392925.2	XP_001602794.1	Cflo_01590	Hsal_03139

SUPPORTING REFERENCES

- S1. R. Li, Y. Li, K. Kristiansen, J. Wang, *Bioinformatics* **24**, 713 (Mar 1, 2008).
- S2. R. Li *et al.*, *Genome Res* **20**, 265 (Feb, 2010).
- S3. I. H. G. S. Consortium, *Nature* **431**, 931 (Oct 21, 2004).
- S4. R. Gil *et al.*, *Proc Natl Acad Sci U S A* **100**, 9388 (Aug 5, 2003).
- S5. J. R. Cole *et al.*, *Nucleic Acids Res* **37**, D141 (Jan, 2009).
- S6. J. A. Russell *et al.*, *Proc Natl Acad Sci U S A* **106**, 21236 (Dec 15, 2009).
- S7. N. Elango, B. G. Hunt, M. A. D. Goodisman, S. V. Yi, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 11206 (Jul 7, 2009).
- S8. J. Wang *et al.*, *Nature* **456**, 60 (Nov 6, 2008).
- S9. R. Li *et al.*, *Genome Res* **19**, 1124 (Jun, 2009).
- S10. J. Gadau, J. Heinze, B. Holldobler, M. Schmid, *Mol Ecol* **5**, 785 (Dec, 1996).
- S11. D. A. Wheeler *et al.*, *Nature* **452**, 872 (Apr 17, 2008).
- S12. E. M. Belle, G. Piganeau, M. Gardner, A. Eyre-Walker, *Gene* **355**, 58 (Aug 1, 2005).
- S13. Z. Zhao, E. Boerwinkle, *Genome Res* **12**, 1679 (Nov 1, 2002).
- S14. I. Keller, D. Bensasson, R. A. Nichols, *PLoS Genet.* **3**, e22 (Feb 2, 2007).
- S15. V. V. Kapitonov, J. Jurka, *Nat Rev Genet* **9**, 411 (May, 2008).
- S16. A. L. Price, N. C. Jones, P. A. Pevzner, *Bioinformatics* **21 Suppl 1**, i351 (Jun, 2005).
- S17. G. Benson, *Nucleic Acids Res* **27**, 573 (Jan 15, 1999).
- S18. K. Sahara, F. Marec, W. Traut, *Chromosome Res* **7**, 449 (Jan 1, 1999).
- S19. J. P. Abad *et al.*, *Mol Biol Evol* **21**, 1613 (Sep, 2004).
- S20. H. Biessmann, F. Kobeski, M. F. Walter, A. Kasravi, C. W. Roth, *Insect Mol Biol* **7**, 83 (Feb, 1998).
- S21. H. Takahashi, S. Okazaki, H. Fujiwara, *Nucleic Acids Res* **25**, 1578 (Apr 15, 1997).
- S22. H. G. S. Consortium, *Nature* **443**, 931 (Oct 26, 2006).
- S23. S. Okazaki, K. Tsuchida, H. Maekawa, H. Ishikawa, H. Fujiwara, *Mol. Cell. Biol.* **13**, 1424 (Mar 1, 1993).
- S24. A. Dahanukar, E. A. Hallem, J. R. Carlson, *Curr Opin Neurobiol* **15**, 423 (Aug, 2005).
- S25. J. Liebig, in *Insect Hydrocarbons: Biology, Biochemistry, and Chemical Ecology.*, G. J. Blomquist, A. G. Bagnères, Eds. (Cambridge University Press, Cambridge, 2010).
- S26. M. Sassoe-Pognetto, J. M. Fritschy, *Eur J Neurosci* **12**, 2205 (Jul, 2000).
- S27. J. Dunlop, X. Morin, M. Corominas, F. Serras, G. Tear, *Curr Biol* **14**, 2039 (Nov 23, 2004).
- S28. C. Trapnell, L. Pachter, S. L. Salzberg, *Bioinformatics* **25**, 1105 (May 1, 2009).
- S29. Q. Pan, O. Shai, L. J. Lee, B. J. Frey, B. J. Blencowe, *Nat Genet* **40**, 1413 (Dec, 2008).
- S30. V. Stolc *et al.*, *Science* **306**, 655 (Oct 22, 2004).
- S31. A. Pires-daSilva, R. J. Sommer, *Nat Rev Genet* **4**, 39 (Jan, 2003).
- S32. G. E. Heimpel, J. G. de Boer, *Annu Rev Entomol* **53**, 209 (2008).
- S33. M. Beye, M. Hasselmann, M. K. Fondrk, R. E. Page, S. W. Omholt, *Cell* **114**, 419 (Aug 22, 2003).
- S34. M. Hasselmann *et al.*, *Nature* **454**, 519 (Jul 24, 2008).
- S35. E. C. Verhulst, L. W. Beukeboom, L. van de Zande, *Science* **328**, 620 (Apr 30, 2010).
- S36. S. Albert, D. Bhattacharya, J. Klaudiny, J. Schmitzova, J. Simuth, *J Mol Evol* **49**, 290 (Aug, 1999).
- S37. M. D. Drapeau, S. Albert, R. Kucharski, C. Prusko, R. Maleszka, *Genome Res* **16**, 1385 (Nov, 2006).
- S38. S. O. Andersen, P. Hojrup, P. Roepstorff, *Insect Biochem Mol Biol* **25**, 153 (Feb, 1995).
- S39. J. E. Rebers, J. H. Willis, *Insect Biochem Mol Biol* **31**, 1083 (Oct, 2001).
- S40. B. Hölldobler, E. O. Wilson, *The Ants.* (Harvard University Press, Cambridge, MA, 1990).
- S41. E. S. Kelleher, W. J. Swanson, T. A. Markow, *PLoS Genet* **3**, e148 (Aug, 2007).
- S42. A. Hefetz, M. S. Blum, *Science* **201**, 454 (Aug 4, 1978).
- S43. N. D. Rawlings, A. J. Barrett, *Methods Enzymol* **248**, 183 (1995).
- S44. H. Neurath, *Science* **224**, 350 (Apr 27, 1984).
- S45. J. H. Law, P. E. Dunn, K. J. Kramer, *Adv Enzymol Relat Areas Mol Biol* **45**, 389 (1977).
- S46. J. D. Evans *et al.*, *Insect Mol Biol* **15**, 645 (Oct, 2006).

- S47. D. A. Veal, J. E. Trimble, A. J. Beattie, *J Appl Bacteriol* **72**, 188 (Mar, 1992).
- S48. Y. Oba, M. Ojika, S. Inouye, *Gene* **329**, 137 (Mar 31, 2004).
- S49. M. Tatar *et al.*, *Science* **292**, 107 (Apr 6, 2001).
- S50. S. Ogg *et al.*, *Nature* **389**, 994 (Oct 30, 1997).
- S51. D. E. Wheeler, N. Buck, J. D. Evans, *Insect Mol Biol* **15**, 597 (Oct, 2006).
- S52. S. D. Taverna, H. Li, A. J. Ruthenburg, C. D. Allis, D. J. Patel, *Nat. Struct. Mol. Biol.* **14**, 1025 (Nov 1, 2007).
- S53. R. Bonasio, E. Lecona, D. Reinberg, *Semin Cell Dev Biol* **21**, 221 (Apr, 2010).
- S54. C. Birchmeier, S. Sharma, M. Wigler, *Proc Natl Acad Sci U S A* **84**, 9270 (Dec, 1987).
- S55. D. R. Nassel, *Prog Neurobiol* **68**, 1 (Sep, 2002).
- S56. M. Corona *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7128 (Apr 24, 2007).
- S57. E. A. Hallem, A. Dahanukar, J. R. Carlson, *Annu Rev Entomol* **51**, 113 (2006).
- S58. W. M. Winston, C. Molodowitch, C. P. Hunter, *Science* **295**, 2456 (Mar 29, 2002).
- S59. X. Belles, *Annu Rev Entomol* **55**, 111 (Jan, 2010).
- S60. C. Lucas, M. B. Sokolowski, *Proc Natl Acad Sci U S A* **106**, 6351 (Apr 14, 2009).
- S61. R. Li *et al.*, *Nature* **463**, 311 (Jan 21, 2010).
- S62. A. Smit, R. Hubley, P. Green. (1996-2004).
- S63. J. Jurka *et al.*, *Cytogenet Genome Res* **110**, 462 (2005).
- S64. T. Wicker *et al.*, *Nat Rev Genet* **8**, 973 (Dec, 2007).
- S65. J. A. Bailey *et al.*, *Science* **297**, 1003 (Aug 9, 2002).
- S66. E. S. Lander *et al.*, *Nature* **409**, 860 (Feb 15, 2001).
- S67. M. D. Adams *et al.*, *Science* **287**, 2185 (Mar 24, 2000).
- S68. E. Birney, M. Clamp, R. Durbin, *Genome Res* **14**, 988 (May, 2004).
- S69. S. Richards *et al.*, *Nature* **452**, 949 (Apr 24, 2008).
- S70. J. H. Werren *et al.*, *Science* **327**, 343 (Jan 15, 2010).
- S71. Q. Xia *et al.*, *Science* **306**, 1937 (Dec 10, 2004).
- S72. T. I. A. G. Consortium, *PLoS Biol* **8**, e1000313 (2010).
- S73. M. Stanke, S. Waack, *Bioinformatics* **19 Suppl 2**, ii215 (Oct, 2003).
- S74. I. Korf, *BMC Bioinformatics* **5**, 59 (May 14, 2004).
- S75. C. G. Elsik *et al.*, *Genome Biol* **8**, R13 (Jan 1, 2007).
- S76. E. M. Zdobnov, R. Apweiler, *Bioinformatics* **17**, 847 (Sep, 2001).
- S77. I. L. Hofacker, P. F. Stadler, *Bioinformatics* **22**, 1172 (May 15, 2006).
- S78. P. Jiang *et al.*, *Nucleic Acids Res* **35**, W339 (Jul, 2007).
- S79. X. Wang *et al.*, *Bioinformatics* **21**, 3610 (Sep 15, 2005).
- S80. H. Li *et al.*, *Nucleic Acids Res.* **34**, D572 (Jan 1, 2006).
- S81. R. C. Edgar, *Nucleic Acids Res* **32**, 1792 (2004).
- S82. D. Posada, K. A. Crandall, *Bioinformatics* **14**, 817 (1998).
- S83. J. P. Huelsenbeck, F. Ronquist, *Bioinformatics* **17**, 754 (Aug, 2001).
- S84. T. De Bie, N. Cristianini, J. P. Demuth, M. W. Hahn, *Bioinformatics* **22**, 1269 (May 15, 2006).
- S85. A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer, B. Wold, *Nat Methods* **5**, 621 (Jul 1, 2008).
- S86. B. R. Zeeberg *et al.*, *Genome Biol* **4**, R28 (2003).
- S87. M. Abramoff, P. Magelhaes, S. Ram, *Biophotonics International* **11**, 36 (2004).
- S88. J. Wilhelm, A. Pingoud, M. Hahn, *Nucleic Acids Res.* **31**, e56 (May 15, 2003).
- S89. D. Takai, P. A. Jones, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 3740 (Mar 19, 2002).