

Supplementary Information

Figure S1. **Sequencing depth distribution across the assembled *A. echinator* genome.** Sequencing depth was obtained by mapping the sequence reads back to the genome assembly using SOAPaligner. The average sequencing depth is $123\times$, and the peak coverage $107\times$. More than 96% of the assembly has $20\times$ or more coverage.

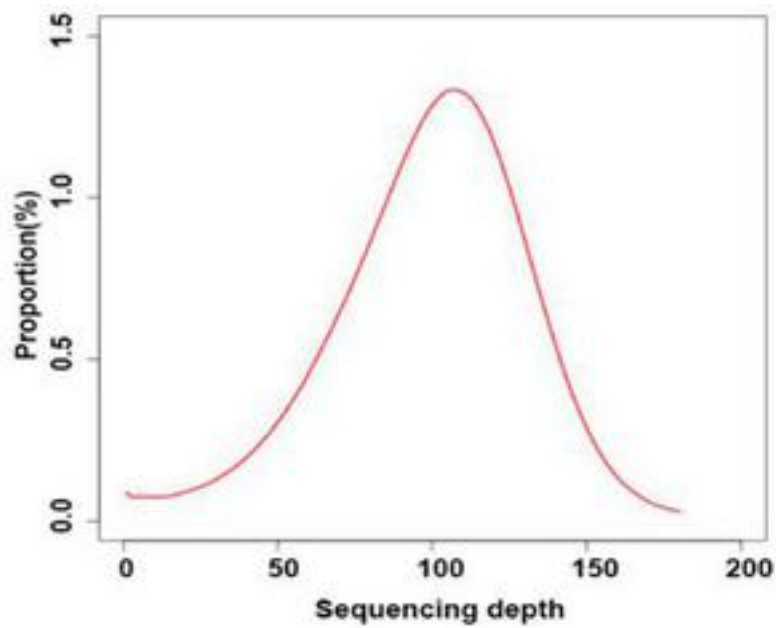


Figure S2. **Genomic phylogenetic tree of Hymenoptera.** Phylogenetic tree of six Hymenoptera with sequenced genomes (*Drosophila* included as an outgroup) constructed with PhyML from an alignment of four-fold degenerate sites from 1995 gene families with only one member from each species: *D. melanogaster* (D. mel), *A. echinator* (A. ech), *S. invicta* (S. inv), *C. floridanus* (C. flo), *H. saltator* (H. sal), *A. mellifera* (A. mel) and *N. vitripennis* (N. vit).

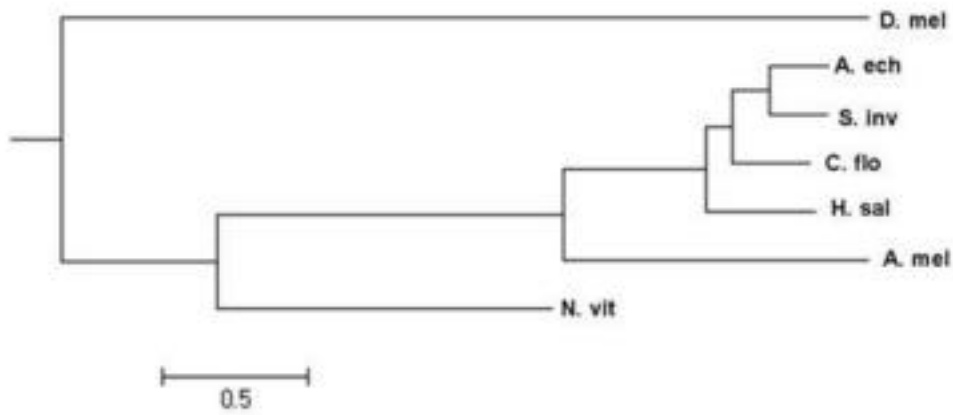


Figure S3. **Flow diagram of gene family reconstruction methods.** An overview of the methods used for constructing gene families (the Treefam pipeline: Li et al. 2006), for the generation of the phylogenetic tree for seven insect species (Figure S2), and for modeling the contracted and expanded gene families (CAFE, De Bie et al. 2006).

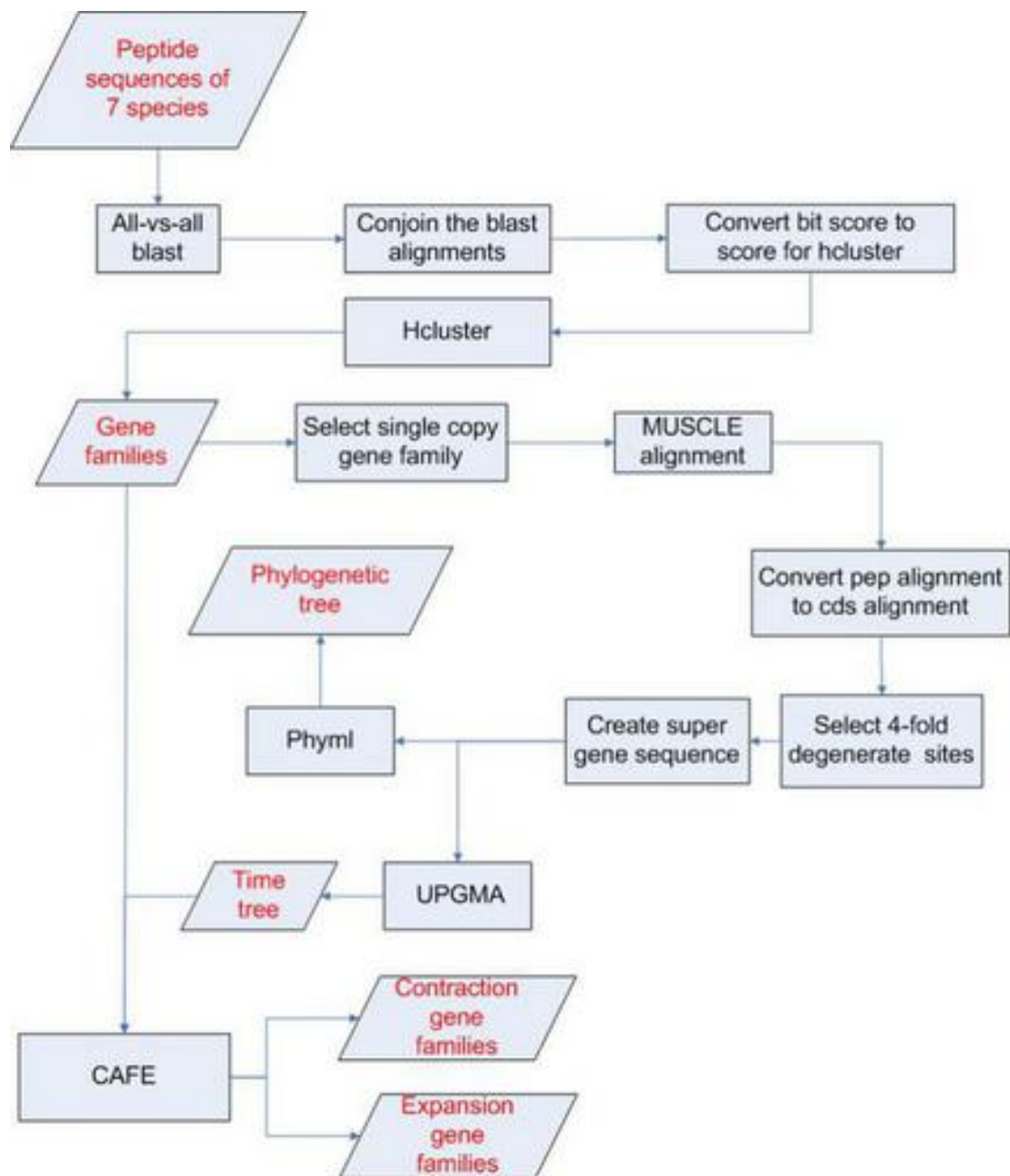
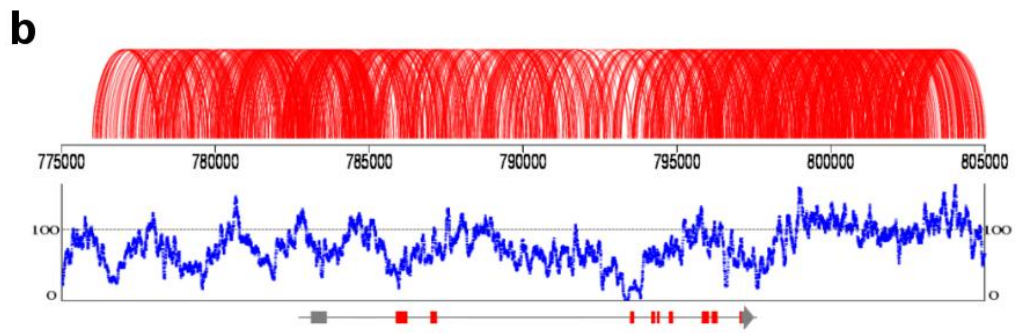
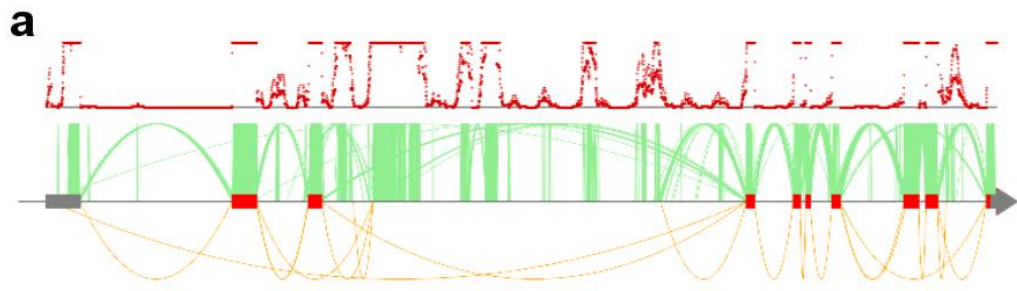


Figure S4. **Gene structure, expression, and splicing of the *feminizer* gene homolog in *A. echinatio*.** (A) Transcriptome analysis for the single *feminizer* homolog (EGI61367), with the red top profile showing the expression level of the gene, the green lines depicting connections between paired-end transcriptome reads, the red boxes indicating the protein coding exons of the gene model, and the gray boxes denoting UTRs. The yellow lines connect the junction sites of the exons as inferred from the paired-end transcriptome. The existence of expressed regions and splice junctions besides those included in the gene model most likely represent alternative exons, and strongly suggests that the gene is alternatively spliced. (B) Alignment of paired-end genomic reads across the *feminizer*-homologous locus, with red curves joining the two ends of all the mapped paired-end reads. The blue line shows the sequencing depth across the region and the dashed grey line represents the genomic average. The grey arrow at the bottom denotes the direction and structure of the *feminizer*-homologous locus, with grey blocks being UTRs, and red blocks protein coding exons. If two adjacent and highly similar loci had been erroneously merged, we might expect ambiguous mate pairs and higher than average coverage for this locus. However, the region appears to be well assembled, with a coverage somewhat below, rather than above average.



Supplementary Tables

Table S1. **DNA sequence data generated for the assembly of the *A. echinator* genome.** Columns show the length of the inserts in each insert library, the length of the reads, the amount of raw and filtered data for each library, and the fold coverage based on an estimated total genome size of 313 Mb.

Data	Average Read Length	Raw data (Mb)	Filtered data (Mb)	× coverage
0.5 Kbp insert	100	10680	9532	30.45
0.8 Kbp insert	100	13152	9687	30.95
2 Kbp insert	90	12400	7245	23.15
5 Kbp insert	90	12738	8116	25.93
10 Kbp insert	90	11708	3900	12.46
Total		60678	38480	122.94

Table S2. **Assembly statistics for the genome of *A. echinator*.** N50 - N90 statistics denote that 50 - 90 % of the entire assembly is contained in contigs or scaffolds that are equal to or longer than this value. Also shown are the numbers of contigs or scaffolds at these thresholds.

Assembly stats	Contig Size (bp)	Contig Number	Scaffold Size (bp)	Scaffold Number
N90	12808	5127	169172	339
N80	24715	3528	337092	217
N70	36487	2566	564056	149
N60	48373	1871	853565	106
N50	62705	1340	1094267	75
Longest	464283		5247136	
Total Size	292083337		299573819	

Table S3. Comparison of key genome assembly statistics for four ant genomes. The *A. echinator* numbers are from this study, the *C. floridanus* and *H. saltator* numbers from Bonasio et al. 2010, and the *S. invicta* numbers from Wurm et al. 2011. No official estimate of the repeat content in the *S. invicta* assembly is available, but it is believed to be substantially higher than in other ant genomes (Wurm et al. 2011).

Species	Assembly size (Mb)	Estimated genome size (Mb)	N50 Scaffold length (Mb)	GC content (%)	repeat content (%)
<i>A. echinator</i>	300	313	1.1	34	28
<i>S. invicta</i>	353	484	0.7	36	-
<i>C. floridanus</i>	238	240	0.6	34	15
<i>H. saltator</i>	297	330	0.6	45	27

Table S4. Overview of repeat predictions. DNA elements, Long Interspersed Nuclear Elements (LINE), Short Interspersed Nuclear Elements (SINE), Long Terminal Repeat elements (LTR), Satellite, Minisatellite, Microsatellite, Repeats classified in other classes (Other), and Repeats that could not be classified by RepeatMasker (Unknown).

Type	Length (bp)	Percent of genome
DNA	35661774	11.99
LINE	7848306	2.64
SINE	286228	0.10
LTR	29702493	9.98
Satellite	859484	0.29
Minisatellite	4018582	1.35
Microsatellite	1809277	0.61
Other	3348	0.00
Unknown	23396158	7.86
Total	82516295	27.74

Table S5. **Non-coding RNAs.** The types and number of non-coding RNA genes identified in the *A. echinator* genome for the different categories of RNA.

Type	Copy	Average Length (bp)	Total Length (bp)
miRNA	93	81.5	7580
tRNA	316	74.8	23637
18S rRNA	27	154.2	4164
28s rRNA	4	99.5	398
5.8S rRNA	1	45	45
5S rRNA	26	74.4	1935
Total rRNA	58	112.8	6542
CD-box snRNA	9	157.9	1421
HACA-box snRNA	1	157	157
splicing snRNA	19	152.0	2887
Total snRNA	29	154.0	4465

Table S6. **Bacterial sequences in the *A. echinator* assembly (pre-filtering).** The number and total length of contigs/scaffolds of bacterial origin in the *A. echinator* assembly, and the number of genes on these contigs/scaffolds. "Duplicate" refers to the total number of genes that could be aligned with each other at >90% identity for over >80% of the sequence. Almost all of the scaffolds/contigs of bacterial origin stemmed from intracellular *Wolbachia* bacteria, and the existence of duplicate copies of many of these genes that are normally single-copy, suggest that the sequenced *A. echinator* individuals harbored two strains of *Wolbachia*, consistent with earlier findings (Van Borm et al. 2003). We did not find any sequences belonging to the *Actinobacteria* genus, which include the *Pseudonocardia* ectosymbionts of *A. echinator* (Currie et al. 1999, Cafaro and Currie 2005, full references below). However, as males would normally carry negligible amounts of these bacteria, this is not surprising.

Genus	#contigs/scaffolds	Total length (bp)	#genes	#duplicate genes
<i>Wolbachia</i>	751	2,083,157	1785	423
<i>Bacillus</i>	4	604	0	0
Total	755	2,083,661	1785	423

Currie CR, Scott JA, Summerbell RC, Malloch D. 1999. Fungus-growing ants use antibiotic-producing bacteria to control garden parasites. *Nature* **398**(6729): 701-704.

Cafaro MJ, Currie CR. 2005. Phylogenetic analysis of mutualistic filamentous bacteria associated with fungus-growing ants. *Can J Microbiol* **51**(6): 441-446.

Table S7. **Gene ontology (GO) enrichment analyses.** GO enrichment in *A. echinator* specific and Myrmicine specific gene families, and in gene families where *A. echinator* homologs appear to be lost. GO classes: Molecular function (MF), Biological Process (BP), Cellular Component (CC).

GO_ID	GO_Term	GO Class	P	Adjusted P
<i>A. echinator</i> specific (all genes from <i>A. echinator</i> -specific gene families + singletons)				
GO:0046914	transition metal ion binding	MF	0.0002	0.0304
Myrmicine specific (gene families present in <i>A. echinator</i> and <i>S. invicta</i> , but absent in <i>C. floridanus</i> and <i>H. saltator</i>)				
GO:0008270	zinc ion binding	MF	1.0×10^{-9}	2.0×10^{-7}
GO:0046914	transition metal ion binding	MF	0.000001	0.00005
GO:0046872	metal ion binding	MF	0.00003	0.0009
GO:0005175	CD27 receptor binding	MF	0.0001	0.0021
GO:0016765	transferase activity, transferring alkyl or aryl (other than methyl) groups	MF	0.0003	0.004
GO:0005488	binding	MF	0.0006	0.0071
GO:0003676	nucleic acid binding	MF	0.0014	0.0140
GO:0043065	positive regulation of apoptosis	BP	0.0003	0.0040
GO:0006260	DNA replication	BP	0.0040	0.0338
GO:0000775	chromosome, centromeric region	CC	9.0×10^{-7}	0.00004
GO:0005694	chromosome	CC	0.0014	0.0140
GO:0005622	intracellular	CC	0.0041	0.0338
Lost in <i>A. echinator</i> (gene families absent in <i>A. echinator</i> , but present in <i>S. invicta</i> , and present in either <i>C. floridanus</i> or <i>H. saltator</i> , or both)				
GO:0008270	zinc ion binding	MF	0.000002	0.0001
GO:0005198	structural molecule activity	MF	0.00001	0.0005
GO:0003735	structural constituent of ribosome	MF	0.00002	0.0006
GO:0046914	transition metal ion binding	MF	0.00003	0.0007
GO:0003676	nucleic acid binding	MF	0.00005	0.0009
GO:0004198	calcium-dependent cysteine-type endopeptidase activity	MF	0.0043	0.0396
GO:0015074	DNA integration	BP	0.00002	0.0006
GO:0006414	translational elongation	BP	0.00004	0.0008
GO:0006310	DNA recombination	BP	0.00004	0.0008
GO:0006412	translation	BP	0.0014	0.0170
GO:0000245	spliceosome assembly	BP	0.0018	0.0195
GO:0016012	sarcoglycan complex	CC	2.0×10^{-7}	0.00001
GO:0005840	ribosome	CC	0.00002	0.0006

Table S8. **Gene family expansions and contractions in *A. echinator***. Gene families that are significantly ($P < 0.001$) expanded or contracted in *A. echinator* compared to three other ants and the honeybee, as determined by the CAFE method (De Bie et al. 2006). The number of genes in each species is shown, as well as the Interpro (IPR) domains assigned to genes in the family. Redundant IPR domain matches implying similar functions have been removed. Gene families discussed in the main text are highlighted in yellow. *A. echinator* (A.ech), *S. invicta* (S.inv), *C. floridanus* (C.flo), *H. saltator* (H.sal), *A. mellifera* (A.mel).

famid	A.ech	S.inv	C.flo	H.sal	A.mel	expansion/ contraction	IPR
188	2	10	6	2	1	contraction	IPR013120; Male sterility, NAD-binding.
479	11	20	22	10	5	contraction	IPR001128; Cytochrome P450.
557	3	15	31	22	6	contraction	IPR004117; Olfactory receptor, Drosophila.
558	4	16	12	14	5	contraction	IPR004117; Olfactory receptor, Drosophila.
561	9	27	29	6	1	contraction	IPR004117; Olfactory receptor, Drosophila. IPR001254; Peptidase S1/S6, chymotrypsin/Hap.
569	3	8	11	10	17	contraction	IPR004117; Olfactory receptor, Drosophila. IPR002132; Ribosomal protein L5.
804	1	11	8	9	1	contraction	IPR008906; HAT dimerisation.
1747	9	22	21	19	3	contraction	IPR007087; Zinc finger, C2H2-type.
2168	8	12	11	11	7	contraction	IPR001356; Homeobox.
2379	9	15	14	16	11	contraction	IPR002213; UDP-glucuronosyl/UDP-glucosyltransferase.
554	11	2	2	1	1	expansion	IPR004117; Olfactory receptor, Drosophila.
2579	8	4	4	6	4	expansion	IPR010800; Glycine rich.
2662	9	5	4	5	3	expansion	IPR000834; Peptidase M14, carboxypeptidase A. IPR003341; Cysteine rich repeat, tripleX.
3063	7	4	3	3	3	expansion	IPR001440; Tetratricopeptide TPR-1.
3970	4	2	3	1	1	expansion	IPR001320; Ionotropic glutamate receptor.
4184	4	1	1	1	1	expansion	IPR000266; Ribosomal protein S17.
5049	16	2	3	2	3	expansion	IPR007863; Peptidase M16, C-terminal. IPR011765; Peptidase M16, N-terminal.
5486	3	1	1	2	1	expansion	IPR000589; Ribosomal protein S15.
8431	6	3	3	1	1	expansion	IPR004299; Membrane bound O-acyl transferase, MBOAT.
10761	5	1	1	6	1	expansion	IPR010347; Tyrosyl-DNA phosphodiesterase. IPR019406; Zinc finger, C2H2, APLF-like.

Table S9. **Gene family expansions and contractions in the ant subfamily Myrmicinae.** Gene families that are significantly expanded or contracted in the two representatives of the ant subfamily Myrmicinae (*A. echinator* and *S. invicta*), compared to two other ants and the honeybee, as determined by the CAFE method (De Bie et al. 2006). The number of genes in each species is shown as well as IPR domains assigned to genes in the family. Redundant IPR domain matches implying similar functions have been removed. *A. echinator* (A.ech), *S. invicta* (S.inv), *C. floridanus* (C.flo), *H. saltator* (H.sal), *A. mellifera* (A.mel).

famid	A.ech	S.inv	C.flo	H.sal	A.mel	expansion/ contraction	IPR
553	5	3	26	14	8	contraction	IPR004117; Olfactory receptor, Drosophila.
557	3	15	31	22	6	contraction	IPR004117; Olfactory receptor, Drosophila.
8200	1	2	8	7	1	contraction	IPR001440; Tetratricopeptide TPR-1
2119	5	8	3	3	3	expansion	IPR013525; ABC-2 type transporter.
3501	3	5	1	1	1	expansion	IPR000092; Polyprenyl synthetase.
3531	5	8	1	1	1	expansion	IPR004119; Protein of unknown function DUF227. IPR015897; CHK kinase-like.
4745	8	12	1	1	2	expansion	IPR013604; 7TM chemoreceptor.

Table S10. **Protein domains associated with detoxification pathways.** The number of genes encoding each domain was predicted with InterproScan as described in the Methods section. *A. echinator* (A.ech), *S. invicta* (S.inv), *C. floridanus* (C.flo), *H. saltator* (H.sal), *A. mellifera* (A.mel), *N. vitripennis* (N.vit), *D. melanogaster* (D.mel).

Domain	IPR id	A.ech	S.inv	C.flo	H.sal	A.mel	N.vit	D.mel
Cytochrome P450	IPR001128	73	124	132	95	60	96	87
Glutathione S-transferases								
Glutathione S-transferase, N-terminal	IPR004045	19	10	14	14	12	20	39
Glutathione S-transferase, C-terminal	IPR004046	17	13	14	16	10	21	39
Glutathione S-transferase/ chloride channel, C-terminal	IPR017933	20	13	15	16	11	24	41
Carboxy/cholinesterases								
Carboxylesterase, type B	IPR002018	28	39	43	28	29	48	35
Alpha/beta hydrolase fold-1	IPR000073	27	30	31	23	19	43	41
UDP-glucosyltransferase	IPR002213	13	22	15	25	12	22	35

Table S11: Alignment of argininosuccinate synthase from several hymenopteran species.

A. mellifera (A.mel), *N. vitripennis* (N. vit), *H. saltator* (H.sal), *C. floridanus* (C.flo), *S. invicta* (S.inv), *A. echinator* (A.ech). Stop-codons in the latter sequence are marked with #. This sequence is found on contig7953 (AEVX01007953.1: 78476-80084 +)

```

CLUSTAL 2.0.12 multiple sequence alignment

XP_396374.2_A.mel      -----MSETEKKVVLAYSGGLDTSCILLWLKEKGYQVIAY 35
XP_001603273.1_N.vit  MPVNSTLVRQFVKSTMEERKKVILAYSGGLDTSCILLWLEKGYDVIAY 50
EFN88795.1_H.sal      -----MSDSKKKVILAYSGGLDTSCILFWLREKGFQVVAY 35
EFN69689.1_C.flo      -----MADLKKKVILAYSGGLDTSCILFWLKKKGFQVIAY 35
EFZ22280.1_S.inv      -----MSDSKRKVILAYSGGLDTSCILLWLKEKGYQVIAY 35
Pseudo_gene_A.ech     -----

XP_396374.2_A.mel      MANVGQEEDFDAARKKALKIGATKVVIEDLREVFLSSFVWPAIACGLIYE 85
XP_001603273.1_N.vit  VANIGQDEDFEAAREKALSIGAKKVIIEDLREAFVTSYVWPAVRSGLLYE 100
EFN88795.1_H.sal      VANIGQNELDVAVRKKALKIGAVKVVIDDVKEAFVTSYVWPAVACGLLYE 85
EFN69689.1_C.flo      VANIGQDEDFDAVREKALKIGAIVKVVIDDLKEVFTSYVWPAVACGLLYE 85
EFZ22280.1_S.inv      VANIGQKEDFDAIRDKALKVAVKVVIDDVREVFVTSYVWPAVACGLLYE 85
Pseudo_gene_A.ech     -----LLYK 4
                                                                *:*:

XP_396374.2_A.mel      DRYLLGTAIARPCISEGLVRLAKGENAHIIAHGATGKGNDQVRFELNCYS 135
XP_001603273.1_N.vit  GRYLLGTSLARPCISEGLIRAARSENAMTISHGATGKGNDQVRFELSCYT 150
EFN88795.1_H.sal      GRYLLGTSIARPCISQGLIRIARAEGAGIIAHGATGKGNDQVRFELSCYS 135
EFN69689.1_C.flo      GRYLLGTSVARPCISQGLVKLAKAEGANIIAHGATGKGNDQVRFELSCYS 135
EFZ22280.1_S.inv      GRYLLGTAIARPCISQGLIKVAKAEGASVIAHGATGKGNDQVRFELNCYS 135
Pseudo_gene_A.ech     GRYLL-AYLTSPCISQSLIKAIAKADSSRMEQ--LEKWNDQIRFELNCYN 51
                                                                .**** : :: ****:.*:: :.: : : : * ****:****.**.

XP_396374.2_A.mel      LCPDIQILAPWREEEFYTRFSGRTDLLKYAQQNGIPVSVTPKEPWSTDAN 185
XP_001603273.1_N.vit  LLPAVQILAPWREPEFFNRFQGRPDLLLEYAKKNKIPVSATPKPEWSTDAN 200
EFN88795.1_H.sal      LCPDIKILAPWREHEFYTRFQGRPDLLLEYARQNGIPTSATPKPEWSTDAN 185
EFN69689.1_C.flo      LYPDITVLAPWREREFYKFKGRSDLLEYARQNDIPISATPKPEWSTDAN 185
EFZ22280.1_S.inv      LYPEITILAPWREVEFYTKFQGRSGLLEYARQNGIPIISATPKPEWSTDAN 185
Pseudo_gene_A.ech     LYFKITILAPWRKIEFYKFKGKSDLLEYA#QNGILIFATAKELWNTDK- 100
                                                                * : :*****: **: :*.*:..**:* * * .*.** **.*

```

XP_396374.2_A.mel	LLHVSYESGILENPNALVPKEIYKMTDRDPLCSPTPEPEEIEIKFERGYPS	235
XP_001603273.1_N.vit	LLHISYESGVLENPANQAPSDLCKMTNPNQDAPDEPLDVEIYFGKGYPTF	250
EFN88795.1_H.sal	LVHISYESGVLENPANPAPRGLYKMTTDPMSAAAAEAEIEIDFKHGYP	235
EFN69689.1_C.flo	LVHISYESGVLENPATPAPKELYKMTVDPINSCLDAEEIEISFAKGYPI	235
EFZ22280.1_S.inv	LVHISYESGILENPAIPAPKGLYNMTTDPDTSALAEAEIEINFKKGYP	235
Pseudo_gene_A.ech	LVHISYESRILEDPAIPAPKGNKMIIDPIYSTLE-EEIEISFKKGYP	149
	*:**** *:* . * : * : : : : * * : * *	
XP_396374.2_A.mel	VERLKNNETFLTPLEIMQCLNEVGNLHGIGRIDIVENR-----	273
XP_001603273.1_N.vit	VENLHTHQVYKSLPAIIEFLNNVGGEGHIGRIDIVENR-----	288
EFN88795.1_H.sal	VRNVKSGRTLDAPLAIIEYLNEVGGAHGIGRVDIVENR-----	273
EFN69689.1_C.flo	VKSLKDNKIYDTPLAIEYLNKIGGSHGIGRIDIVENR-----	273
EFZ22280.1_S.inv	VKNLKNKIFNAPLAIIEYLNEVGGAHGIGRIDIVENR-----	273
Pseudo_gene_A.ech	VKSLKDR--KIPLAIEYLNIEIG-AH--GRIDIVENRYRLKLLMKLFFM	193
	*. : : * * * : : * * * : * * * : * * * * * * * * *	
XP_396374.2_A.mel	-----FIGLKSRGIYESPGATILYKAHQDLEIFIL	303
XP_001603273.1_N.vit	-----FIGLKSRGIYESPGAKIVHEAHQDLEVYLL	318
EFN88795.1_H.sal	-----YIGLKSRGVYESPGAKILHVAHQDLEVFAL	303
EFN69689.1_C.flo	-----YIGLKSRGIYESPGAKILHVAHQDLEVFIL	303
EFZ22280.1_S.inv	-----YIGLKSRGIYESPGVNILEVAHRDLEVFTL	303
Pseudo_gene_A.ech	KFFYIFYKIFMKCIKLTITIFVQFQSRGIYESSDVNILEVVH#DLKVFTF	243
	: : : : * * * : * * * . . . * : . * * * : : :	
XP_396374.2_A.mel	DREVLRIKSYLTSKMSDYVYNGFWFSPDFVFRNSILYSQKYVNGTVRLK	353
XP_001603273.1_N.vit	DKEVLRRLKSYLADKMSDYVYNGLWFSPECDFARESIALTQKYVTGSVIVR	368
EFN88795.1_H.sal	DREVLRVKSYLTDKMSDYVYNGFWFSPFCNFVRESILLSQRYVNGTVRLQ	353
EFN69689.1_C.flo	DREILRIKSYLTKMSDYVYNGFWFSPCECFVRESIQHSQKYVNGVTRLQ	353
EFZ22280.1_S.inv	DREVLRVKTYLTDKMSDYVYNGFWFSPFCDFVRECIQHSQRHVNGTVRVH	353
Pseudo_gene_A.ech	DREILRVK--LTDKIYQITFNGLFN--FWNTILECIQHSQRYVNDTVRLQ	289
	*:***** * : . : : : . : * * : : * . * : :	
XP_396374.2_A.mel	LYKGNVSVLSRHSKTSLYNQSLVSMQVQDFEPEDVGGFIRTAQYRLKEF	403
XP_001603273.1_N.vit	LYKGSVSVIVARKSDVSLYNEGLVSMQVQDFEPEDAGGFIRIQAALRLKEY	418
EFN88795.1_H.sal	LYRGNVMAIGRSSEVSLYNENLVSMQVQDFEPADATGFIRTAQALRLKEF	403
EFN69689.1_C.flo	LYKGNVMAIGRSSEVSLYNENLVSMQVQDFEPADATGFIRIQAALRLREF	403
EFZ22280.1_S.inv	LYKGMAMAIARSEISLYNQTLVSMQVQDFEPTDATGFIRMQAALRLKEF	403

Pseudo_gene_A.ech	LYK GKAMAIARSSEVSLYNQILISMNVQGDYR--NATEFIRTQALKLKEF 337
	:* . . :.* *. *** :*:** :*:*. :. *** ** :*:**:
XP_396374.2_A.mel	NRFK--EQCDI 412
XP_001603273.1_N.vit	QRFK--EQHKK 427
EFN88795.1_H.sal	HRFS--KQRDV 412
EFN69689.1_C.flo	HRFNSNKQHNV 414
EFZ22280.1_S.inv	HRFN--KQHN- 411
Pseudo_gene_A.ech	YRFN--KQHNV 346
	** . :* .

Table S12. **Orthologs of *A. mellifera* seminal fluid proteins in *A. echinator*.** The seminal fluid proteins of *A. mellifera* described in Baer et al. 2009b and their orthologous genes in *A. echinator*. Protein IDs from BeeBase (<http://hymenoptera.genome.org/beebase/>) are given for *A. mellifera*, GenBank IDs for *A. echinator*.

<i>A. mellifera</i>	<i>A. echinator</i>	Protein description
GB10139	EGI65384	Glutamine synthetase 2
GB10192	EGI62089	Circadian clock protein
GB10275	EGI66955	Tubulin
GB10441	EGI70839	Aminopeptidase
GB10467	EGI65745	Aspartate aminotransferase 2
GB10584	EGI59502	Glucocerebrosidase
GB10973	EGI58962	Arginine kinase
GB11056	EGI64907	Phosphoglycerate kinase
GB11665	EGI57614	Chitinase
GB11698	EGI65218	CLIP-domain serine protease
GB11983	EGI64832	Carboxypeptidase
GB11987	EGI62847	E3 ubiquitin-protein factor
GB12546	EGI66977	Acid phosphatase
GB12573	EGI69480	Citrate synthase
GB12716	EGI60214	Unknown function
GB13065	EGI63752	Glucose dehydrogenase
GB13380	EGI68921	Profilin
GB13401	EGI60280	Aldehyde dehydrogenase
GB13438	EGI61971	Dynein
GB13626	EGI69344	α -Mannosidase
GB13687	EGI66473	Apolipoprotein
GB14019	EGI59025	Spermine oxidase
GB14634	EGI58234	β -Hexosaminidase
GB14710	EGI60008	Serine carboxy Peptidase
GB14758	EGI61913	Heat shock protein 90
GB14852	EGI70201	Heat shock protein 70
GB15044	EGI68836	Clusterin
GB15543	EGI58485	Lactate/sulfolactate/malate/dehydrogenase
GB15662	EGI60204	Dehydrogenase/reductase
GB16429	EGI67470	Glucose-6-phosphate isomerase
GB16448	EGI70234;EGI66394	Annexin IX
GB16492	EGI62461	Tubulin
GB16552	EGI58677	Lysosomal thiol reductase
GB16568	No hit	Cytochrome c oxidase VB subunit
GB16589	EGI66968	L-2-Hydroxyglutarate dehydrogenase
GB16619	EGI63203	Metallo-endopeptidase (Neprilysin 2)
GB16738	No hit	3-Hydroxyacyl-CoA dehydrogenase
GB17218	EGI62820	5'-Nucleotidase/apyrase

GB17311	EGI64427	ATPase accessory protein
GB17366	EGI67856	Glucose transporter
GB17473	EGI64718	Triosephosphate isomerase
GB17495	No hit	Malate dehydrogenase
GB17542	EGI60275	Succinate dehydrogenase B
GB17681	EGI67274	Actin-5C
GB17780	EGI65623	Cysteine rich secretory allergen
GB17852	EGI69843	Lysosomal β -galactosidase
GB17864	EGI65101	4-Hydroxybutyrate CoA transferase
GB17900	EGI64546;EGI64547	Fucosidase, α -L-2
GB18819	EGI70267	Antennal-specific protein 3c
GB18862	EGI66271	Multiple inositol polyphosphate phosphatase
GB18955	EGI67689	Phospholipid hydroperoxide glutathione peroxidase
GB19075	EGI63757;EGI58029	Secreted unknown function protein
GB19378	EGI68950	Sodium-dependent phosphate transporter
GB19380	EGI61696	Thioredoxin peroxidase
GB19729	No hit	Cytochrome c
GB19745	EGI69239	Transferrin
GB19857	EGI57313	Tubulin α -1 chain

Table S13. **Orthologs of *A. mellifera* spermathecal fluid proteins in *A. echinator*.** The spermathecal fluid proteins of *A. mellifera* described in Baer et al. 2009a and their orthologous genes in *A. echinator*. Protein IDs from BeeBase (<http://hymenoptergenome.org/beebase/>) are given for *A. mellifera*, GenBank IDs for *A. echinator*.

<i>A. mellifera</i>	<i>A. echinator</i>	Protein description
GB10467	EGI65745	Aspartate aminotransferase 2 precursor
GB10973	EGI58962	Arginine kinase
GB15049	No hit	Delta-1-pyrroline-5-carboxylate synthetase
GB15171	EGI63287	Ornithine aminotransferase precursor
GB16218	EGI65722	Proline oxidase
GB17641	EGI66290	Alanine aminotransferase 2
GB18844	EGI63319	Glutamate oxaloacetate transaminase 1
GB10133	EGI68526;EGI68528;EGI70961	Superoxide dismutase
GB10498	No hit	Peroxiredoxin
GB12029	EGI59687	Glyoxalase domain-containing protein
GB14972	EGI70194	Thioredoxin reductase
GB15855	EGI64147	Thioredoxin-2
GB18955	EGI67689	Phospholipid hydroperoxide glutathione peroxidase
GB19380	EGI61696	Thioredoxin peroxidase 1
GB30268	No hit	Glutathione s transferase S1
GB12586	EGI67714	Protein disulfide-isomerase precursor
GB12447	No hit	CAP; adenylate cyclase-associated protein 1
GB13596	EGI58074	ATP synthase
GB14791	EGI63505	ATP synthase subunit
GB15291	No hit	ATP synthase gamma subunit
GB16485	EGI62205	ATP synthase D chain
GB10989	EGI68169	Vacuolar ATPase catalytic subunit A
GB11380	EGI57447	Vacuolar H ⁺ ATPase 44 kDa C subunit
GB12913	EGI68957	Vacuolar proton pump E subunit
GB13499	EGI62608	Vacuolar ATPase subunit G
GB15226	EGI61088	Vacuolar ATPase subunit D 1
GB17480	EGI68544	Vacuolar ATPase subunit H
GB17499	EGI62319	ADP/ATP translocase
GB19171	EGI65100	Vacuolar ATPase 55 kDa B subunit
GB20017	EGI70849	Endoplasmic reticulum ATPase
GB10355	No hit	Melittin
GB10695	EGI60975	Pyruvate kinase
GB10992	EGI68476	ATP citrate lyase isoform A
GB11056	EGI64907	Phosphoglycerate kinase isoform 1
GB11461	EGI57893	UTP-glucose-1-phosphate uridylyltransferase
GB12488	EGI59293	Aconitase
GB12573	EGI69480	Citrate synthase
GB12741	EGI68216	Aldehyde dehydrogenase

GB12949	EGI69482	6-Phosphogluconate dehydrogenase
GB13058	EGI69850	Dihydroxyacetone kinase 2
GB13237	EGI65049	Phosphogluconate mutase
GB13882	No hit	L-lactate dehydrogenase
GB13955	EGI59833	N-acetyltransferase 5
GB14517	EGI69605	Isocitrate dehydrogenase
GB14798	EGI66496;EGI66497	Glyceraldehyde-3-phosphate dehydrogenase 2
GB14803	EGI59295	Alpha;alpha-trehalase
GB15039	EGI65919	Enolase
GB15052	EGI59081	Phosphoglyceromutase
GB15463	No hit	Aldolase
GB15543	EGI58485	Malate/L-lactate dehydrogenases
GB15619	EGI66180	Transketolase-like
GB15888	No hit	Carbonic anhydrase
GB16429	EGI67470	Glucose-6-phosphate isomerase
GB16464	EGI64085	malate dehydrogenase
GB16951	EGI60294	Malic enzyme
GB17113	EGI61891	Phosphofructokinase
GB17473	EGI64718	Triosephosphate isomerase 1
GB18109	EGI65078	Aldose reductase (NADP+)
GB18727	EGI65281	Malate dehydrogenase
GB19030	EGI65076	Aldo/keto reductase family protein
GB19387	EGI65069	Hexokinase A, isoform A
GB19460	EGI61646	Aldolase, isoform F
GB11665	EGI57614	Chitinase-like protein
GB11876	EGI67804	LDLa domain containing chitin binding protein
GB16986	EGI65076	S Endochitinase precursor
GB10397	EGI63414	Alpha-crystallin
GB10800	EGI61078	T-complex protein 1
GB10836	EGI66445	HSP70
GB14758	EGI61913	Heat shock protein 90
GB14852	EGI70201	Heat shock protein 8 isoform 1
GB15016	EGI62314;EGI70210	Heat shock protein cognate 3
GB17056	EGI68160;EGI64953	Cyclophilin 1
GB18662	No hit	Alpha-crystallin, small HSP
GB18969	EGI60183	Heat shock protein 60
GB12818	EGI63566;EGI63563	Histone 2A
GB14548	EGI64312	Deoxyribonuclease II
GB16515	EGI64145	ATP dependent DNA helicase
GB19247	EGI59103	Elongation factor 2, isoform 1
GB16568	No hit	Cytochrome c oxidase subunit
GB19293	EGI62567	Cytochrome c
GB19729	No hit	Cytochrome c
GB11059	EGI68222	Retinoid- and fatty-acid binding protein
GB15044	EGI68836	Phosphatidylethanolamine-binding protein

GB14639	EGI58101	Major royal jelly protein 8
GB16324	EGI58610	Major royal jelly protein 9
GB12951	EGI64087	14-3-3-like protein
GB15202	EGI70857	DJ-1; neuroprotective transcriptional co-activator
GB15582	EGI64087	14-3-3 epsilon
GB16178	EGI57927	Neuropeptide Y receptor
GB16716	EGI64583	Leucine-rich repeat-containing protein
GB16072	EGI63266	Iron regulatory protein 1B
GB10536	No hit	Odorant binding protein 14
GB19662	EGI67974	Juvenile hormone binding protein
GB19745	EGI69239	Transferrin
GB10009	EGI58698;EGI67764	Tubulin alpha-1 chain
GB10091	No hit	Cuticlin-1 precursor
GB10122	EGI65405	Tubulin, beta, 2
GB10275	EGI66955;EGI66957	Tubulin isoform B
GB10514	EGI67764	alpha tubulin
GB11282	EGI68594	Moesin isoform D
GB11920	EGI66956	Tubulin
GB12614	No hit	Actin
GB13049	EGI66955;EGI66957	Tubulin, beta, 2
GB13229	EGI57949	PDZ and LIM domain protein
GB13999	EGI60934;EGI60932	Vitellogenin
GB15794	EGI57482	Cuticlin-1 precursor
GB16448	EGI70234;EGI66394	Annexin IX
GB17673	EGI67012	Talin-1
GB17681	EGI67274	Actin-5C isoform 1
GB18917	EGI57258	Cofilin/actin-depolymerizing factor homolog
GB12113	EGI58834	Porin
GB14012	EGI63001	Phosphate carrier
GB16577	EGI60893	Sialin, inorganic phosphate cotransporter
GB11987	EGI65106	Unknown
GB12562	EGI70188;EGI70190;EGI59981	Hypothetical protein
GB13778	EGI68315	Unknown
GB14970	No hit	Muscle-specific protein 300
GB15662	EGI60204;EGI60205	Unknown
GB17311	EGI64427	Unknown
GB17500	No hit	Hypothetical protein
GB19255	EGI57644	Osiris 14 CG1155-PA
GB30569	EGI69055	Hypothetical protein

Table S14. **Core set of neuropeptide genes.** This set has, so far, been regarded as the core set of neuropeptide genes found in *Apis mellifera* (honeybee), *Nasonia vitripennis* (parasitic wasp), *Drosophila melanogaster* (fruit fly), *Aedes aegypti* (yellow fever mosquito), *Bombyx morix* (silkmoth), *Tribolium castaneum* (flour beetle) and *Acyrtosiphon pisum* (pea aphid) (Hauser et al. 2010). However, in all ants, including *Acromyrmex echinator* (first column), the RYamide neuropeptide gene is absent, meaning that the RYamide neuropeptide gene from now on belongs to the variable set (see Tables S14 and S17). Abbreviations: AKH, adipokinetic hormone; AST, allatostatin; CCAP, crustacean cardio-active peptides; DH (Calc.-like), calcitonin-like diuretic hormone; DH (CRF-like), corticotropin releasing factor-like diuretic hormone; EH, eclosion hormone; ETH, ecdysis triggering hormone; ILP, insulin-like peptide; ITP, ion transport peptide; sNPF, short neuropeptide F. Modified from Hauser et al. 2010, with permission.

neuropeptide	<i>Acromyrmex</i>	<i>Apis</i>	<i>Nasonia</i>	<i>Droso-phila</i>	<i>Aedes</i>	<i>Bombyx</i>	<i>Tribolium</i>	<i>Acyrtosiphon</i>
AKH	1	1	1	1	1	2	2	1
AST-C	1	1	1	1	1	1	1	1
AST-CC	1	1	1	1	1	1	1	1
Bursicon- α	1	1	1	1	1	1	1	1
Bursicon- β	1	1	1	1	1	1	1	1
CCAP	1	1	1	1	1	1	1	1
CCHamide-1	1	1	1	1	1	1	1	1
CCHamide-2	1	1	1	1	1	1	1	1
DH (Calc.-like)	1	1	1	1	1	1	1	1
DH (CRF-like)	1	1	1	1	1	1	1	1
EH	1	1	1	1	5	1	1	3
ETH	1	1	1	1	1	1	1	1
ILP-B	1	1	1	5	6	38	2	7
ITP	1	1	1	1	1	1	1	1
Myosuppressin	1	1	1	1	1	1	1	1
Pyrokinin	1	1	1	1	1	1	1	1
RYamide	0	1	1	1	1	1	1	1
SIFamide	1	1	1	1	1	2	1	1
sNPF	1	1	1	1	2	1	1	1
Tachykinin	1	1	1	1	1	1	1	1
	19	20	20	24	30	59	22	28

Table S15. **Variable set of neuropeptide genes.** The variable set of neuropeptide genes found in *Apis mellifera* (honeybee), *Nasonia vitripennis* (Jewel wasp), *Drosophila melanogaster* (fruit fly), *Aedes aegypti* (yellow fever mosquito), *Bombyx mori* (silkworm), *Tribolium castaneum* (flour beetle) and *Acyrtosiphon pisum* (pea aphid) (Hauser et al. 2010). Abbreviations: ACP, adipokinetic hormone/corazonin-like neuropeptide; ADF, antidiuretic factor; AST, allatostatin; GPA2, glycoprotein hormone A2; GPB5, glycoprotein hormone B5; ILP, insulin-like peptide; NPF, neuropeptide F; NPLP, neuropeptide-like precursor; PDF, pigment dispersing factor; PTTH, prothoracicotropic hormone. Modified from Hauser et al. 2010, with permission.

neuropeptide	<i>Apis</i>	<i>Nasonia</i>	<i>Drosophila</i>	<i>Aedes</i>	<i>Bombyx</i>	<i>Tribolium</i>	<i>Acyrtosiphon</i>
ACP	no	yes	no	yes	yes	yes	no
ADF-b	no	no	no	no	no	yes	no
Allatotropin	no	no	no	yes	yes	yes	yes
AST-A	yes	yes	yes	yes	yes	no	yes
AST-B	no	no	yes	yes	yes	yes	yes
Capa	yes	no	yes	yes	yes	yes	yes
Corazonin	yes	yes	yes	yes	yes	no	no
FMRFamide	yes	no	yes	yes	yes	yes	yes
GPA2	no	no	yes	yes	yes	yes	yes
GPB5	no	no	yes	yes	yes	yes	yes
ILP-A	no	no	yes	yes	no	yes	no
ILP-C	yes	yes	yes	yes	no	yes	yes
Inotocin	no	yes	no	no	no	yes	no
Kinin	no	no	yes	yes	yes	no	yes
Neuroparsin	yes	yes	no	yes	yes	yes	no
NPF	yes	yes	yes	yes	yes	no	yes
NPLP-1	yes	no	yes	yes	yes	yes	yes
NPLP-2	yes	no	yes	no	no	no	no
NPLP-3	yes	no	yes	no	no	no	no
NPLP-4	no	no	yes	no	no	no	no
Orcokinin	yes	yes	no	yes	yes	no	yes
PDF	yes	yes	yes	yes	yes	no	no
Proctolin	no	no	yes	no	no	yes	yes
PTTH	no	yes	yes	yes	yes	yes	no
RYamide	yes	yes	yes	yes	yes	yes	yes
Sex peptide	no	no	yes	no	no	no	no
Sulfakinin	yes	no	yes	yes	yes	yes	no

Table S16. **The presence or absence of neuropeptide genes in four ant genomes.** The absence of a neuropeptide gene is always corroborated by the absence of its receptor gene with the exception of tachykinin in *Harpegnathos saltator*, where the tachykinin gene is absent, but the tachykinin receptor is present. We assume that this is due to sequencing gaps. For neuropeptide abbreviations, see Tables S14 and S15. For prohormone sequences, see Table S14.

neuropeptide	<i>Acromyrmex echinator</i>	<i>Solenopsis invicta</i>	<i>Camponotus floridanus</i>	<i>Harpegnathos saltator</i>
ACP	no	no	no	no
ADF-b	no	no	no	no
AKH	yes	yes	yes	yes
Allatotropin	no	no	no	no
AST-A	yes	yes	yes	yes
AST-B	no	no	no	no
AST-C	yes	yes	yes	yes
AST-CC	yes	yes	yes	yes
Bursicon- α	yes	yes	yes	yes
Bursicon- β	yes	yes	yes	yes
Capa	no	no	no	no
CCAP	yes	yes	yes	yes
CCHamide-1	yes	yes	yes	yes
CCHamide-2	yes	yes	yes	yes
Corazonin	yes	yes	yes	yes
DH (Calc.-like)	yes	yes	yes	yes
DH (CRF-like)	yes	yes	yes	yes
EH	yes	yes	yes	yes
ETH	yes	yes	yes	yes
FMRFamide	no	no	no	no
GPA2	no	no	no	no
GPB5	no	no	no	no
ILP-A	no	no	no	no
ILP-B	yes	yes	yes	yes
ILP-C	yes	yes	yes	yes
Inotocin	yes	yes	yes	yes
ITP	yes	yes	yes	yes
Kinin	no	no	no	no
Myosuppressin	yes	yes	yes	yes
Neuroparsin	yes	yes	yes	yes
NPF	yes	yes	yes	yes
NPLP-1	yes	yes	yes	yes
NPLP-2	no	no	no	no
NPLP-3	no	no	no	no
NPLP-4	no	no	no	no
Orcokinin	yes	yes	yes	yes
PDF	yes	yes	yes	yes
Proctolin	no	no	no	no

PTTH	yes	yes	yes	yes
Pyrokinin	yes	yes	yes	yes
RYamide	no	no	no	no
Sex peptide	no	no	no	no
SIFamide	yes	yes	yes	yes
sNPF	yes	yes	yes	yes
Sulfakinin	yes	yes	yes	yes
Tachykinin	yes	yes	yes	yes

Table S17. Ant neuropeptide precursor sequences.

Alphabetical list of all neuropeptide precursor sequences identified in the genomes of *Acromyrmex echinator* (Aech), *Solenopsis invicta* (Sinv), *Camponotus floridanus* (Cflo) and *Harpegnathos saltator* (Hsal). The putative signal peptides (predicted by SignalP) are boxed, the putative active neuropeptides or protein hormones (based on similarity to neuropeptides or protein hormones identified in other insects) are marked in yellow. Green indicates putative basic cleavage sites flanking the putative neuropeptides; glycines used for amidation are shown in blue. Dots indicate missing N- or C-termini. Underlined cysteine residues are proposed to be involved in cysteine bridges. GenBank accession numbers are given in parentheses. When protein IDs are not available, accession numbers and coordinates refer to nucleotide sequences (contigs) instead. The gene orientation in a contig is indicated by + or -.

>Aech AKH (EGI65089.1)
MSMSLLGIATFLFLILVADGEAQLNFSTGWGQGKRSQDTRIRSSNMDCSPQGASLEQLLKLYSFIQMEAQ
RILDCQTLNK

>Sinv AKH (AEAQ01023630.1: 1396-1590, 2657-2698 +)
MSVSLVALTFLFLILVANGKAQLNFSTGWGQGKRSQNMRIKSSSMDCSSQDASLEQLLKLYSFIQIEAQ
RILDCQTLNK

>Cflo AKH (EFN65319.1)
MPMSLFALAFFLFLILVSDGEAQLNFSPGWGQGKRSQDMRIRPSNTDCSSQSALLEQLLKLYTFIQIEAQ
RILDCQNLNK

>Hsal AKH (EFN77871.1)
...RRAPVSLLTVAFLFCLILALSVEAQLTYSTDWGKRSQNVRMKSDCPSRSTSLLEQLLKLYTFVQIEAQ
RILDCQNLN

>Aech Ast-A (EGI57352.1)
MKTTTSLITMRIIMFYLLSVVGRAVAIEEAASSSLHIQQLHPLLNMEYEEPIKKAYIAEYKRLPLYTF
GIGKRWIDNNEDKRSRQFSFGIGKRLRNYDFGIGKRNNEYHPLNLDYFDNMGDYQSHEDNSNDFMENKR
SNHQFGFGIGKRVWKLATGETTVSGRRLNDAIVPKYWFSILAKELENEENLNQ

>Sinv Ast-A (AEAQ01012245.1: 3932-4171, 5589-5873 +)
...SRTSLITIRIIIFYLLSVVGQSTAAIQETSSSSLHI PRFHPLLNLEFEEPEKKSZIAEYKRLPLYNF
GIGKRMNSDEDKRNRQFSFGIGKRLRDFRFGIGKRNGYHPFSLDFPADNIEGYQSRDNSNDFMEDKRSQ
HFGFGLGKRGAWKLASEEMAERRLNDDPIVSKFLFGLGK...

>Cflo Ast-A (EFN68211.1)
MKTSSLIAMRLIIFYLLSVVGRSTAAVEEAPASSLHI PRLNPLSSNLEYDEPSEKRAYAYISEYKRLPL
YNFGIGKRWIDNSEDKRTRPFSFGIGKRLRDYRFGIGKRNSGYRPLGMDFSVDNMFHSREDNLDFFID
DKRGGQPFSGIGKRGWKLPMEGEMAVSGRRLNDVVGPKYLLGLGKGLSE...

>Hsal Ast-A (EFN80332.1)
MLCLLSVVGESTAAMEDMPSSSLHMPRLNPLLNHVEYEEPESEKRSYAYVSEYKRLPLYNFGIGKRWVDD
NEDKRTRPFSFGIGKRLRDYRFGIGKRNSHPLNLDYLPADNLEAYHSREDNADDLMEEKRSNQPFSFGI
GKRGWKLAGATARRPNADVVAAPRYLLSLGKIGEDEEM...

>Aech Ast-C (EGI60677.1)
MLSVQSTVALALMLFVLAEYSAAMPTIDKDKGRFLNNVNLVDDGSDKVLNMNYLFTKQIVKHLRNQMN
VNDLQRKRNFWRQCSLNAVACFGK

>Sinv Ast-C (AEAQ01021842.1: 60294-60461, 69956-70072 -)
MLSIRSVVTVALMLFVFAEYTGAMPTIDKDKDRFLKNMNLVDADDGGIETALMNYLLTKQIVKRLHNQL
DISDDFERKRKRWQLCSVNAIACFGK

>Cflo Ast-C (EFN70796.1)
MSSVRNIAALALVLLVLAEWSAAMPTTDKDKERLLNTVDLIDDDGSIETALINYLFTKQIVKRLRSQLD
IGDLQRKRSYWKQCAFNAVSCFGK

>Hsal Ast-C (EFN89652.1)
MPSVRSALALVLFVFAEWSRAMPAIDKDNKERILNAVDLMDDDGNIETALINYLFTKQIVKRLRNQL
DIGDLQKRKRSYWKQCAFNAVSCFGK

>Aech Ast-CC (AEVX01015845.1: 118433-118516 +)
...MDLQRRALAQSELFWR_CYFNALACFKKK

>Sinv Ast-CC (AEAQ01017772.1: 2979-3056 -)
...MDLKRKGHKSDPWL_CYFNALACFKKK

>Cflo Ast-CC (EFN62371.1)
...MDLQRRGQAKGQTKGQVAFWR_CYFNAVTCFKRK

>Hsal-Ast-CC (AEAC01026692.1: 14333-14416 -)
...MDLQRRGQKKSQVYWR_CYFNAVTCFRRK

>Aech bursa (AEVX01012178.1: 36000-36164, 37776-37898, 38738-38884 +)
...VINGLLIICLLYGATEAIVGVDE_CQATRVIHFLQYPG_CVPKPIPSYACRGR_CSSYLQVSGSKMWQMERS
CMCCQESGEREASVSLFC_PRAKPGEKKFRKMVTKAPLDCMCRP_CTSVEEYAIIPQEIAGFADEGPFPTTS
AHFRRASGLQ

>Sinv bursa (EFZ11126.1)
...VINGLVIICLLMYGGTEAIVGVDE_CQATRVIHFLQYPG_CVPKPIPSYACRGR_CSSYLQVSGSKMWQME
RSCMCCQESGEREASVSLFC_PRAKPGEKKFRKVIITKAPLECMCRP_CTSVEEYAIIPQEIAGFADES PFT
TSAHFRRSSGLP

>Cflo bursa (EFN71945.1)
MYRCRSKLLRYFAINGLLMCLLYGSAAEIVGVDE_CQATRVIHFLQYPG_CVPKPIPSYACRGR_CSSYLQV
SGSKMWQMERSCMCCQESGEREASVSLFC_PRAKPGEKKFRKMVTKAPLECMCRP_CTSVEEYAIIPQEIA
GFADEGPFATSAHFRRSSGLQ

>Hsal bursa (EFN80834.1)
...DECQATRVIHFLQYPG_CVPKPIPSYACRGR_CSSYLQVSGSKMWQMERSCTCCQISGEREASVSLFC
PRAKPGEKKFRKVIITKAPLECMCRP_CTSVEEYAIIPQEIAGFADEGPFATSAHFRRSPGLQ

>Aech bursb (AEVX01012178.1: 33170-33379, 33505-33672 +)
MLVVPKSLLVLVVYVVTISDAEITEIDDES_CETLQSEIRITKDEYDELAHLRRTCSGDVSVTKCEGFCNS
QECYCCRESYLKERFVILNHCYDTHGIKLIGTDDEAMEIKIQEPAECKCICKGDFAR

>Sinv bursb (EFZ20687.1)
MLVAPLLVLMTYVVTISDAEITEVDDDES_CETLQSEIRITKDEYDEIGRLRRTCSGDVSVIKCEGFCNSQV
QPSVVTGFSKECYCCRESYLKERFVVLNQC_YDADGIKLIDTDEAMEIKIREPAECKCICKGDLAR

>Cflo bursb (EFN71946.1)
MLVSPKSLILVIYATVNEAEKTEMKIDDES_CETLQSEVRITKDEYDEIGRLRRTCSGDVSVTKCEGFC
NSQECYCCRESYLKERFVILNQC_YDADGIKLIGIDAEIMEIKIREPAECKCICKGDLAR

>Hsal bursb (EFN80833.1)
MSLFSKSLILAVYAAINDAQVDDDEGCETLQTEVRITKDEYDELTRLRRTCSGEISVSKCEGFCNSRM
QPSIVTTGFSKECYCCRESYLKERPVILSQ_YDADGIRLIGTDDEMEIKIREPAECKCICKNDLAR

>Aech_CCAP (AEVX01013372.1: 44416-44666, 44940-45006 -)
...LLNALNAEKTIKSKRPF_CNAFTGCGRFVSEERRMKKGLSTLQDLSDFEVTNNNVQLPVSLYRAMLDA
AKQNVWVMDHEAYDYRLQQIPRIYLSGQMPVRNTME

>Sinv_CCAP (AEAQ01008009.1: 16528-16767, 17214-17282 -)
...LLNGLNAEKTIKAKRPF_CNAFTGCGRFISEEKREKKKGPSTLQDLLNFEVTSNNVQYPVSLYKALL
NAAQNAWKTMDREEYDYRSQQLPIYLSGQMPPLRDTV

>Cflo_CCAP (AEAB01004141.1: 1119-1187, 1432-1620 +)
...LLDALMMENTMKPKRPF_CNAFTGCGRYVSEEKREKKKEPRISTLLHLLRLLNTSKQNTWNTIDRE
NYQLQQMPQVYVSNRMPLRD

>Hsal_CCAP (EFN89452.1)
...LLDALTTERTMKPKRPF_CNAFTGCGRKRSFSASIPTLRDSQDFRANDNVRLPVSLYRALLLNAAKQN
TWAIDREEYDYRLRRIPQIYLPQQTPLRDTLES

>Aech_CHa1 (AEVX01009652.1: 35784-35927, 51613-51648 -)
...AETCLHFGNS_CWAHGKRGNIQGVPTMHTLTAKALLNGFPQDNIVSSSKVQWILSRLLIIRQ...

>Sinv CCHa1 (AEAQ01004246.1: 1419-1556, 1676-1705 -)
 ...**CLSYGNACWGAHGKR**SDIRDKHTMRILTLNNGFPQNGGIIPSSKAQWLSRFNTGQ...

>Cflo CCHa1 (EFN68229.1)
 ...**GSCLSYGHSCWGAHGKR**SDAPAIRLLTAKSLLSKSPQKNIVHSSKAQWILSRLITGQ...

>Hsal CCHa1 (AEAC01006396.1: 41381-41542,41697-41729 -)
 ...**TCLNIGHSCWGAHGKR**SDVQDAAPALRILAAKALLSELPKNGIASSSSSSSSKAQWLLSRLIAGQ...

>Aech CCHa2 (AEVX01006932.1: 3917-3970 -)
 ...AG**GCAAFGHTCYGGYGKR**...

>Sinv CCHa2 (AEAQ01004242.1: 6541-6594 +)
 ...AD**NCMSFGHTCFGAHGKR**...

>Cflo CCHa2 (EFN68230.1)
 ...**GCASFHSCFCGGHGKR**...

>Hsal CCHa2 (EFN86674.1)
 ...SAD**GCSIFGHSCFCGGHGKR**...

>Aech corazonin (EGI58954.1)
MATKCVFVFILLSLVISTVFC**QTFQYSRGWTNGKR**SDFPNSLGISNPGDEHFINGEFKRLKMLVYGNVD
 EQPLLMHCDVFDNGKLRKLGHADNYASQLREKQNDGNY

>Sinv corazonin (EFZ09618.1)
MTEKRLFVLFVLSLIIISTVLC**QTFQYSRGWTNGKR**SSLVPNSLDVPNSMDERFTGEPRTVKMLLYGDLN
 EQPLLIHCDVMDKFKKFLHTDN

>Cflo corazonin (EFN63812.1)
MLVLFVLSLVVSCALC**QTFQYSRGWTNGKR**SNFPAEISALGYDRFTNGELKRLKMLIYGSADQPLLIH
 CDFVDKLLKFAHTDN...

>Hsal corazonin (EFN88292.1)
 ...**IFVFLI**LSLVMSAVFG**QTFQYSRGWTNGKR**SEFPNSAEISNSISDEGFTSNELKRLKMLVHGNVDEQ...

>Aech_DH (EGI66870.1)
 ..VSSLTADAI PRN**KRGLDLGLNRGYS**SGSQAAKHMGLAAANYAGGP**GRRRR**SEQA

>Sinv_DH (EFZ22024.1)
MQPKMTVLCTLLTFIVVVAISSLTAEAVPHSHESYWDQDDIDRDEFVFEFLSRLSRAVMNRPEMESS**KR**
GLDLGLSRGFSSGSQAAKHMGLAAANYAGGP**GRRRR**S

>Cflo_DH (EFN61187.1)
 ...SS**KRGLDLGLSRGFS**SGSQAAKHMGLAAANYAGGP**GRRRR**SEQ

>Hsal_DH (EFN76310.1)
 ...ST**KRGLDLGLSRGFS**SGSQSAKHMGLAAANYAGGP**GRRRR**SEQ

>Aech_DH/CRF (EGI59165.1)
 ...DEMLNSENIFGSDVMRT**KRIGSL**SIVNNLDVLRQV**LLELARRKALQDQ**RQVEENRRF**LESV****GKR**SVS
 NADRIVRSSMNNERSATSDRNEWTEENPLFRELQDDRTVINQCNNPVSRLAEMGIEIVHRSEMRLIQ
 CMTDLTGAHNVPPK

>Sinv_DH/CRF (EFZ17990.1)
 ...DEMLNSENDLGSVMRT**KRIGSL**SIVNNLDVLRQV**LLELARRKALQDQ**RQVDENRRF**LETI****GKR**S
 VSDANRIVRSGVKNNRERSTASSRNEWIEQNPLFHKSQDDRTVRI

>Cflo_DH/CRF (EFN70881.1)
MILLGILASTTIIIGLTSS**APLSSYERRD**VSDNRPKIFLLMDERIPELENEILGNELGSDVTRT**KRIGSL**
SIVNNLDVLRNRVM**LELARRKQERDQ**RQ**IEENRRFLENI****GKR**SVPVSDAGKIVRSDKSRNDRNQPLQFN
 RIEWIEEEDPLFRGSQDDQMTRVQANELRLL

>Hsal_DH/CRF (EFN88255.1)
MILLGILTSTTIIIGLTSS**TSLSSYERRD**LLADRPKVFLLDVDRTPELENEMLDSSNDPSAVVVRT**KRIG**
SLSIVNSLDVLRQM**LLELARRKALQDQ**QLINTNYQ**FLDTV****GKR**SVPEYGADTGRITTIQRQHVESRD
 RSAASDRATSKMQQNWFEEENDPVFRESQDDQMTRANELHLL

>Aech EH (EGI68318.1)
MPSLSNRIMVLLIMVFAILCFTVSTNAERNIGVCIRNCAQCCKMFGVYFMGQKCADFCMKYKGLIPDC
EDEYSIRPFLQVAEYDY

>Sinv EH (AEAQ01025664.1: 1118-1154, 1240-1381 +)
...TTSTTDAERNIGVCIRNCAQCCKMFGTYFMGQKCADFCVKYKGLIPDCEDEFSIRPFL

>Cflo EH (EFN61914.1)
...GVCIRNCAQCCKMFGSFFIGQKCADYCIKYKGRFVDCEDEFSIQPFL

>Hsal EH (EFN81201.1)
...LGVCIRNCAMCRDMYGTYFKVQKCADFCVKYKGLIPDCEDEDSIRLFL

>Aech ETH (EGI60034.1)
...FLVGAVIMPLLICENTIAEEVPAFFLKIAKIPTLPRVGRSGRFEDFFYKAEKHIPRIGRSNQHKQSNN
PLLLQDEQETYSDLTKRRIDYPSKG

>Sinv ETH (AEAQ01008054.1: 2868-3053, 3179-3254, 3325-3412 +)
...FLIGIVTLLVCQNITRADEVPAFFLKIAKIPTLPRVGRSGRFEDFFYKAEKHIPRIGRSNPQSINPTS
QDEQETYSDLTKRRIEYPKVEEWTWQNFPLAIEGPRELWRALAGYSRD

>Cflo ETH (EFN63630.1)
...SLLGVIMMLICQNVAKADEVPAFFLKIAKIPTLPRVGRSGRFEDFFYKAEKHIPRIGRNIQDQODIHY
GLSKRRIDYPLKADEWSWQNFPLAIEGPKELWRTLAGYSRDGDE

>Hsal ETH (EFN79403.1)
...LFAGVITLLIYQDIARAEEVPAFFLKIAKVPNVPRVGRSGKFEDFFYKAEKHIPRIGRNSQLNDPLPL
QNEQETYSDLTKRRIDYTAKVAQDAWSWQNFPLAIEGPKELWRTLALYARDDSD

>Aech ILP-B (EGI63846.1)
MSANRLNVLVTLMLLAVALLVTEGNAQVDGYPOFSSKRSVAVSAPQKYCGKLSNALQIICDGVNSMFK
KSGQAREEANVDSRDSNFYVRTQEMEMADYPIAYEFPFPLPRARARGMLDGRFAGRRYRRQGRGIHEE
CCINPCTINELTSYCGGRGSASE

>Sinv ILP-B (EFZ13592.1)
MSANRLNVLVTLMLLAVALLVTEGNAQVDGYLQFNPKRSAVSSPQKYCGKLSNALQIICDGVNSMFK
KSGQDFPPQNKRIAHRIHRINGNEEESFTTLKSNFLNWCVEVYHRHYRFFVSEMEMADYPLAYDISPYLP
PFLSRARARGMLDGRFAGRRYRRESRGIHEECCINGCTINELTSYCP

>Cflo ILP-B (EFN61735.1)
MSTNHLNVFVSLMIAIAFLVSESRNAQANGYPOFNQKPAQVQKYCGKLSNALQIVCDGVNSMFKKS
LDQEMEMADYPFAYDSPFPLPRARARGMLNGPFAGRRYRRQSRGIHEECCLNACTISELSSYCA

>Hsal ILP-B (EFN78996.1)
MSAYRLSVIVTLVMAVAFLMPESGNAQSDGSYALKWSMNVQRYCGRRLSNALQTVCTGVYNNMFKKNV
QQEENKTQQEKSHFVDPGTRRSNMYSLEEVSEYPLNYERPSLFI SGMGATQFINRYRGSRVRRARGIYE
ECCLNACTYNELSTYCGPQQ

>Aech ILP-C (EGI57346.1)
MMIRSSDRAMNEAVLLTALIFLSVLYAVDA DNVI FKKSHQRMQKLC SRKLSDALQVMCRDRGYNEPFYS
NEDESRIIDPGPLVVECCYHQCTYEQMEQYCKPLPAEKRIDSRDDVIDLSYIANLPHSTTKALLEQHSQ
TEM DYAGDAIKRKVDDLKRGRRHGKSGRNNDGECKGKADAKKRHRGRHRCRRRRLECRAGKVLRSNA
KPLDNKFVTPLITDEPTSFETI

>Sinv ILP-C (EFZ16905.1)
MLVRSRGHAMIVAMLLIALVFLCSLYAADAQSNFRRSHIRMQKLC SRKLSDALYLI CRERGYNEPFSYSN
EDEPRVDPGPGLVDECCYHQCTLGMLEAYCKPLPEEKRAGLRDDVM...

>Hsal ILP-C (EFN80340.1)
MVQRASKAGTVLLATFLLLSVLHIIDS VPLKSLRSGTLRLCSRNLSDALYLVC RERGYNSYSYSDDDDET
QVDNPGPLVDECCYHACS YEQLERYCKPI PGEKQDES RDVIDGTYISVRMPFMPKDP SKENLRSEMDYT
DGTKKR KVNSMKRRRHRGKGRK DAGECKGKAAAKRRHHGGHYRCRHRCLECRRTGKMKLALILMALVA
SSKTESTNFGSTRVENPFVILDNI VDFDEMKG MVEKHWSSISKQSLKLFIEFSILTESKAFSDFIKQ
IKSDNFQQIFKAYNPNI R KILM

>Cflo ILP-C (EFN61235.1)
 MTDGDRVTCKARPVFLFVLVLLSILHIADSQVTLRKSHIRMQKLCRSRLSDALYLVCRRERGYNEPFSYS
 GEDEPRVDPGPGLVEECCYHSCSYQLERYCKPLPEERHDVIGESYMGSVVNLPPFSMTSSEQRSGTEHV
 GTKKRVKVDLKRERHRGKGGRNIDDECKGKPDVKKRHRSRHRCRRRRLECRAGKVSRRKLLANQF
 VTSSPTNKPTSSPTIS

>Aech inotocin (EGI58692.1)
 MLKELIVFASLIIFLSYA~~CLITNCPRG~~GKRSDIASLKTVIRECPSCGPNHLGQCFGPYICCGPSIGCFIG
 TPETFRCRKESLYTRPCIAGYAMCRGKTARCASDGI CCSQASCHMDTSCKISDVGNDRNLDDNVNVLIP
 GNEVSNEILQ

>Sinv inotocin (EFZ22443.1; only C-terminal part)
 ...CPSCGPNRLGQCFGPHICCGPSIGCFIGTPETYRCRKESLYTRPCIAGYAMCHGNTARCASNGICCSQ
 ...

>Cflo inotocin (AEAB01028364.1: 8-133 -, AEAB01028362.1: 78-281 -)
 MLKQLVICASLIIFLSHA~~CLIVNCPRG~~GKRSDIASFLKTVTRECPTCGPNHMGQCFGPRICCGPNIGCFI
 GTPETYRCRKESLYTKPCIAGYAMCRGNTARCAANGICCSQ

>Hsal inotocin (EFN79183.1)
 MLRELVVVFASLIIFLSYA~~CLITNCPRG~~GKRSDIIPSLGTVTRECPCGPNHLGQCFGPHICCGPTIGCFI
 GTPETYRCRKESPYARPCIAGYAMCRGNTARCATNGICCSQDSCHMDTSCRISDVVSNDRKMDADLSAI
 LSSNEASHEIIQ

>Aech ITP (EGI68809.1)
 MSILTWSLTLILLISSCIDLGADAASLSGHPLG~~KRSFFDIOCKGVYDKSIFARLDRI~~~~CEDCYNLFREPQL~~
~~HMLCKKCEFTTDYFKGCLDVLLLTDEVGKIQMWIKQLHGADPGV~~

>Sinv ITP (EFZ14169.1)
 MYRQONTQSSSNDDEFAMFAPSAAYRSSCSTLPSSMSTSSSSSSSSSSSSSSSRSSPAGSSCPLLASVLAW
 SLTLILLISSCIGLGADAASLSGHPLG~~KRSFFDIOCKGVYDKSIFARLDRI~~~~CEDCYNLFREPQLHMLCNE~~
~~TISQNVITAIILVSRASSAAKTGVNNSILKCEFTTDYFKGCLDVLLLTDEVGKIQMWIKQLHGADPG~~

>Cflo ITP (EFN64772.1)
 MHRQQDVQSSSSNGEFAMYPASAVYHSSCSTSTPSSSSRSSTTCPFLLSIFTWSLTFLLISSCIGLGADA
 AAMNAHSLG~~KRSFFDIOCKGVYDKSIFARLDRI~~~~CEDCYNLFREPQLHMLCKKCEFTTDYFKGCLDVLLLT~~
~~TDEVGKIQRQECFSTEYFTSCMQVLLLEDEKETLQEMARYLGRKK~~

>Hsal ITP (EFN85846.1)
 MYPSAAYHSSHSTLPSSTLSASSSSSPASPSSLPSPRPLSASSSPILLSVLWLSLALLLISSCINLT
 DARTLNHGPLS~~KRSFFDIOCKGVYDKSIFARLDRI~~~~CEDCYNLFREPQLHMLCKKNCFTTDYFKGCLDVLL~~
~~LLSDEVEKIQMWIKQLHGADPGV~~

>Aech MS (EGI69517.1)
 MSSTIMILVSVTTMAILSGEALAGSPGISCNPGFLELPPRLRKCMAFARIWDVRDMNDFIDDKEYRE
 NLPYDRAV~~KRQDVDHIFLRF~~~~GKRR~~

>Sinv MS (EFZ21935.1)
 MMSSSLMILVSVTTMAVLSGEALALPPQCNSLLELPPRLRKCVAIARIWDIKLNNFVDDRGNREN
 LPRYDRGV~~KRQDVDHVFLRF~~~~GKRR~~

>Cflo MS (EFN71919.1)
 MSPTLMILISITTTMAILSGESFGAMPAQCNSFLEELPPRLRKCVAIARIWDAREMNDVDDREYREN
 LPRYDSSV~~KRQDVDHVFLRF~~~~GKRR~~

>Hsal MS (EFN78200.1)
 MISPTLMVLVSVTTMAIISGEAFAALPAQCNSPGFLELPPRLRKCVTIARIWDANDMNDVDDRDLPLP
 RYNNGV~~KRQDVDHVFLRF~~~~GKRR~~

>Aech neuroparsin (AEVX01014486.1: 28640-28765, 29233-29373 +)
 ...YRIDKCHAHPAIAQRHEPRPLCVGCGDKCEQCEFGYTTSTLC...GPGHICGGPSDSWGVCGDGLICSCN
 RCTGCSVDSLTCFANTCLPHQS

>Sinv neuroparsin (AEAQ01013249.1: 26551-26691, 27247-27393 -)
 ...FRIDKCHAHPAIAQRQYQAISLRCGCGDQCDCQCEYGTIISTLCGVKECLKGPGERCGGPSEIEGVCGEG
 MMCNHCNKICGSSINFCKTDDYCLPH

>Cflo neuroparsin (EFN74097.1)
...HPTSRREQEVRPALCVGCGTECDKCKFGFVISAICGIAECRRGPGDICTGGPSEAWGVCGDGLICSCNRC
AGCSLDSLVCFANHAACLPHQSLESGLHDLDFERFSPQFDRVA

>Hsal neuroparsin (EFN83436.1)
...SSQCTGCGNECDKCKYGVTVSALCGIEECRRQGPGEYCGGPSQSWGVCGEGMFCICDRVCVGSVDFLT
CFTKSCLPHQS

>Aech NPF (AEVX01008543.1: 18111-18266, 19039-19233 -)
MRTRMNPICFFCCLLVTTIIGITITCAEPEPMTRLTRPPQISNSNVQLQRYADSVKNFYHMYGKSRHGKR
SDAAPLMELNATWETLKM IQDAQRQNEQRRQENIRQNKEQILLRDFPNLDANEYTLVDVTRSGSLSDII
NKYYNDVQ

>Sinv NPF (AEAQ01015448.1: 1903-2142, 2766-2946 -)
MRTRMNLDRFLCCLLIVTVIEMTIACAELPARLARSDETIDSDISLQKFVDCLRNRYIYTKPRHGKRS
DAMPMMELNATWETLKM IQDAQRQNEQRRQENIRQNKEQILLRDFPSSDADEYTSYDATRSGSLPDVID
KYYNDV

>Cflo NPF (EFN68519.1)
MRRAWIMSLAYFSFCLFTVIGTIVQAEPESMARPTRPKVFGSPEELKKYADLVRDYYYLIGKARYGKRGD
MAPMSTELNNSWETLRMIQDARRQNEQQRQEKMRQSKQVLFSDLQSSDIDKHTSHVATRPGSLSDMIG
NYIDVQ

>Hsal NPF (EFN77166.1)
MWTRMSLVRFLGCLFLVAVVGTNAYA EPEPMARPTRPKAIASPEELERYFDLVRDYYSLRRARYGKRS
GDAAPRTLELNATWETLKTILDAQRQNEQRRRLGKNKPTTNEQEIAYRDYHGLDADRHASRVAARPGYLL
DVAGRYYYDDVQ

>Aech_NPLP1 (EGI58293.1)
MAHLKALSSEENDDAEICIPSDIYLELLRDLTIRGHLVTRDLIPESEKRSLATLAKNDDLPFIVQERED
DGDDEEKRSINISSDNLDELVRALMGDEYRRRTSGIYDYQVKLDPELIEYPLDKRNVGALARDFALPTG
RRHIGSVLRDYNTMSGKRNI GSLARQSMPLPMSGKRNVASLARDSMLPQNGKRNVAALARDSSLPYGKRY
LGSLVRNGGYPVRYGDEGKRNIASLARNADWPGFMKRGGTAGVGRMIARVLRHGRSLNDHEAPSEPLD
LQQLIKQDHNNGNDAKENVAGNWPTLFSISEELDDSRAKNRSNRRMDAATSQTRHKRQIDFSDEYPLPV
VQNTNILDYEDMIEALADHYPNTEKRFGLGSSPEMPADSGYTEVLQPIKWEDPADGTSNYPPLASAKTR
SLRPFIPKTRYLQSLHRLHGDCRHGFKRFLLPDIDNFLRTSNRIVSNSM

>Sinv_NPLP1 (EFZ21681.1)
...NHNDMAHLKALSSEENDDAEICIPSSVYLELFRDPALRGHLSVMARTHKLPEFPGRLLDDEIDSRLDI
PESEKRSLATLAKNDDLPVTIQERDDDDGDDDEEKRSIVPRHDTDEIMRSLWDEYRRRMSGLYDYLDAD
LDPEVDEYPLDKRNVGSLARDFALPTGRRHIGSVARDYGLPSGKRNI GSLARQRMPLPGKRNVASLAR
YSNLPTGKRNVAALARDSSLPYGKRYLGSLVRNGYPVRYDSDSKRNVASLARNADWPDFMKRGGGAPG
RMMVRVLRHGRSLSDDHETRSDDLQQLIRQGHNDGNDAKANVAGNWQLAPFIVAEEELDESKAKNRS
NRRIDAATSQTRHKRQIDFTDEYALPVVQNTNMLDYEDMLEALADHYPNAEKRFMGK

>Cflo_NPLP1 (EFN73979.1)
...NHDDMVHLKALSSEENEDFEICIPGGTYLELFRNPVLRSHLSAMERSYKFPGRFLSEEVDSDRLIPESE
EKRSLATLAKNGDLPITIQERESDNDDEEKRSASSSDNVGSLHELFEVSYRRRTPEIYDYLTEQDPEV
LEYSPLDKRNVGSLARDFALPTGRRHIASVARDHGLPSGKRNVGSLARQSMPLPSGKRNVASLARYYMLP
QSGKRNVAALARDSSLPYGKRYLGSLARSYPTRDYDEGKRSIASLARSQWPSVAKRGRMTSGRIMA
RVLNRRYGRSLSDREAPSEPLDLQQLIRQGNSEGKENEWQATPFTVSEDLDEGKAKNRSNRRIEASQT
RHKRQIDFSDEYPLPVMQNNMLDYEDMMEA IADHYPNAEKRFMGQTVSQPPLRVFRRVLFYLDVMSTV
FCRTCQVPEPHLLQTEKIYHIYVCYQQIYQPHHVKDNPNNTNEPTTMYNHVYQKCSMPSSSSTHGIIT
VVRENPADGSSSNYTPDASTRSLRPIFTPKTRYLQSLHGDCRHGFKRFLLPDIDNFLRTSNSHIAPRS
M

>Hsal_NPLP1 (EFN87936.1)
...KALSSEESDDAEICIPSGVYLELFRDPALRGHLSVMARTHKLPEFPGRLLDDSEEI EPRDVI SELEKRS
SLATLAKNDDLPVTIQERRSDRDDEEKRSHPSDNRGALISALLAEYRRRVPEYYDFLSELDPEVDEYT
LDKRN VGLARDFALPTGRRHIASVARDYGLPSGKRNI GALARQSMPLPLNGKRNVASLARYYMLPQADK
RNVAALARESSLPYGKRYLGSLARSYPTRDYDEGKR SVASLARNADWPAFAKRGSTPGRIMATMRV
LNRHGRSLNDDREARNNEPLDLQQLIRQGHGDEEAKEDVAEWQTPPFVLDLELDTKAKNRSNRKMDV
ASQTRHKRQIDFSDEYPLPVVQNAVFDYEDMIDAFADHYPNAEKRFMACCELWTLSGTGSVPMPADS
GYPEVLRTKKEAVFRDITKARMMNFREN PADGSTSHYTSDTSTRSLRPLFTPKTRYLQSLHGDCRHGFK
RFLLPD VDDFLRPNRSRIAPRSM

>Aech orcokinin (EGI63916.1)
...NFDEIDRSVDFHFSKRNIDEIDTAFDSLKRNFEIDRAGWGGFKRLNNYL TDR

>Sinv orcokinin (EFZ16485.1)
 ...NFDEIDRSPFDHFSKRNIIDEIDTAFDSFFKRFNFDEIDRASWDGFVKR FNNYLADRQR

>Cflo orcokinin (EFN74772.1)
 ...NFDEIDRSVDFRFPKRNIIDEIDTAFDSFFKRFNFDEIDRVGWNGFVKR LNNYLADRQR

>Hsal orcokinin (EFN80782.1)
 ...NFDENDRSMFDRFSKRNIIDEIDTAFDSFFKRFNFDEIDRVGWSGFVKR LDNYLADRQR

>Aech PDF (EG158539.1)
 MANYVRLAVIVAIIGIVRGESNELEIDRNILKLNLPYGRRLDNELELVRLMLVAPRLCHPKRNSEL
 INSLGLPKNMHNACKRK

>Sinv PDF (EFZ20211.1)
 MANYVRFVVFVFAIIGMVCGESSEIEDVDKNI FLNLPYGRGLDSELQVLKMLLAPRLCHPKRNSELI
 NSLLGLPKNMHNACK

>Cflo PDF (EFN61958.1)
 MANYITIAIIVGIVCGQALSVEDVDRNLELNLNLPYGRGLDSELQVLARLMLAAPRFCHPKRNSELINSL
 GLPKNMHNACK

>Hsal PDF (EFN79783.1)
 MANYARCGVAVVIVLGLFCGQVFG ELEDADRNVQDMSIRFGRGLDSELQVLARLLLAPRFCHPKRNSEI
 INSLGLPKNMHNACK

>Aech PK (EG164203.1)
 MIVTGNFVRRTTTTICILAMLLCLGSRASGEYEAREMGLNGGSNDGSCIEGKCMKR TTQDITSGMWFGR
 LGKRRR SDEKQEV SPEIEVLANALDGVRAWITIPANDKRQPSQFTPRLGRGSEDLPSYGDSHEVDED
 DRLLPPIFTARLGRFPWIPSPRLGRQLRNVLRKL

>Sinv PK (ACL35348.1)
 MIVTRNSVNRATIVCIMAMLLCLGSRASGEYESREIGSNGGSSESRSPSNDFGSCIDGKCIKR TSQDIA
 SGMWFGPRLGKRYKSDEKQELSSEIEILANALDGVRAWITIPASDKRQPQFTPRLGRGSGEDLSYGDA
 YEVEDDDHPLFVPRLGRRLPWIPSPRLGRQLRNVLRKL

>Cflo PK (EFN67760.1)
 MIVTGNPVCAIALLLCLVFRASGEYELEMSSGGSNDGRSPSNDFGSCIDGKCTKR TTTTQESGISSGMW
 FGPRLGKRHKSNEKQQINPEIEMLVNALDQPMRWTVITIPANEKRQPTQFTPRLGRGSEEKFIYSDAT
 DRNEIDEDDPLFTPRLGRVPWIPSPRLGRQSRVSRKI

>Hsal PK (EFN90009.1)
 ...MSNGESSDRLPGNFGSCTDGKCIKR TTQBITSGMWFGRGRRRSGERPEVGSPPDIEALANALDG
 ARWTLVKIPANDKRQPTQFTPRLGRGSEDLFSYAPGDALDRSEENDHPSPLFAPRLGRRLPWAPSPR
 LGRQLRNALRKM

>Aech PTTH (EG163917.1)
 ...VMIYGLSAESTEEVLSLGRWKEQVIAPEFLLNSREDISESSRDAFLFEDNFRPKGLNEIKR ITNTENV
 GIRLQPRLVTRSLQCTCETQYEIRDLGDGHYPRYLTTSSCVPKACLKFNKSCRLLYIVVLSQRELS
 LNDDRYSHSDSLQETPLPEALRHKWQLKPMKIPVACVSMG

>Sinv PTTH (EFZ16482.1)
 ...VMIHGLLAEGTTEEVLSIGGRWVDQVIGPEFLLDNREDVSEDNRNTFLYEDKRTFRPEGLSEVKR ITGA
 EDIGLQPRLVTRSLSCICETEYEIRDLGEGRYPRYLTTSSCKPKACLKFNKSCRLLYYMHVLSQRDL
 SELSDDRYSDNNEFPETLLPETLRHKWQLKPMRIPVACVSA

>Cflo PTTH (EFN74770.1)
 MKLLIILCVMVHGLLAEGPGQVLWKEQVVAPEFLLDDREDIASNRNAFFYEDKRSFRPEGLGEQVKRIAG
 AEDVGLQPRLVTRSLQCTCETEYERNLGEGHYPRYLTTSHCKPKACQNKFNKSCRLLYYKVVHLSQRDL
 NGLSDDRYSDSETETPLPEALRHKWQLKPMKIPVACVPATG

>Hsal PTTH (AEAC01016543.1: 18032-18271 +, AEAC01016544.1: 484-618,
 4332-4481 +)
 ...VMVHGLSAEPPGQVLSIGRWKEQVVPDFLLGDEHEEIPGGSRSAYGEKRNYPPEGLAEQVKR IAGA
 EDVGLQPRLVTRSLQCTCETEYEMRDLGEGHYPRYLTTASRCKPKACQSKFNKSCRLLYYHVLVLSQRDLGGLN
 DEHYSEENGLPEAPLPEALRHKWQLKPMKIPVACVPAT

>Aech SIFa (AEVX01013385.1: 198896-199024, 199440-199529 +)
MVSIRLILTLAVIMVAFVISANAAYKKPPFNGSIFGKR SNTVTDYEVTSRALSAICEVASETCTAWLSH
QESN

>Sinv SIFa (AEAQ01008053.1: 44860-44988 +, AEAQ01008054.1: 227-316 +)
MASIRLILTLAVIMVVFVISANAAYKKPPFNGSIFGKR SSTVTDYEITSRALSAICEVASETCTAWLSH
QDSN

>Cflo SIFa (AEAB01024202.1: 4014-4103, 4580-4708 -)
MVSIRLTFALAIVAIIFAFSVDAAYKKPPFNGSIFGKR SNTMTDYEFTSRALSAICEVASETCTAWMSR
QESN

>Hsal SIFa (AEAC01018952.1: 15958-16047, 16520-16648 -)
MVSVRALVLAVVMVILALSVDAAFRKPPFNGSIFGKR SSTVTDYDVTSRSLSTICEVASETCTAWLLR
QESN

>Aech SK (EGI60712.1)
MNLTLALSCMIAVMWLF CGRCKA APGLITGIHRNLGKHPHIRGYAVEGLVKDLVEDLVGEDEDSLRLSK
RQQLDDYGHMRF GKR FNGDDEEYGHSLR GNIE

>Sinv SK (EFZ19004.1)
MNLILGLSCMMAVMWLF CGRCKA APGFTGIHRSLGNRPHIRGYAVEGLVKDLVEDLVGEDEDILRLSK
RQQLDDYGHMRF GKR SNGDDEEF GHPRF GNIE

>Cflo SK (EFN67877.1)
MNLTFALSCMIAVVWLF CGKCKA APRLIPSIHRNLGNPHIYGYAVEGLVKDLVEDLVGEDEDNRLSK
RQQLDDYGHMRF GKR SNEDEEYGHLLRF GNIEK

>Hsal SK (EFN83738.1)
MNLTLALACMIAIMSLLCGRCKA APGRLIASRHNLENHPHIHGHAMEELMRDFLEDLVGEDEDGLRLSK
RQHLGDYDHMRF GKRRANDDDEYGRSRFS R NTE

>Aech sNPF (EGI59536.1)
MYAKYCAAFILVVVIVGLVDA TENYMDYGEEMA EKAPAENIHELKLLLQRNALDNVFGGIPLEHLMI
RKSQRSPSLRLRFGR SGQHISAGALPRPLGAVAAAGYDDNN

>Sinv sNPF (AEAQ01008190.1: 9662-9712, 9774-9959, 15183-15278 -)
MYAKRCAAFVLFVVVIVGLVDA TENYVDYGEEMA EKAPAENIHELKLLLQRNTLDNGFGDIPPOQL
ILRKSLR SALAAGHLRYGR SGSQFSGFYCLRMFKGKKKHKNS

>Cflo sNPF (EFN66516.1)
MSAMYAKRCAALVLLVVTVGLVNA TENYMDYGEEMA EKTPAENIHELKLLLQRNTLDNAGFGGIPLEH
LMIRKSQRSPSLRLRFGR SGPHVSARALPRPMGAVAGYDDNN

>Hsal sNPF (EFN85447.1)
MDARRITGLLAFVIVGLASG VENYMDYSDEPTDKTMDWINELYRLLMQRRAPEDNFFGTPLEHIMTRK
SQRSPSLRLRFGR SGPHTLGILPRPMGGPEAGYDDKN

>Aech TK (EGI65606.1)
MLFSSVLF LAVVTSSSFA EESSNDAASA KRAPMGFQMRGKK DLIPTIAEHNELSKRTLNVFQDKDSSA
SEIEDNLLHDEFD KRAPMGFQMRGKK DYLIPTDFEDSYFLEDYN KRAPMGFQMRGKK AILEDEYKRA
PMGFQMRGKK SLEEMLSIEIKRAALGFYGTGRKK TYVFEYPODYEKRLLEMDQDMHNKIKEEWEKRA
PMGFQMRGKK ALYDEIEELEKRALMGFQMRGKK DGFENYIDYIDDPDMDFD KRAPMSFQMRGKKD
TDKRAPMGFQMRGKR SVGQRFDPNMNFGSLNEYQGRSNRRHILASCQVE KRSPFRYFEMRGKK NPRWE
LRGMFVGVGRGKK WANAPYEDNSPFIRVLDNTERIGMDGDSPTTLGNRLADSQLISNIYVAVLFITSFIF
HVYLMAKYACSIYQVSSNLSERRERKINFFSFHTNVQNIFFSAFYMQCIIFKNLC

>Sinv TK (EFZ11618.1)
MLISSIFFLVVVTSSSFA EESSSSDAASA KRAPMGFQMRGKK DLIPTIAEHNELSKRTLNVFQDKDSS
ASDIEDNLLRDELDKRAPMGFQMRGKK DYLIPTDFEDSLREDYE KRAPMGFQMRGKK IIVKLQVLSDI
EKRAAMGFYGTGRKK TYVFEYPODYEKRLLEMDQDMHDMKEFPAEWE KRALMGFQMRGKK TLLDEE
LEKRALMGFQMRGKK DGYENYVDYVDPDMDFD KRAPMGFQMRGKK DTDKRAPMGFQMRGKR SVGQ
RFEPGMDFGTRPFHEYQGMSSDRRALASCQVE KRSPFRYDMRGKK NPRWELRGMFVGVGRGKK WARAPY
EDNSPFIRLFDNTERIGVDGDSPTLGN

>Cflo TK (EFN66667.1)

MLINSVLFLAIWSSSFAEESSSSDATSNKRAAMGFQDMRGKNLIPTSLEHNKLSKRTLMDFQDNKDSN
APDIEDNLSHEFEKRAPMGFQGMRGKKGYLTPDFEDSYFRDEKRAPMGFQGMRGKKNVSDDDYYKRAPM
GFQGMRGKKSLEEVLGEIEKKAAMDYYDTRDKKTYVFEYPEDYEKRLLASIRGKLKEFPMEWEKRAPMG
FQGMRGKKSLLDEIEELEKRTIMGFQGMRGKKNALENYIDYLDPMDFDKRAPMGFQGMRGKKSDDSKR
APMGFQGMRGKRNTRQRFDTGINFNIRSSNEYQGTNRRNALASCQLEKRSPPRYFEMRGKKNPRWELR
GMFVGVRGKKNWATAPPYEDDSPFISVFDNTERIGVDGDS PAILGNSIS

Table S18. **The variable set of neuropeptide genes in ants compared to other insects.** The variable set of neuropeptide genes found in ants (the four ants from Table S13 are merged into one column), *Apis mellifera* (honeybee), *Nasonia vitripennis* (parasitic wasp), *Drosophila melanogaster* (fruit fly), *Aedes aegypti* (yellow fever mosquito), *Bombyx morix* (silkmoth), *Tribolium castaneum* (flour beetle) and *Acyrtosiphon pisum* (pea aphid). For neuropeptide abbreviations see Tables S14 and S15. Modified from Hauser et al. 2010, with permission.

neuropeptide	Ants	<i>Apis</i>	<i>Nasonia</i>	<i>Drosophila</i>	<i>Aedes</i>	<i>Bombyx</i>	<i>Tribolium</i>	<i>Acyrtosiphon</i>
ACP	no	no	yes	no	yes	yes	yes	no
ADF-b	no	no	no	no	no	no	yes	no
Allatotropin	no	no	no	no	yes	yes	yes	yes
AST-A	yes	yes	yes	yes	yes	yes	no	yes
AST-B	no	no	no	yes	yes	yes	yes	yes
Capa	no	yes	no	yes	yes	yes	yes	yes
Corazonin	yes	yes	yes	yes	yes	yes	no	no
FMRFa	no	yes	no	yes	yes	yes	yes	yes
GPA2	no	no	no	yes	yes	yes	yes	yes
GPB5	no	no	no	yes	yes	yes	yes	yes
ILP-A	no	no	no	yes	yes	no	yes	no
ILP-C	yes	yes	yes	yes	yes	no	yes	yes
Inotocin	yes	no	yes	no	no	no	yes	no
Kinin	no	no	no	yes	yes	yes	no	yes
Neuroparsin	yes	yes	yes	no	yes	yes	yes	no
NPF	yes	yes	yes	yes	yes	yes	no	yes
NPLP-1	yes	yes	no	yes	yes	yes	yes	yes
NPLP-2	no	yes	no	yes	no	no	no	no
NPLP-3	no	yes	no	yes	no	no	no	no
NPLP-4	no	no	no	yes	no	no	no	no
Orcokinin	yes	yes	yes	no	yes	yes	no	yes
PDF	yes	yes	yes	yes	yes	yes	no	no
Proctolin	no	no	no	yes	no	no	yes	yes
PTTH	yes	no	yes	yes	yes	yes	yes	no
RYamide	no	yes	yes	yes	yes	yes	yes	yes
Sex peptide	no	no	no	yes	no	no	no	no
Sulfakinin	yes	yes	no	yes	yes	yes	yes	no