Genomics, transcriptomics, and peptidomics of neuropeptides and protein hormones in the red flour beetle *Tribolium castaneum*

Bin Li,[1] Reinhard Predel,[2] Susanne Neupert,[2] Frank Hauser,[3] Yoshiaki Tanaka,[4] Giuseppe Cazzamali,[3] Michael Williamson,[3] Yasuyuki Arakane,[5] Peter Verleyen,[6] Liliane Schoofs,[6] Joachim Schachtner,[7] Cornelis J.P. Grimmelikhuijzen,[3] and Yoonseong Park[1,8]

[1]*Department of Entomology, Kansas State University, Manhattan, Kansas 66506-4004, USA;* [2]*Institute of Zoology, University of Jena, D-07743 Jena, Germany;* [3]*Center for Functional and Comparative Insect Genomics and Department of Cell Biology and Comparative Zoology, Institute of Biology, University of Copenhagen, DK-2100 Copenhagen, Denmark;* [4]*National Institute of Agrobiological Science, Division of Insect Science, Tsukuba, Ibaraki 305-8634, Japan;* [5]*Department of Biochemistry, Kansas State University, Manhattan, Kansas 66506-4004, USA;* [6]*Department of Animal Physiology and Neurobiology, University of Leuven, BE-3000 Leuven, Belgium;* [7]*Department of Animal Physiology, University of Marburg, D-35032 Marburg, Germany*

Neuropeptides and protein hormones are ancient molecules that mediate cell-to-cell communication. The whole genome sequence from the red flour beetle *Tribolium castaneum*, along with those from other insect species, provides an opportunity to study the evolution of the genes encoding neuropeptide and protein hormones. We identified 41 of these genes in the *Tribolium* genome by using a combination of bioinformatic and peptidomic approaches. These genes encode >80 mature neuropeptides and protein hormones, 49 peptides of which were experimentally identified by peptidomics of the central nervous system and other neuroendocrine organs. Twenty-three genes have orthologs in *Drosophila melanogaster*. Sixteen genes in five different groups are likely the result of recent gene expansions during beetle evolution. These five groups contain peptides related to antidiuretic factor-b (ADF-b), CRF-like diuretic hormone (DH37 and DH47 of *Tribolium*), adipokinetic hormone (AKH), eclosion hormone, and insulin-like peptide. In addition, we found a gene encoding an arginine-vasopressin-like (AVPL) peptide and one for its receptor. Both genes occur only in *Tribolium* and not in other holometabolous insects with a sequenced genome. The presence of many additional osmoregulatory peptides in *Tribolium* agrees well with its ability to live in very dry surroundings. In contrast to these extra genes, there are at least nine neuropeptide genes missing in *Tribolium*, including the genes encoding the prepropeptides for corazonin, kinin, and allatostatin-A. The cognate receptor genes for these three peptides also appear to be absent in the *Tribolium* genome. Our analysis of *Tribolium* indicates that, during insect evolution, genes for neuropeptides and protein hormones are often duplicated or lost.

[Supplemental material is available online at www.genome.org.]

Multicellular organisms use signaling molecules for cell-to-cell communication. Important among these signaling molecules are peptides and protein hormones which are produced in endocrine cells or neurons as larger precursors. These precursors (prepropeptides) are cleaved and further modified to yield mature peptides that are secreted into the extracellular environment. Peptides exert their action by binding to membrane receptors, mostly being G-protein coupled receptors (GPCRs), although some of them are receptor tyrosine kinases.

Studies of a number of insect species have provided invaluable information for understanding the function and the evolution of neuropeptides. Earlier studies on insect neuropeptides have used large physiological model species (i.e., locust, cockroach, and moth), and these have provided the groundwork for identifying the active signaling molecules. Further characterization of the functions of neuropeptides has been provided by re-

cent genetic studies in *Drosophila melanogaster*, examining the genetic null mutants and cell ablations of specific peptidergic cells (McNabb et al. 1997; Park et al. 2002a, 2003; Kim and Rulifson 2004; Isabel et al. 2005; Kim et al. 2006).

The earliest traceable representatives of ancestral neuropeptides and endocrine protein hormones date back to the most primitive metazoans, such as cnidarians (Grimmelikhuijzen et al. 2002). Furthermore, various neuropeptide signaling systems found in different metazoan taxa show considerable similarities, suggesting a common ancestor. For example, the neuropeptide arginine vasopressin (AVP) has in the course of evolution been conserved in both vertebrates (including mammals) and invertebrates including insects (as shown in this study), suggesting a common ancestor before the split of proto- and deuterostomia. Insulin and insulin-like growth factors are also highly conserved hormones in animals (Wu and Brown 2006).

Ortholog neuropeptide genes show a high degree of divergence in their overall amino acid sequences while only small portions of the genes have been highly conserved, namely, those regions coding for mature peptides or even only the motif within

the peptide sequence that is required for biological activity (Liu et al. 2006a). Understanding of the evolution of the peptidergic signaling molecules has thus been hampered by the discontinuity of available data on evolutionarily related taxa, as well as by incomplete data within a particular species.

Another powerful approach to gain insight in the evolution of peptidergic signaling systems is based on the phylogeny of the cognate receptors for the neuropeptide ligands. G-protein coupled receptors are large transmembrane proteins that have been well conserved and that carry more informative sequences allowing the evolutionary analysis of the respective genes. This phylogenetic analysis, which of course assumes that ligands and cognate receptors co-evolve, allows evolutionary grouping of insect neuropeptides (Park et al. 2002b; Hauser et al. 2006a,b, 2007).

New techniques have recently emerged to identify the genes encoding neuropeptides. These new approaches include bioinformatic tools to predict the genes encoding neuropeptides from whole genome sequences and expressed sequence tag (EST) libraries, and direct detection of the processed mature peptides, using mass spectrometry (MS), also known as peptidomics (Hewes and Taghert 2001; Baggerman et al. 2002; Riehle et al. 2002; Predel et al. 2004; Hummon et al. 2006; Liu et al. 2006b). Upon identification of putative peptide encoding genes from the genome sequence, the transcription of these predicted genes can be confirmed by reverse-transcription PCR (RT-PCR) and by EST data. The amino acid sequence of the mature peptides obtained by peptidomic approaches can be compared with the predicted sequences, thereby revealing the mature peptide after post-translational processing of the prepropeptide. As such, genomics, transcriptomics, and peptidomics approaches mutually support and supplement each other and lay the foundation for further functional investigations of the neuropeptides.

Genomic studies in the fruit fly, malaria mosquito, and honeybee have been highly successful and have identified 31, 32, and 36 neuropeptide genes, respectively (Hewes and Taghert 2001; Riehle et al. 2002; Hummon et al. 2006; Predel and Neupert 2007). These initial genomics surveys have been supported and elaborated by peptidomics data (Baggerman et al. 2002; Hummon et al. 2006; Predel and Neupert 2007). In the current paper, we have made a large collaborative effort to survey *Tribolium* neuropeptides at the genomic, transcriptomic, and peptidomic levels. In addition to providing large data sets on *Tribolium* neuropeptides, the present paper also gives important insights on the evolution of insect neuropeptides.

Results and Discussion

The current version of the genome sequence for *Tribolium* covers >95% of the total genome with >16,500 genes in the computerized annotation (Tribolium Genome Sequencing Consortium, in prep.). Homology-based searches of neuropeptide and protein hormone genes identified 41 genes encoding >80 mature peptides carrying the typical signatures of neuropeptides. *Tribolium* EST database searches and RT-PCR on a subset of genes confirmed the existence of transcripts for 20 of these genes (Table 1; Supplemental Data 1). By using different mass spectrometric methods, the presence of 71 mature/processed peptides (with truncated and repetitive sequences), derived from 20 genes, was detected (Supplemental Data 2). Altogether, the expression of 30 neuropeptide genes was confirmed using these methods. Table 1 also shows the putative cognate *Tribolium* GPCRs for these peptides

(Hauser et al. 2007) as well as the homologous neuropeptides from other holometabolous insects.

Mass spectrometric analyses

In insects, the majority of the known bioactive neuropeptides contain a carboxy-terminal amide group (derived from a glycine residue preceding the cleavage sites in the precursor) which is important for their bioactivity and biostability (Eipper et al. 1993). In this context, it is noteworthy that our mass spectrometric analyses of brain and other nervous tissues revealed the expression of all predicted amidated neuropeptides from the genes encoding FMRFamides, tachykinin-related peptides (TKRPs), myosuppressin, allatostatin-B, short neuropeptide F, diuretic hormones (DH31 and DH47), adipokinetic hormone 2, SIFamide, ecdysis triggering hormone, and sulfakinins (Supplemental Data 1 and 2). In some cases, small peaks corresponding to truncated neuropeptides or other cleavage products from these prepropeptides were observed.

There are two genes in the upper group of Table 1, AKH-1 and AKH-3, for which we could not detect the predicted amidated neuropeptides by MS. Furthermore, a putative pyrokinin with the C-terminal WFGPRLamide sequence (also named CAPA-pyrokinin, which is pyrokinin encoded by the capa-gene [Predel and Wegener 2006]; or pyrokinin-1, which is pyrokinin binding to the pyrokinin-1 receptor [Cazzamali et al. 2005]), was one of the peptides not detected in the mass spectrometric analyses (Fig. 1), whereas this peptide is present in nearly all other insects that were studied. The *Tribolium* pyrokinin gene encodes a second pyrokinin with the WFGPRLamide motif (Supplemental Data 1) that was also not detectable in our mass spectrometric analysis.

Conserved and orthologous neuropeptide genes

Most *Tribolium* neuropeptides and protein hormone sequences have clear counterparts in other insect species (Table 1), which can be identified by conserved sequence motifs, implying that parts of their genes may be under strong evolutionary pressure to be conserved.

Also the structure of the prepropeptide is often highly conserved in terms of the location of the immature peptides. For example, in the prepropeptide of adipokinetic hormone, proctolin, and SIFamide (Supplemental Data 1), the short neuropeptide sequences (5–12 amino acid residues long) are located immediately after the N-terminal signal peptide and are tailed by longer associated peptides (~40–120 amino acid residues long). In the case of ecdysis triggering hormone (ETH), the two peptides are consecutively located after the signal peptide (Supplemental Data 1). This *ETH* gene structure is conserved in other insects as well (Park et al. 1999; Zitnan et al. 2003).

Multiple repeated similar peptide sequences separated by cleavage sites are typical of other prepropeptides, including FMRFamide, allatostatin-B, CAPA, pyrokinin, and TKRPs with six, six, four, five, and eight repeats, respectively, in *Tribolium* (Supplemental Data 1). The numbers of repeats among the orthologs in different insect species may vary. For example, the allatostatin-B (also referred to as myoinhibitory peptide MIP), has six repeats in *Tribolium* but five, four, and thirteen repeats in *D. melanogaster*, *Anopheles gambiae*, and *Bombyx mori*, respectively.

The conservation of neuropeptide gene structure is extended to the conservation of alternative splicing patterns. For the ion transport peptide and the DH genes, for example, mutu-

Table 1. Neuropeptide and protein hormones identified by searching genome sequences of four holometabolous insects

Name	T. castaneum			D. melanogaster[d]	A. gambiae[e]	A. mellifera[f]
	GLEAN no.	Putative receptor[a]	mRNA[b]/peptide[c]			
Neuropeptides (alphabetical order)						
Adipokinetic hormone 1 (AKH)	GLEAN_16057	Tc46/47	√ / −	AKH-like (CG1171)	XP_001238167	GB30028
AKH-2	AKH-2	Tc46/47	√ / √	AKH-like (CG1171)	XP_001238167	GB30028
AKH-3	AKH-3	Tc46/47	—		XP_563757	
Allatostatin B (myoinhibitory peptide, MIP)	GLEAN_07661	Tc54	√ / √	Myoinhibiting peptide	XP_316799	ND
Allatostatin C	GLEAN_05428	Tc52	− / (√)	Allatostatin-C	XP_001238143	ND
Allatotropin	GLEAN_06446		√ / √	ND	AgamAT1, 2	ND
Calcitonin-like diuretic hormone (DH31)	GLEAN_04987	Tc70	√ / √	Diuretic hormone 31	XP_321755.2	GB17064
CAPA (periviscerokinin, PVK)	GLEAN_08429	Tc33	− / √	CG15520	XP_566030	GB30386
Crustacean cardioactive peptide (CCAP)	GLEAN_01088	Tc43/44	√ / −	CCAP	XP_318812	GB18160
Ecdysis triggering hormone (ETH)	GLEAN_09466	Tc29	√ / √	CG18105	XP_308702	GB15713
FMRFamide	GLEAN_07769	Tc26	− / √	CG2346	XP_556154	GB30418
Myosuppressin (MS)	GLEAN_01469	Tc21	− / √	CG6440	XP_321650	GB14259
Proctolin	GLEAN_15137	Tc25	—	CG7105	AgPT	ND
Pyrokinin (PK)	GLEAN_04636	Tc30/31/32	− / √	CG6371	XP_307885	GB18468
Short neuropeptide F (sNPF)	sNPF	Tc57/58	− / √	CG13968	ABD96048	GB18048
SIFamide	GLEAN_08645	Tc38	− / √	SIFamide	XP_308708	GB12452
Sulfakinin (SK)	Sk	Tc34/35	− / √	CG18090	AAR03495	GB15923
Tachykinin-related peptide (TKRP)	GLEAN_05685	Tc40/41	− / √	CG14734	XP_319161	GB12290
Protein hormones (polypeptides)						
Bursicon-alpha subunit	GLEAN_13832	Tc48	√ / −	burs	bursicon	GB11959
Bursicon-beta subunit	GLEAN_13831	Tc48	√ / −	pburs	XP_313804	GB19117
Eclosion hormone (EH)	EH	—	√ / −	Eclosion hormone	XP_001230805	GB19466
Eclosion hormone-like peptide	GLEAN_00178	—	√ / −	Eclosion hormone	XP_001230805	GB19466
Glycoprotein hormone alpha-2 (GPA2)	GLEAN_11941	Tc49/50	√ / −	AAX38184.1	XP_317164	ND
Glycoprotein hormone beta-5 (GPB5)	GLEAN_12548	Tc49/50	—	CG40041	XP_555160	ND
Insulin-like peptide 1	GLEAN_08479		− / √	DmIlp 1–6	AgIlp 1–4, 6, 7	GB17332
Insulin-like peptide 2	GLEAN_07035		—	DmIlp 1–6	AgIlp 1–4, 6, 7	GB10174
Insulin-like peptide 3	GLEAN_00934		—	DmIlp 1–6	AgIlp 1–4, 6, 7	ND
Insulin-like peptide 4	Inlp4		—	Ilp7	AgIlp5	ND
Ion transport peptide (ITP)-a and -b	ITP-a and -b	—	√ / −	ITP	XP_313928	XP_001120062
Neuroparsin A	GLEAN_13258	—	—	ND	XP_311039	GB30220
Prothoracicotropic hormone (PTTH)	PTTH	—	—	Drm PTTH	XP_555854	ND
Novel osmoregulatory neuropeptides						
Arginine-vasopressin-like (AVPL)	GLEAN_06626	Tc45	√ / (√)	ND	ND	ND
Antidiuretic factor b-1 (ADF)	GLEAN_09955	—	—	ND	ND	ND
ADF b-2	GLEAN_09956	—	√ / −	ND	ND	ND
ADF b-3	GLEAN_09957	—	√ / −	ND	ND	ND
ADF b-4	GLEAN_09958	—	—	ND	ND	ND
ADF b-5	GLEAN_09959	—	—	ND	ND	ND
Corticotropin releasing factor like-DH (DH37/47)	GLEAN_02243/DH47	Tc65/66	√ / √	Diuretic hormone	XP_001230569	CAI45289.1
Other putative neuropeptide genes						
Apis-NVP-like	GLEAN_05571	—	√ / √	ND	ND	XP_397558.2
Apis-ITG-like	GLEAN_08660	—	√ / √	CG8216	XP_319743	GB14512
Neuropeptide-like precursor-1 (NPLP)	GLEAN_06787	—	√ / √	Neuropeptide-like precursor	XP_311578	GB12943
Genes not detected						
Pigment dispersing factor (PDF)		Tc69	—	PDF	XP_315791	GB30123
Neuropeptide F (NPF)		Tc59	—	Neuropeptide F	XP_315165	GB16364
Corazonin		—	—	Corazonin	XP_001238800	GB17245
Insect kinin		—	—	Leucokinin	Leucokinins-KIN	ND
Allatostatin-A		—	—	CG13633	XP_313511	GB30252
Orcokinin		—	—	ND	XP_320317	GB11752

[a]Data from Hauser et al. 2007.
[b]Evidence for transcriptions is either from EST or RT-PCR for partial or full-length mRNA.
[c]Detection of any peptide fragment for the annotated sequence is counted as positive for the gene. Parentheses in the evidence for peptides are the ones confirmed only by matching mass.
[d]*Drosophila melanogaster* gene names are given for the peptide name, CG number, or GenBank accession numbers when available.
[e]*Anopheles gambiae* gene names are given for GenBank Accession number or peptide name (Riehle et al. 2002) when available.
[f]*Apis mellifera* gene names are given for annotation number (Hummon et al. 2006) or GenBank accession number.
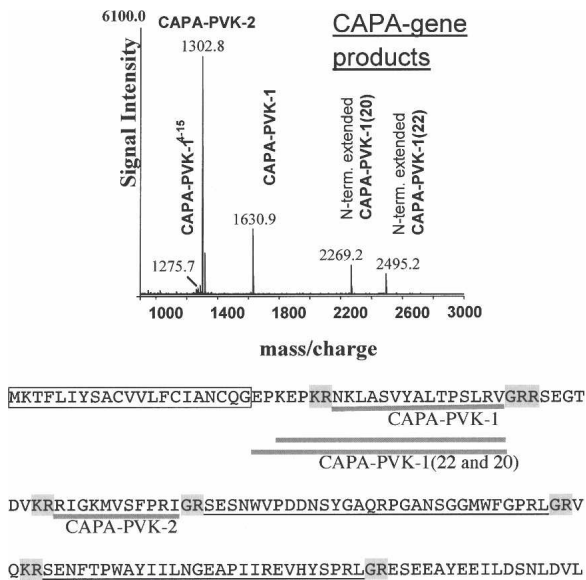
Figure 1. Representative MALDI-TOF mass spectrum of a single anterodorsal ganglionic sheath preparation of abdominal ganglion-2 (neurohemal release site of CAPA-peptides) of an adult *Tribolium*. The direct mass spectrometric screening of nervous tissues may provide reproducible semiquantitative data. Note the different abundances of the expressed *capa* peptides. CAPA-PVK-1, which is present with a number of truncated/extended forms, has a lower signal intensity than CAPA-PVK-2; CAPA-pyrokinin is not detectable at all. *Lower* panel shows the annotated sequence with thick gray lines for the peptides determined in the mass spectrometric analysis as shown in the *upper* panel. Underlines are for putative mature peptides containing C-terminal -PRX motifs, and the letters with gray backgrounds are canonical signals for amidations and/or basic cleavages. Putative signal peptide is boxed.

ally exclusive ways of alternative splicing are conserved in several insect orders (Dai et al. 2007).

Insulin-like peptides and neuroparsin

Four genes encoding insulin-like peptides (ILP) were identified in the *Tribolium* genome and assigned to three evolutionary groups: one to ILP-A, two to ILP-B, and one to ILP-C (Fig. 2).

The largest group, ILP-B, contains highly diverse genes including recently duplicated genes in mosquitoes (AgILP1 and AgILP7, and AgILP3 and AgILP6 in Fig. 2). Phylogenetic analysis within the group was not successful because of the highly diverged sequences resulting in a low statistical support for most of the clades. However, a strict conservation of the cysteine motif CCxxxC (also typical for vertebrate insulins) was found in the A chain. The ILP-B group includes the multiple copies of bombyxin-like peptides, which have originally been isolated from the silkworm brain (Nagasawa et al. 1986) based on their stimulating activity of adult development of debrained pupae (Nagasawa et al. 1980). Two closely related *Tribolium* ILP genes, TcILP1 and TcILP2, were identified and classified within this group.

By contrast, the ILP-A and -C groups are characterized by the CCxxxxC-motif, with four amino acids between the cysteine residues. All insect ILP-A members display high sequence homology, particularly of the A- and B-chain regions. The ILP-A group has not been identified in the honeybee nor in the silkworm genomes. The ILP-C group is further characterized by a very short or absent C-peptide between the A- and B-chains and a single putative mono- or dibasic cleavage signal between the A- and B-

chains (Fig. 2). The ILP-C group comprises mammalian insulin-like growth factors (i.e., IGF-1b of human P05019) and a large number of ILPs (named INS) in *Caenorhabditis elegans* (Duret et al. 1998; Li et al. 2003). *Tribolium* ILP3 and ILP4 are members of the ILP-A and ILP-C groups, respectively (Fig. 2).

Neuroparsins are structurally similar to the mosquito ovary ecdysteroidogenic hormone and to the vertebrate insulin-like growth factor binding proteins (Schoofs et al. 1997; Brown et al. 1998). We identified one copy of neuroparsin in the *Tribolium* genome (Fig. 3). In locusts, multiple isoforms (four or more) and multiple functions have been described for this family of proteins (Janssen et al. 2001; Claeys et al. 2003), including an antidiuretic function (Fournier et al. 1994), an involvement in phase transition (Claeys et al. 2005), and a function for neural growth (Vanhems et al. 1990). In the honeybee, a neuroparsin-like protein has been described as queen brain-selective protein-1 (GenBank accession no. Q1T786), implying a potential role of neuroparsin in cast determination, possibly by antagonizing the insulin pathway, similar to the function of vertebrate insulin-like growth factor binding proteins. The ancestral neuroparsin gene appears to have undergone considerable changes particularly in the lineage of higher dipteran insects, similar to the divergence observed in the ovary ecdysteroidogenic hormone in mosquitoes (Brown and Cao 2001). The neuroparsin gene is absent in *Drosophila*.

Novel osmoregulatory neuropeptide genes in *Tribolium*

The *Tribolium* genome revealed three groups of osmoregulatory neuropeptide genes: One gene encoding an arginine vasopressin-like peptide (AVPL), five genes encoding antidiuretic factors (ADFs), and one gene giving rise, by alternative splicing, to two corticotropin releasing factor (CRF)-like diuretic hormones DH37 and DH47.

The gene encoding AVPL is absent in all other holometabolous insects with a sequenced genome (Table 1). In addition, an AVPL GPCR has been identified in *Tribolium*, which is also not found in other holometabolous insects (Hauser et al. 2007), further supporting the presence of AVPL signaling pathways in *Tribolium* and their absence in the other mentioned insects. The AVPL gene has a remarkably conserved overall structure. The N-terminal signal peptide is followed by AVPL and immediately thereafter by a neurophysin-like sequence containing the conserved 14 cysteine amino acid residues (Fig. 4; Supplemental Data 1). In mammals, neurophysin is thought to play an essential role in the proper cellular processing of pro-AVP and to bind to the mature AVP peptide (de Bree and Burbach 1998). As the neurophysin sequence is conserved in *Tribolium* (Fig. 4), its function may be conserved as well.

The first evidence for an AVPL in insects was revealed by an antibody raised against vertebrate AVP, which stained neurosecretory cells in the CNS of locusts (Proux and Rougon-Rapuzzi 1980; Proux et al. 1987). This locust AVPL was later purified from neuroendocrine extracts, and its diuretic activity was shown by using the isolated locust Malpighian tubules in the Ramsey assay (Proux et al. 1987). This function was later questioned by an independent study (Coast et al. 1993). The absence of the *avpl* gene in the honeybee, which is the most ancestral order of the holometabolous insects (Savard et al. 2006), indicates that the *avpl* gene has probably been lost in other holometabolous branches at least two times. However, AVP-related peptides have been described in a number of other invertebrates (Fig. 4) and play a role in reproductive behavior and osmoregulation (Fujino
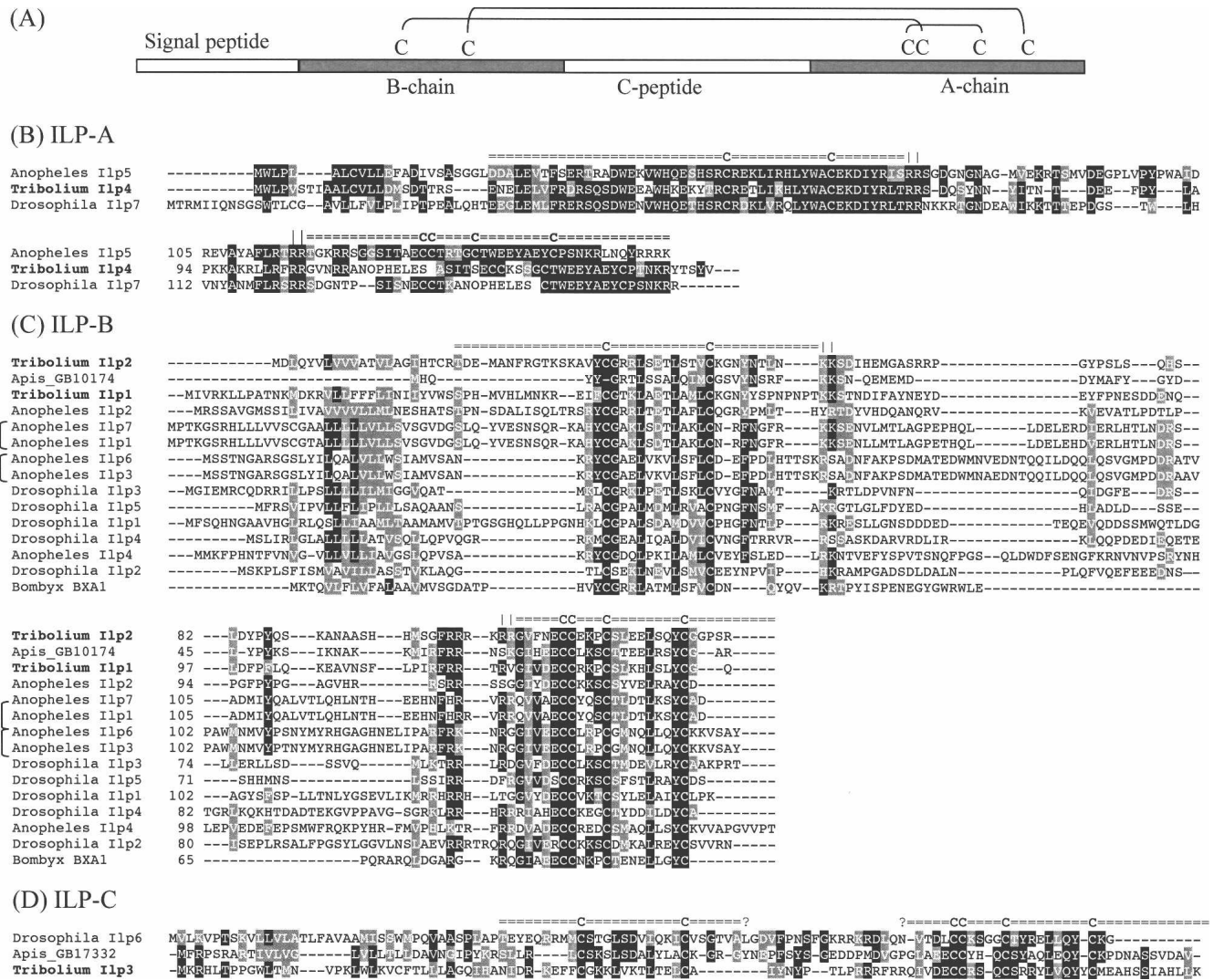
(A)

Signal peptide   C   C   CC   C   C
                B-chain   C-peptide   A-chain

(B) ILP-A

```
                              ===========================C=========C========||
Anopheles Ilp5    ----------MWLPI----ALCVLLEFADIVSASGGLDDALEVTFSERTRADWEKVWHQESHSRCREKLIRHLYWACEKDIYRISRRSGDGNGNAG-MVEKRTSMVDEGPLVPYPWAID
Tribolium Ilp4    ----------MWLPVSTIAALCVLLDMSDTTRS-----ENELELVPRDRSQSDWEEAWHKEKYIRCRETLIKHLYWACEKDIYRLTRRS-DQSYNN--YITN-T--DEE--FPY---LA
Drosophila Ilp7   MTRMIIQNSGSWTLCG--AVLLFVLPLIPTPEALQHTEEGLEMLFRERSQSDWENVWHQETHSRCRDKLVRQLYWACEKDIYRLTRRNKKRTGNDEAWIKKTTTEPDGS---TW---LH

                      ||============CC====C========C============
Anopheles Ilp5    105 REVAYAFLRTRRTGKRRSGGSITAECCTRTGCTWEEYAEYCPSNKRLNQYRRRK
Tribolium Ilp4     94 PKKAKRLLRFRRGVNRRANOPHELES ASITSECCKSSGCTWEEYAEYCPTNKRYTSYV---
Drosophila Ilp7   112 VNYANMFLRSRRSDGNTP--SISNECCTKANOPHELES CTWEEYAEYCPSNKRR-------
```

(C) ILP-B

```
                              ============================C=========C===========||
Tribolium Ilp2    ------------MDIQYVLVVVATVLAGIHTCRADE-MANFRGTKSKAVYCGRRLSETLSTVCKGNYNTLN-----KKSDIHEMGASRRP----------------GYPSLS---QIS--
Apis_GB10174      --------------------------HQ-------------------YY-GRTLSSALQIMCGSVYNSRF-----KKSN-QEMEMD------------------DYMAFY---GYD--
Tribolium Ilp1    --MIVRKLLPATNKYDKRVLLFFFLINIIYVWSSPH-MVHLMNKR--EIECGTKLAETLAMLCKGNIYSPNPNPTKKSTNDIFAYNEYD------EYFPNESDDENQ--
Anopheles Ilp2    ---MRSSAVGMSSIIIVAVVVVLMLNESHATSIPN-SDALISQLTRSRYCGRLTFTLAFLCQGRYPMLT---HYRTLYVHDQANQRV---------------KVEVATLPDTLP--
Anopheles Ilp7    MPTKGSRHLLLVVSCGAALLLLLVLLSVSGVDGSLQ-YVESNSQR-KAHYCGAKLSFTLAKLCN-RENGFR-----KKSENVLMTLAGPEPHQL------LDELERDIERLHTLNDFS--
Anopheles Ilp1    MPTKGSRHLLLVVSCGTALLLLLVLLSVSGVDGSLQ-YVESNSQR-KAHYCGAKLSFTLAKLCN-RENGFR-----KKSPNLLMTLAGPETHQL------LDELEHDVERLHTLNDFS--
Anopheles Ilp6    ----MSSTNGARSGSLYILQALVLLWSIAMVSAN------------KIYCCAELVKVLSFLCD-EFPDIHTTSKRSAPNFAKPSDMATEDWMNVEDNTQQILDQOIQSVGMPDDRATV
Anopheles Ilp3    ----MSSTNGARSGSLYILQALVLLWSIAMVSAN------------KIYCCAELVKVLSFLCD-EFPDIHTTSKRSAPNFAKPSDMATEDWMNAEDNTQQILDQOIQSVGMPDDRAAV
Drosophila Ilp3   --MGIEMRCQDRRIILPSLLLLILMIGGIQAT-------------MRLCGRTLPETLSKLCVYGFNAIT-----KRTLDPVNFN------------------QIDGFE---DIS--
Drosophila Ilp5   ----------MFRSVIPVLLFLIPLLLSAQAANS-------------LRACGPALMPLMRVACPNGPFNSIF---AIKRGTLGLFDYED-----------------HIPADLD---SSE--
Drosophila Ilp1   --MFSQHNGAAVHGIRLQSLLTAAMLTAAMAMVTPTGSGHQLLPPGNHKLCPALSPAMDVICPHGFNTIP----RKRESLLGNSDDDED-----------TEQEVQDDSSMWQTLDG
Drosophila Ilp4   ---------MSLIRIGLALLLLATVSQILQPVQGR-----------RKMCGEALIQALDVICVNGFTRRVR---RSSASKDARVRDLIR----------------KLQQPDEDIEQETE
Anopheles Ilp4    ---MMKFPHNTFVNIVG-VLLVLLIAIGSIQPVSA-----------KRYCGDQLPKILAMLCVEYESLED----LIKNTVEFYSPVTSNQFPGS--QLDWDFSENGFKRNVNVPSIYNH
Drosophila Ilp2   --------MSKPLSFISMIAVIILASSTYKLAQG-----------TLCSEKLNEVLSMVCEEYNPVIP----IKRAMPGADSDLDALN----------PLQFVQEFEEEDNS--
Bombyx BXA1       --------------MKTQVIFLVFAEAAIMVSGDATP----------HVYCGRRLATMLSFVCDN-----CYQV--KRTPYISPENEGYGWRWLE---------------------
```

```
                              ||=====CC===C========C============
Tribolium Ilp2    82  ---IDYPYQS---KANAASH--HISGFRRR---KRRGVFNECCEKPCSLEELSQYCCGPSR-----
Apis_GB10174      45  ---I-YPYKS---IKNAK-----KIIRFRR---NSKGIHEECCLKSCTTEELRSYCG---AR-----
Tribolium Ilp1    97  ---IDFPISLQ---KEAVNSF---LPIIFRR---TRVGIVDECCRKPCSIKHLSLYCG---Q-----
Anopheles Ilp2    94  ---PGFPYPG---AGVHR---------RSRR---SSGGIYDECCKKSCSYVELRAYCD---------
Anopheles Ilp7    105 ---ADMIYQALVTLQHLNTH---EEHNFHR---VRRQVVAECCYQSCTIDTLKSYCAD---------
Anopheles Ilp1    105 ---ADMIYQALVTLQHLNTH---EEHNFHR---VRRQVVAECCYQSCTIIDTLKSYCAD---------
Anopheles Ilp6    102 PAWINMVVYPSNYMYRHGAGHNELIPAFRK---NRGGIVEECCLRPCGWNQLLQYCKKVSAY----
Anopheles Ilp3    102 PAWINMVVYPTNYMYRHGAGHNELIPAFRK---NRGGIVEECCLRPCGWNQLLQYCKKVSAY----
Drosophila Ilp3   74  --LIERLLSD----SSVQ------MLKTRR---LRDGVFDECCLKSCTMDEVLRYCAAKPRT----
Drosophila Ilp5   71  ----SHHMNS-------------ISSIRR----DFRGVVDSCCRKSCSFSTLRAYCDS--------
Drosophila Ilp1   102 ---AGYSFSP-LLTNLYGSEVLIKLRFHRRH---LTGGVYDECCVKTCSYLELAIYCLPK-------
Drosophila Ilp4   82  TGRIKQKHTDADTEKGVPPAVG-SGRKLRR---HRRRIAHECCKEGCTYDDILDYCA---------
Anopheles Ilp4    98  LEPVEDETEPSMWFRQKPYHR-FMYPKLKTR--FRRDVAEECCREDCSMAQLLSYCKVVAPGVVPT
Drosophila Ilp2   80  ---ISEPLRSALFPGSYLGGVLNSIAEVRRRTRORQGTVERCCKKSCDAKALREYCSVVRN-----
Bombyx BXA1       65  ------------------PQRARQIDGARG--KRQGIAEECCNKPCTENELLGYC----------
```

(D) ILP-C

```
                              ==========C=========C=======?        ?=====CC====C========C===========
Drosophila Ilp6   MVLIKVPISKVLLVLATLFAVAAIISEWYPQVAASPIAPLEYEQIRMNCSTGLSDLVQKIGVSGTVALGDVFPNSFGKRRKRDLQN-VDDLCCKSGGCTYRELLQY-CKG-----------
Apis_GB17332      -MFRPSRARTIVLVG------LVLLKLIDAVNGIPYKRSILR----LCSKSLSDALYLACK-GR-GYNEPISYS-GEDDPMDVGPGIAPECCYH-QCSYAQLEQY-CKPDNASSVDAY-
Tribolium Ilp3    -MKRHLTPPGWLTYN--VPKIWLKVCFTLILAGOIHANIDR-KEFFCGKKLVKTLTELCA--------IMYNYP--TLPRRRFRRQIVDECCRS-QCSRRYLVQYYCMEAHSSIAHLIK
```

Figure 2. Structure of typical preproinsulin-like peptides (*A*) and sequence alignment of insulin-like peptide from *Tribolium castaneum*, *Apis mellifera*, *Anopheles gambiae*, *Drosophila melanogaster*, and one *Bombyx mori* Bombyxin (BBXA1, P33718) (*B–D*). Conserved cysteines are marked by "C" on the *top* of putative A- and B-chains (double lines, =) with canonical dibasic cleavage sites as vertical bars (|) and unknown cleavage sites as question marks (?). *Tribolium* and *Apis* sequences are from their genome database, *Drosophila* genes are from FlyBase, and *Anopheles* sequences are from Riehle et al. (2002).

et al. 1999; Kanda et al. 2005; Levoye et al. 2005); comparable functions have been attributed to vasopressin/oxytocin in mammals.

The other group of peptides, ADFs, was known only from a very closely related beetle, *Tenebrio molitor*, which, like *Tribolium*, also lives in very dry ecosystems. In *T. molitor*, two ADFs (ADFa and ADFb) inhibit diuretic activity of the Malpighian tubules (Eigenheer et al. 2002, 2003; Wiehart et al. 2002). Surprisingly, ADFs show high degrees of sequence identity to the C terminus

of a cuticle protein from *T. molitor* (Baernholdt and Anderson 1998). In the *Tribolium* genome, we identified five putative ADF-b genes, homologous to each other, each containing a highly conserved ADF-b at the C terminus of the putative gene products (Fig. 5). These five ADF-b genes are all clustered within 10 kb in linkage group 7 and show similar gene structures, suggesting that they arose by recent gene expansions. It is possible that genes encoding ADF-b only exist in drought-tolerant coleop-
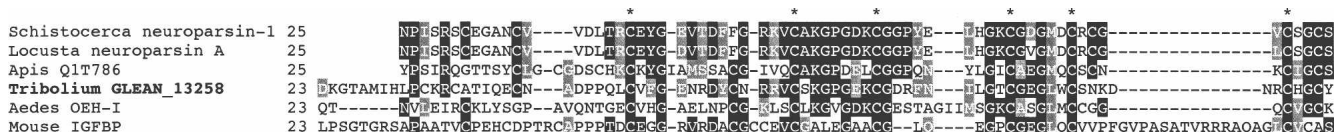
```
                                 *            *     *         *   *                    *
Schistocerca neuroparsin-1  25  NPISRSCEGANCV----VDLTRCEYG-EVTDRFG-RKVCAKGPGDKCGGPYE---IHGKCGDGMDCRCG----------------VCSGCS
Locusta neuroparsin A       25  NPISRSCEGANCV----VDLTRCEYG-DVTDRFG-RKVCAKGPGDKCGGPYE---LHGKCGVGMDCRCG----------------LCSGCS
Apis Q1T786                 25  YPSIRQGTTSYCIG-G-CGDSCHBombyxin IAMSSACG-IVQCAKGPDILCGGPQN---YLGICAEGKQCSCN----------------KCIGCS
Tribolium GLEAN_13258      23  DKGTAMIHLPCKRCATIQECN---ADPPQLCVFG-ENRDYCN-RRVCSKGPGEKCGDREN---ILGTCGEGIWCSNKD----------NRCHGCY
Aedes OEH-I                 23  QT------NVIEIRCKLYSGP---AVQNTGPCVHC-AELNPCG-KLSCLKGVGDKCGESTAGII SGKCASGLMCCGG----------QCVGCK
Mouse IGFBP                 23  LPSGTGRSAPAATVCPEHCDPTRCAPPPTDCEGG-RVRDACGCCEVCGALEGAACG--LE----EGPCGEGLQCVVPFGVPASATVRRRAQAGLCVCAS
```

Figure 3. Sequence alignment of insect neuroparsin and related sequences. The sequences are from *Schistocerca gregaria* neuroparsin precursor 1 (AC38869), *Locusta migratoria* neuroparsin A precursor (P10776), *Apis mellifera* queen brain-selective protein-1 (Q1T786), *Aedes aegypti* ovarian ecdysteroidogenic hormone I (AAD00823), and *Mus musculus* insulin-like growth factor binding protein 5 protease (Mm IGFBP). Strictly conserved cysteines are marked with asterisks (*) on the *top* of the sequence alignment.
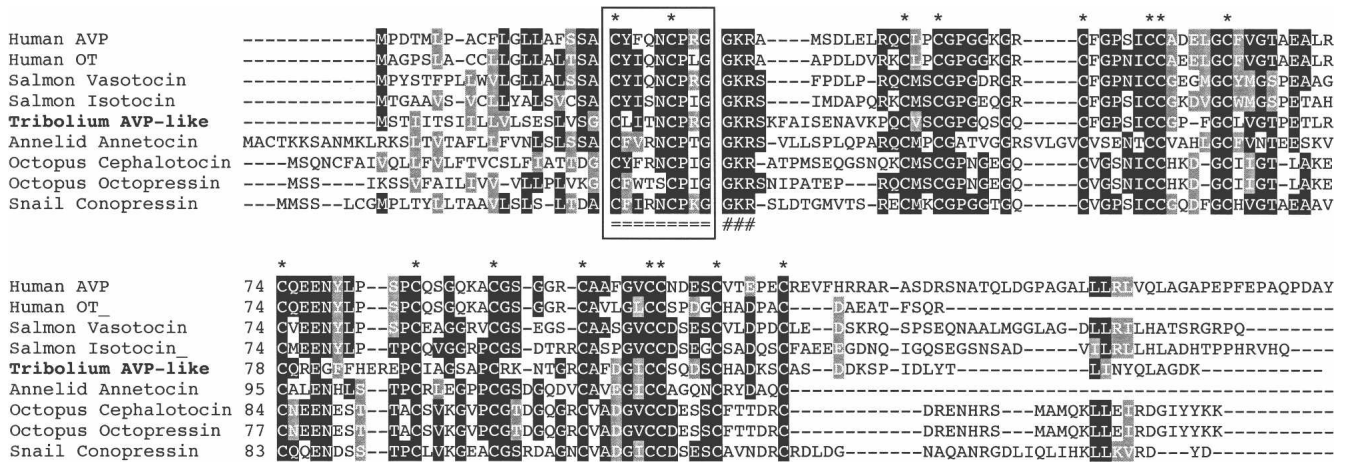
```
                        *         *                              *   *              *     **        *
Human AVP             -----------MPDTMIP-ACFLGLLAFSSA CYFQNCPRG GKRA----MSDLELRQCLPCGPGGKGR-----CFGPSICCADELGCFVGTAEALR
Human OT              -----------MAGPSIA-CCLLGLLALTSA CYIQNCPLG GKRA--APDLDVRKCLPCGPGGKGR-----CFGPNICCAEELGCFVGTAEALR
Salmon Vasotocin     -----------MPYSTFPLLWVLGLLALSSA CYIQNCPRG GKRS--FPDLP-RQCMSCGPGDRGR-----CFGPNICCGEGMGCYMGSPEAAG
Salmon Isotocin      -----------MTGAAVS-VCLLYALSVCSA CYISNCPIG GKRS----IMDAPVCMSCGPGEQGR-----CFGPSICCGKDVGCWNGSPETAH
Tribolium AVP-like   -----------MSTIITSIILLVLSESLVSG CLIINCPRG GKRSKFAISENAVKPQCVSCGPGQSGQ-----CFGPSICCGP-FGCLVGTPETLR
Annelid Annetocin    MACTKKSANMKLRKSLTVTAFLLFVNLSLSSA CFVRNCPTG GKRS-VLLSPLQPARQCMPCGATVGGRSVLGVCVSENTCCVAHLGCFVNTEESKV
Octopus Cephalotocin ----MSQNCFAIVQLIFYLFTVCSLFIATTDG CYFRNCPIG GKR-ATPMSEQGSNQKCMSCGPNCEGQ-----CVGSNICCHKD-GCIIGT-LAKE
Octopus Octopressin  ----MSS-----IKKSVFAILIVV-VLIPLVKG CFWTSCPIG GKRSNIPATEP---RQCMSCGPNGEGQ-----CVGSNICCHKD-GCIIGT-LAKE
Snail Conopressin    ---MMSS--LCGMPLTYLLTAAVLSLS-LTDA CFIRNCPKG GKR-SLDTGMVTS-RECMKCGPGGTGQ-----CVGPSICCQLFGCHVGTAEAAV
                                                     =========  ###
```

```
                        *           *       *         *   **       *
Human AVP            74 CQEENYLP--SPCQSGQKACGS-GGR-CAAFGVCCNDESCVTEPECREVFHRRAR-ASDRSNATQLDGPAGALLRLVQLAGAGAPEPFEPAQPDAY
Human OT             74 CQEENYLP--SPCQSGQKACGS-GGR-CAVLGLCCSPDGCHADPAC----DAEAT-FSQR-----------------------------------
Salmon Vasotocin    74 CVEENYLP--SPCEAGGRVCGS-EGS-CAASGVCCDSESCVLDPDCLE--DSKRQ-SPSEQNAALMGGLAG-DLLRLLHATSRGRPQ--------
Salmon Isotocin     74 CMEENYLP--TPCQVGGRPCGS-DTRR-CASPGVCCDSEGCSADQSCFAEEEGDNQ-IGQSEGSNSAD-----VILRLLHLADHTPPHRVHQ----
Tribolium AVP-like  78 CQREGLFHEREPCIAGSAPCRK-NTGR-CAFDGICCSQDSCHADKSCAS--DDKSP-IDLYT----------LINYQLAGDK-----------
Annelid Annetocin   95 CALENHLS--TPCRLEGPPCGSDGQDVCAVEGICCAGQNCRYDAQC---------------------------------------------
Octopus Cephalotocin 84 CNEENEST--TACSVKGVPCGTDGQGRCVADGVCCDESSCFTTDRC------------DRENHRS---MAMQKLLEIRDGIYYKK---------
Octopus Octopressin 77 CNEENEST--TACSVKGVPCGTDGQGRCVADGVCCDESSCFTTDRC------------DRENHRS---MAMQKLLEIRDGIYYKK---------
Snail Conopressin   83 CQQENDSS--TPCLVKGEACGSRDAGNCVADGICCDSESCAVNDRCRDLDG-------NAQANRGDLIQLIHKLLKVRD---YD---------
```

Figure 4. Sequence alignment for the arginine-vasopressin (AVP) and oxytocin prepropeptide family. Sequences in the box are for mature peptides with canonical amidation and dibasic sequences (#). Strictly conserved 16 cysteines are marked with asterisks (*) on the *top* of the sequence alignment. GenBank accession nos. for the sequences are AAB86629, AAL92860, BAA01736, BAD12146, ABN79655, BAA36458, BAC82436, BAC82435, and Q00945.

teran insects, since we could not find ADF-b in other holometabolous insects with a sequenced genome. We were unable to identify the ADF-a sequence in the *Tribolium* genome. The ADF peptides were not detected in the mass spectrometric analyses.

There are two diuretic hormone genes in *Tribolium*, one yielding DH31 and one yielding two different, but homologous CRF-like diuretic hormones, DH37 and DH47 (Supplemental Data 1). DH37 and DH47 are encoded by two separate exons that are alternatively spliced to a common 5′ exon encoding the signal peptide (Fig. 6). This alternative splicing is not found in *Drosophila*, where there is only one CRF-like diuretic hormone DH44 (Cabrero et al. 2002). In *Tribolium*, DH37 and DH47 are correlated with two GPCRs that are found in the clade with the DH44 receptor of *D. melanogaster* (Hauser et al. 2007), suggesting that there may be two separate CRF-like DH signaling pathways in the beetle.

The amino acid sequence of *Tribolium* DH47 is identical to that of the previously isolated DH47 from *T. molitor* (Furuya et al.

1998). *T. molitor* also contains a DH-37-like peptide sharing an amino acid sequence identity of 73% with DH37 from *Tribolium* (Furuya et al. 1995). This suggests that *T. molitor* has two CRF-like DH signaling systems, similar to *T. castaneum*. Thus, compared to other insects with a completely sequenced genome (*A. gambiae, Apis mellifera, B. mori,* and *D. melanogaster*), *Tribolium* has several additional osmoregulatory peptides (AVP, ADFs, and one extra copy of DH) that might be important for the animal to survive in its dry habitats.

Putative neuropeptide genes in *Tribolium*

The NPLP1 gene, a gene displaying all typical hallmarks of a neuropeptide precursor gene, was first identified in *D. melanogaster* by a peptidomics analysis (Baggerman et al. 2002) and by immunocytochemistry (Verleyen et al. 2004). However, its physiological role is unknown, so far. The orthologous NPLP1 gene in
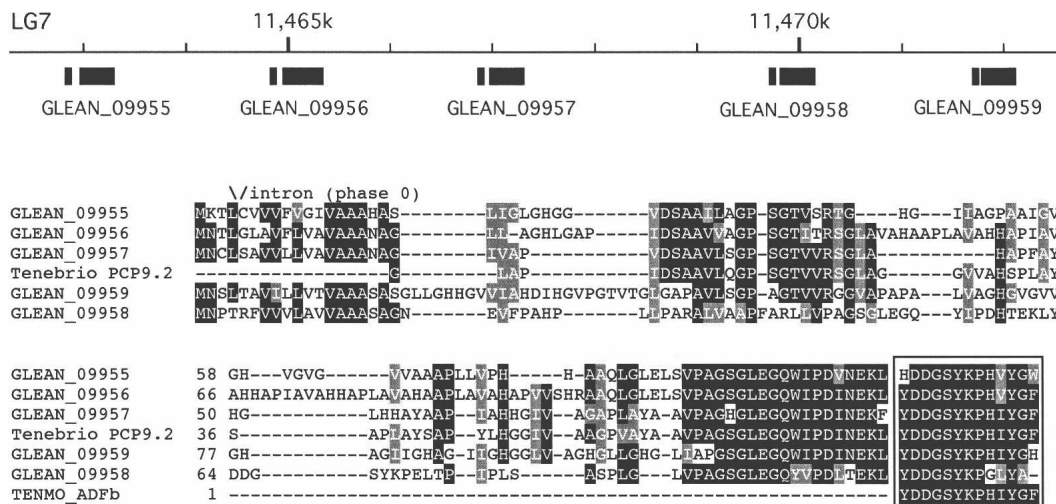
```
LG7            11,465k                        11,470k
        ▮█          ▮█          ▮█              ▮█              ▮█
    GLEAN_09955   GLEAN_09956   GLEAN_09957   GLEAN_09958   GLEAN_09959
```

```
                          \/intron (phase 0)
GLEAN_09955    MKTLCVVVFVGIVAAAHAS--------LIGLGHGG-------VDSAAILAGP-SGTVSRTG----HG---IIAGPAAIGV
GLEAN_09956    MNTLGLAVFLVAVAAANAG--------LI-AGHLGAP-----IDSAAVVAGP-SGTITRSGLAVAHAAPLAVAHHAPIAV
GLEAN_09957    MNCLSAVVLLVAVAAANAG--------IVAP----------VDSAAVLSGP-SGTVVRSGLA----------HAPFAY
Tenebrio PCP9.2 -----------------------------G--------LAP----------IDSAAVLQGP-SGTVVRSGLAG-----GVVAHSPLAY
GLEAN_09959    MNSLTAVTLLVTVAAASASGLLGHHGVVIAHDIHGVPGTVTGLGAPAVLSGP-AGTVVRGCVAPAPA---LVAGHCVGVV
GLEAN_09958    MNPTRFVVVLAVVAAASAGN-------EVFPAHP-------LLPARALVAAPFARLLVPAGSCLEGQ---YIPDHTEKLY
```

```
GLEAN_09955    58 GH---VGVG------VVAAAPLLVPH-----H-AAQLGLELSVPAGSGLEGQWIPDVNEKL HDDGSYKPHVYGW
GLEAN_09956    66 AHHAPIAVAHHAPLAVAHAAPLAVAHAPVVSHRAAQLGLELSVPAGSGLEGQWIPDINEKL YDDGSYKPHVYGF
GLEAN_09957    50 HG-----------LHHAYAAP----IAHHGIV--AGAPLAYA-AVPAGHGLEGQWIPDINEKF YDDGSYKPHIYGF
Tenebrio PCP9.2 36 S-------------APLAYSAP---YLHGGIV--AAGPVAYA-AVPAGSGLEGQWIPDINEKL YDDGSYKPHIYGF
GLEAN_09959    77 GH-----------AGIIGHAG-IIGHGGLV-AGHGLLGHG-LIAPGSGLEGQWIPDINEKL YDDGSYKPHIYGF
GLEAN_09958    64 DDG----------SYKPELTP---IPLS------ASPLG---LVPAGSGLEGQYVPDLTEKL YDDGSYKPGLYA-
TENMO_ADFb      1 ------------------------------------------------------------- YDDGSYKPHIYGF
```

Figure 5. Structure of the genomic region containing five clustered genes for *Tribolium castaneum* ADF-b1 to -5 (GLEAN_09955–09959; *upper* panel) and sequence alignment of the five ADF-b with *Tenebrio molitor* ADF-b (P83109) and cuticle protein TmPCP9.2 (Baernholdt and Anderson 1998). The putative mature peptides located at the carboxy termini are indicated by a box. The conserved intron position is marked on the *top* of the sequence alignment. Note that all five ADF-b genes are predicted to have small introns at the homologous position in the signal peptides with phase 0.
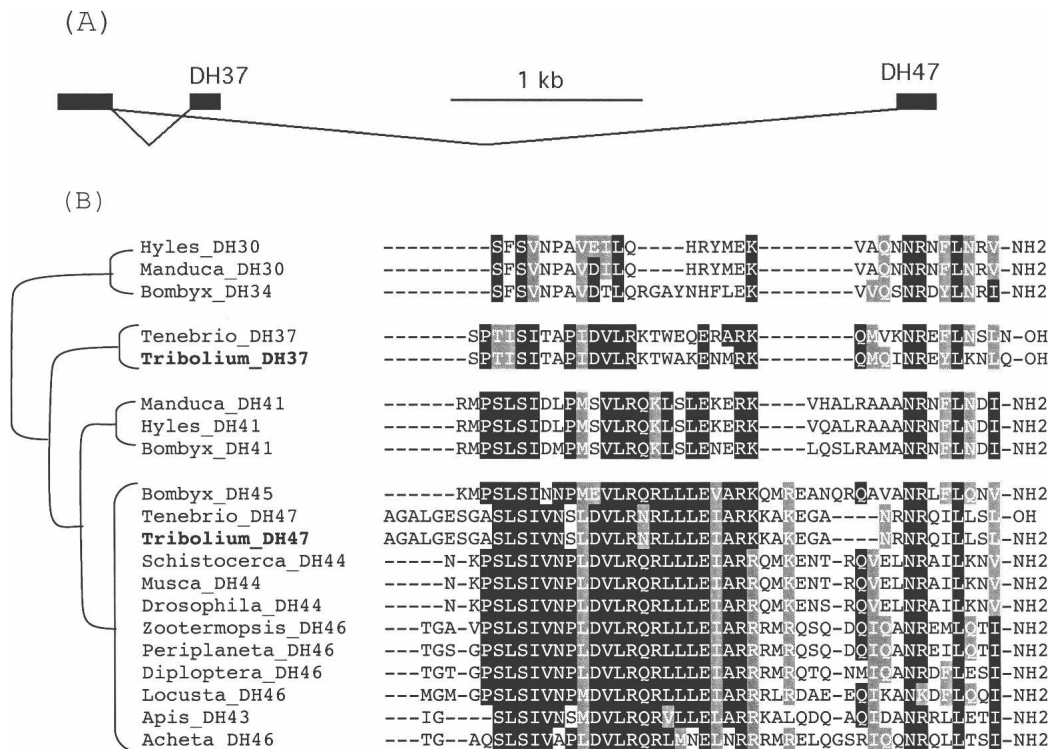
Figure 6. Sequence alignment and phylogenetic relationships of insect corticotropin releasing factor (CRF)-like diuretic hormones (DH). The hypothetical tree implies possible evolutionary relationships among the DHs from different species of insects. Sequences are from *Hyles*_DH30 (P82015), *Manduca*_DH30 (P24858), *Tenebrio*_DH37 (P56618), *Tribolium*_DH37 (GLEAN_02243), *Manduca*_DH41 (AAB59200.1), *Hyles*_DH41 (P82014), *Tenebrio*_DH47 (P56619), *Tribolium*_DH47, *Schistocerca*_DH44 (P67801), *Musca*_DH44 (P67800), *Drosophila*_DH44 (NP_649922.2), *Zootermopsis*_DH46 (P82707), *Periplaneta*_DH46 (P41538), *Diploptera*_DH46 (P82373), *Locusta*_DH46 (AAB19827.1), *Apis*_DH43 (CAI45289.1), and *Acheta*_DH46 (P23834). The three *Bombyx* DHs were from the GenBank genome database. Amino acid numbers are not indicated because many have only mature peptide sequences in the GenBank.

Tribolium encodes a large number of putative neuropeptides without similarities to each other or to other known neuropeptides (Supplemental Data 1). MS revealed six mature NPLP peptides (Supplemental Data 2).

Several peptides identified in a recent peptidomics study of the honeybee have been tentatively named after their partial N-terminal peptide sequences, i.e., NVP-containing, IDL-containing, and ITG-containing peptides (Hummon et al. 2006). Analysis of the *Tribolium* genome also revealed the genes encoding NVP-, IDL-, and ITG-like peptides. Both NVP-like and ITG-like cDNA sequences, were also found in the EST database of *Tribolium*. MS analysis yielded seven processed peptides from the NVP-like peptide precursor and one peptide from the ITG-like precursor (Supplemental Data 1 and 2). Their function and status as neuropeptides remain elusive.

Expansions of neuropeptide genes in *Tribolium*

Expanded neuropeptide genes in the *Tribolium* genome are those coding for ADF-b, adipokinetic hormones (AKH), eclosion hormones, and insulin-like peptides. Obviously, some of these genes involve recent gene expansions (i.e., five clustered genes for ADF-b), whereas others could have been from ancestral gene duplications. As described earlier, the gene encoding the CRF-like diuretic hormones (DH37/47) has not been duplicated, but two different transcripts are generated by alternative splicing. A similar alternative splicing occurs in the gene encoding the ion-transporting peptides ITP-a and ITP-b.

Three genes for AKH-like peptides were found in the *Tribolium* genome. Predicted prepropeptides have common structures with signal peptides for secretion directly followed by the immature peptides and an additional C-terminal conserved region surrounding two conserved cysteines. The two *Tribolium* AKHs—TcAKH1 and TcAKH2—are similar to each other, while another containing TcAKH3 is more divergent and has mild homology with the sequences found in mosquito species (Fig. 7). Sequence similarity among different insect AKH prepropeptides suggests that there are at least two distinct ancestral forms. The *Drosophila* genome contains only one AKH gene copy, which implies the loss of at least one lineage in *Drosophila*. AKHs of insects are usually synthesized in the glandular portion of the corpora cardiaca (CC) but not in the central nervous system itself. Direct profiling of the CC by MALDI-MS as well as analyses of CC-extracts by ESI-MS revealed the expression of AKH-2, whereas AKH-1 and AKH-3 could not be observed.

Eclosion hormone (EH) is known for its involvement in ecdysis behavior (Truman 2005). In contrast to other insects where only one *Eh* gene per genome appears to be common, the *Tribolium* genome contains two genes, one encoding EH that is highly conserved with the previously known EHs in other species, and another encoding an EH-like peptide (Fig. 8).

Neuropeptide genes lacking in *Tribolium*

Several neuropeptide genes are apparently lacking in the *Tribolium* genome. These include neuropeptide F, pigment dispers-
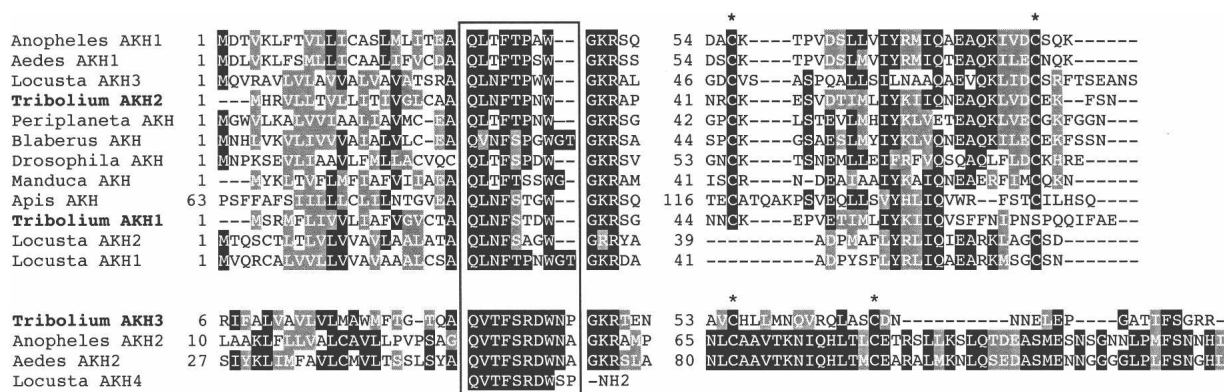
Figure 7. Sequence alignment of adipokinetic hormones (AKH). The mature peptides are in the box and followed by the glycine required for amidation and dibasic cleavage sites. Conserved cysteines are marked by asterisks (*) on the *top* of the alignment. Note that the middle of the sequence is omitted. Sequences are from *Anopheles*_AKH1 (ABD43194.1), *Aedes*_AKH1(EAT35885.1), *Locusta*_AKH3 (P19872), *Periplaneta*_AKH (AAV41425.1), *Blaberus*_AKH (Q17128), *Drosophila*_AKH (NP_523918.1), *Manduca*_AKH (P67788), *Apis*_AKH (from Hummon et al. 2006), *Locusta*_AKH2 (P08379), *Locusta*_AKH1 (P55319), *Anopheles*_AKH2 (ABD60145.1), *Aedes*_AKH2 (EAT37004.1), and *Locusta*_AKH4 (Siegert 1999).

ing factor (PDF), ADF-a, orcokinin, corazonin, kinin, and allato-statin-A. Interestingly, for the latter three, their cognate GPCRs also have not been found in the *Tribolium* genome (Hauser et al. 2007), supporting the lack of these peptidergic signaling systems in the beetle. For the other peptide genes, however, some of the "absence" could have been caused by missing sequences in the current version of the genome assembly or by highly diverging sequences not detected by our homology-based searches. Again, the loss of these signaling systems, together with the gains discussed earlier, shows that the *Tribolium* endocrine system is quite different from that of other insects with a sequenced genome.

Methods

Gene identification

The whole genome sequence of *Tribolium castaneum*, version 2.0, was used for the homology searches at the Web site of the Human Genome Sequence Center, Baylor College of Medicine (http://www.hgsc.bcm.tmc.edu/blast/blast.cgi?organism= Tcastaneum), or in the local BLAST by using Blosum62 or PAM30 for searching the short matching sequences. Prediction of the gene structure and open reading frame was done in GENBOREE (http://www.genboree.org) that contains the GLEAN2 predictions incorporating the results from multiple gene-prediction software programs and by manual correction. The homology search was followed by a prediction of the gene structure. The signal peptide in the N-terminal was predicted by the SIGNALP server (Bendtsen et al. 2004) (http://www.cbs.dtu.dk/services/ SignalP/). Identifying the N-terminal signal peptide sometimes overrode the largest open reading frame of the gene in the pre-

diction. Thus, the second translation initiation site is preferred in some cases (Supplemental data 1).

Mining EST data was done using the NCBI non-redundant database and trace archives (http://www.ncbi.nlm.nih.gov/ Traces/trace.cgi). RT-PCR to confirm the presence of transcripts was applied using the primers designed on the highly conserved regions in the gene predictions. Details of primer information will be published elsewhere. Multiple sequence alignment was performed in ClustalW (http://www.ebi.ac.uk/clustalw/) (Thompson et al. 1994) with manual adjustment. Phylogenetic analysis was done using PAUP 4.08 (Swofford 2000). Box shading in the sequence alignments was made with inverted fonts, with black (identical) and gray (similar) for 50% majority rules in the alignment.

The transcript of each gene was confirmed by either reverse transcription PCR or EST sequences (http://www.ncbi.nlm.nih. gov/dbEST/index.html; Park et al. 2007). List of the primers used in the RT-PCR is in Table 2.

Sample preparation for MS

Different parts of the CNS and peptide release sites (retrocerebral complex, thoracic perisympathetic organs, abdominal perisympathetic organs, Inka cells) were first directly analyzed on a MALDI-TOF mass spectrometer. The resulting ion signals were compared with the theoretical masses of the predicted peptides. A number of ions could be fragmented on the matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) MS, but most of the fragmentations were subsequently done using an ESI-Q-TOF MS and extracts of 30 brains, 30 thoracic ganglia, and 50 corpora cardiaca, respectively.
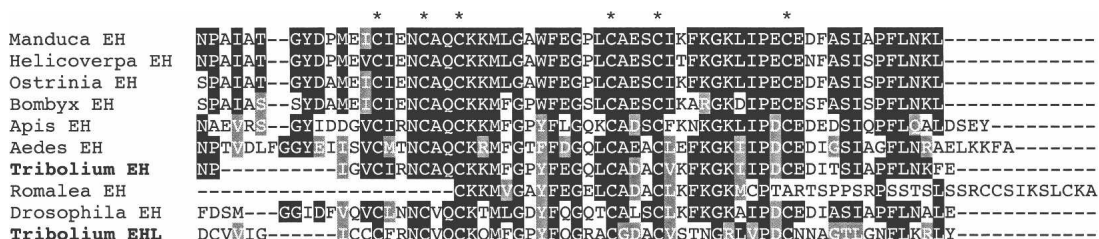

Figure 8. Sequence alignment of eclosion hormones (EH). The six conserved cysteines are marked by asterisks (*) on the *top* of the alignment. Sequences are from *Manduca*_EH (P11919), *Helicoverpa*_EH (AAV69026.1), *Ostrinia*_EH (ABG66962.1), *Bombyx*_EH (P25331), *Apis*_EH (XP_001122120.1), *Aedes*_EH (EAT41662.1), *Romalea*_EH (AAD28480.1), and *Drosophila*_EH (CAA51050.1).

Table 2. List of primers used for RT-PCR to confirm the transcript of predicted neuropeptide and protein hormone gene

Gene name	Forward primer (5′ to 3′)	Reverse primer (5′ to 3′)
AKH-1	GGGTGTGTACCGCTCAGTTA	GGCGAATTGGGAATATTGAA
AKH-2	ATCACCATTGTGGGTTTGTG	TCACAGTCCACTAATTTCTG
Allatostatin-B	CTGCGGGTAAACCACCAC	TTGTCCCACTTGCTCCTCTT
AVP-like	CGTGAGGGCTTTTTCCACGA	TTTATCGTCACTGGCACAGC
Bursicon-alpha subunit	AAATCCAGGTTTCGGGGGCT	GAAACCGAAGCTTCCCGCTC
Bursicon-beta subunit	ACAGAGAATATGCAATGGGG	GAGACGCACGCCGTCAGGGT
CCAP	ATGTGGGCGCAAACGCTCGA	GGCAGGGACAGCAGCGTCCT
Calcitonin-like diuretic hormone (DH31)	TCCCGAATATCTCCTCCAGA	TATGCTTCTTCTTCGCTTCG
Corticotropin releasing factor like-DH (DH37)	TGGAGCCCGTTGATGTAGCT	CGCATGTTCTCTTTTGCCCA
Corticotropin releasing factor like-DH (DH-47)	TGGAGCCCGTTGATGTAGCT	AACGTCCAACGAATTCACAA
EH	GAAATTTCCTCGTCCTTTTG	GTTCAAAAGGGCGCAATCG
ETH	GCTGCCAAAAATGTGCCCAG	AACCGACGGATCATCCCCCA
ITP-a	ATGAATTACCGATCATCCAA	CTAAATTAAGTTAGTGTGTTAG
ITP-b	ATGAATTACCGATCATCCAA	TTACTTCGATAAATATTCAATCA

Dissection of nervous tissues was performed as described for *D. melanogaster* (Predel et al. 2004). Since in *T. castaneum* the ventral nerve cord does not develop median nerves, release sites of putative FMRFamides, CAPA-peptides, and ETHs were studied by immunocytochemistry prior to the experiments (not shown). Dissected tissues were transferred with the help of a glass capillary to a stainless steel sample plate for MALDI-TOF MS or into a chilled solution of 5 µL of methanol/water/trifluoroacetic acid (90:9:1, v/v/v) for electrospray ionization quadrupole time-of-flight MS (ESI-Q-TOF MS).

MALDI–TOF MS

Dissected nervous tissues were dried on the sample plate and subsequently rinsed with water to reduce salt contamination. Matrix solution (α-cyano-4-hydroxycinnamic acid dissolved in methanol/water) was pumped onto the dried preparations over a period of ~5 sec using a Nanoliter injector (World Precision Instruments). Each preparation was allowed to dry and then covered with pure water for a few seconds which was removed by cellulose paper. At least five preparations each were prepared for mass spectrometric analyses which were performed as described in Predel et al. (2007).

ESI–Q–TOF MS

Following the dissection procedure, 50 µL of 0.1% TFA was added to the 5-µL methanol/water/trifluoroacetic acid solution. The extract was sonicated and centrifuged, and the methanol was evaporated from the supernatant. The resulting aqueous supernatant was then loaded onto an activated and equilibrated home-made microcolumn (Luna C18 material [10 µm]; Phenomenex). Subsequent mass spectrometric analyses were performed as described in Predel et al. (2004).

References

Baernholdt, D. and Anderson, S.O. 1998. Sequence studies on post-ecdysial cuticular proteins from pupae of the yellow mealworm, *Tenebrio molitor*. *Insect Biochem. Mol. Biol.* **28:** 517–526.

Baggerman, G., Cerstiaens, A., De Loof, A., and Schoofs, L. 2002. Peptidomics of the larval *Drosophila melanogaster* central nervous system. *J. Biol. Chem.* **277:** 40368–40374.

Bendtsen, J.D., Nielsen, H., von Heijne, G., and Brunak, S. 2004. Improved prediction of signal peptides: SIGNALP 3.0. *J. Mol. Biol.* **340:** 783–795.

Brown, M.R. and Cao, C. 2001. Distribution of ovary ecdysteroidogenic hormone I in the nervous system and gut of mosquitoes. *J. Insect Sci.* **1:** 3.

Brown, M.R., Graf, R., Swiderek, K.M., Fendley, D., Stracker, T.H., Champagne, D.E., and Lea, A.O. 1998. Identification of a steroidogenic neurohormone in female mosquitoes. *J. Biol. Chem.* **273:** 3967–3971.

Cabrero, P., Radford, J.C., Broderick, K.E., Costes, L., Veenstra, J.A., Spana, E.P., Davies, S.A., and Dow, J.A. 2002. The Dh gene of *Drosophila melanogaster* encodes a diuretic peptide that acts through cyclic AMP. *J. Exp. Biol.* **205:** 3799–3807.

Cazzamali, G., Torp, M., Hauser, F., Williamson, M., and Grimmelikhuijzen, C.J.P. 2005. The *Drosophila* gene CG9918 codes for a pyrokinin-1 receptor. *Biochem. Biophys. Res. Commun.* **335:** 14–19.

Claeys, I., Simonet, G., Van Loy, T., De Loof, A., and Vanden Broeck, J. 2003. cDNA cloning and transcript distribution of two novel members of the neuroparsin family in the desert locust, *Schistocerca gregaria*. *Insect Mol. Biol.* **12:** 473–481.

Claeys, I., Simonet, G., Breugelmans, B., Van Soest, S., Franssens, V., Sas, F., De Loof, A., and Vanden Broeck, J. 2005. Quantitative real-time RT-PCR analysis in desert locusts reveals phase dependent differences in neuroparsin transcript levels. *Insect Mol. Biol.* **14:** 415–422.

Coast, G.M., Rayne, R.C., Hayes, T.K., Mallet, A.I., Thompson, K.S., and Bacon, J.P. 1993. A comparison of the effects of two putative diuretic hormones from *Locusta migratoria* on isolated locust Malpighian tubules. *J. Exp. Biol.* **175:** 1–14.

Dai, L., Zitnan, D., and Adams, M.E. 2007. Strategic expression of ion transport peptide gene products in central and peripheral neurons of insects. *J. Comp. Neurol.* **500:** 353–367.

de Bree, F.M. and Burbach, J.P. 1998. Structure–function relationships of the vasopressin prohormone domains. *Cell. Mol. Neurobiol.* **18:** 173–191.

Duret, L., Guex, N., Peitsch, M.C., and Bairoch, A. 1998. New insulin-like proteins with atypical disulfide bond pattern characterized in *Caenorhabditis elegans* by comparative sequence analysis and homology modeling. *Genome Res.* **8:** 348–353.

Eigenheer, R.A., Nicolson, S.W., Schegg, K.M., Hull, J.J., and Schooley, D.A. 2002. Identification of a potent antidiuretic factor acting on beetle Malpighian tubules. *Proc. Natl. Acad. Sci.* **99:** 84–89.

Eigenheer, R.A., Wiehart, U.M., Nicolson, S.W., Schoofs, L., Schegg, K.M., Hull, J.J., and Schooley, D.A. 2003. Isolation, identification and localization of a second beetle antidiuretic peptide. *Peptides* **24:** 27–34.

Eipper, B.A., Milgram, S.L., Husten, E.J., Yun, H.Y., and Mains, R.E. 1993. Peptidylglycine alpha-amidating monooxygenase: A multifunctional protein with catalytic, processing, and routing domains. *Protein Sci.* **2:** 489–497.

Fournier, B., Guerineau, N., Mollard, P., and Girardie, J. 1994. Effects of two neuronal antidiuretic molecules, neuroparsin and 5-hydroxytryptamine, on cytosolic free calcium monitored with indo-1 in epithelial and muscular cells of the African locust rectum.

Biochim. Biophys. Acta **1220:** 181–187.

Fujino, Y., Nagahama, T., Oumi, T., Ukena, K., Morishita, F., Furukawa, Y., Matsushima, O., Ando, M., Takahama, H., Satake, H., et al. 1999. Possible functions of oxytocin/vasopressin-superfamily peptides in annelids with special reference to reproduction and osmoregulation. *J. Exp. Zool.* **284:** 401–406.

Furuya, K., Schegg, K.M., Wang, H., King, D.S., and Schooley, D.A. 1995. Isolation and identification of a diuretic hormone from the mealworm *Tenebrio molitor. Proc. Natl. Acad. Sci.* **92:** 12323–12327.

Furuya, K., Schegg, K.M., and Schooley, D.A. 1998. Isolation and identification of a second diuretic hormone from *Tenebrio molitor. Peptides* **19:** 619–626.

Grimmelikhuijzen, C.J.P., Williamson, M., and Hansen, G.N. 2002. Neuropeptides in cnidarians. *Can. J. Zool.* **80:** 1690–1702.

Hauser, F., Cazzamali, G., Williamson, M., Blenau, W., and Grimmelikhuijzen, C.J.P. 2006a. A review of neurohormone GPCRs present in the fruitfly *Drosophila melanogaster* and the honey bee *Apis mellifera. Prog. Neurobiol.* **80:** 1–19.

Hauser, F., Williamson, M., Cazzamali, G., and Grimmelikhuijzen, C.J.P. 2006b. Identifying neuropeptide and protein hormone receptors in *Drosophila melanogaster* by exploiting genomic data. *Brief. Funct. Genom. Proteom.* **4:** 321–330.

Hauser, F., Cazzamali, G., Williamson, M., Park, Y., Li, B., Tanaka, Y., Predel, R., Neupert, S., Schachtner, J., Verleyen, P., et al. 2007. A genome-wide inventory of neurohormone GPCRs in the red flour beetle *Tribolium castaneum. Front. Neuroendocrinol.* (in press).

Hewes, R.S. and Taghert, P.H. 2001. Neuropeptides and neuropeptide receptors in the *Drosophila melanogaster* genome. *Genome Res.* **11:** 1126–1142.

Hummon, A.B., Richmond, T.A., Verleyen, P., Baggerman, G., Huybrechts, J., Ewing, M.A., Vierstraete, E., Rodriguez-Zas, S.L., Schoofs, L., Robinson, G.E., et al. 2006. From the genome to the proteome: Uncovering peptides in the *Apis* brain. *Science* **314:** 647–649.

Isabel, G., Martin, J.R., Chidami, S., Veenstra, J.A., and Rosay, P. 2005. AKH-producing neuroendocrine cell ablation decreases trehalose and induces behavioral changes in *Drosophila. Am. J. Physiol. Regul. Integr. Comp. Physiol.* **288:** R531–R538.

Janssen, T., Claeys, I., Simonet, G., De Loof, A., Girardie, J., and Vanden Broeck, J. 2001. cDNA cloning and transcript distribution of two different neuroparsin precursors in the desert locust, *Schistocerca gregaria. Insect Mol. Biol.* **10:** 183–189.

Kanda, A., Satake, H., Kawada, T., and Minakata, H. 2005. Novel evolutionary lineages of the invertebrate oxytocin/vasopressin superfamily peptides and their receptors in the common octopus (*Octopus vulgaris*). *Biochem. J.* **387:** 85–91.

Kim, S.K. and Rulifson, E.J. 2004. Conserved mechanisms of glucose sensing and regulation by *Drosophila* corpora cardiaca cells. *Nature* **431:** 316–320.

Kim, Y.J., Zitnan, D., Galizia, C.G., Cho, K.H., and Adams, M.E. 2006. A command chemical triggers an innate behavior by sequential activation of multiple peptidergic ensembles. *Curr. Biol.* **16:** 1395–1407.

Levoye, A., Mouillac, B., Riviere, G., Vieau, D., Salzet, M., and Breton, C. 2005. Cloning, expression and pharmacological characterization of a vasopressin-related receptor in an annelid, the leech *Theromyzon tessulatum. J. Endocrinol.* **184:** 277–289.

Li, W., Kennedy, S.G., and Ruvkun, G. 2003. daf-28 encodes a *C. elegans* insulin superfamily member that is regulated by environmental cues and acts in the DAF-2 signaling pathway. *Genes & Dev.* **17:** 844–858.

Liu, F., Baggerman, G., Schoofs, L., and Wets, G. 2006a. Uncovering conserved patterns in bioactive peptides in Metazoa. *Peptides* **27:** 3137–3153.

Liu, F., Baggerman, G., D'Hetog, W., Verleyen, P., Schoofs, L., and Wets, G. 2006b. In silico identification of new secretory peptide genes in *Drosophila melanogaster. Mol. Cell. Proteomic.* **5:** 510–522.

McNabb, S.L., Baker, J.D., Agapite, J., Steller, H., Riddiford, L.M., and Truman, J.W. 1997. Disruption of a behavioral sequence by targeted death of peptidergic neurons in *Drosophila. Neuron* **19:** 813–823.

Nagasawa, H., Guo, F., Zhong, X.C., Xia, B.Y., Wang, Z.S., Qui, X.J., Wei, D.Y., Chen, E.I., Wang, J.Z., Suzuki, A., et al. 1980. Large-scale purification of prothoracicotropic hormone of the silkworm (*Bombyx mori*). *Sci. Sin.* **23:** 1053–1060.

Nagasawa, H., Kataoka, H., Isogai, A., Tamura, S., Suzuki, A., Mizoguchi, A., Fujiwara, Y., Suzuki, A., Takahashi, S.Y., and Ishizaki, H. 1986. Amino acid sequence of a prothoracicotropic hormone of the silkworm *Bombyx mori. Proc. Natl. Acad. Sci.* **83:** 5840–5843.

Park, Y., Zitnan, D., Gill, S.S., and Adams, M.E. 1999. Molecular cloning and biological activity of ecdysis-triggering hormones in *Drosophila melanogaster. FEBS Lett.* **463:** 133–138.

Park, Y., Filippov, V., Gill, S.S., and Adams, M.E. 2002a. Deletion of the ecdysis-triggering hormone gene leads to lethal ecdysis deficiency. *Development* **129:** 493–503.

Park, Y., Kim, Y.J., and Adams, M.E. 2002b. Identification of G protein-coupled receptors for *Drosophila* PRXamide peptides, CCAP, corazonin, and AKH supports a theory of ligand–receptor coevolution. *Proc. Natl. Acad. Sci.* **99:** 11423–11428.

Park, J.H., Schroeder, A.J., Helfrich-Forster, C., Jackson, F.R., and Ewer, J. 2003. Targeted ablation of CCAP neuropeptide-containing neurons of *Drosophila* causes specific defects in execution and circadian timing of ecdysis behavior. *Development* **130:** 2645–2656.

Park, Y., Aikiins, J., Wang, L.J., Beeman, R.W., Oppert, B., Lord, J.C., Brown, S.J., Lorenzen, M.D., Richards, S., Weinstock, G.M., et al. 2007. Analysis of transcriptome data in the red flour beetle *Tribolium castaneum. Insect Biotech. Mol. Biol.* (in press).

Predel, R. and Neupert, S. 2007. Social behavior and the evolution of neuropeptide genes: Lessons from the honeybee genome. *BioEssays* **29:** 416–421.

Predel, R. and Wegener, C. 2006. Biology of the CAPA peptides in insects. *Cell. Mol. Life Sci.* **63:** 2477–2490.

Predel, R., Wegener, C., Russell, W.K., Tichy, S.E., Russell, D.H., and Nachman, R.J. 2004. Peptidomics of CNS-associated neurohemal systems of adult *Drosophila melanogaster*: A mass spectrometric survey of peptides from individual flies. *J. Comp. Neurol.* **474:** 379–392.

Predel, R., Eckert, M., Pollák, E., Molnár, L., Scheibner, O., and Neupert, S. 2007. Peptidomics of identified neurons demonstrates a highly differentiated expression pattern of FXPRLamides in the neuroendocrine system of an insect. *J. Comp. Neurol.* **500:** 498–512.

Proux, J. and Rougon-Rapuzzi, G. 1980. Evidence of vasopressin-like molecule in migratory locust. Radioimmunological measurements in different tissues: Correlation with various states of hydration. *Gen. Comp. Endocrinol.* **42:** 378–383.

Proux, J.P., Miller, C.A., Li, J.P., Carney, R.L., Girardie, A., Delaage, M., and Schooley, D.A. 1987. Identification of an arginine vasopressin-like diuretic hormone from *Locusta migratoria. Biochem. Biophys. Res. Commun.* **149:** 180–186.

Riehle, M.A., Garczynski, S.F., Crim, J.W., Hill, C.A., and Brown, M.R. 2002. Neuropeptides and peptide hormones in *Anopheles gambiae. Science* **298:** 172–175.

Savard, J., Tautz, D., Richards, S., Weinstock, G.M., Gibbs, R.A., Werren, J.H., Tettelin, H., and Lercher, M.J. 2006. Phylogenomic analysis reveals bees and wasps (Hymenoptera) at the base of the radiation of Holometabolous insects. *Genome Res.* **16:** 1334–1338.

Schoofs, L., Veelaert, D., Vanden Broeck, J., and De Loof, A. 1997. Peptides in the locusts, *Locusta migratoria* and *Schistocerca gregaria. Peptides* **18:** 145–156.

Siegert, K.J. 1999. Locust corpora cardiaca contain an inactive adipokinetic hormone. *FEBS Lett.* **447:** 237–240.

Swofford, D.L. 2000. *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods)*, Version 4. Sinauer Associates, Sunderland, MA.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22:** 4673–4680. doi: 10.1093/nar/22.22.4673.

Truman, J.W. 2005. Hormonal control of insect ecdysis: Endocrine cascades for coordinating behavior with physiology. *Vitam. Horm.* **73:** 1–30.

Vanhems, E., Delbos, M., and Girardie, J. 1990. Insulin and neuroparsin promote neurite outgrowth in cultured locust CNS. *Eur. J. Neurosci.* **2:** 776–782.

Verleyen, P., Baggerman, G., Wiehart, U., Schoeters, E., Van Lommel, A., De Loof, A., and Schoofs, L. 2004. Expression of a novel neuropeptide, NVGTLARDFQLPIPNamide, in the larval and adult brain of *Drosophila melanogaster. J. Neurochem.* **88:** 311–319.

Wiehart, U.I., Nicolson, S.W., Eigenheer, R.A., and Schooley, D.A. 2002. Antagonistic control of fluid secretion by the Malpighian tubules of *Tenebrio molitor*: Effects of diuretic and antidiuretic peptides and their second messengers. *J. Exp. Biol.* **205:** 493–501.

Wu, Q. and Brown, M.R. 2006. Signaling and function of insulin-like peptides in insects. *Annu. Rev. Entomol.* **51:** 1–24. Zitnan, D., Zitnanova, I., Spalovska, I., Takac, P., Park, Y., and Adams, M.E. 2003. Conservation of ecdysis-triggering hormone signalling in insects. *J. Exp. Biol.* **206:** 1275–1289.

Zitnan, D., Zitnanova, I., Spalovska, I., Takac, P., Park, Y., and Adams, M.E., 2003. Conservation of ecdysis-triggering hormone signalling in insects. *J. Exp. Biol.* **206:** 1275–1289.