Insect Biochemistry and Molecular Biology 86 (2017) 9–19

Contents lists available at ScienceDirect

Insect Biochemistry and Molecular Biology

journal homepage: www.elsevier.com/locate/ibmb

DINeR: Database for Insect Neuropeptide Research

Joseph G.C. Yeoh [a, 1], Aniruddha A. Pandit [a, 1], Meet Zandawala [b], Dick R. Nässel [b], Shireen-Anne Davies [a], Julian A.T. Dow [a, *]

[a] Institute of Molecular, Cell and Systems Biology, College of Medical, Veterinary and Life Sciences, University of Glasgow, G12 8QQ Glasgow, Scotland, UK
[b] Department of Zoology, Stockholm University, S-10691 Stockholm, Sweden

ARTICLE INFO

Article history:
Received 27 January 2017
Received in revised form
3 May 2017
Accepted 8 May 2017
Available online 11 May 2017

Keywords:
Insect neuropeptide
Database
DINeR
Neuroendocrinology
Neuropeptide hormones
Neuropeptide signalling

ABSTRACT

Neuropeptides are responsible for regulating a variety of functions, including development, metabolism, water and ion homeostasis, and as neuromodulators in circuits of the central nervous system. Numerous neuropeptides have been identified and characterized. However, both discovery and functional characterization of neuropeptides across the massive Class Insecta has been sporadic. To leverage advances in post-genomic technologies for this rapidly growing field, insect neuroendocrinology requires a consolidated, comprehensive and standardised resource for managing neuropeptide information.

The Database for Insect Neuropeptide Research (DINeR) is a web-based database-application used for search and retrieval of neuropeptide information of various insect species detailing their isoform sequences, physiological functionality and images of their receptor-binding sites, in an intuitive, accessible and user-friendly format. The curated data includes representatives of 50 well described neuropeptide families from over 400 different insect species. Approximately 4700 FASTA formatted, neuropeptide isoform amino acid sequences and over 200 records of physiological functionality have been recorded based on published literature. Also available are images of neuropeptide receptor locations. In addition, the data include comprehensive summaries for each neuropeptide family, including their function, location, known functionality, as well as cladograms, sequence alignments and logos covering most insect orders. Moreover, we have adopted a standardised nomenclature to address inconsistent classification of neuropeptides.

As part of the H2020 nEUROSTRESSPEP project, the data will be actively maintained and curated, ensuring a comprehensive and standardised resource for the scientific community. DINeR is publicly available at the project website: http://www.neurostresspep.eu/diner/.

1. Introduction

Neuropeptides and neuropeptide hormones are synthesised by and released from neurons or neuroendocrine cells to trigger a physiological response. In insects, neuropeptides play an important role in coordinating complex homeostatic processes, such as development, metabolism, mating, water and ion homeostasis, reproduction, aggression and are also known to act as neuromodulators in circuits of the central nervous system (Caers et al., 2012; Nassel and Winther, 2010; Schoofs et al., 2017; Terhzaz et al., 2015). Since the discovery of the first insect neuropeptide, proctolin, in the American cockroach (Starratt and Brown, 1975), insect neuroendocrinology has progressed rapidly.

Neuropeptides are produced from larger precursor proteins which are known as prepropeptides (Fig. 1). Prepropeptides comprise of a signal peptide (which directs the protein to the secretory pathway), progenitors of mature peptides (the biologically-active peptides), spacer peptides (peptide fragments with no known biological function and non-conserved sequences) and cleavage sites (monobasic and dibasic) (Fig. 1). A useful website for predicting prepropeptide cleavage sites is Neuropred (Southey et al., 2006) (http://stagbeetle.animal.uiuc.edu/cgi-bin/neuropred.py). About 50 neuropeptide precursor-encoding genes are known in each insect species, with some species having a larger complement of precursors than others (Hauser et al., 2010). Each precursor can give rise to one or more mature neuropeptides or peptide

Abbreviations: DH31, diuretic hormone 31; DINeR, Database for Insect Neuropeptide Research.

* Corresponding author.
E-mail address: Julian.Dow@glasgow.ac.uk (J.A.T. Dow).

[1] Joseph G.C. Yeoh and Aniruddha A. Pandit contributed equally.

http://dx.doi.org/10.1016/j.ibmb.2017.05.001

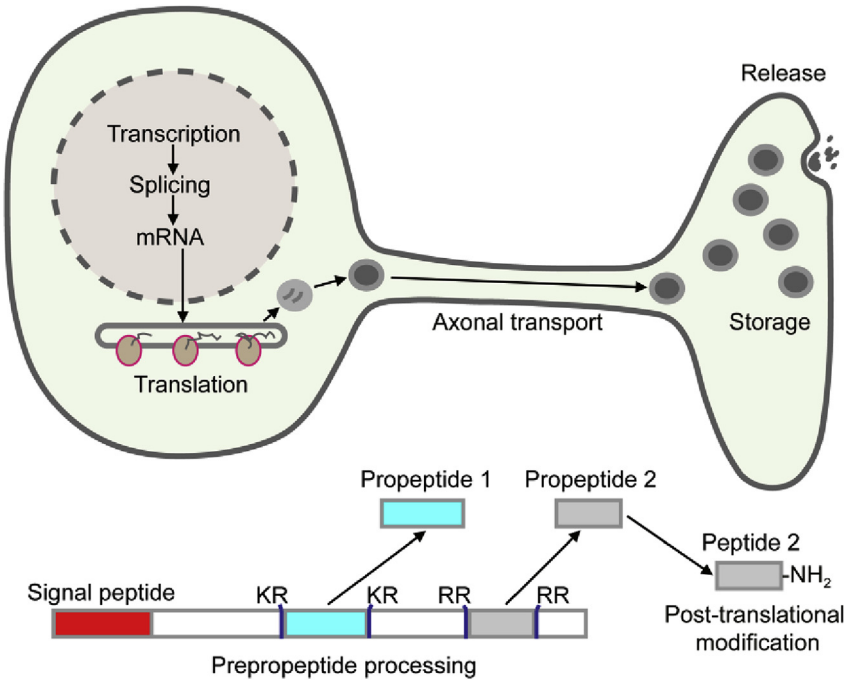

Fig. 1. Neuropeptide production starts in the nucleus and ends in the dense core vesicle. Neuropeptides are produced as part of larger precursor proteins, known as prepropeptides, which are encoded in the genome. These can give rise to one or several bioactive peptides. Neuropeptide encoding genes are transcribed in the nucleus. After splicing, mRNA is translated on ribosomes and with the aid of a signal peptide the immature prepropeptide is incorporated in the secretory pathway and ends up in vesicles. As the vesicles are transported to the axon termination the precursor is processed. The bioactive peptides are each surrounded by mono- or dibasic cleavage sites such as KR or RR shown here, that direct peptidases to enzymatically liberate the peptides. The white boxes represent non-conserved sequences (spacing regions) between peptide progenitors. Finally, posttranslational modifications may occur, such as C-terminal amidation (-NH_2) shown here. The mature neuropeptides are stored in dense core vesicles in the axon termination. A depolarization of the axon termination followed by Ca^{2+} influx triggers release of the peptide. This figure was redrawn and modified from Fig. 11.1 in Nässel and Larhammar (2013).

hormones. The number of mature peptides produced from a given precursor can vary from one insect species to another. These can either be (1) a set of very similar peptides with partly conserved sequences and thus similar receptor activation properties, or (2) in some cases peptides with distinct sequences and functions. Examples of the former are tachykinin-related peptides, AstBs and FMRFamides that exist in multiple closely related forms in the precursors. Examples of precursors containing peptides with distinct sequences and functions (bind distinct receptors) are those of CAPA/Pyrokinin, NPLP1 and vasopressin (Nassel and Winther, 2010; Stafflinger et al., 2008). There are also examples of peptides with similar sequences being produced by paralogs and splice variants. Insulin-like peptides are encoded by up to 38 paralogous genes in the moth *Bombyx mori* and 8 genes in *Drosophila* (Mizoguchi and Okamoto, 2013). The orcokinin gene in insects produces two different neuropeptide precursors by alternative splicing: orcokinin A and orcokinin B (Jiang et al., 2015; Sterkel et al., 2012). A typical prepropeptide and its biosynthesis and processing is shown in Fig. 1. At present, there are at least 50 well-described neuropeptide families identified from numerous species across different insect Orders. However, the wealth of neuropeptide information has generated a problem.

The naming of insect neuropeptide families has created confusion in the literature. Traditionally, neuropeptides were, in many cases, named after their first described function. However, neuropeptides can have multiple functions and thus, the same neuropeptide family might attract several names. For example, the first allatostatins identified (from *Diploptera punctata*) were named because they inhibited juvenile hormone production (Woodhead et al., 1989). Lorenz et al. (1995) then found similar neuropeptides with similar inhibitory properties in crickets, but with slightly different functional groups. Thus, this new group was named cricket type allatostatin or allatostatin B (AstB), and the original allatostatins were designated allatostatin A (AstA) or FGLa type allatostatin (FGLa/AST). However, a few years earlier, AstB orthologs were independently identified in the migratory locust, *Locusta migratoria,* found to have myoinhibiting properties, and named Myoinhibitory Peptide (or MIP) (Schoofs et al., 1991). Additionally, the allatostatin B/MIP peptides have also been shown to be the ancestral ligands for the sex peptide receptor (Kim et al., 2010; Poels et al., 2010). It is thus possible for researchers to encounter the same peptide family in entirely different contexts, and be unaware of the functional pleiotropy. Moreover, additional families of peptides with allatostatic activity, 'AstC' (also known as *Manduca* type allatostatin or PISCF/AST) and 'AstCC' (Kramer et al., 1991; Veenstra, 2009), have been found, further complicating the nomenclature. It is important to note here that although all the three allatostatins (AstA-AstC) may be found in a species, so far only one has been shown to display allatostatic properties in that species (Coast and Schooley, 2011; Nassel and Winther, 2010).

There is also a need to ensure that the different isoforms from the same species can be correctly identified and curated, as well as different isoforms in other species (interspecific isoform). The diuretic hormone, DH31, shows remarkable conservation throughout the insects. All DH31 sequences are 31 amino acids in length and the full sequence is important for DH31 to function (Zandawala, 2012). However, other neuropeptides show a higher degree of variability, both within and between species. For example, eight different kinin amino acid sequences have been found in the Madeiran cockroach, *Leucophaea maderae* (Holman

et al., 1986a, 1986b, 1987a, 1987b), while *Drosophila* has one (Terhzaz et al., 1999) and *Rhodnius prolixus* may have up to 12 kinin sequences (Te Brugge et al., 2011). Except for the C-terminal pentapeptide, kinin sequences show diversity in length and amino acid composition (Te Brugge et al., 2011; Terhzaz et al., 1999).

As a rapidly growing field in which more than 5800 papers (Google Scholar) have been published, insect neuropeptide endocrinology requires a comprehensive and standardised resource for managing neuropeptide information. Identification of insect neuropeptides has been facilitated by inexpensive next-generation sequencing, for example in the context of the i5k project (Poelchau et al., 2015); and sensitive mass-spectrometric peptidomic surveys (Audsley et al., 2015; Baggerman et al., 2002; Li et al., 2008; Predel et al., 2007) resulting in an explosion in available insect neuropeptide data. However, there is no dedicated online resource for insect neuroendocrinology. Existing databases emphasise amino acid sequences from higher animals or do not attempt to capture functional annotation. 'Neuropeptides' (http://www.neuropeptides.nl) is a single table resource, which lists information on gene families, mainly for vertebrate neuropeptides (Burbach, 2010). 'NeuroPedia' is a downloadable database of neuropeptide sequence and spectra data libraries which aid in identifying neuropeptides from mass spectrometry, available at http://proteomics.ucsd.edu/Software/NeuroPedia/(Kim et al., 2011). However, it does not curate arthropod neuropeptides. Another neuropeptide website, 'NeuroPep' (http://isyslab.info/NeuroPep/home.jsp), contains 21 arthropod neuropeptide families, less than half those known (Wang et al., 2015).

DINeR (http://www.neurostresspep.eu/diner/) is a web-based database application dedicated to insect neuropeptide information of multiple insect species detailing their peptide sequences, physiological functionality and images of their receptor-binding sites. Where useful, data from ancestral hexapods, and details of receptors, are also included. The database aims to fulfil the role of a primary resource for insect neuroendocrinology research by providing continuously curated, up-to-date and accurate information to the scientific community. There is no requirement for users to register or login to access DINeR.

2. Materials and methods

2.1. Data collection

2.1.1. Neuropeptide amino acid sequences

Insect neuropeptide amino acid sequences were mined manually from NCBI entries and research articles. Keywords such as the insect species or neuropeptide family name were used to query the Protein database in NCBI or PubMed. Only published neuropeptide sequences were added to DINeR. Unannotated genome or SRA data were not added. The amino acid sequence of each neuropeptide was then manually added to the database. Other information, such as post-translational modifications to the neuropeptide, information regarding the species which it was discovered and calculations of its molecular weight sequences were also added.

2.1.2. Neuropeptide function and images

The functions of insect neuropeptides were mined manually from research articles. Where possible, standard Gene Ontology terms from QuickGO are used to describe the neuropeptide function (Binns et al., 2009). Images from receptor-binding assays were extracted from research articles with appropriate permissions from authors.

2.2. Database and web application overview

The web application for DINeR is built using a Real-time Web Framework for the PERL programming language known as Mojolicious (http://mojolicious.org/). Mojolicious, as a web framework, is gaining popularity and is widely used since it provides many of the required modules for web application development such as RESTful routes, templates and session management tools as well as having a web development kit which can be used for many different kinds of applications. The main programming language used for the web application is PERL (https://www.perl.org/).

For the front end of the web application, Bootstrap 3 was implemented (http://getbootstrap.com/) along with the Mojolicious templates. This is one of the most popular HTML, CSS and Javascript frameworks used for developing intuitive, responsive, user-friendly and mobile-ready web applications. Jquery Javacripts libraries (https://jquery.com/) were also incorporated for enhanced features such as dropdown selection, validation of forms, and display of tabular results among others.

For the database, PostGreSQL (https://www.postgresql.org/) is used, which is also a popular open source object-relational database system. The database is used for data storage, organisation of data in a relational manner and for input and retrieval of data using SQL queries. Please see Fig. 2 for an illustration of the schema of the database structure.

The entire web application has been developed on Heroku (https://www.heroku.com/). Heroku is a cloud "Platform as a Service" or PaaS. Heroku supports Mojolicious and PERL and provides an environment for seamless development and deployment of web applications.

2.3. Bioinformatics

2.3.1. Cladograms and sequence alignment

The amino acid sequences from each neuropeptide family were grouped into the different insect orders and saved as FASTA files. These were then imported into CLC Genomics Workbench 9 (https://www.qiagenbioinformatics.com/). The FASTA files were used to create alignments for displaying conservation using the

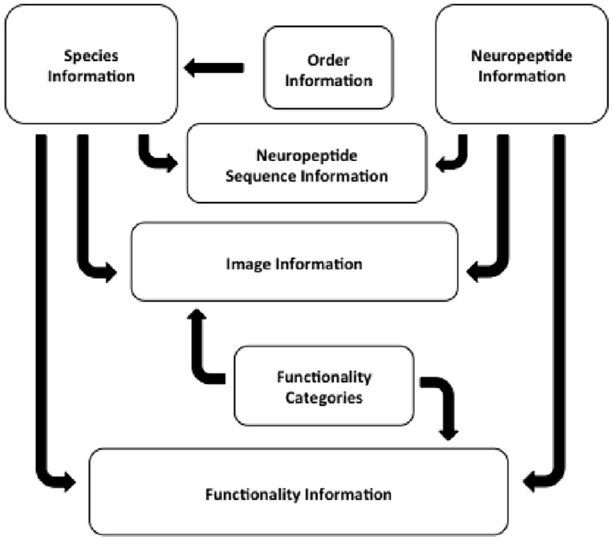

Fig. 2. Basic Database Relationship Schema. A basic representation of the relationships between the various tables in the database. For a more detailed explanation of the tables, please refer to the Supplementary Document.

default settings with the exception of the value of the End Gap Cost, which was set to "Cheap". A master FASTA file containing all the sequences from the whole neuropeptide family was also created. This file was used to generate the cladogram trees using the default settings.

2.3.2. Sequence logos

The individual insect order alignment files generated above were exported into Weblogo 3.4 (Crooks et al., 2004; Schneider and Stephens, 1990). Sequence logos were generated using default settings with the exception of units, the value of which were changed to "probability" and the colour scheme changed to "chemistry". Alignments and Seqlogos for insect orders with less than 5 sequences are not generated.

3. Results

3.1. User interface and functionality of the web application

3.1.1. Information search page

The information search form is composed of four search options (Supplementary Document Figure S2). Users may use any of the five fields to perform a (i) species search, (ii) common name search, (iii) order search (iv) neuropeptide family search, or (v) functionality search. These fields are in the form of dropdowns where multiple options can be chosen for the search (except for common name search where users can type in the search parameter). Users can scroll through the dropdown options and select relevant options or use the auto-complete functionality for the same. A combination of selections across the five fields can also be used to query the database. This allows the user to customize the search functionality according to their interests.

3.1.2. Information search results page

The results from the search are displayed as different sections; General Information, Neuropeptide Isoform Information, Functionality Information and the Image Results. Each of these sections has been described in further detail below (Please see Supplementary Document for figures).

The General Information section (Figure S3a) displays information relating to the insect. The insect order, common name, human impact as well as availability of its genome sequence are displayed. Where available, a link to the genome database is provided.

Figure S3b shows the Neuropeptide Isoform Information section, which contains data about the amino acid sequences. The first three columns display the neuropeptide isoform name, the neuropeptide family and the species in which it was discovered. Next is the amino acid sequence. Any modifications to the N and C termini are displayed in the form of a pyroglutamate prefix (p-) or an amide suffix (-amide) respectively. The next column shows the calculated molecular weight of the amino acid sequence. Each amino acid sequence is available for download in FASTA format. Clicking on the "F" button in the FASTA sequences column will open a new webpage, with the selected amino acid sequence in FASTA format. Selecting the "RF" button will open a new webpage listing all the amino acid sequences in that particular neuropeptide family. The last column is the link to the GenBank entry or research article.

The third table (Figure S3c) is the Functionality Information table. This table mainly lists published studies on the functions of each neuropeptide. The first two columns display the neuropeptide isoform name and the neuropeptide family name. The third column shows the species used in that particular study. More often, a neuropeptide isolated from a species is tested on a different but related or model insect species. The Functionality Category column shows the action of the neuropeptide in GO terms. The link in the QuickGO reference column directs the user to the EMBL QuickGO webpage (19), where they can obtain further information relating to the GO terms. A brief description of the effect of the neuropeptide is listed in the Functionality Description column. Finally, the last column in this section is the link to the published article, facilitating the user to retrieve the original source of experimental data.

The last section is the Image Results. Images obtained from *ex vivo* receptor-binding assays (23) are displayed in Figure S3d. Links are provided to the original research articles from where the images were obtained. Together with the Functionality Information table, this provides a comprehensive understanding on the target tissue or location of receptors of the neuropeptides.

A date of when the database was last updated is also displayed at the lower right hand corner.

3.1.3. Insect neuropeptide page

Additional to the search results, there are also information pages on each of the Insect neuropeptide families (http://www.neurostresspep.eu/diner/insectneuropeptides). Users can obtain more information for each insect neuropeptide family. The insect neuropeptide webpage contains detailed information on the original discovery; site of production and a description of the neuropeptide function (Please see Supplementary Document Figure S4). Links to suggested reviews are also shown to assist the keen user in obtaining more information. A list of additional references is also provided.

3.2. Neuropeptide nomenclature

Several inconsistencies for the neuropeptide names were found in the literature. In order to standardise classification of neuropeptides, DINeR adopts the nomenclature for naming insect neuropeptide families proposed by Coast and Schooley (2011) with a few exceptions. A full list of the neuropeptide families listed in DINeR is found in Table 1. Other synonyms and the abbreviation used for the neuropeptide family are also listed.

The naming of interspecific isoforms is also adopted from the nomenclature proposed by Coast and Schooley (2011). In brief, to distinguish between interspecific isoforms, a five letter code is used. The first three letters from the genus and the first two letters of the species name are combined. *Drosophila simulans* is abbreviated to 'Drosi' and is distinguishable from *Drosophila suzukii*, 'Drosu'. However, of the 425 insect species listed in DINeR, 8 species would have the same five letter code. For these particular species, additional letters from the species name were added to distinguish between them. For example, *Drosophila mojavensis* was abbreviated to 'Dromoj' while *Drosophila montana* was abbreviated to 'Dromon'. The full list can be found in Table S1 in the Supplementary Information. It should be noted that this aspect of nomenclature will remain dynamic; as data are added from new species, it will become necessary to extend the number of letters used for each code. However, the database will always also display the full species name for disambiguation.

Within a species, a neuropeptide can have more than one isoform (intraspecific isoform). Either Latin or Arabic numerals were used to differentiate intraspecific isoforms, based on the order that they are encoded in the genome. As an example, the six different allatostatin B isoforms from the red flour beetle, *Tribolium castaneum*, are named 'Trica-AstB-1' through to 'Trica-AstB-6'.

Table 1
Nomenclature of neuropeptide family and frequency in DINeR. The abbreviation and synonyms for each neuropeptide family is also listed.

Neuropeptide	Abbreviation	Synonym(s)	No. Of Records
Adipokinetic hormone/ Corazonin related Peptide	ACP		34
Anti-diuretic Factor	ADF		5
Adipokinetic Hormone	AKH	Red Pigment Concentrating Hormone (RPCH); Hypertrehalosemic Hormone (HrTH)	421
Allatostatin A	AstA	Cockroach-type allatostatin, FGLa-related allatostatin (FGLa/AST)	328
Allatostatin B	AstB	Cricket-type allatostatin, Myoinhibitory peptide (MIP), myoinhibiting peptide, W(X_6)Wamide	233
Allatostatin C	AstC	Moth-type allatostatin, Manduca allatostatin, PISCF-related allatostatin (PISCF/AST)	51
Allatostatin CC	AstCC		37
Allatostatin CCC	AstCCC	Allatostatin triple C	4
Allatotropin	AT		37
Bursicon	Burs	Bursicon alpha	58
Calcitonin	Cal		11
Capability/CAP2b	CAPA	Capability 2B peptides; Periviscerokinins; CAPA-pyrokinin	347
Crustacean Cardio-Active Peptide	CCAP	CAP2a	36
CCHamide	CCHa		63
CNMamide	CNMa		18
Corazonin	Crz		140
Diuretic Hormone 31	DH31	Calcitonin-like diuretic hormone	110
Diuretic Hormone 44	DH44	Diuretic peptide II; diuresin; Corticotropin releasing-factor (CRF)-related diuretic hormone	89
Eclosion hormone	EH		52
Elevenin	Ele		22
Ecdysis-triggering hormone	ETH		52
FMRFamide	FMRFa	Extended FMRFamides	268
GPA2	GPA2		13
GPB5	GPB5		13
Insulin-like Peptide	ILP	Insulin-related peptide and relaxin, Insulin-like growth factor (IGF-like)	175
Ion transport peptide	ITP	Crustacean hyperglycaemic hormone-related ion transport peptide (CHH/ITP)	80
Kinin	K	Leucokinin; Myokinin; Insectakinin	224
Limostatin	Lst		6
Myosuppressin	MS	FMRFamide-related peptides, Dromyosuppressin	46
Neuroparsin and Ovary Ecdysteroidogenic Hormone (OEH)	NP		14
Neuropeptide F	NPF		77
Neuropeptide-like precursor	NPLP		124
Natalisin	NTL		129
Orcokinin	OK		296
Pheromone Biosynthesis Activating Neuropeptide	PBAN	FXPRLamide; Hugin; PGN; suboesophageal neuropeptides; Diapause hormone; FXPRLamide	45
Diapause hormone	PBAN-DH	FXPRLamide	34
Partner of bursicon	Pburs	Bursicon beta	11
Pigment-dispersing factor	PDF	Pigment-dispersing hormone (PDH)	32
Pre-ecdysis triggering hormone	PETH		2
Pyrokinin	PK	CAPA-PK; FXPRLamide	394
Proctolin	Proc		31
Prothoracicotropic hormone	PTTH		13
RY amide	RYa		74
SIFamide	SIFa	LFamide	44
Sulfakinin	SK	LSK, DSK	192
Short neuropeptide F	sNPF	Head peptide, FMRFamide-related peptides (FaRPs)	70
Sex peptide	SP		1
Trissin	Tris		46
Tachykinin-related peptide	TRP	Tachykinins, neurokinins	175
Vasopressin	VPL	Arginine vasopressin-like peptide (AVLP); inotocin	5

3.3. Number of insect neuropeptide records in DINeR

A total of 4782 insect neuropeptide records were obtained and deposited into this initial version of DINeR (Fig. 3). The number of insect species in DINeR is 425, which covers 23 out of 31 insect orders. These 23 insect orders contain 98.5% of species abundance in the Class Insecta.

Fig. 4 shows the top 20 insects with the most neuropeptide records available. Of these insects, 8 of them were from the order Diptera, three each from Hemiptera and Hymenoptera, two each from Lepidoptera and Orthoptera, and one each in Coleoptera and Isoptera.

3.4. Examples in use: insect DH31 and kinins

DH31 (a conserved 31 amino acid neuropeptide) and kinin were chosen as examples because these are well studied insect neuropeptides, with a significant volume of sequence information from multiple insect species in different insect orders (Terhzaz et al., 1999; Zandawala, 2012). An important functionality of DH31 is stimulation of fluid transport by Malpighian tubules (Zandawala, 2012). Although kinin plays a similar role, this functionality has been lost in *Rhodnius prolixus* (Bhatt et al., 2014; Te Brugge et al., 2002) while kinin signalling seems to have been lost from the order Coleoptera (Halberg et al., 2015).

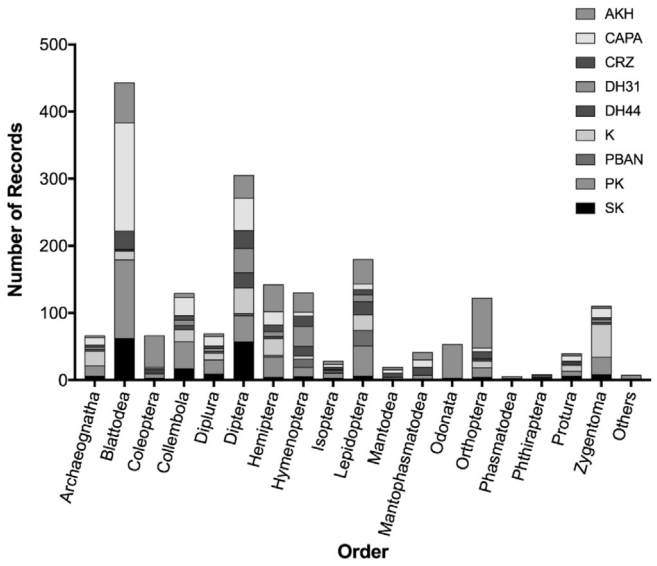

Fig. 3. Number of records per insect order with neuropeptide families of interest. Well-studied insect orders such as Diptera, Blattodea, Lepidoptera, Hemiptera and Orthoptera are well represented in the database. Neuropeptide families shown are Adipokinetic Hormone (AKH), Capability/CAP2b (CAPA), Corazonin (CRZ), Diuretic Hormone 31 (DH31), Diuretic Hormone 44 (DH44), Kinin (K), Pheromone Biosynthesis Activating Neuropeptide (PBAN), Pyrokinin (PK) and Sulfakinin (SK).

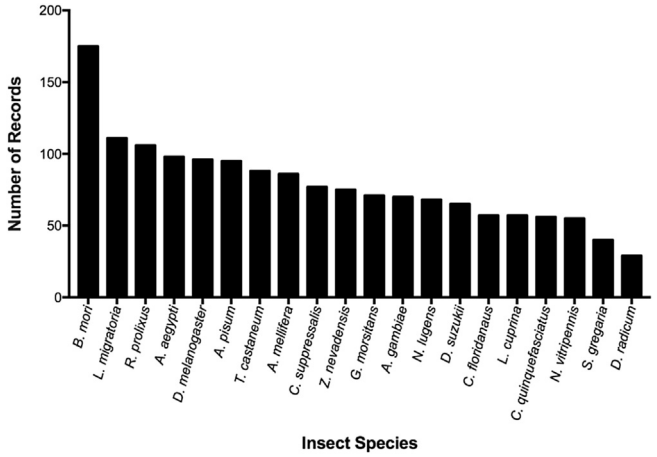

Fig. 4. Model insect species and the number of records in DINeR. The full repertoire of neuropeptides for these 20 insect species and others is available from DINeR.

Although Coast and Schooley (2011) proposed changing the neuropeptide name from DH31 to calcitonin-like diuretic hormone (CT-DH), a new insect neuropeptide family was discovered (Veenstra, 2014) which showed a higher similarity to vertebrate calcitonin and was thus named Calcitonin. To avoid unnecessary confusion, the historical neuropeptide name DH31 (Furuya et al., 2000) would be more appropriate and is therefore used in DINeR.

Insect DH31 amino acid sequences are highly conserved. The length of DH31 is the same in all insects (Zandawala, 2012). Additionally, the amino acid residue in each position is present in most of the species within the insect order, as evidenced by the alignments and sequence logos (Fig. 5). Most of the amino acids are also conserved between insect orders. The results from Diptera, Hemiptera, Hymenoptera and Lepidoptera are shown.

Fig. 6 shows the cladogram generated by using all 89 insect DH31 sequences. Each branch is shown with the insect order, species name and isoform. The dipteran and lepidopteran sequences are branched together, forming a monophyletic group.

Fig. 7 shows the alignments and sequence logos according to insect orders for kinin. Across the insect orders, a clearly visible conservation at the C-terminal with the consensus FX_1X_2WG is present. However, there is more variation in the kinin sequences as compared to DH31 (Fig. 7). This is due to insects having multiple kinins, and overall length or sequence is not conserved except for the C-terminal FX_1X_2WG sequence. Additionally, Serine is present at a higher frequency compared to other amino acids at positions X_1 or X_2, or both. The length of the kinin sequence varies between the insect Orders. As an example, the kinin sequences found in Lepidoptera, Orthoptera and Blattodea are short (less than 10 except for Locmi-K-V) while Hymenoptera, Hemiptera and Diptera have both short and longer sequences. Within Diptera, all *Drosophila* species studied so far have a single, fifteen amino acid kinin sequence. As a result, the cladogram, despite containing more sequences than DH31, does not contain any monophyletic groups and sequences from different insect orders are present in the same branch with other insect orders (Fig. 8).

4. Discussion

Insect neuropeptides show only a distant resemblance to those of mammals, but show clear similarity across the insects, despite a large phylogenetic distance across this huge and diverse class. It is thus appropriate to establish an insect-specific resource. It is particularly timely and important to generate such a database now, as there is a perfect storm of high-throughput technologies (next-generation sequencing and peptidomics), growing interest in neuropeptide signalling in model organisms such as *Drosophila*, and from researchers that might not have previous endocrinological experience. Therefore, the full repertoire of many neuropeptides for several model insect species including *D. melanogaster, B. mori, T. castaneum, A. pisum, A. mellifera* and *L. migratoria*, have been added to the database.

There are at least 50 described and well-known insect neuropeptide families. The presentation of data in DINeR allows visualisation of insect neuropeptide sequences, allowing the user to quickly look for regions of high amino acid conservation. Caution is advised for any interpretation of phylogenetic relationship in the cladogram. The cladograms can be viewed as similarity trees where neuropeptides with similar sequences are clustered together, but it is not a representation of evolutionary relationship among the insect neuropeptides.

How comprehensive does the database need to be? We acknowledge that DINeR can never be utterly comprehensive. DINeR can be considered as a useful introductory tool in the field of insect neuroendocrinology. DINeR will be continuously updated to keep in tandem with the rapidly growing field of insect neuroendocrinology. However, the cladograms of well-studied neuropeptide families already contain enough representatives, and the conservation of sequence is sufficiently high, that one could predict the likely sequence of a neuropeptide in a given species with reasonable confidence. For this reason, the database can be considered to be reasonably mature and useful.

Further enhancements to DINeR will include an advanced search parameter option, search by molecular weight, a BLAST tool for searching the database and a multiple sequence alignment tool, as well as creation of a secure data submission system. This system, using an Information Submission Form, will allow users to securely submit their own data in an appropriate and standard format. An administrator will perform a stringent verification process of the submitted data prior to making it publicly available. The aim of

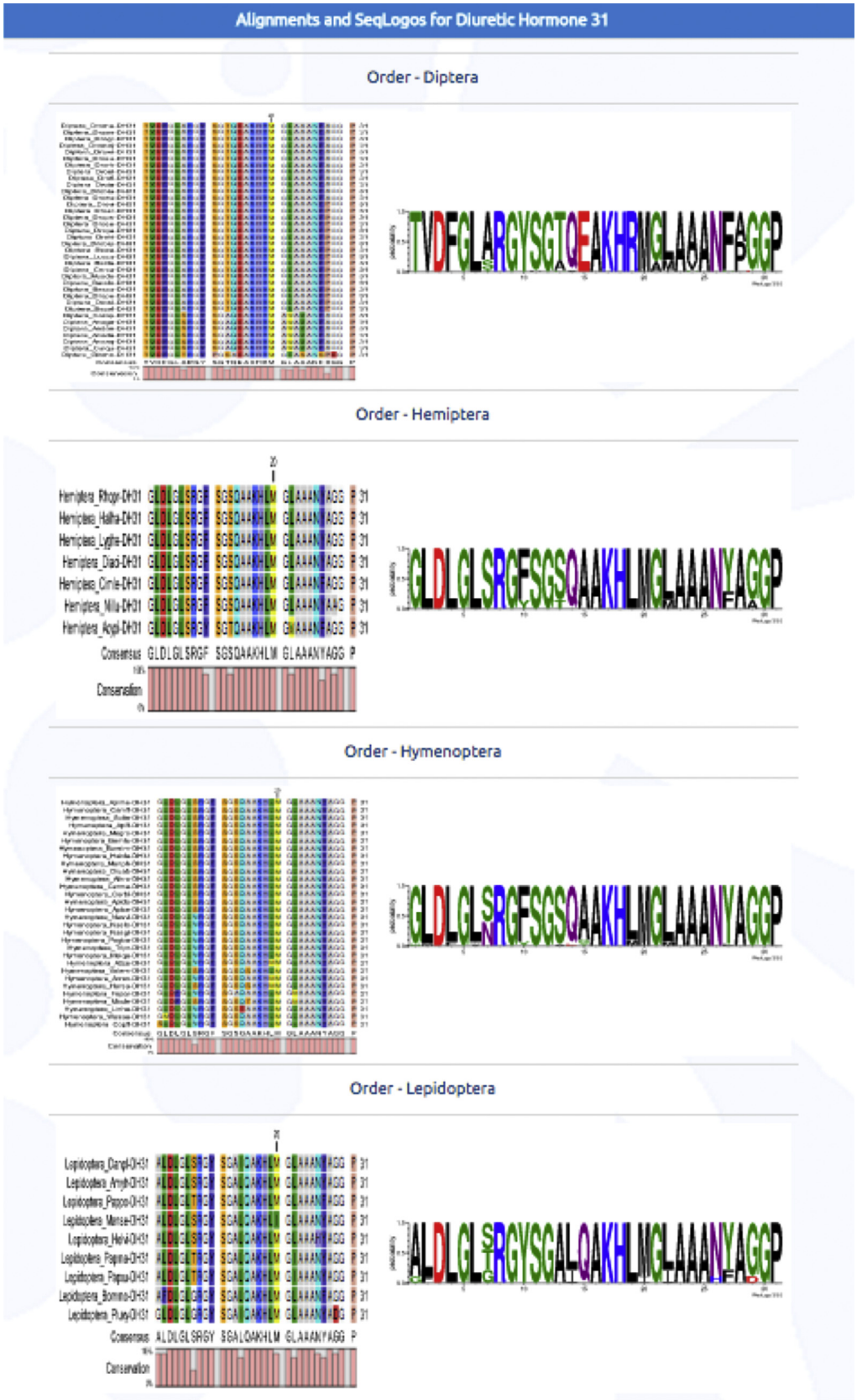

Fig. 5. Alignments and Seqlogos for DH31. Alignments and SeqLogos display the conserved regions among the various sequences for that particular order per neuropeptide.

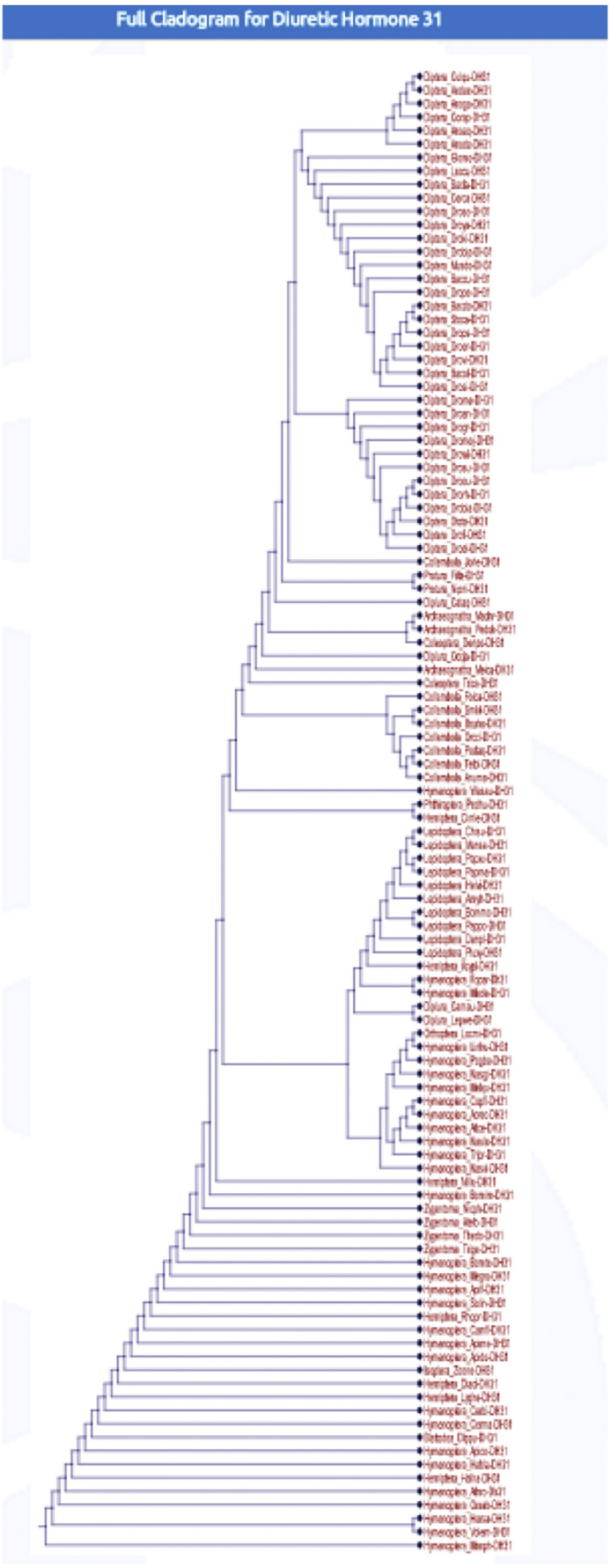

Fig. 6. Full Cladogram for DH31. Cladograms represent the evolutionary relationship of the neuropeptide between species.

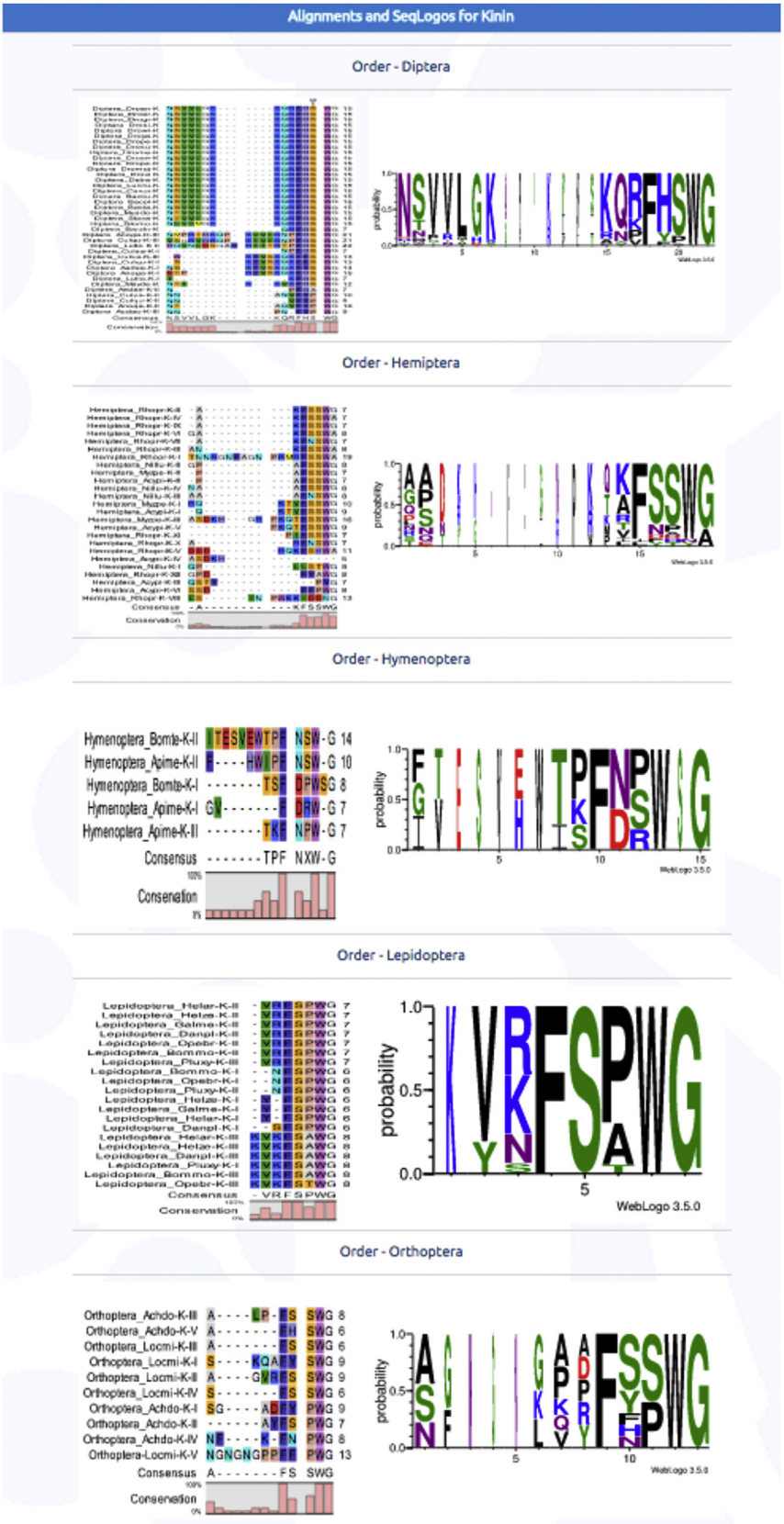

Fig. 7. Alignments and Seqlogos for Kinin. Alignments and SeqLogos display the conserved regions among the various sequences for that particular order per neuropeptide.

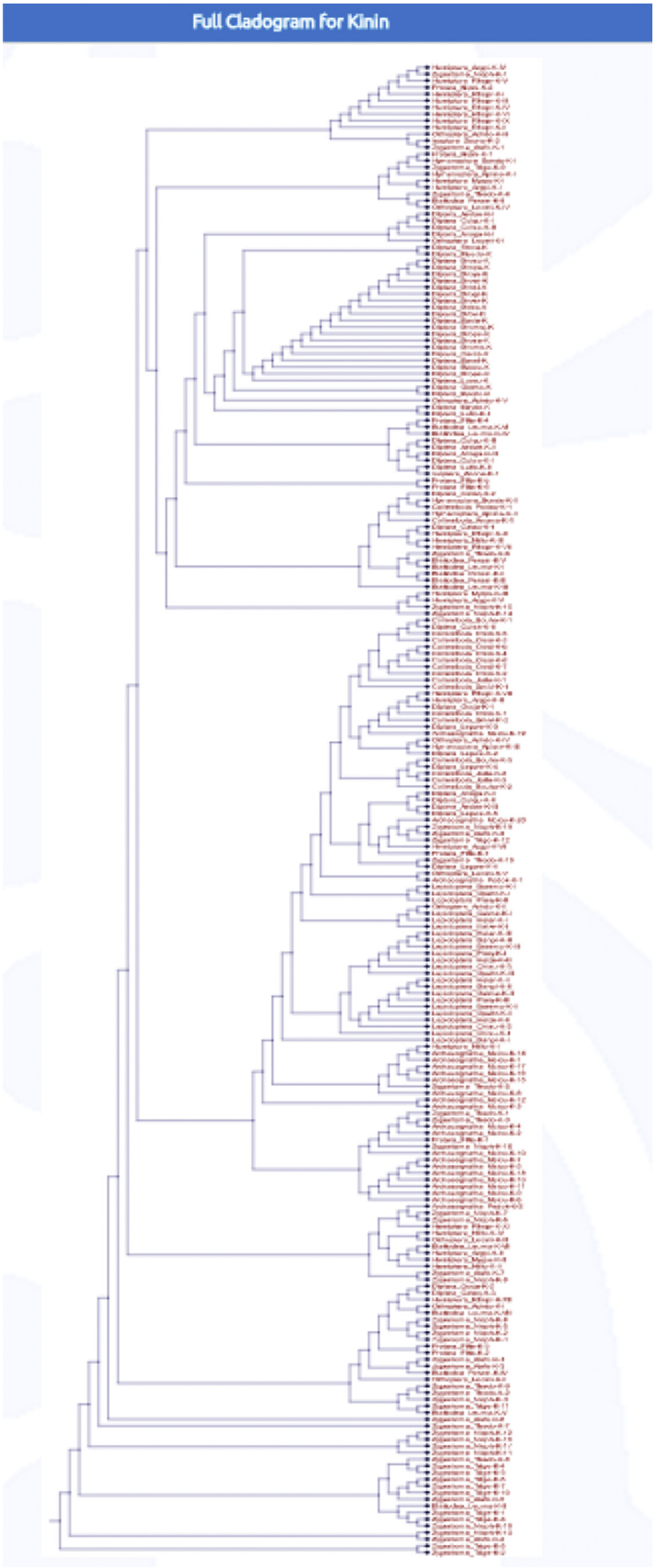

Fig. 8. Full Cladogram for Kinin. Cladograms represent the similarity of the neuropeptide between species.

these and other features will be to improve the utility and resourcefulness of this database. For optimal use, the recommended web browsers are Google Chrome, Firefox or Safari. DINeR may also be used on other current standard web browsers such as IE8 or above. More data will also be added as it becomes available. DINeR is integrated into the official nEUROSTRESSPEP website (http://www.neurostresspep.eu/) and is available for use by the public without restriction.

The impact of this database to the research community should not be underestimated and we believe DINeR will prove to be a useful tool in the field of entomology, insect neuroendocrinology as well as integrated pest management.

Funding

This work was supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 634361.

Conflict of interest

None declared.

Acknowledgements

We thank Dr Lucy Alford for useful discussions and Liam Gallen for input into data curation.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.ibmb.2017.05.001.

References

Audsley, N., Down, R.E., Isaac, R.E., 2015. Genomic and peptidomic analyses of the neuropeptides from the emerging pest, *Drosophila suzukii*. Peptides 68, 33–42.
Baggerman, G., Cerstiaens, A., De Loof, A., Schoofs, L., 2002. Peptidomics of the larval *Drosophila melanogaster* central nervous system. J. Biol. Chem. 277, 40368–40374.
Bhatt, G., da Silva, R., Nachman, R., Orchard, I., 2014. The molecular characterization of the kinin transcript and the physiological effects of kinins in the blood-gorging insect, *Rhodnius prolixus*. Peptides 53, 148–158.
Binns, D., Dimmer, E., Huntley, R., Barrell, D., O'Donovan, C., Apweiler, R., 2009. QuickGO: a web-based tool for Gene Ontology searching. Bioinformatics 25, 3045–3046.
Burbach, J.P.H., 2010. Neuropeptides from concept to online database. Eur. J. Pharmacol. 626, 27–48. www.neuropeptides.nl.
Caers, J., Verlinden, H., Zels, S., Vandersmissen, H.P., Vuerinckx, K., Schoofs, L., 2012. More than two decades of research on insect neuropeptide GPCRs: an overview. Front. Endocrinol. 3, 151.
Coast, G.M., Schooley, D.A., 2011. Toward a consensus nomenclature for insect neuropeptides and peptide hormones. Peptides 32, 620–631.
Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E., 2004. WebLogo: a sequence logo generator. Genome Res. 14, 1188–1190.
Furuya, K., Milchak, R.J., Schegg, K.M., Zhang, J., Tobe, S.S., Coast, G.M., Schooley, D.A., 2000. Cockroach diuretic hormones: characterization of a calcitonin-like peptide in insects. Proc. Natl. Acad. Sci. U. S. A. 97, 6469–6474.
Halberg, K.A., Terhzaz, S., Cabrero, P., Davies, S.A., Dow, J.A.T., 2015. Tracing the evolutionary origins of insect renal function. Nat. Commun. 6, 6800.
Hauser, F., Neupert, S., Williamson, M., Predel, R., Tanaka, Y., Grimmelikhuijzen, C.J.P., 2010. Genomics and peptidomics of neuropeptides and protein hormones present in the parasitic wasp *Nasonia vitripennis*. J. Proteome Res. 9, 5296–5310.
Holman, G.M., Cook, B.J., Nachman, R.J., 1986a. Isolation, primary structure and synthesis of 2 neuropeptides from *Leucophaea maderae* - members of a new family of cephalomyotropins. Comp. Biochem. Physiol. C-Pharmacology Toxicol. Endocrinol. 84, 205–211.
Holman, G.M., Cook, B.J., Nachman, R.J., 1986b. Primary structure and synthesis of 2 additional neuropeptides from *Leucophaea maderae* - members of a new family of cephalomyotropins. Comp. Biochem. Physiol. C-Pharmacology Toxicol. Endocrinol. 84, 271–276.
Holman, G.M., Cook, B.J., Nachman, R.J., 1987a. Isolation, primary structure and synthesis of leukokinin-VII and leukokinin-VIII - the final members of this new family of cephalomyotropic peptides isolated from head extracts of *Leucophaea maderae*. Comp. Biochem. Physiol. C-Pharmacology Toxicol. Endocrinol. 88, 31–34.
Holman, G.M., Cook, B.J., Nachman, R.J., 1987b. Isolation, primary structure, and synthesis of leukokinin-V and leukokinin-VI - myotropic peptides of *Leucophaea maderae*. Comp. Biochem. Physiol. C-Pharmacology Toxicol. Endocrinol. 88, 27–30.
Jiang, H., Kim, H.G., Park, Y., 2015. Alternatively spliced orcokinin isoforms and their functions in *Tribolium castaneum*. Insect Biochem. Mol. Biol. 65, 1–9.
Kim, Y., Bark, S., Hook, V., Bandeira, N., 2011. NeuroPedia: neuropeptide database and spectral library. Bioinformatics 27, 2772–2773.
Kim, Y.-J., Bartalska, K., Audsley, N., Yamanaka, N., Yapici, N., Lee, J.-Y., Kim, Y.-C., Markovic, M., Isaac, E., Tanaka, Y., Dickson, B.J., 2010. MIPs are ancestral ligands for the sex peptide receptor. Proc. Natl. Acad. Sci. U. S. A. 107, 6520–6525.
Kramer, S.J., Toschi, A., Miller, C.A., Kataoka, H., Quistad, G.B., Li, J.P., Carney, R.L., Schooley, D.A., 1991. Identification of an allatostatin from the tobacco hornworm *Manduca sexta*. Proc. Natl. Acad. Sci. U. S. A. 88, 9458–9462.
Li, B., Predel, R., Neupert, S., Hauser, F., Tanaka, Y., Cazzamali, G., Williamson, M., Arakane, Y., Verleyen, P., Schoofs, L., Schachtner, J., Grimmelikhuijzen, C.J.P., Park, Y., 2008. Genomics, transcriptomics, and peptidomics of neuropeptides and protein hormones in the red flour beetle *Tribolium castaneum*. Genome Res. 18, 113–122.
Lorenz, M.W., Kellner, R., Hoffmann, K.H., 1995. A family of neuropeptides that inhibit juvenile-hormone biosynthesis in the cricket, *Gryllus bimaculatus*. J. Biol. Chem. 270, 21103–21108.
Mizoguchi, A., Okamoto, N., 2013. Insulin-like and IGF-like peptides in the silkmoth *Bombyx mori*: discovery, structure, secretion, and function. Front. Physiol. 4, 217.
Nässel, D.R., Larhammar, D., 2013. Neuropeptides and peptide hormones. In: Galizia, C.G., Lledo, P.M. (Eds.), Neurosciences - from Molecule to Behavior: a University Text Book. Springer, Berlin, pp. 213–237.
Nassel, D.R., Winther, A.M.E., 2010. *Drosophila* neuropeptides in regulation of physiology and behavior. Prog. Neurobiol. 92, 42–104.
Predel, R., Neupert, S., Russell, W.K., Scheibner, O., Nachman, R.J., 2007. Corazonin in insects. Peptides 28 (1), 3–10.
Poelchau, M., Childers, C., Moore, G., Tsavatapalli, V., Evans, J., Lee, C.-Y., Lin, H., Lin, J.-W., Hackett, K., 2015. The i5k Workspace@NAL - enabling genomic data access, visualization and curation of arthropod genomes. Nucleic Acids Res. 43, D714–D719.
Poels, J., Van Loy, T., Vandersmissen, H.P., Van Hiel, B., Van Soest, S., Nachman, R.J., Vanden Broeck, J., 2010. Myoinhibiting peptides are the ancestral ligands of the promiscuous *Drosophila* sex peptide receptor. Cell. Mol. Life Sci. 67, 3511–3522.
Schneider, T.D., Stephens, R.M., 1990. Sequence logos - a new way to display consensus sequences. Nucleic Acids Res. 18, 6097–6100.
Schoofs, L., Holman, G.M., Hayes, T.K., Nachman, R.J., Deloof, A., 1991. Isolation, identification and synthesis of locustamyoinhibiting peptide (LOM-MIP), a novel biologically-active neuropeptide from *Locusta migratoria*. Regul. Pept. 36, 111–119.
Schoofs, L., De Loof, A., Van Hiel, M., 2017. Neuropeptides as regulators of behavior in insects. Annu. Rev. Entomol. 62, 35–52.
Southey, B.R., Amare, A., Zimmerman, T.A., Rodriguez-Zas, S.L., Sweedler, J.V., 2006. NeuroPred: a tool to predict cleavage sites in neuropeptide precursors and provide the masses of the resulting peptides. Nucleic Acids Res. 34, W267–W272 (Web Server issue).
Stafflinger, E., Hansen, K.K., Hauser, F., Schneider, M., Cazzamali, G., Williamson, M., Grimmelikhuijzen, C.J.P., 2008. Cloning and identification of an oxytocin/vasopressin-like receptor and its ligand from insects. Proc. Natl. Acad. Sci. U. S. A. 105, 3262–3267.
Starratt, A.N., Brown, B.E., 1975. Structure of pentapeptide proctolin, a proposed neurotransmitter in insects. Life Sci. 17, 1253–1256.
Sterkel, M., Oliveira, P.L., Urlaub, H., Hernandez-Martinez, S., Rivera-Pomar, R., Ons, S., 2012. OKB, a novel family of brain-gut neuropeptides from insects. Insect Biochem. Mol. Biol. 42, 466–473.
Te Brugge, V., Paluzzi, J., Neupert, S., Nachman, R., Orchard, I., 2011. Identification of kinin-related peptides in the disease vector, *Rhodnius prolixus*. Peptides 32, 469–474.
Te Brugge, V., Schooley, D., Orchard, I., 2002. The biological activity of diuretic factors in *Rhodnius prolixus*. Peptides 23, 671–681.
Terhzaz, S., O'Connell, F.C., Pollock, V.P., Kean, L., Davies, S.A., Veenstra, J.A., Dow, J.A.T., 1999. Isolation and characterization of a leucokinin-like peptide of *Drosophila melanogaster*. J. Exp. Biol. 202, 3667–3676.
Terhzaz, S., Teets, N.M., Cabrero, P., Henderson, L., Ritchie, M.G., Nachman, R.J., Dow, J.A.T., Denlinger, D.L., Davies, S.A., 2015. Insect capa neuropeptides impact desiccation and cold tolerance. Proc. Natl. Acad. Sci. U. S. A. 112, 2882–2887.
Veenstra, J.A., 2009. Allatostatin C and its paralog allatostatin double C: the arthropod somatostatins. Insect Biochem. Mol. Biol. 39, 161–170.
Veenstra, J.A., 2014. The contribution of the genomes of a termite and a locust to our understanding of insect neuropeptides and neurohormones. Front. Physiol. 5, 1–22.
Wang, Y., Wang, M., Yin, S., Jang, R., Wang, J., Xue, Z., Xu, T., 2015. NeuroPep: a comprehensive resource of neuropeptides. Database-the J. Biol. Databases Curation 2015 article ID bav038.
Woodhead, A.P., Stay, B., Seidel, S.L., Khan, M.A., Tobe, S.S., 1989. Primary structure of 4 allatostatins - neuropeptide inhibitors of juvenile-hormone synthesis. Proc. Natl. Acad. Sci. U. S. A. 86, 5997–6001.
Zandawala, M., 2012. Calcitonin-like diuretic hormones in insects. Insect Biochem. Mol. Biol. 42, 816–825.