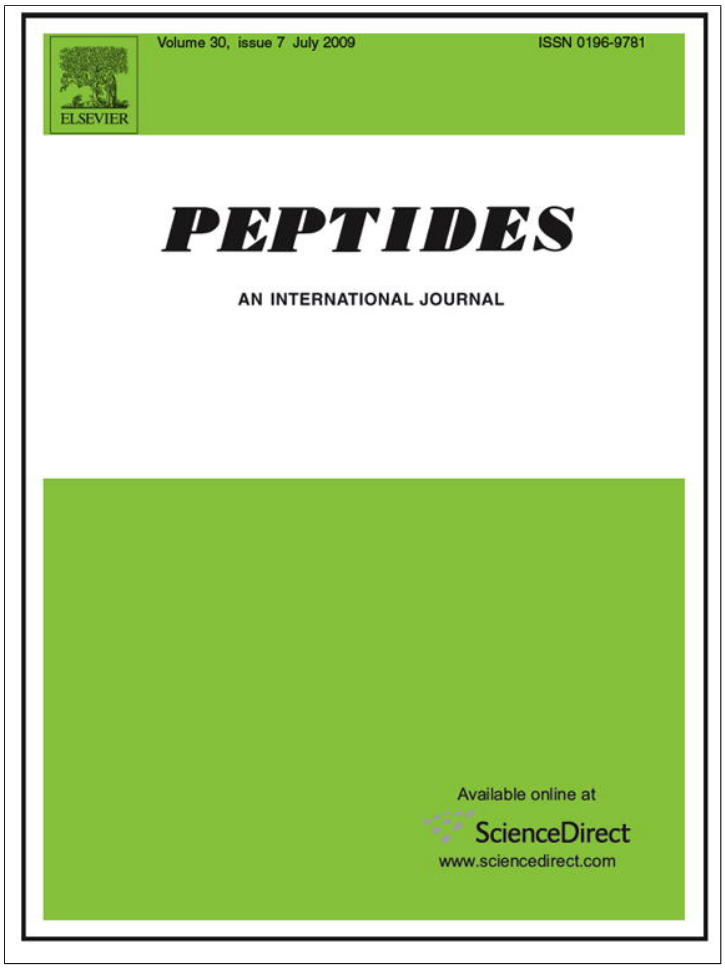

Peptides 30 (2009) 1241–1248

Contents lists available at ScienceDirect

Peptides

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

FGLamide Allatostatin genes in Arthropoda: Introns early or late?

Francisco Martínez-Pérez [a,b], William G. Bendena [c], Belinda S.W. Chang [a], Stephen S. Tobe [a,*]

[a] Department of Cell and Systems Biology, University of Toronto, Toronto, ON, Canada
[b] Department of Genetics and Molecular Biology, CINVESTAV, Mexico, D.F., Mexico
[c] Department of Biology, Queen's University, Kingston, ON, Canada

ARTICLE INFO

Article history:
Received 28 February 2009
Received in revised form 1 April 2009
Accepted 2 April 2009
Available online 10 April 2009

Keywords:
Allatostatins
Arthropoda
Gene structure
Introns
Late model

ABSTRACT

FGLamide allatostatins are invertebrate neuropeptides which inhibit juvenile hormone biosynthesis in Dictyoptera and related orders and also show myomodulatory activity. The FGLamide allatostatin (AST) gene structure in Dictyoptera is intronless within the ORF, whereas in 9 species of Diptera, the FGLamide AST ORF has one intron. To investigate the evolutionary history of AST intron structure, (intron early versus intron late hypothesis), all available Arthropoda FGLamide AST gene sequences were examined from genome databases with reference to intron presence and position/phase. Three types of FGLamide AST ORF organization were found: intronless in *I. scapularis* and *P. humanus corporis*; one intron in *D. pulex*, *A. pisum*, *A. mellifera* and five *Drosophila* sp.; two introns in *N. vitripennis*, *B. mori* strains, *A. aegypti*, *A. gambiae* and *C. quinquefasciatus*. The literature suggests that for the majority of genes examined, most introns exist between codons (phase 0) which may reflect an ancient function of introns to separate protein modules. 60% of the FGLamide AST ORFs introns were between the first and second base within a codon (phase 1), 28% were between the second and third nucleotides within a codon (phase two) and 12% were phase 0. As would be required for correct intron splicing consensus sequence, 84% of introns were in codons starting with guanine. The positioning of introns was a maximum of 9 codons from a dibasic cleavage site. Our results suggest that the introns in the analyzed species support the intron late model.

1. Introduction

The FGLamide allatostatins (ASTs) are neuropeptides which are potent inhibitors of juvenile hormone (JH) biosynthesis in cockroaches [44], termites [45] and crickets [24], but do not have any known effect on JH production in insects of any other order [35]. They may also serve as neurotransmitters or neuromodulators in the central and peripheral nervous systems, and/or as hormones when released from endocrine cells in the gut [4]. *In vitro* assays have demonstrated that FGLamide ASTs have a potent inhibitory effect on spontaneous contractions of various regions of the blowfly gut [6]. In crustaceans, the FGLamide ASTs can stimulate production of some intermediates in the JH biosynthetic pathway [20]. The FGLamide ASTs have variable amino acid numbers: 6–18 in insects and 6–20 in crustaceans [13]. The N-terminal amino acid sequence is variable, but the C-terminal core sequence Y/FXFGL/I/V-amide is essential for the inhibitory effect on JH biosynthesis, and it is present in all FGLamide ASTs [15]. The FGLamide ASTs are synthesized as a prepropeptide in which the number of FGLamide ASTs processed into mature peptides is species-specific [13,36]. In the precursor, each FGLamide AST is flanked by a C-terminal glycine required for amidation immediately followed by a KR/RR, potential endoproteolytic cleavage site [2,36,39,40]. The intron and exon organization of the FGLamide AST open reading frame (ORF) have been established for insects. In 6 species from Dictyoptera, the FGLamide AST ORF does not have introns [2], whereas in 7 species of Diptera from the genus *Drosophila*, *Calliphora vomitoria* and *Lucilia cuprina*, the gene has two introns [4,7]. These differences in the intron number raise several questions: what were the origins of introns in the FGLamide AST ORF of *Drosophila* species? Do other FGLamide AST ORFs from Arthropoda have introns?

In recent years, two opposite points of view about the origin of introns in eukaryotic cells have been proposed: the intron early hypothesis – the introns are extremely ancient structures, dating to before the divergence between the three eukaryotic kingdoms [23,31,33,34]; the intron late hypothesis – the introns have arisen more recently during early eukaryotic evolution [23,31,33,34]. Understanding of the evolution of spliceosomal introns, with respect to the dynamics of both intron loss and gain, has increased exponentially over the past few years, as a result of the release of many genome sequences of major eukaryotic lineages [14].

* Corresponding author at: Department of Cell and Systems Biology, University of Toronto, 25 Harbord Street, Toronto, ON M5S 3G5, Canada. Tel.: +1 416 978 3517; fax: +1 416 978 3522.
E-mail address: stephen.tobe@utoronto.ca (S.S. Tobe).

0196-9781/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.peptides.2009.04.001

Studies of individual genes and genomes have revealed that some intron positions have been conserved through eukaryotic evolution [14]. The intron position in any ORF can occur in three different positions: phase zero—between two codons, phase one—between the first and second nucleotide of a codon and phase two—between the second and third nucleotide of a codon [23,28]. The position of the introns could be ancient, for example, when these have been conserved in at least two eukaryotic kingdoms (animal, fungi and plants), or in at least two distant animal lineages (e.g. nematodes, arthropods, and vertebrates), whereas an intron can be considered recent if the intron phase and position are specific for an eukaryotic lineage [23,34].

The existence of protosplice sites has been addressed directly by examining the context of introns inserted within codons encoding amino acids that are conserved in all eukaryotes and that, accordingly, are not subjected to selection for splicing efficiency [1,41,43]. Evidence has been presented that introns are either predominantly inserted into specific protosplice sites, which have the exon|intron consensus sequence (A/C)AG|GT, or are inserted randomly but preferentially fixed at such sites [1,41,43]. For example, a preference for phase one introns occurs when a glycine codon (GGN) is localized near codons for dibasic proteolytic sites in secreted proteins [1,41,43]. The aim of the present work was to determine if the introns in the ORF of the FGLamide AST gene are recent or ancient as revealed in the Arthropoda genome databases.

2. Material and methods

2.1. Genome data base

FGLamide AST genes and mRNA were obtained via keyword searches for: *Apis mellifera*, *Nasonia vitripennis*, *Acyrthosiphon pisum*, *Pediculus humanus corporis*, *Bombyx mori* strain p50T, *B. mori* variety Dazao, *Drosophila ananassae*, *Drosophila grimshawi*, *Drosophila melanogaster*, *Drosophila mojavensis*, *Drosophila pseudoobscura*, *Drosophila simulans*, *Drosophila yakuba*, *Drosophila erecta*, *Drosophila persimilis*, *Drosophila sechellia*, *Drosophila virilis*, *Drosophila willistoni*, *Aedes aegypti*, *Anopheles gambiae*, *Culex quinquefasciatus* and *Ixodes scapularis* (Arachnida) These sequences were obtained from the following databases: National Center for Biotechnology Information, Fly Base [42] and Human Genome Sequencing Center. The FGLamide AST gene and mRNA from *Daphnia pulex* was obtained from Daphnia pulex V1.0 from the Joint Genome Institute.

2.2. In silico determination of FGLamide AST gene and mRNA in genome data base

To determine FGLamide AST gene structure not reported in some genome databases, BLAST GENOME was used with FGLamide allatostatin cDNA from *D. punctata* and *D. melanogaster*. With the sequences obtained, the exon–intron organization was established with Softberry programs (http://linux1.softberry.com/berry.phtml). The conceptual translation and the sequence edition for each gene were done with the software Translate from EXPASY (http://ca.expasy.org/) and Molecular Kit tools (http://www.vivo.colostate.edu/molkit/index.html).

2.3. FGLamide AST gene and mRNA in database construction

For all comparative analysis of FGLamide AST ORF structures, a database was constructed with the sequences obtained and FGLamide AST ORFs reported previously from: *Diploptera punctata*, *Periplaneta americana*, *Blattella germanica*, *Blaberus craniifer*, *Blatta orientalis*, *Supella longipalpa*, *C. vomitoria*, *L. cuprina*, *D. ananassae*, *D. grimshawi*, *D. melanogaster*, *D. mojavensis*, *D. pseudoobscura*, *D. simulans* and *D. yakuba* [2,4,7].

2.4. Phase and intron position

The intron phase from each gene was determined using alignments of the gene and mRNA with: ClustalW [38]. To determine the intron position between the precursors, multiple alignments were built using ClustalW (with gap 0.05 and window 9). The alignment was edited with GENEDOC software [27]. The phylograms were edited with MEGA [37,19].

2.5. Phylogenetic analyses

FGLamide AST precursor proteins were aligned using ClustalW [21,38] and subjected to phylogenetic analyses using maximum likelihood and Bayesian methods [11,32]. Maximum likelihood phylogenetic methods were implemented in the program PHYML 3.0 [11,12] using the LG amino acid replacement matrix [22]. For the likelihood analyses, bootstrapping methods were used to assess the degree of confidence in nodes of the phylogeny [8]. Bayesian inference was performed in MrBayes 3.1.2 [32] using model jumping among fixed-rate amino acid models, with all of the AST sequence data in a single partition. A single Markov chain Monte Carlo run was performed, with four chains (three heated and one cold) for 1 million generations, sampling trees (and parameters) every 1000 generations. Stationarity was assessed using Tracer 1.4 [30] and the first 25% of trees discarded as burn-in. Remaining trees were taken as representative of the posterior probability distribution.

3. Results

3.1. Identification of FGLamide AST ORFs and mRNA in genome databases from Arthropoda

The C-terminus consensus to FGLamide ASTs from Arthropoda was found in all genome databases analyzed. In the majority of sequences, the gene and mRNA were reported by each Genome Project as FGLamide AST; however, in *P. humanus corporis* and five *Drosophila* sp., some of the FGLamide AST ORFs had different names, whereas in *N. vitripennis*, *A. pisum*, *B. mori* strain *p50T* and *B. mori* strain *Davao*, the exon and intron organization were not previously established (Table 1).

3.2. FGLamide AST ORF organization in Arthropoda

Three different forms of FGLamide AST ORF organization were found: (1) the ORF has no introns, and therefore the mRNA for the encoded prepropeptide originated from a single exon. This pattern was found in 2 species: the tick *I. scapularis* in which the precursor has four potential copies of FGLamide ASTs, and the insect *P. humanus corporis* with 6 copies FGLamide AST copies. (2) The second type of FGLamide AST ORF organization shows one intron and two exons. The genera that possess this organization had differing FGLamide AST number. In this gene structure, the first exon codes for the majority of FGLamide ASTs, whereas the second exon gave one or two FGLamide ASTs. This organization was observed in the crustacean *D. pulex* and in the insects *Drosophila* sp., *C. vomitoria*, *L. cuprina* and *A. pisum*, whereas *A. mellifera* had a different organization, with one FGLamide AST in the first exon and the rest in the second exon. (3) In this third type of organization, the ORF has two introns and three exons. This pattern was found in *N. vitripennis*, *B. mori strain p50T*, *B. mori strain Davao*, *A. aegypti*, *A. gambiae* and *C. quinquefasciatus*. The first exon encoded one or two FGLamide ASTs. The second exon possessed codons for the amino terminus for one or more FGLamide ASTs and the last exon had codons for the carboxyl terminus of one or more FGLamide ASTs, depending on the species (Table 2).

Table 1
FGLamide AST gene and mRNA from Arthropoda obtained from Genome Projects.

Organism	Gene	mRNA	Reported as
Arachnida			
Ixodes scapularis	DS971562	DS971562	Bombystatin
Crustacea			
Daphnia pulex	NCBI_GNO_0600073	NCBI_GNO_0600073	Allatostatin A
Insecta			
Hymenoptera			
Apis mellifera	NW_001253130	NW_001253130	Allatostatin
Nasonia vitripennis	DS265633	This work	Allatostatin[a]
Hemiptera			
Acyrthosiphon pisum	EQ118335	This work	Allatostatin[a]
Neoptera			
Pediculus humanus corporis	DS235203	DS235203	Hypothetical protein
Lepidoptera			
Bombyx mori strain *p50T*	Exon 1: 46724731; Exon 2: 46629065; Exon 3: 46792320	NM_001043571	Allatostatin[a]. The exons are in three different accession numbers.
Bombyx mori variety, Davao	CH386506	NM_001043571	Allatostatin[a]
Dictyoptera			
Diploptera punctata	[2]	DPU00444	Allatostatin
Periplaneta americana	[2]	X91029	Allatostatin
Blattella germanica	[2]	[2]	Allatostatin
Blaberus craniifer	[2]	[2]	Allatostatin
Blatta orientalis	[2]	[2]	Allatostatin
Supella longipalpa	[2]	[2]	Allatostatin
Diptera			
Callíphora vomitoria	[7]	[7]	Allatostatin
Lucila cuprina	[7]	[7]	Allatostatin
Drosophila ananassae	[4]	[4]	Allatostatin
Drosophila grimshawi	[4]	[4]	Allatostatin
Drosophila melanogaster	[4]	AF263923; NM_079765	Allatostatin
Drosophila mojavensis	[4]	[4]	Allatostatin
Drosophila pseudoobscura	[4]	[4]	Allatostatin
Drosophila simulans	[4]	[4]	Allatostatin
Drosophila yakuba	[4]	[4]	Allatostatin
Drosophila erecta	AAPQ01006808	XM_001981909	Transcript GG12326-RA
Drosophila persimilis	AAIZ01003146	XM_002020181	Transcript GL13866-RA
Drosophila sechellia	AAKO01001047	XM_002032516	Transcript GM23467-RA
Drosophila virilis	AANI01017106	XM_002052990	Transcript GJ23562-RA
Drosophila willistoni	AAQB01009413	XM_002073224	Transcript GK13247-RA
Aedes aegypti	CH477340	AAU66841	Gene: conserved hypothetical protein mRNA preproallastostatin.
Anopheles gambiae	NT_078266	XM_313511	Allatostatin
Culex quinquefasciatus	DS232003	XM_001849994	Preproallatostatin

[a] Gene structure proposed in this work.

3.3. Intron phases of FGLamide AST gene in Arthropoda

The introns in the FGLamide AST ORFs were present as all three phases although the majority was phase one (60%), with phase two at 28% and phase zero, 12%. The phase of the introns showed a pattern of localization within the codons. The vast majority, 84%, of the introns were before or after guanine, whereas 16% were localized before or after cytosine (Table 3).

3.4. Intron position in FGLamide AST gene in Arthropoda

In general, all the introns for the FGLamide AST ORFs were within nine codons of the codons for KR, which corresponds to a major proteolytic cleavage site. The precursor alignment of the FGLamide AST ORFs showed five intron clusters with different conserved positions: the first and second from 9 and 3 *Drosophila* species, respectively, the third from *C. vomitoria* and *L. cuprina*, the fourth from *A. mellifera* and *N. vitripennis*, and the fifth which corresponds to the 3 species from the family Culicidae. *B. mori*, *A. pisum* and *D. pulex* introns do not show conserved positions with respect to other species (Fig. 1).

3.5. Phylogeny of FGLamide AST precursors in Arthropoda

The phylogeny revealed a relationship between the number of FGLamide ASTs and the number of introns. Species from the derived order Diptera with one or two introns possessed 4 or 5 FGLamide ASTs, whereas more basally diverging species such as Dictyoptera and *P. humanus corporis* lacked introns but had the largest number of FGLamide ASTs-13 or -14 and -6 respectively. It is noteworthy that in the tick, *I. scapularis*, the AST ORF did not possess introns and had 4 FGLamide ASTs whereas the crustacean *D. pulex* had one intron and 6 FGLamide ASTs. A similar pattern was observed in *A. pisum* which had one intron and 8 FGLamide AST (Fig. 2).

4. Discussion

This is the first report in which the evolution of introns of the FGLamide AST gene of different Arthropods was studied. The FGLamide AST ORF was analyzed and determined from genome databases. This strategy allowed us to establish FGLamide AST ORF organization without cloning. However, in some genome data-

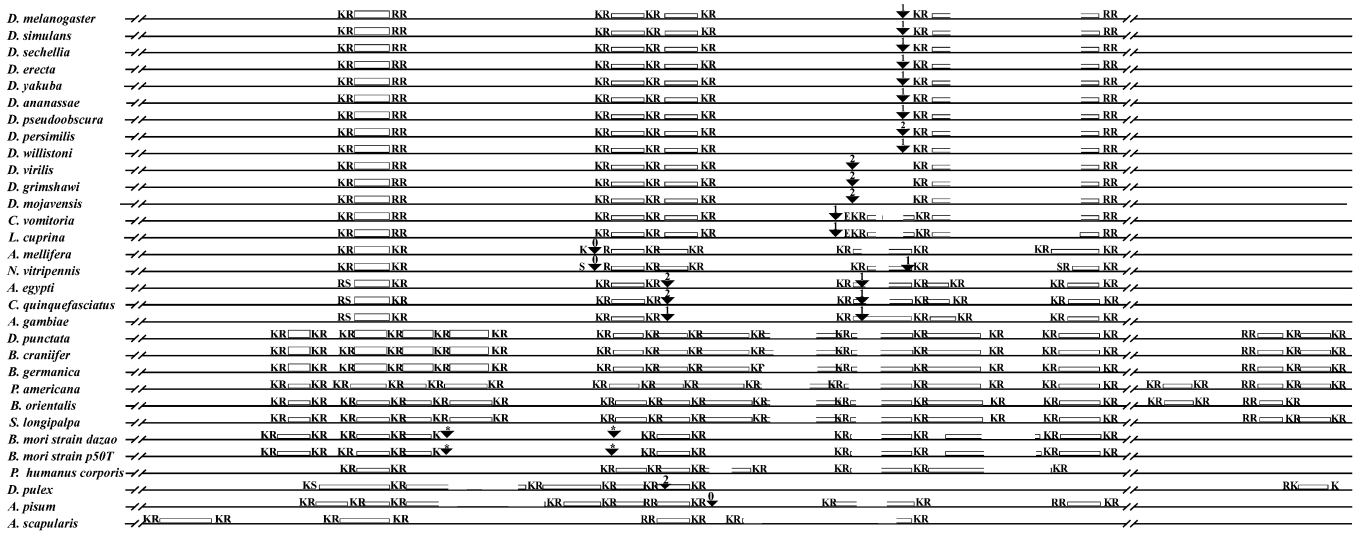

Fig. 1. Intron position in FGLamide AST ORFs. The graphic corresponds to the amino acid alignment from the precursors of FGLamide AST. The position of each intron is indicated by a black arrow head. The number at the top of each arrow represents intron phase. The position in the precursor of each FGLamide AST is indicated in white box. The dibasic sites for the proteolytic cleavage are shown with their respective amino acids.

bases, the FGLamide AST ORF had been identified with different nomenclature as in *D. erecta*, *D. persimilis*, *D. sechellia*, *D. virilis*, *D. willistoni* and *P. humanus corporis*. In other genomes in which the FGLamide AST ORF was unknown, it was necessary to employ other *in silico* approaches to establish the FGLamide AST gene structure, e.g. *N. vitripennis* and *A. pisum*. Although the *in silico* method is a powerful approach to determine gene structure, it is limited, particularly when the genome has not been completely sequenced or some sequences possess gaps. This problem was observed in the FGLamide AST gene from *B. mori*, in which we could determine the structure of the first and third exons but not the second exon, because there were gaps in one database, and in another database, the gene sequences was distributed across three GenBank accession numbers. Moreover, the genome of the beetle *Tribolium castaneum* was searched extensively but we were unable to find any putative FGLamide AST gene in the database. However, this result does not confirm the absence of the FGLamide gene because we found several gaps in the genome (data not shown). With the genome databases obtained, we determined that the tick *I. scapularis* and the louse *P. humanus corporis* do not have introns as was previously reported for the FGLamide AST genes in Dictyopteran species [2] using inverse PCR. Using *in silico* analysis and molecular biology approaches, the FGLamide AST gene from seven *Drosophila* species [4] as well as *C. vomitoria* and *L. cuprina* [7] revealed that the gene had two introns, the first in the 5′UTR (untranslated region) and the second within the ORF, between the third and fourth FGLamide AST. Similarly, we found that five *Drosophila* species and *A. mellifera* and *A. pisum* and one crustacean, *D. pulex* had one intron within the ORF. Moreover, we determined for first time the presence of two introns within the ORF of the FGLamide AST gene in three Dipteran species, as well as *N. vitripennis* and *B. mori*.

Table 2
FGLamide Allatostatin gene organization with one intron.

Organism	Exon A	Intron A	Exon B	Gene[a]	mRNA[a]
D. pulex	FGGNPTGDPNLNIYSFGLGKR 1 TSRSYSINPYSFGLGKR 2 GGNAKSYPQQIPYSFGLGKR 3 NPTKYNFGLGKR 4 P D 5	Phase2 88 bp. Codon 368 Arg AG|A	FGFGLGKR 5 LPVYNFGLGK 6	NCBI_GNO_0600073	NCBI_GNO_0600073
A. mellifera	LPVYNFGIGKR 1	Phase 0 1,033 bp. Codon 368 Lys-Arg AAG|AGA	GRDYSFGLGKR 2 RQYSFGLGKR 3 GRQPYSFGLGKR 4 PNDMLSQRYHFGLGKR 5	NW_001253130	NW_001253130
A. pisum	AHKQYGFGLGKR 1 LYRQYEFGLGKR 2 SASKQYGFGLGKR 3 AALKQYEFGLGKR 4 ASPTFYSFGLGRR 5 ASPQYSFGLGKR 6	Phase 0 3,174 bp. Codon 368 Arg-Val AGA|GTG	TADDMGHGQRFAFGLGKR 7 ARLQYGFGLGKR 8	EQ118335	This work
C. vomitoria	VERYAFGLGRR 1 AYTYTNGGNGIKRLPVYNFGLGKR 2 ARPYSFGLGKR 3	Phase 1 67 bp Codon 144 Asap G|AT	NRPYSFGLGKR 4 DPLNEERRANRYGFGLGRR 5	[7]	[7]
L. cuprina	VERYAFGLGRR 1 AYTYTNGGNGIKRLPVYNFGLGKR 2 ARPYSFGLGKR 3	Phase 1 63 bp Codon 144 Asap G|AT	NRPYSFGLGKR 4 DPLNEERRANSYGFGLGRR 5	[7]	[7]
D. ananassae	MERYAFGLGRR 1 LPVYNFGLGKR 2 SRPYSFGLGKR 3	Phase 1 50 bp Codon 132 Val G|TG	TTRPQPFNFGLGRR 4	[2]	[2]
D. erecta	VERYAFGLGRR 1 LPVYNFGLGKR 2 SRPYSFGLGKR 3	Phase 1 60 bp. Codon 129 Val G|TG	TTRPQPFNFGLGRR 4	AAPQ01006808	XM_001981909
D. grimshawi	MERYAFGLGRR 1 LPVYNFGLGKR 2 SRPYSFGLGKR 3	74 bp. Phase 2 Codon 129 Gly GG|C	TTRPQPFNFGLGRR 4	[2]	[2]
D. melanogaster	VERYAFGLGRR 1 LPVYNFGLGKR 2 SRPYSFGLGKR 3	Phase 1 61 bp. Codon 129 Val G|TG	TTRPQPFNFGLGRR 4	NT_033777.2	AF263923, NM_079765
D. mojavensis	MERYAFGLGRR 1 LPVYNFGLGKR 2 SRPYSFGLGKR 3	Phase 2 61 bp. Codon 139 Gly GG|C	TTRPQPFNFGLGRR 4	[2]	[2]

Table 2 (*Continued*)

Organism	Exon A	Intron A	Exon B	Gene[a]	mRNA[a]
D. persimilis	VERYAFGLGRR 1 LPVYNFGLGKR 2 SRPYSFGLGKR 3	64 bp. Phase 2 Codon 131 Ala GC|A	TTRPQPFNFGLGRR 4	AAIZ01003146	XM_002020181
D. pseudoobscura	VERYAFGLGRR 1 LPVYNFGLGKR 2 SRPYSFGLGKR 3	64 bp. Phase 1 Codon 131 Val G|TC	TTRPQPFNFGLGRR 4	[2]	[2]
D. sechellia	VERYAFGLGRR 1 LPVYNFGLGKR 2 SRPYSFGLGKR 3	60 bp. Phase 1 Codon 129 Val G|TG	TTRPQPFNFGLGRR 4	AAKO01001047	XM_002032516
D. simulans	VERYAFGLGRR 1 LPVYNFGLGKR 2 SRPYSFGLGKR 3	61 bp. Phase 1 Codon 129 Val G|TG	TTRPQPFNFGLGRR 4	[2]	[2]
D. virilis	VERYAFGLGRR 1 LPVYNFGLGKR 2 SRPYSFGLGKR 3	Phase 2 53 bp. Codon 134 Gly GG|C	TTRPQPFNFGLGRR 4	AANI01017106	XM_002052990
D. willistoni	VERYAFGLGRR 1 LPVYNFGLGKR 2 SRPYSFGLGKR 3	135 bp. Phase 1 Codon 134 Ala G|CT	TTRPQPFNFGLGRR 4	AAQB01009413	XM_002073224
D. yakuba	VERYAFGLGRR 1 LPVYNFGLGKR 2 SRPYSFGLGKR 3	74 bp. Phase 1 Codon 128 Val G|TG	TTRPQPFNFGLGRR 4	[2]	[2]

FGLamide Allatostatin gene organization with two introns.

	Exon 1	Intron 1	Exon 2	Intron 2	Exon 3	Gene	mRNA
A. aegypti	PKYNFGLGKR 1 LPHYNFGLGKR 2	Phase 2 57 bp. Codon 368 Arg AG|G	AS 3	Phase 1 8,412 bp. Codon 512 368 Ala G|CT	AYRYHFGLGKR 3 RVYDFGLGKR 4 LPNRYNFGLGKR 5	CH477340	AAU66841
A. gambiae	PKYNFGLGKRR 1 LPHYNFGLGKR 2	Phase 1 59 bp Codon 109 Ser T|CC	TASGNGAG 3	Phase 1 85 bp Codon 166 Ser T|CT	AYRYHFGLGKR 3 RAYDFGLGKR 4 LPNRYNFGLGKR 5	NT_078266	XM_313511
C. quinquefasciatus	PKYNFGLGKR 1 LPQYNFGLGKR 2	Phase2 128 bp. Codon 110 Arg AG|G	AS 3	Phase 1 85 bp Codon 154 Gly G|GA	RYHFGLGKR 3 RNYNFGLGKR 4 LPNRYNFGLGRR 5	DS232003	XM_001849994
N. vitripennis	LPIYQFGLGKR 1	Phase 0 216 bp. Codons 125-126 Ser-Arg AGC|CGA	RSQPFSFGLGKR 2 TRPYSFGLGKR 3 TGGFNFGL 4	Phase 1 139 bp Codon 203 Gly G|GT	KR 4 DKYLFGLGKR 5	DS265633	This Work
B. Mori	SPQYDFGLGKR 1 LPVYNFGLGKR 2	Phase ND >180 bp. Codons Arg	YYVACSQSRPYLFGLGKR	Phase ND >180 bp. Codons Arg	ARPYSFGLGKR ARMYSFGLGKR ARSYSFGLGKR LSSKFNFGLGKR QRDMHRFSFGLGKR	Exon 1: 46724731; Exon 2: 46629065; Exon 3: 46792320	NM_001043571
B. mori strain Dazao	SPQYDFGLGKR 1 LPVYNFGLGKR 2	Phase ND >1000 bp. Codons Arg	YYVACSQSRPYLFGLGKR	Phase ND >1000 bp. Codons Arg	ARPYSFGLGKR ARMYSFGLGKR ARSYSFGLGKR LSSKFNFGLGKR QRDMHRFSFGLGKR	CH386506	NM_001043571

| = intron position within of codon.
[a] GenBank accession number.

Table 3
FGLamide AST intron phases in Arthropoda.

Organism	AA	Codon intron	Phase	Size[a]
A. mellifera	Lys|Arg	AAG|AGA	Zero	1033
A. pisum	Arg|Val	AGA|GTG	Zero	3174
N. vitripennis I1	Ser|Arg	AGC|CGA	Zero	216
A. gambiae I1	Ser	T|CC	One	59
A. gambiae I2	Ser	T|CT	One	85
L. cuprina	Asp	G|AT	One	63
C. vomitoria	Asp	G|AT	One	67
D. willistoni	Ala	G|CT	One	135
A. aegypti I2	Ala	G|CT	One	8412
D. persimilis	Ala	GC|A	Two	64
D. pulex	Arg	AG|A	Two	88
A. aegypti I1	Arg	AG|G	Two	57
C. quinquefasciatus I1	Arg	AG|G	Two	128
C. quinquefasciatus I2	Gly	G|GA	One	85
N. vitripennis I2	Gly	G|GT	One	139
D. virilis	Gly	GG|C	Two	53
D. mojavensis	Gly	GG|C	Two	61
D. grimshawi	Gly	GG|C	Two	74
D. ananassae	Val	G|TG	One	50
D. erecta	Val	G|TG	One	60
D. sechellia	Val	G|TG	One	60
D. melanogaster	Val	G|TG	One	61
D. simulans	Val	G|TG	One	61
D. yakuba	Val	G|TG	One	74
D. pseudoobscura	Val	G|TC	One	64

| = intron position. *I1* and *I2*: Intron 1 and Intron 2 respectively.
[a] Size in base pairs.

In all the exons analyzed, at least one FGLamide AST was present, and the majority of the introns showed donor and acceptor sites and a branch sequence characteristic of intron class U2, which are excised from pre-RNA by spliceosomes in eukaryotic nuclei [1]. Interestingly, the principal introns were phase 1 or 2 and they were located near or beside codons encoding Arg or Lys. Furthermore 84% of introns were localized between guanine and other nucleotides. It has been proposed that the nucleotides GG, GC or CC are potential sites for insertion of introns, based on homology with acceptor and donor sites in the protosplice sites [1,41,43]. Although alternative splicing has not been demonstrated for the FGLamide AST gene, it is possible that this phenomenon could occur in species with introns in the ORF as in other eukaryotes [3,18]. As well, the presence of these types of introns in the FGLamide AST gene strongly suggests that some regulation of splicing is necessary to obtain mRNA that encodes the prepropeptide of all FGLamide ASTs.

Two models have been proposed to explain the origin of introns in eukaryotic genes: Model 1—the introns early were present in the last universal common ancestor of pro- and eukaryotes, and that intron loss is responsible for the lack of introns observed in bacteria; and Model 2—the introns late have appeared during the evolution of the eukaryotic lineage, and intron gain is a frequent event in the evolution leading to the gene structures we see today [23,31,33,34].

Studies have shown that intron positions are conserved through eukaryotic evolution [14]. Taking into consideration invertebrate

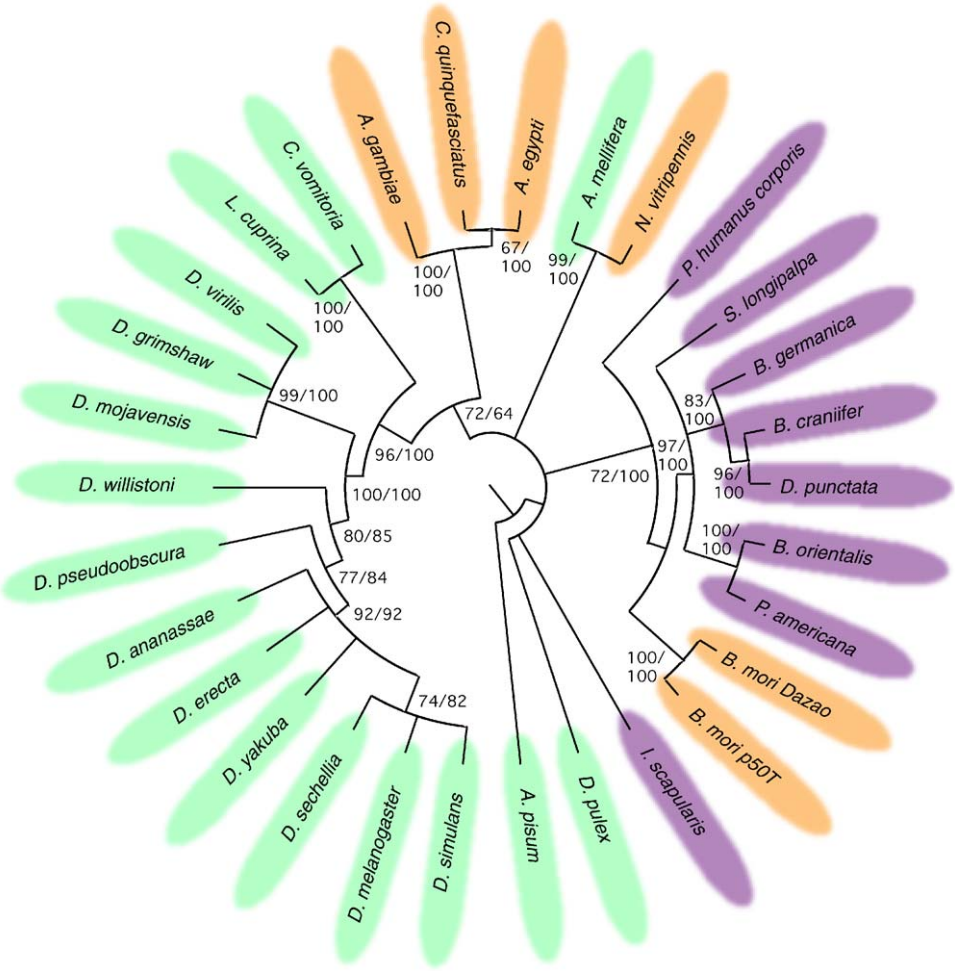

Fig. 2. Maximum likelihood phylogeny of FGLamide AST precursor proteins. Numbers above each node represent likelihood bootstrap percentages (100 replicates) followed by Bayesian posterior probabilities, with all nodes having <60% bootstrap support collapsed. Genes with one intron are marked in green; two introns in orange; zero introns in purple. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

palaeontology records, the order Dictyoptera and the class Arachnida are some of the oldest arthropods [5,10,16,26] and neither group have introns in the FGLamide AST ORF, whereas more recent orders such as Diptera, Lepidoptera, Hemiptera and Hymenoptera [9,10,29] have one or two introns. This might suggest that the presence of introns in insect species is consistent with the introns late model, but the presence of one small intron in the crustacean *D. pulex* which is probably older than Insecta and Arachnida [10] could argue against this suggestion. All the introns analyzed share the position and phase exclusively between genera and species. However, no common intron position was found with respect to *D. pulex*.

Neuropeptide genes in both invertebrates and vertebrates can be intronless in one species, but have one or more introns in the ORF of other species. For example, the ORF of the adipokinetic hormone (AKH) gene is intronless in *Manduca sexta* whereas in *D. melanogaster*, *Schistocerca gregaria* and *Schistocerca nitans*, the ORF has one intron in different positions [25]. Additionally, gonadotropin-releasing hormone II is intronless in the tunicate *Ciona intestinalis* and in the sea lamprey *Petromyzon marinus*, but in other vertebrates, the gene has three introns and four exons [17].

Our results strongly suggest that, independent of classification, the introns in the FGLamide AST gene of the arthropods are recent, and were inserted preferentially in codons Gly or guanine-rich regions near codons for the dibasic proteolytic cleavage site Lys-Arg.

Acknowledgments

Francisco José Martínez Pérez was recipient of postdoctoral fellowship from CONACYT (77390) and he is the member of S.N.I. This research was supported in part by Discovery grants from the Natural Sciences and Engineering Research Council of Canada (SST, BSWC and WGB).

References

[1] Basu MK, Rogozin IB, Koonin EV. Primordial spliceosomal introns were probably U2-type. Trends Genet 2008;24:525–8.
[2] Bellés X, Graham LA, Bendena WG, Ding QI, Edwards JP, Weaver RJ, et al. The molecular evolution of the allatostatin precursor in cockroaches. Peptides 1999;20:11–22.
[3] Ben-Dov C, Hartmann B, Lundgren J, Valcárcel J. Genome-wide analysis of alternative pre-mRNA splicing. J Biol Chem 2008;283:1229–33.
[4] Bowser PRF, Tobe SS. Comparative genomic analysis of allatostatins encoding (*Ast*) genes in *Drosophila* species and prediction of regulatory elements by phylogenetic footprinting. Peptides 2007;28:83–93.
[5] de la Fuente J. The fossil record and the origin of ticks (Acari: Parasitiformes: Ixodida). Exp Appl Acarol 2003;29:331–44.
[6] Duve H, Johnsen AH, Maestro JL, Scott AG, East PD, Thorpe A. Identification of the dipteran Leu-callatostatin peptide family: the pattern of precursor processing revealed by isolation studies in *Calliphora vomitoria*. Regul Pept 1996;67:11–9.
[7] East P, Tregenza K, Duve H, Thorpe A. Identification of the dipteran Leu-callatostatin peptide family: characterisation of the prohormone gene from *Calliphora vomitoria* and *Lucilia cuprina*. Regul Pept 1996;67:1–9.
[8] Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 1985;39:783–91.
[9] Gaunt MW, Miles MA. An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks. Mol Biol Evol 2002;19:748–61.
[10] Glenner H, Thomsen PF, Hebsgaard MB, Sørensen MV, Willerslev E. Evolution. The origin of insects. Science 2006;314:1883–4.
[11] Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 2003;52:696–704.
[12] Guindon S, Lethiec F, Duroux P, Gascuel O. PHYML online—a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res 2005;33:W557–9.
[13] Hult EF, Weadick CJ, Chang BSW, Tobe SS. Reconstruction of ancestral FGLamide-type insect allatostatins: a novel approach to the study of allatostatin function and evolution. J Insect Physiol 2008;54:959–68.
[14] Irimia M, Roy SW. Spliceosomal introns as tools for genomic and evolutionary analysis. Nucleic Acids Res 2008;36:1703–12.
[15] Kai ZP, Ling Y, Liu WJ, Zhao F, Yang XL. The study of solution conformation of allatostatins by 2-D NMR and molecular modeling. Biochim Biophys Acta 2006;1764:70–5.
[16] Kambhampati S. A phylogeny of cockroaches and related insects based on DNA sequence of mitochondrial ribosomal RNA genes. Proc Natl Acad Sci U S A 1995;92:2017–20.
[17] Kavanaugh SI, Nozaki M, Sower SA. Origins of gonadotropin-releasing hormone (GnRH) in vertebrates: identification of a novel GnRH in a basal vertebrate, the sea lamprey. Endocrinology 2008;149:3860–9.
[18] Kim E, Goren A, Ast G. Alternative splicing: current perspectives. BioEssays 2008;30:38–47.
[19] Kumar S, Dudley J, Nei M, Tamura K. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief Bioinform 2008;9:299–306.
[20] Kwok R, Zhang JR, Tobe SS. Regulation of methyl farnesoate production by mandibular organs in the crayfish, *Procambarus clarkii*: a possible role for allatostatins. J Insect Physiol 2005;51:367–78.
[21] Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics 2007;23:2947–8.
[22] Le SQ, Gascuel O. An improved general amino acid replacement matrix. Mol Biol Evol 2008;25:1307–20.
[23] Long M, Rosenberg C, Gilbert W. Intron phase correlations and the evolution of the intron/exon structure of genes. Proc Natl Acad Sci U S A 1995;92:12495–9.
[24] Lorenz MW, Kellner R, Hoffmann KH. Identification of two allatostatins from the cricket, *Gryllus bimaculatus* de geer (Ensifera, Gryllidae): additional members of a family of neuropeptides inhibiting juvenile hormone biosynthesis. Regul Pept 1995;57:227–36.
[25] Martínez-Perez F, Becerra A, Valdes J, Zinker S, Arechiga H. A possible molecular ancestor for mollusk APGWamide, insect adipokinetic hormone, and crustacean red pigment concentrating hormone. J Mol Evol 2002;54:703–14.
[26] Maxmen A, Browne WE, Martindale MQ, Giribet G. Neuroanatomy of sea spiders implies an appendicular origin of the protocerebral segment. Nature 2005;437:1144–8.
[27] Nicholas KB, Nicholas Jr HB, Deerfield DWII. GeneDoc: analysis and visualization of genetic variation. EMBNEW News 1997;4:14.
[28] Patthy L. Intron-dependent evolution: preferred types of exon and introns. FEBS Lett 1987;214:1–7.
[29] Poinar Jr GO, Danforth BN. A fossil bee from early cretaceous Burmese amber. Science 2006;314:614.
[30] Rambaut A, Drummond AJ., Tracer v1.4. 2007. Available from http://beast.bio.ed.ac.uk/Tracer.
[31] Rodríguez-Trelles F, Tarrío R, Ayala FJ. Models of spliceosomal intron proliferation in the face of widespread ectopic expression. Gene 2006;366:201–8.
[32] Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 2003;19:1572–4.
[33] Roy SW, Gilbert W. The evolution of spliceosomal introns: patterns, puzzles and progress. Nat Rev Genet 2006;7:211–21.
[34] Ruvinsky A, Eskesen ST, Eskesen FN, Hurst LD. Can codon usage bias explain intron phase distributions and exon symmetry? J Mol Evol 2005;60:99–104.
[35] Stay B. A review of the role of neurosecretion in the control of juvenile hormone synthesis: a tribute to Berta Scharrer. Insect Biochem Mol Biol 2000;30:653–62.
[36] Stay B, Tobe SS. The role of allatostatins in juvenile hormone synthesis in insects and crustaceans. Ann Rev Entomol 2007;52:277–99.
[37] Tamura K, Dudley J, Nei M, Kumar S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 2007;24:1596–9.
[38] Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence weighing, position specific gap penalties and weighting matrix chose. Nucleic Acids Res 1994;22:4673–80.
[39] Tobe SS, Bendena WG. Allatostatins in the insects. In: Kastin AJ, editor. Handbook of biologically active peptides 2006. Burlington: Elsevier Academic Press; 2006. p. 201–6.
[40] Tobe SS, Bendena WG. The regulation of juvenile hormone production in arthropods. Functional and evolutionary perspectives. Ann N Y Acad Sci 1999;897:300–10.
[41] Tordai H, Patthy L. Insertion of spliceosomal introns in proto-splice sites: the case of secretory signal peptides. FEBS Lett 2004;575:109–11.
[42] Tweedie S, Ashburner M, Falls K, Leyland P, McQuilton P, Marygold S, et al. FlyBase Consortium. FlyBase: enhancing Drosophila gene ontology annotations. Nucleic Acids Res 2009;37:D555–9.
[43] Vibranovski MD, Sakabe NJ, de Souza SJ. A possible role of exon-shuffling in the evolution of signal peptides of human proteins. FEBS Lett 2006;580:1621–4.
[44] Woodhead AP, Stay B, Seidel SL, Khan MA, Tobe SS. Primary structure of four allatostatins: neuropeptide inhibitors of juvenile hormone synthesis. Proc Natl Acad Sci U S A 1989;86:5997–6001.
[45] Yagi KJ, Kwok R, Chan KK, Setter RR, Myles TG, Tobe SS, et al. Phe-Gly-Leu-amide allatostatin in the termite *Reticulitermes flavipes*: content in brain and corpus allatum and effect on juvenile hormone synthesis. J Insect Physiol 2005;51:357–65.