SCIENTIFIC REP♦RTS

OPEN

The Evolutionary History of The Orexin/Allatotropin GPCR Family: from Placozoa and Cnidaria to Vertebrata

María Eugenia Alzugaray[1,2], María Cecilia Bruno[1,2], María José Villalobos Sambucaro[1,2] & Jorge Rafael Ronderos[1]

Peptidic messengers constitute a highly diversified group of intercellular messengers widely distributed in nature that regulate a great number of physiological processes in Metazoa. Being crucial for life, it seem that they have appeared in the ancestral group from which Metazoa evolved, and were highly conserved along the evolutionary process. Peptides act mainly through G-protein coupled receptors (GPCRs), a family of transmembrane molecules. GPCRs are also widely distributed in nature being present in metazoan, but also in Choanoflagellata and Fungi. Among GPCRs, the Allatotropin/Orexin (AT/Ox) family is particularly characterized by the presence of the DRW motif in the second intracellular loop (IC Loop 2), and seems to be present in Cnidaria, Placozoa and in Bilateria, suggesting that it was present in the common ancestor of Metazoa. Looking for the evolutionary history of this GPCRs we searched for corresponding sequences in public databases. Our results suggest that AT/Ox receptors were highly conserved along evolutionary process, and that they are characterized by the presence of the E/DRWYAI motif at the IC Loop 2. Phylogenetic analyses show that AT/Ox family of receptors reflects evolutionary relationships that agree with current phylogenetic understanding in Actinopterygii and Sauropsida, including also the largely discussed position of Testudines.

Cell-cell communication is a basic principle in all organisms, necessary to facilitate the coordination and integration between cell populations, and with their environment. Indeed, integrative mechanisms as nervous and endocrine systems have appeared early along the evolutionary process and play a very important role, regulating many physiological processes in all animal phyla. As it is known, these systems act by mean of messengers which can be basically grouped as hormones and neuromodulators. Among these chemical messengers, peptides constitute a highly diversified group of molecules widely distributed in nature, and regulate a great number of physiological processes in most groups of Metazoa, from cardiac and visceral muscle activity, to more complex phenomena as sleep-wakefulness, and appetite.

Being this family of messengers crucial for life, it would seem that they have appeared in the ancestral group from which Metazoa evolved, and became highly conserved along the evolutionary process. Indeed, peptidic messengers are present in *Hydra sp.* and others members of the phylum Cnidaria[1–4], as well as in *Trichoplax adhaerens*[5–7], a member of the neuron-less animal phylum Placozoa, that also shares a common ancestor with Bilateria.

Peptides act mainly through G-protein coupled receptors (GPCRs), a complex and ubiquitous family of transmembrane molecules. GPCRs are widely distributed in Vertebrata, but also, this family of proteins, have been proved to be present in all metazoan, including Placozoa, Cnidaria, Ctenophora and Porifera, which share a common ancestor with Bilateria; also in Choanoflagellata (a group of unicellular eukaryotes related with metazoans), and even in Fungi[1–3,8–11].

GPCRs are characterized by the presence of seven transmembrane (TM) domains, an extracellular N-terminal and an intracellular C-terminal domains. The transmembrane domains are linked by three extracellular and three

---

[1]Cátedra de Histología y Embriología Animal, Facultad de Ciencias Naturales y Museo, Universidad Nacional de La Plata (FCNyM-UNLP), La Plata, Argentina. [2]Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina. Correspondence and requests for materials should be addressed to J.R.R. (email: jrondero@museo.fcnym.unlp.edu.ar)

1

intracellular loops (for a review see[12,13]). GPCRs are usually grouped in five major families, named *Rhodopsin*, *Frizzled*, *Glutamate*, *Adhesion* and *Secretin*[14]. Among these, the *Rhodopsin* family seems to be the most widely distributed in Metazoa and it is particularly characterized by the existence of a **E/D**R motif associated to the third transmembrane domain (TM III) (i.e. IC Loop 2), which seems to be relevant for the transmission of the message, facilitating the activity of the associated G-proteins[13,14].

A vast number of the *Rhodopsin* family of receptor presents, as a conserved feature, the **E/D**RY/F motif[14,15]. In spite of that, a more limited number show the presence of a Tryptophan (W) instead that a Tyrosine (Y) residue (i.e. **E/D**RW). Among these, we found the receptors corresponding to the Allatotropin (AT) family of peptides[16].

AT is a neuropeptide originally isolated and characterized in insects on the basis of its ability to modulate the synthesis of Juvenile Hormones (JHs) in the gland corpora allata (CA) of the moth *Manduca sexta* (Lepidoptera: Insecta)[17]; and some other holometabolous species like the mosquito *Aedes aegypti*[18,19]. Beyond the first biological function assigned, AT has proved to have multiple functions, including modulation of digestive enzymes secretion, and ion exchange regulation in the digestive system of Lepidotera[20,21]. As a pleiotropic peptide, AT has also shown to be involved in myoregulatory processes, stimulating foregut movements in Lepidoptera[22]; and of the hindgut and midgut of both Chagas' disease vectors *Triatoma infestans* and *Rhodnius prolixus* (Insecta: Hemiptera)[23–25]. Furthermore, AT has proved to have cardioacceleratory functions synergizing the activity of serotonin in these species[24,25]. In spite that AT was originally characterized as a neuropeptide (i.e. secreted by neurons at the central nervous system), it is also secreted by epithelial cells of the Malpighian tubules, and open-type cells at the level of the digestive system, acting in a paracrine and also endocrine way[25–28].

Looking for the evolutionary origin of allatoregulatory peptides, Alzugaray *et al.*[1,2] have suggested that the AT/Ox and AST-C/somatostatin signalling systems are present in *Hydra sp*., a fresh water member of the phylum Cnidaria, playing myoregulatory roles during feeding, and modulating cytosolic Ca^{2+} levels[3]. Indeed, it was suggested that the allatotropic function of this peptides would constitute an insect synapomorphy, and that the ancestral function of these peptides could be myoregulatory[1,29–31].

On the basis of a transcriptomic analysis performed in the CA/corpora cardiaca complex of the silkworm *Bombyx mori* the AT receptor (ATr) was identified[32]. Afterward, the receptor of AT in other species of Lepidoptera as *M. sexta*[33] and *Helicoverpa armigera*[34], as well as in hemimetabolous species as *Rhodnius prolixus* was also characterized[24]. Indeed, it was confirmed that the receptor pertains to the *Rhodopsin* family of GPCRs, sharing a 48% of identity with the orexin receptor of vertebrates in the region comprised between the TMI and TMVII domains[33]. Moreover ATr shares with orexin receptors the characteristic DR**W** motif[16,35].

Orexins (Ox), also named Hypocretins[36], originally identified in neurons located at the level of the hypothalamus in the rat, are two peptides sharing structural characteristics, derived from a same precursor by proteolytic processing[35,36]. Initially related with physiological mechanisms regulating feeding behaviour, the activity of these peptides was posteriorly associated with mechanisms regulating wakefulness and sleep (for a review see[37]), and also with peripheral tissues activities. In fact, the presence of Ox and their receptors in the enteric nervous system, as well as at the level of the mucosa and smooth muscle of the digestive tract of mammals was also shown, suggesting that they also act as myoregulators[38,39].

AT and Ox peptides are structurally different. Interestingly, bioinformatic search doesn´t show the presence of Ox in protostomates as well as AT in Deuterostomata, being possibly that, beyond the similarity between both receptors, Ox has evolved only in Deuterostomata and AT in Protostomata[1,29,30]. In fact, due that homology-based searches are often not sensitive enough to detect precursors of small peptides[5] and the difficulties to look for orthologues at the level of peptides, homologies between signal systems some times are based on their receptors[1,40].

Looking for the evolutionary history of these signalling systems, we decided to go deeper in the analysis of these families of GPCRs (i.e. AT and Ox receptors). Based on fully characterized receptors both in vertebrates as well as in insects, we looked at the GenBank for putative AT/Ox receptors in all metazoan phyla. We have found sequences that might be considered AT/Ox GPCRs in several phyla including, Placozoa, Cnidaria, Mollusca, and Brachiopoda. On the basis of multiple sequence alignment we found motifs that might be considered "signatures" of the AT/Ox family of GPCRs. Phylogenetic analysis suggested that these families of receptors would be present in the ancestor of Metazoa, and that the system was highly conserved along evolutionary process. Moreover, a detailed maximum likelihood (ML) analysis of groups like Actinopterygii and Sauropsida, reflects phylogenetic trees that agree with current understanding of their phylogenetic relationships, including also the largely discussed evolutionary position of Testudines.

Results

The Allatotropin/Orexin receptors ancestral signature. As it is described above, GPCRs are characterized by the presence of the **E/D**R motif associated to the TMIII (i.e. IC Loop 2). Based on fully characterized AT and Ox receptors we looked in the GenBank for sequences in all animal phyla. After the analysis of 392 complete sequences, including N-terminal, C-terminal and the presence of 7 TM domains, we found that the motif **E/D**R**W**YI in the IC Loop 2 can be tracked from Chordata and Arthropoda, to Cnidaria and Placozoa. The most frequent motif found is **DRWYAI**, being present in 374 sequences, including the ancestral species *T. adhaerens* (Placozoa) (Table 1; Supporting Information File 1). The analysis of the rest of the sequences (eighteen), shows that seven of them exhibit only one conservative change, presenting **E**RWYAI corresponding to sequences of phyla pertaining to Lophotrochozoa (i.e. Mollusca, Brachiopoda and Annelida). The comparison of the codons codifying for the asparctic acid (D) and glutamic acid (E), shows that a point mutation at the third position of the codon would be responsible of this conservative change. A particular situation is presented in *H. vulgaris* (Cnidaria: Hydrozoa) in which the Tyrosine (Y) residue is substituted by asparagine (N), being the only sequence analysed showing this conformation (i.e. ERW**N**AI). A point mutation at the first position of the codon should be responsible, and it has previously proposed as a sequence artefact[3].

Phylum	Signature
Placozoa	**D/ERWYAI/V**
Cnidaria	**ERWYAI/V**
Brachiopoda	**ERWYAI**
Annelida	**ERWYAI**
Mollusca	**ERWYAI**
Arthropoda	**DRWYAI**
Chordata	**DRWYAI**

Table 1. Characteristic Allatotropin/Orexin signature located at the interphase between transmembrane domains 3 (TMIII) and the second intracellular loop (IC loop 2) distribution for every phylum analysed.
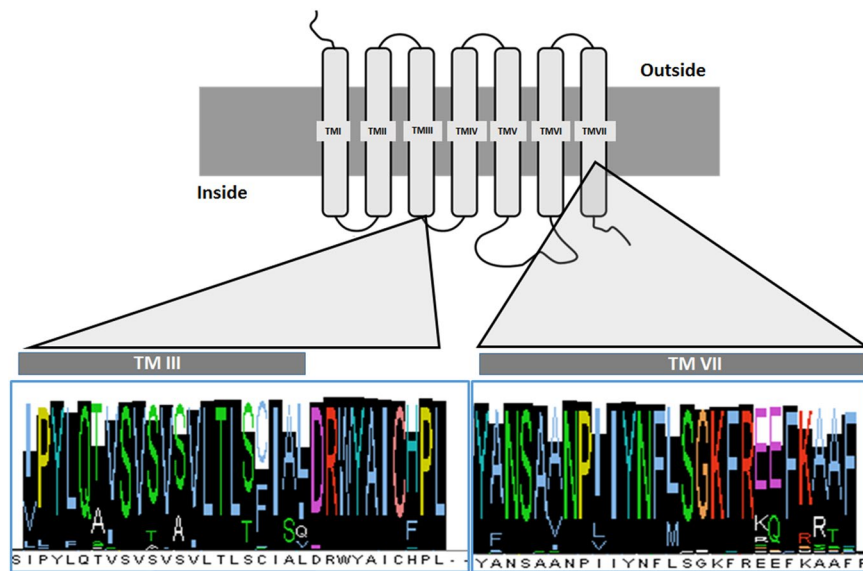

Figure 1. Schematic view of a generalized GPCR showing the two highly conserved domains, and the corresponding consensus after a multiple sequence alignment of sequences pertaining to the Allatotropin/orexin family of receptors. The alignment includes species pertaining to Placozoa, Cnidaria, Arthropoda, Mollusca, Annelida, Brachiopoda and Chordata.

Predicted sequences and general relationships between the animal phyla. As a result of a multiple sequence alignment, it also seems clear that at least two region of the AT/Ox receptor were highly conserved. One comprising the third transmembrane domain and its associated intracellular loop, and the second one comprising the TMVII (Fig. 1).

As a first approach to understand the relationships between the total sequences analysed, a Neighbour-Joining analysis were performed (Fig. 2). The analysis shows that, as might be predicted, Placozoa (two sequences) and Cnidaria (three sequences pertaining to two different species of Anthozoa), clusters together sharing a common ancestor. Interestingly, the only sequence fitting the characteristics of the AT/Ox family of GPCR in *Hydra vulgaris* (Cnidaria: Hydrozoa) is clustered alone as the sister group of Bilateria (Fig. 2).

Despite of several genomes of the phylum Nematoda are fully sequenced, none of the GPCR sequences found in the GenBank showed the **DRWY** motif, suggesting that the AT/Ox system is not present in this phylum. Similar situation was found for the other two groups of Metazoa with uncertain positions as Porifera and Ctenophora.

Mammals is the only group of organisms in which the existence of two different kind of receptors was proved (i.e. Type 1 and Type 2), suggesting that the presence of these two receptors constitutes a synapomorphy of this group. Interestingly, *Lepisosteus oculatus*, pertaining to the group of Lepisosteiformes (with only six extant species), representing together with Halecostomi, the extant groups of Neopterygii, also presents two sequences, sharing the same clade with Type 1 receptor of Mammals. A more detailed analysis performed with ML methodology (see Fig. 3) also shows that these two sequences fit in the same clade, suggesting that the Type 1 Ox receptor appeared at least twice along the evolutionary history of Vertebrata (Fig. 2).

Finally, the three best represented groups (i.e. Arthropoda, and the Type 1 and 2 receptors of Vertebrata) can be recognized at least by a highly conserved motif at the level of the interphase between TMIII and the second intercellular loop (see Table 2 and Supporting Information File 1).

Evolutionary history of orexin receptors in vertebrates. As previously stated, there exist two types of receptors in Vertebrate (i.e. Type 1 and Type 2). A ML analysis clearly divides the groups analysed in two clades based on the Type 1 and Type 2 characteristics (Fig. 3). As we described above, two out of three sequences
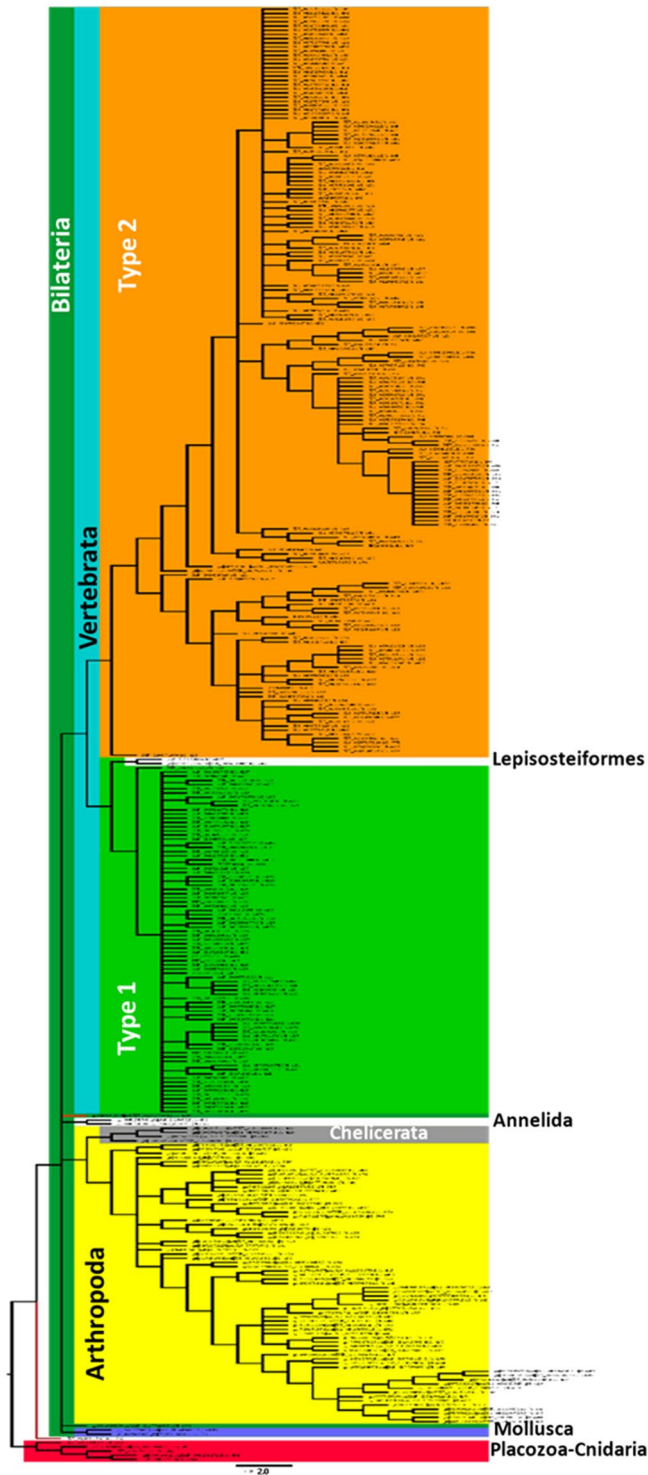
Figure 2. Evolutionary history of the Allatotropin/orexin family of receptors. All the sequences included present the seven transmembrane domains and the corresponding N-terminal and C-terminal domains. The tree was inferred by the Neighbour-Joining method. The cut-off value of replicate trees in which the associated taxa clustered together after a bootstrap test (1000 replicates) was 50%.

predicted for *L. oculatus* are grouped in the same clade of Type 1 receptor of Mammals. The other one (accession number XP_006638920) is grouped as a Type 2 receptor in the Actinopterygii clade (Fig. 3).

Regarding the Type 2 group, beyond that the Sarcopterygii are not grouped as a clade, showing Coelacanthimorpha, Amphibia, and the rest of tetrapoda a common ancestor with Actinopterygii and Chondrichthyes, the more represented groups (i.e. Mammals, Acitnopterygii and Sauropsida) are well defined as monophyletic groups (Fig. 3).
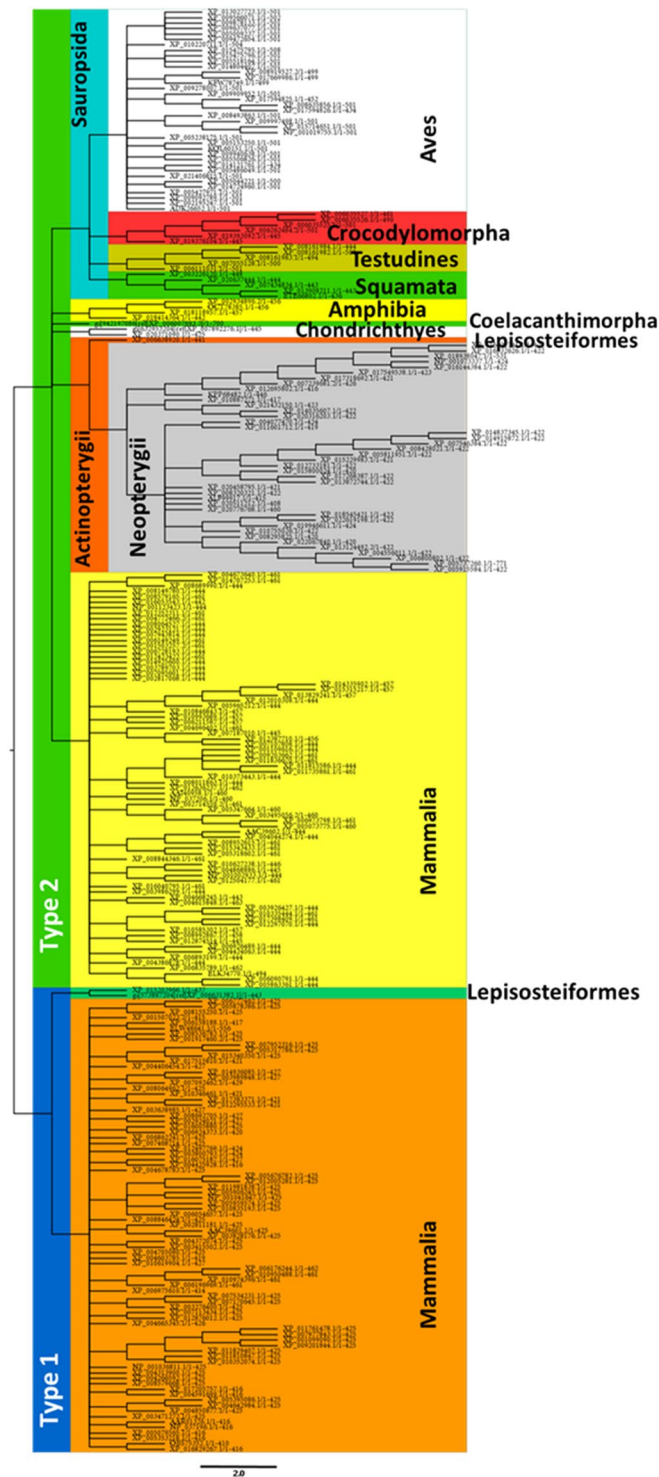
Figure 3. Phylogenetic relationships of Vertebrata. The tree was inferred by the Maximum Likelihood method. The cut-off value of replicate trees in which the associated taxa clustered together after a bootstrap test (1000 replicates) was 50%. Note that both kind of orexin receptors (Type 1 and Type 2) group independently. Type 2 receptor is present in all the groups of vertebrates included in the analysis. Type 1 receptor is only present in mammals with the exception of Lepisosteiformes (Actinopteygii: Neopterygii), suggesting that this kind of receptor could have appeared more than once along the evolution of Vertebrata.

Sauropsida.   As a first attempt to further understand the evolutionary history of the Ox receptor family, we decided to go deeper in the analysis of two groups of vertebrates well represented in our sample, as Sauropsida and Actinopterygii are, looking also for signatures motifs for every group analysed. In fact, after a detailed analysis of

Vertebrata	Type 1	TMIII – IC Loop 2	FIALDRWYAICHPL
Vertebrata	Type 2	TMIII – IC Loop 2	CIAL/QDRWYAICHPL
Arthropoda		TMIII-IC Loop 2	FISI/L/VDRWYAIC
Chondrichthyes		N - Terminal	ADYDDEFI
Teleostei - Coelacanthiformes		TMII – EC Loop 1	CLPASLVVDITET
Cyprinodontiformes	Poeciliidae	N - Terminal	YPAHGGNDTGSR
Cyprinodontiformes	Poeciliidae	N - Terminal	WTDYLHPKEYEW
Cyprinodontiformes	Poeciliidae	IC Loop 3	QRNWRTIQCS
Cyprinodontiformes	Rivulidae	IC Loop 3	RTLRCSAQT
Cyprinodontiformes		N - Terminal – TMI	YLHPKEYEWVLIVAYI
Cyprinodontiformes		C - Terminal	DNLSRLSDQ
Cichliformes		IC Loop 3	IKCSAPTPGP
Cichliformes		N - Terminal	LSSGHLPNSTELHVHPTL
Cichliformes		C – Terminal	RRIRTRTRTDSRKSLSTQVHNV
Protacanthopterygii		C – Terminal	KFRAEFKA
Protacanthopterygii	Salmoniformes	TMIV	SILLIWGVSC
Protacanthopterygii	Salmoniformes	IC Loop 1	KNHHMRTVTNCF
Protacanthopterygii	Salmoniformes	IC Loop 1	CEERWGADV
Protacanthopterygii	Salmoniformes	IC Loop 3	TSSVLQRKRT
Protacanthopterygii	Salmoniformes	EC Loop 4	FKYTNSRETVY
Cypriniformes	Cyprinidae	IC Loop 3	QCSAHAVGS
Osteoglossiformes - Ostarioclupeomorpha		EC Loop 3 – TMVII	NRET/AVYAWFT
Squamata	Serpentes	TMV – IC Loop3	APLCLMVLAYLQIFQKLWCQQ
Squamata	Iguania	TMV – IC Loop3	YMAPLCLMVLAYLQIFQKLWC
Testudines		C – Terminal	TNMSTLPANG
Testudines		IC Loop 3	PLPSLAQPR
Archosauria - Testudines		C – Terminal	ASTESRKSLTTQISNFDN
Crocodylomorpha		N - Terminal	NWSSIPELNE
Crocodylomorpha		N - Terminal	PSTDYDDEEFLRYL
Crocodylomorpha		IC Loop 3	IVQRKWKPLQFSAQP
Crocodylomorpha		C – Terminal	CGIHHHQD
Aves		N- Terminal – TMI	YEWALIAGYIVVFIVA
Aves - Passeriformes		C – Terminal	TSNIDEAM
Aves - Passeriformes	Pipridae	C – Terminal	VLNPSKSME
Aves - Passeriformes	Pipridae	C – Terminal	MTVSAEDTLN
Aves - Passeriformes	Pipridae	C – Terminal	LAEHVVLTN
Aves - Passeriformes	Paridae	C – Terminal	LSEQVALSNV

Table 2. Putative signatures motifs and their location along the primary structure of the protein for AT/Ox GPCRs in different taxonomic groups. Note that most of the signatures are located at the level of the C-terminal domain (29.7%) and the N-terminal domain (21.6%).

the alignments for each group, we could find signature motifs, that once blasted in the GenBank, remitted specifically to most of the groups under study (Table 2).

A ML analysis of Sauropsida shows two well supported clades conformed, one of them by Lepidosauromorpha species, including those corresponding to Iguania and Serpentes, traditionally grouped in the order Squamata, and the second one, conformed by Archosauria and Testudines (Fig. 4). Regarding Squamata, sequences in the TMV – IC Loop 3 seems to be characteristic, showing Serpentes and Iguania the motifs **APLCLMVLAYLQIFQKLWCQQ** and **YMAPLCLMVLAYLQIFQKLWC** respectively (Table 2).

As would be expected, Archosauria presents a well-defined phylogenetic pattern involving Crocodylomorpha, with species representing the three extant groups (i.e. Gavialoidea; Alligatoroidea and Crocodyloidea) and Aves. The clade including Crocodylomorpha seems to be characterized for four different signature motifs; two located at the N-terminal domain, one corresponding to the C-terminal, and a 4th one in the IC Loop 3 (see Table 2). With respect to the birds, a sequence located in the interphase between N-terminal and TMI would act as a signature (Table 2).

The clade corresponding to Aves, currently accepted as members of Coelurosauria (Dinosauria: Saurischia), shows the sequence **YEWALIAGYIVVFIVA** in the interphase N-terminal – TMI, fully conserved (Table 2). With respect to the phylogenetic relationships, the main groups are represented and grouped as well, including Paleognathae (*Tinamus guttatus*), and Neognathae which in fact form two well supported clades including Galloanserae and Neoaves (Fig. 4). Moreover, the two groups of Galloanserae are represented by four species pertaining to different genus, grouped in the expected clades. In fact, *Anser cygnoides* and *Anas platythynchos* (Anseriformes), and *Coturnix japonica* and *Gallus gallus* (Galliformes) form two monophyletic groups.
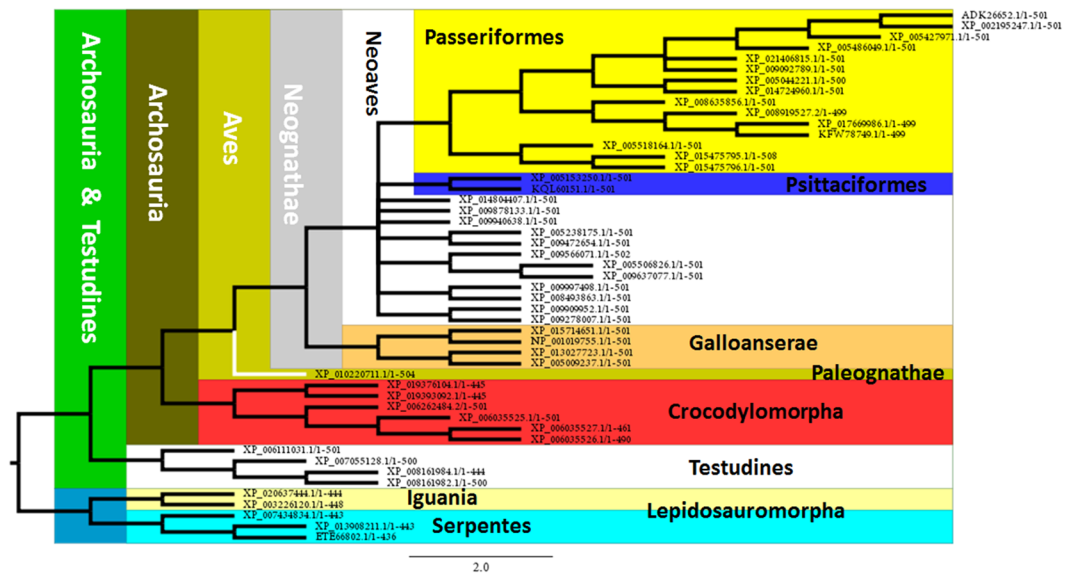
Figure 4. Maximum Likelihood analysis of Sauropsida. The phylogeny is clearly represented showing Lepidosauromorpha as the sister group of Archosauria. The main groups of Aves are also represented. Two orders of Neoaves (Passeriformes and Psittasiformes) are recognized. Furthermore, in Passeriformes, the best represented group, two families can be recognized by signature motifs. Testudines appears as the sister group of Archosauria in agreement with the current accepted hypothesis that recognize them as Diapsida, resembling also the currently proposed group of Archelosauria.

Regarding the Neoaves, only two currently recognized orders, Psittaciformes (represented by two species) and Passeriformes, are well defined (Fig. 4). Passeriformes represented by 16 sequences, would be recognized by the sequence **TSNIDEAM** at the C-terminal domain. Moreover, two families in this group, Pipridae and Paridae, would also be identified by signatures at the level of the C-terminal domain (Table 2).

The last point to analyse is the position of turtles which phylogenetic position have been largely discussed. Our analyses shows the clade of Testudines, represented by species pertaining to three different families, as the sister group of Archosauria (Crocodylomorpha + Aves). Indeed, the sequence **ASTESRKSLTTQISNFDN** corresponding to the C-terminal domain, identify the Archosauria-Testudines clade (Fig. 4, Table 2).

Actinopterygii. Regarding to Actinopterygii (represented by species corresponding only to Neopterygii), the ML analyses of Type 2-like receptor, present them as a well-supported clade, sharing a common ancestor with Chondrichthyes which are characterized by the presence of the **ADYDDEFI** motif at the level of the N-terminal (Fig. 5, Table 2). As expected, the sequence corresponding to Type 2 receptor of Lepisosteiformes appears as the sister group of Halecostomi (Fig. 5). With respect to Halecostomi, only sequences corresponding to Teleostei was found. Amiiformes, one of the extant group is not represented in our samples. Teleostei, the more diversified group, represented by numerous species that can be grouped in 11 different clades (see tolweb.org for reference) is represented by 6, including Osteoglossomorpha, Ostariophysi, Clupeomorpha, Salmoniformes, Esociformes and Acanthomorpha (Fig. 5). Similarly to other studies, Osteoglossomorpha (Scleropages formosus) appears as the sister group of the clade that includes Ostarioclupeomorpha (Ostariophysi and Clupeomorpha) and Euteleostei (Protacanthopterygii and Neoteleostei).

The other two clades of Teleostei (i.e. Ostarioclupeomorpha and Euteleostei) share a common ancestor. The first one, involves one Clupeomorpha species appearing as the sister group of Otophysi, which is well represented by three out of four recognized orders (Characiformes, Siluriformes and Cypriniformes) (Fig. 5). Indeed, Characiformes and Siluriformes are grouped in a clade as expected by previous phylogenetic studies, being the sister group of Cypriniformes. Regarding Euteleostei, the two main clades appear as sister groups; Protacanthopterygii (which could be characterized by the presence of the **KFRAEFKA** motif in the C-terminal), including Esociformes (*Esox lucius*) and Salmoniformes. Salmoniformes are represented by three species of two different genus: *Salmo salar*, *Oncorhynchus mykiss* and *O. kisutch*. Moreover, the two species of the genus *Oncorhynchus* are recognized as a clade (Fig. 5). Regarding Salmoniformes, our analyses show the existence of 5 different motifs that might be considered as signatures (see Table 2).

With respect to Neoteleostei, a total of 28 sequences were analysed, pertaining all of them to the clade of Percomorpha (Acanthopterygii), corresponding to: Pleuronectiformes (2), Gasterosteiformes (1), Synbranchiformes (1), Tetraodontiformes (1), Beloniformes (1), Cyprinodontiformes (10) and 12 species corresponding to the non-monophyletic traditional "Perciformes". The members of two families traditionally considered as members of the order Perciformes, as Pomacentridae (represented by two species) and Cichlidae (five species), are well grouped as individual clades. Indeed, the clade of Cichlidae, currently considered as the Order Cichliformes[41] might be identified by three different motifs located at the N-terminal, C-terminal, and IC Loop 3 (Fig. 5, Table 2). Finally, other well represented group is Cyprinodontiformes, characterized by the presence of
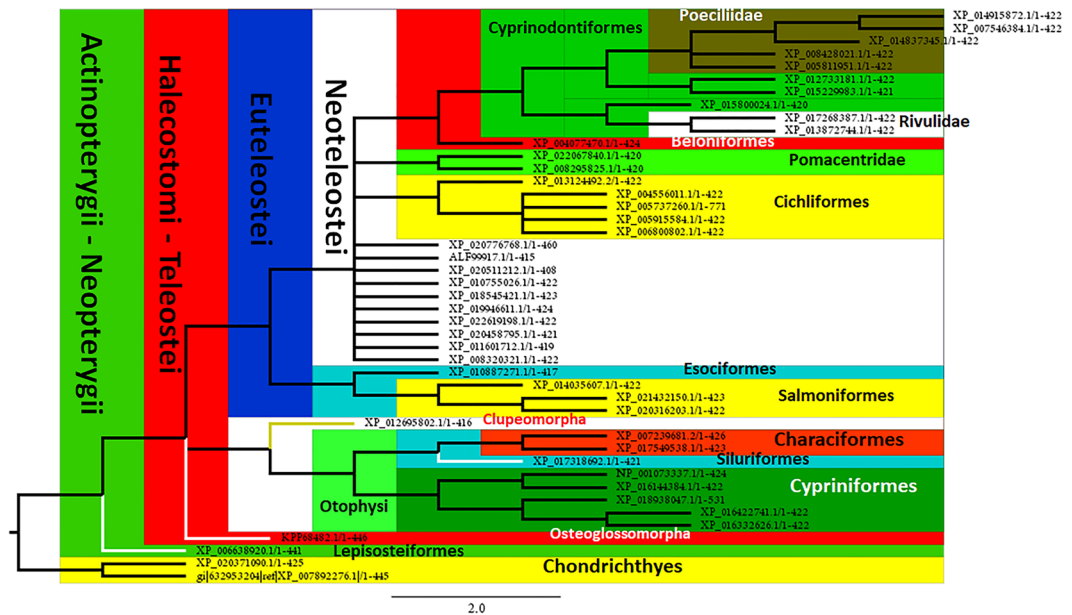
Figure 5. Analysis of Maximum Likelihood of sequences of Orexin receptor corresponding to Actinopterygii. All the species pertain to Neopterygii being represented the two extant groups (Lepisosteiformes and Halecostomi), which appear as the sister group of Chondrichthyes. Currently proposed groups are clearly represented at higher taxonomic levels. The analysis also recognizes taxa at lower levels including families defined by characteristic motifs that might be considered as signatures.

DNLSRLSDQ motif at the C-terminal domain, including 10 sequences corresponding to five different families, being Rivulidae (2) and Poeciliidae (5 species) those best represented. Interestingly both of them are grouped as individual clades (Fig. 5), being characterized by the **RTLRCSAQT** (Rivulidae) and **QRNWRTIQCS** motifs (Poeciliidae). Regarding Poeciliidae, two more motifs might be characteristics at the level of the N-terminal domain (Table 2).

Discussion

As it is known, GPCRs are widely distributed in nature, being associated with the regulation of a great number of physiological mechanisms. As they are engaged with critical processes it is not rare that they were conserved along the evolutionary processes, being appeared early in the evolution. Indeed, SWSI (short-wavelength sensitive opsin), another member of the GPCR family of proteins which is involved in light signal transduction, has proved to be a potential phylogenetic marker in Vertebrata, showing phylogenetic relationships congruent with the evolution of this group at both high and low taxonomic levels[42].

As we stated above, Allatotropin is a peptidic messenger originally characterized by its ability to stimulate the synthesis of Juvenile Hormones in the moth *M. sexta*[17], a highly derived function due that Juvenile Hormones are only present in insects. Originally characterized as a neuropeptide, it was also proved to be secreted by epithelial cell populations[23,25–28]. AT has shown to be pleiotropic, being associated to the regulation of a multiplicity of physiological processes as digestive enzyme secretion and ion exchange regulation[20,21], as well as, the immune response in mosquitoes[43]. Moreover, it also acts as a myoregulatory peptide, modulating the visceral musculature at different levels of the gut, and also as a cardioregulatory peptide[22–27,44]. Regarding Ox peptides, they were originally characterized in mammals, being secreted by neurons located in the hypothalamus. Originally associated with feeding behaviour[35,36], they also act on sleep-wakefulness, being involved in the neurological disorder known as narcolepsy[45,46]. Furthermore, Ox peptides are also related with mechanisms regulating the activity and differentiation of the brown adipose tissue[47], a derived function in view of that this kind of tissue is only present in mammals[48]. Moreover, despite that they were originally characterized as neuropeptides (i.e. secreted by neurons) similarly to AT, the Ox peptides are also secreted by epithelial cell populations[49,50]. Both families of peptides, have proved to be present in other groups related with those in which were originally characterized. Indeed, while its function was not analysed, the presence of AT in other groups of Arthropoda, as Crustacea, Myriapoda and Chelicerata was suggested[51–53]. The presence of Ox peptides in other groups of vertebrata were also proved. In fact, it was shown that Ox has an orexigenic effect on the bullfrog larvae[54]. A similar effect was demonstrated in the goldfish *Carassius auratus*, in which Ox peptides stimulate both, feeding behaviour and food intake[55]. Despite of that the activity as a feeding behaviour modulator might be absent in Sauropsida[56], it was proved that Ox is involved as sleep/wakefulness modulator in birds, playing an important role in the behaviour associated to vigilance[57]. In amphibians, beyond that no experiment about the sleep/wakefulness activity were performed, the distribution of the orexinergic fibers suggests that this function would be conserved[58]. The same was proved in the zebrafish (*Danio rerio*) in which the overexpression of orexins induces an insomnia-like behaviour, promoting locomotion and inhibiting rest[59]. Beyond the complex functions described above, Ox peptides have also

been related with other functions as those related with visceral muscle activity modulation. In fact, like AT, the presence of Ox receptors in the gut, and their activity as myoregulators of smooth muscle cells was also proved[39]. Furthermore, it was recently shown that Ox peptides also act on cardiomyocytes, increasing the shortening of these cells in rats and humans[60].

Regarding AT, as we described above the existence of AT-like peptides was also proposed in other groups of Protostomata. Moreover, the treatment with AT induces muscle contraction at the level of the digestive system in Platyhelminthes[30]. We have previously shown that GPCRs are present in a variety of Metazoa, including *T. adhaerens*, the multicellular organism pertaining to the neuron-less phylum, Placozoa[1]. Moreover, the finding of two predicted sequences exhibiting motifs that may be considered as signatures of the AT/Ox family of GPCRs, are shown in this study. Furthermore, studies in our laboratory suggest that in *Hydra sp.*, ATr would be present (Cnidaria: Hydrozoa)[1–3]. In fact, these studies suggest the existence of an Allatotropin/Orexin homologous system that would acts as myoregulator, controlling the movements associated with the capture and digestion of the prey[1–3]. Beyond the multiplicity of processes regulated by AT and Ox peptides, some of them corresponding to derived functions, both peptides are involved in mechanisms controlling visceral muscle contractions from Cnidaria to Vertebrata, suggesting that this signalling system have appeared associated to feeding in the common ancestor of Metazoa.

AT/Ox GPCRs are characterized by the presence of a Tryptophan (W) instead of a Tyrosine (Y) associated to the **E/D**R motif in the IC Loop 2[16]. Our results show, that the AT/Ox family of GPCRs may be defined by the presence of the **E/DRWYAI** motif, present in 381 out of 392 sequences analysed, covering most of the Metazoa phyla, and that might be considered as a signature of the family. Interestingly, any convincing sequence showing this characteristic motif was found nor in Ctenophora neither in Porifera. The lack of the AT/Ox family of GPCR in those phyla, might be a biological phenomenon, or perhaps an artefact. In fact, beyond the great quantity of information about genomic and transcriptomic sequencing, it may be assumed that it is still perfectible. Moreover, the phylogenetic positions and the evolutionary relationships between Ctenophora, Porifera and the rest of the metazoan groups is still controversial[8,61]. Furthermore, regarding GPCRs, it was already suggested that the Porifera *Rhodopsin* family has not orthologous relationship with the ones found in the rest of Metazoa[11].

Regarding Vertebrata two different groups were found. Interestingly, they are not defined by their phylogenetic relationships, but by the kind of the protein constituting the receptor (Type 1 and Type 2 receptor). One of these groups (i.e. Type 2) is represented in all the groups including, Chondrichthyes, Actinopterygii, Sauropsida and Mammalia, and might be defined for the presence of the **CIAL/QDRWYAICHPL** motif. On the other hand, with the exception of Lepisosteiformes (Actinopterygii: Neopterygii), Type 1 receptor is exclusively expressed in Mammalia (defined by the **FIALDRWYAICHPL** motif). In fact, in Lepisosteiformes, three different sequences were found; two of them are grouped in all the analysis performed with the Type 1 receptor of mammals showing also the **FIALDRWYAICHPL** motif in the interphase between TMIII and the IC loop 2. Beyond these two sequences, a third one (grouped as Type 2 receptor), shows a phylogenetic position according to the current assumption, as the sister group of Halecostomi. The existence of two kind of Ox receptors might be considered as a synapomorphy of Mammalia. The presence of the Type 1-like receptor in Lepisosteiformes would be suggesting that this receptor had appeared more than once along the evolution of Vertebrata.

As a way to further understand the evolutionary history of this family of receptors, we decided to go deeper in the analysis of Type 2-like receptor phylogenetic relationships in two groups of Vertebrata (Sauropsida and Actinopterygii). In both of them, our results show that the sequences phylogenetic relationships are mostly in agreement with current hypothesis about their phylogeny. As an example, a group of species of Neoteleostei (i.e. *Oreochromis niloticus*, *Maylandia zebra*, *Neolamprologus brichardi*, *Haplochromis burtoni* and *Pundamilia nyererei*), traditionally considered as the Cichlidae family pertaining to the order Perciformes (currently considered as polyphyletic), are still grouped as a clade, that in fact is now considered as the order Cichliformes[41]. Another interesting point is that related with the order Cyprinodontiformes. This group represented by 10 species pertaining to five different families, are well defined as independent groups, being the two families represented by two or more species (i.e. Poeciliidae and Rivulidae) grouped as monophyletic groups sharing a common ancestor with the rest of the species of the order. Indeed, these two families might be recognized by signatures located at the N-terminal and IC Loop 3.

Other interesting subject is related with the phylogeny of Sauropsida and the evolutionary position of turtles (Testudines). The phylogenetic position of turtles was largely controversial, as they were traditionally considered as an order pertaining to the group of Anapsida (having no temporal fenestrae in their skull). Traditional studies based on paleontological and morphological characters positioned them as the only extant group of Anapsida being the sister group of Diapsida (a clade that includes Lepidosauromorpha and Archosauria). Based on both paleonthologycal and molecular phylogeny, the evolutionary relationships of Testudines was revisited, considering them as the sister group of Lepidosauromorpha, or as the sister group of Archosauria (Aves and Crocodylomorpha) (for a review see[62]). The finding of a stem-turtle from the middle Triassic finally positioned turtles as a member of Diapsida[63,64]. In agreement with previous molecular studies[65–68], our results, based on sequences pertaining to three different families, place Testudines as the sister group of Archosauria, sharing the **ASTESRKSLTTQISNFDN** motif at the C-terminal domain. Indeed the existence of a new group including Testudines and Archosauria named Archelosauria was recently proposed[66].

Finally, our results show the existence of numerous motifs that might be considered as signatures for several of the groups analysed, being hypothetically possible to test them both as phylogenetical markers at both higher and lower taxonomic levels.

Methods

Data retrieval.   Sequences corresponding to Vertebrate and Insecta AT/Ox GPCRs were searched in protein database of the National Center for Biotechnology Information (NCBI) at https://www.ncbi.nlm.nih.gov/pubmed, and by protein BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome) on the basis of already annotated sequences in the Non-redundant protein sequences database. All the selected sequences were checked for the presence of the characteristic seven transmembrane domains using the TMHMM Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/). The presence of the **E/D**R**W** domain at the IC Loop 2 associated to TMIII was also verified. The sequences were then aligned using the Clustal Omega algorithm for multiple sequence alignment (http://www.ebi.ac.uk/Tools/msa/clustalo/) and further analysed by the JalView 2.7[69]. Only those sequences presenting the seven TMs and the **E/D**R**W** domains, were included.

Sequence analysis and alignment.   Based on the alignment of the full set of sequences a search for motifs that might be considered as signatures in the AT/Ox family was performed. Once established at least one probable signature a search in different phyla including Bilateria and non-bilateria groups as Cnidaria and Placozoa were done. Each sequence were analysed looking for both, the presence of the seven transmembrane domains pattern and the presence of the **E/D**R**W** motif. The phyla in which probable GPCRs associated to the AT/Ox family were found are:

Placozoa, Cnidaria, Arthropoda, Mollusca, Annelida, Brachiopoda and Chordata (see Supporting Information File 1).

Phylogenetic analysis.   Finally, the analysis of evolutionary relationships between sequences, except for the one corresponding to Fig. 1 (Neighbor-Joining), was performed using the ML method based on the Poisson correction model, including a 1000 replicates bootstrap analysis, with a 50% cut-off for condensed tree by the use of Mega 6.06 software[70]. The trees were then edited by the use of FigTree software (http://tree.bio.ed.ac.uk/software/figtree/).

The basic evolutionary relationships between groups are referred to *Tree of Life web Project* (http://tolweb.org/tree/)[71].

Search for signatures.   Once the alignments were performed, we look manually for conserved motifs in different groups. The putative signatures were then blasted (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Only those sequences presenting motifs covering the total length of the query blasted, showing %100 of identity were selected as putative signatures.

Data Availability

All the sequences analysed are in the Supplementary File 1. The datasets generated and/or analysed during the current study are available from the corresponding author on reasonable request.

References

1. Alzugaray, M. E. *et al.* Allatotropin: An ancestral myotropic neuropeptide involved in feeding. *Plos One.* **8**, e77520 (2013).
2. Alzugaray, M. E., Hernández-Martínez, S. & Ronderos, J. R. Somatostatin signaling system as an ancestral mechanism: Myoregulatory activity of an Allatostatin-C peptide in *Hydra. Peptides.* **82**, 67–75 (2016).
3. Alzugaray, M. E. & Ronderos, J. R. Allatoregulatory-like systems and changes in cytosolic Ca²⁺ modulate feeding behavior in. *Hydra. Gen. Comp. Endocrinol.* **258**, 70–78 (2018).
4. Grimmelikhuijzen, C. J. P. & Hauser, F. Mini-review: The evolution of neuropeptide signaling. *Regul. Pept.* **177**, S6–S9 (2012).
5. Nikitin, M. Bioinformatic prediction of *Trichoplax adhaerens* regulatory peptides. *Gen. Comp. Endocrinol.* **212**, 145–155 (2015).
6. Schuchert, P. *Trichoplax adhaerens* (Phylum Placozoa) has cells that react with antibodies against the neuropeptide RFamine. *Acta Zool.* **74**, 115–117 (1993).
7. Smith, C. L. *et al.* Novel cell types, neurosecretory cells, and body plan of the early-diverging metazoan *Trichoplax adhaerens. Curr. Biol.* **24**, 1565–1572 (2014).
8. Jékely, G., Paps, J. & Nielsen, C. The phylogenetic position of ctenophores and the origin(s) of nervous systems. *EvoDevo.* **6**, 1 (2015).
9. Johnson, T. A. *et al.* Identification of the first marine-derived opioid receptor "balanced" agonist with a signaling profile that resembles the endorphins. *ACS Chem. Neurosci.* **8**, 473–485 (2017).
10. Krishnan, A., Almén, M. S., Fredriksson, R. & Schiöth, H. B. The origin of GPCRs: identification of mammalian like Rhodopsin, Adhesion, Glutamate and Frizzled GPCRs in Fungi. *PLoS ONE.* **7**, e29817 (2012).
11. Krishnan, A. *et al.* The GPCR repertoire in the demosponge *Amphimedon queenslandica*: Insights into the GPCR system at the early divergence of animals. *BMC Evol. Biol.* **14**, 1–16 (2014).
12. Bockaert, J. & Pin, J. P. Molecular tinkering of G protein-coupled receptors: an evolutionary success. *EMBO J.* **18**, 1723–1729 (1999).
13. Millar, R. P. & Newton, C. L. The year in G protein-coupled receptor research. *Mol. Endocrinol.* **24**, 261–274 (2010).
14. Fredriksson, R., Lagerström, M. A., Lundin, L. & Schiöth, H. B. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol. Pharmacol.* **63**, 1256–1272 (2003).
15. McCorvy, J. D. & Roth, B. L. Structure and function of serotonin G protein-coupled receptors. *Pharmacol. Ther.* **150**, 129–142 (2015).
16. Verlinden, H. *et al.* The pleiotropic allatoregulatory neuropeptides and their receptors: A mini-review. *J. Insect Physiol.* **80**, 2–14 (2015).
17. Kataoka, H. *et al.* Identification of an allatotropin from adult *Manduca sexta. Science.* **243**, 1481–1483 (1989).
18. Hernández-Martínez, S., Mayoral, J. G., Li, Y. & Noriega, F. G. Role of juvenile hormone and allatotropin on nutrient allocation, ovarian development and survivorship in mosquitoes. *J. Insect Physiol.* **53**, 230–234 (2007).
19. Li, Y., Unnithan, G. C., Veenstra, J. A., Feyereisen, R. & Noriega, F. G. Stimulation of JH biosynthesis by the corpora allata of adult female *Aedes aegypti in vitro*: effect of farnesoic acid and *Aedes* allatotropin. *J. Exp. Biol.* **206**, 1825–1832 (2003).
20. Lee, K. Y., Horodyski, F. M. & Chamberlin, M. E. Inhibition of midgut ion transport by allatotropin (Mas-At) and *Manduca* FRLFamides in the tobacco hornworm *Manduca sexta. J. Exp. Biol.* **201**, 3067–3074 (1998).
21. Lwalaba, D., Hoffmann, K. H. & Woodring, J. Control of the release of digestive enzymes in the larvae of the fall armyworm, *Spodoptera frugiperda. Arch. Insect Biochem. Physiol.* **73**, 14–29 (2009).

22. Duve, H., Audsley, N., Weaver, R. J. & Thorpe, A. Triple co-localisation of two types of allatostatin and an allatotropin in the frontal ganglion of the lepidopteran *Lacanobia oleracea* (Noctuidae): Innervation and action on the foregut. *Cell Tissue Res.* **300**, 153–163 (2000).

23. Santini, M. S. & Ronderos, J. R. Allatotropin-like peptide released by Malpighian tubules induces hindgut activity associated to diuresis in the Chagas disease vector *Triatoma infestans* (Klug). *J. Exp. Biol.* **210**, 1986–1991 (2007).

24. Villalobos Sambucaro, M. J. *et al.* Allatotropin modulates myostimulatory and cardioacceleratory activities in *Rhodnius prolixus* (Stal). *PLoS ONE.* **10**, e0124131 (2015).

25. Sterkel, M., Riccillo, F. L. & Ronderos, J. R. Cardioacceleratory and myostimulatory activity of allatotropin in *Triatoma infestans* (Klüg). *Comp. Biochem. Physiol. A. Mol. Integr. Physiol.* **155**, 371–377 (2010).

26. Santini, M. S. & Ronderos, J. R. Allatotropin-like peptide in Malpighian tubules: Insect renal tubules as an autonomous endocrine organ. *Gen. Comp. Endocrinol.* **160**, 243–249 (2009a).

27. Santini, M. S. & Ronderos, J. R. Daily variation of an allatotropin-like peptide in the Chagas disease vector *Triatoma infestans* (Klug). *Biol. Rhythm Res.* **40**, 299–306 (2009b).

28. Riccillo, F. L. & Ronderos, J. R. Allatotropin expression during the development of the fourth instar larvae of the kissing-bug *Triatoma infestans* (Klüg). *Tissue Cell.* **42**, 355–359 (2010).

29. Adami, M. L., Damborenea, C. & Ronderos, J. R. Expression of a neuropeptide similar to allatotropin in free living Turbellaria (Platyhelminthes). *Tissue Cell.* **43**, 377–383 (2011).

30. Adami, M. L., Damborenea, C. & Ronderos, J. R. An allatotropin-like neuropeptide in *Mesostoma ehrenbergii* (Rhabdocoela, Platyhelminthes). *Zoomorphology.* **131**, 1–9 (2012).

31. Elekonich, M. M. & Horodyski, F. M. Insect allatotropins belong to a family of structurally-related myoactive peptides present in several invertebrate phyla. *Peptides.* **24**, 1623–1632 (2003).

32. Yamanaka, N. *et al.* Neuropeptide receptor transcriptome reveals unidentified neuroendocrine pathways. *PLoS ONE* **3**, e3048 (2008).

33. Horodyski, F. M. *et al.* Isolation and functional characterization of an allatotropin receptor from *Manduca sexta*. *Insect Biochem. Mol. Biol.* **41**, 804–814 (2011).

34. Zhang, F. *et al.* Isolation functional characterization of allatotropin receptor from the cotton bollworm, *Helicoverpa armígera*. Peptides, https://doi.org/10.1016/j.peptides.2017.11.019 (2017).

35. Sakurai, T. *et al.* Orexins and orexin receptors: a family of hypothalamic neuropeptides and G protein-coupled receptors that regulate feeding behavior. *Cell.* **92**, 573–585 (1998).

36. de Lecea, L. *et al.* The hypocretins: hypothalamus-specific peptides with neuroexcitatory activity. *Proc. Natl. Acad. Sci. USA* **95**, 322–327 (1998).

37. Kukkonen, J. P. Physiology of the orexinergic/hypocretinergic system: a revisit in 2012. *Am. J. Physiol. Cell Physiol.* **301**, 2–32 (2013).

38. Näslund, E., Ehrström, M., Ma, J., Hellström, P. M. & Kirchgessner, A. L. Localization and effects of orexin on fasting motility in the rat duodenum. *Am. J. Physiol. Gastrointest. Liver Physiol.* **282**, G470–G479 (2002).

39. Squecco, R., Garella, R., Luciani, G., Francini, F. & Baccari, M. C. Muscular effects of orexin A on the mouse duodenum: mechanical and electrophysiological studies. *J. Physiol.* **589**, 5231–5246 (2011).

40. Jékely, G. Global view of the evolution and diversity of metazoan neuropeptide signaling. *Proc. Natl. Acad. Sci. USA* **110**, 8702–8707 (2013).

41. Betancur-R, R. *et al.* Phylogenetic classification of bony fishes. *BMC Evol. Biol.* **17**, 162 (2017).

42. van Hazel, I., Santini, F., Müller, J. & Chang, B. S. W. Short-wavelength sensitive opsin (SWS1) as a new marker for vertebrate phylogenetics. *BMC Evol. Biol.* **6**, 97 (2006).

43. Hernández-Martínez, S. *et al.* Allatotropin: A pleiotropic neuropeptide that elicits mosquito immune responses. *Plos One.* **12**, e0175759 (2017).

44. Duve, H., East, P. & Thorpe, A. Regulation of lepidopteran foregut movement by allatostatins and allatotropin from the frontal ganglion. *J. Comp. Neurol.* **413**, 405–416 (1999).

45. Chemelli, R. M. *et al.* Narcolepy in orexin knockout mice: Molecular genetics of sleep regulation. *Cell.* **98**, 437–451 (1999).

46. Ling, L. *et al.* The sleep disorder canine narcolepsy is caused by a mutation in the hypocretin (Orexin) receptor 2 gene. *Cell.* **98**, 365–376 (1999).

47. Sellayah, D., Bharaj, P. & Sikder, D. Orexin is required for brown adipose tissue development, differentiation, and function. *Cell Metab.* **14**, 478–490 (2011).

48. Mezentseva, N. V., Kumaratilake, J. S. & Newman, S. A. The brown adipocyte differentiation pathway in birds: An evolutionary road not taken. *BMC Biol.* **6**, 1–13 (2008).

49. Dall'Aglio, C., Pascucci, L., Mercati, F., Boiti, C. & Ceccarelli, P. Localization of the orexin system in the gastrointestinal tract of fallow deer. *Acta Histochem.* **114**, 74–78 (2012).

50. Valiante, S. *et al.* Expression of orexin A and its receptor 1 in the human prostate. *J. Anat.* **222**, 473–480 (2013).

51. Christie, A. E. *et al.* Neuropeptidergic signaling in the American Lobster Homarus Americanus: New insights from high-throughput nucleotide sequencing. *PLoS ONE.* **10**, 1–29 (2015).

52. Christie, A. E. In silico characterization of the neuropeptidome of the Western black widow spider Latrodectus hesperus. *Gen. Comp. Endocrinol.* **210**, 63–80 (2015).

53. Christie, A. E. Neuropeptide discovery in Symphylella vulgaris (Myriapoda, Symphyla): In silico prediction of the first myriapod peptidome. *Gen. Comp. Endocrinol.* **223**, 73–86 (2015).

54. Shimizu, S., Nakamachi, T., Konno, N. & Matsuda, K. Orexin A enhances food intake in bullfrog larvae. *Peptides.* **59**, 79–82 (2014).

55. Volkoff, H., Bjorklund, J. M. & Peter, R. E. Stimulation of feeding behavior and food consumption in the goldfish, Carassius auratus, by orexin-A and orexin-B. Stimulation of feeding behavior and food consumption in the goldfish, Carassius auratus, by orexin-A and orexin-B. *Brain Res.* **846**, 204–209 (1999).

56. Tachibana, T. & Tsutsui, K. Neuropeptide control of feeding behavior in birds and its difference with mammals. *Front. Neurosci.* **10**, 1–13 (2016).

57. da Silva, E. S. *et al.* Behavioral and metabolic effects of central injections of orexins/hypocretins in pigeons (Columba livia). *Regul. Pept.* **147**, 9–18 (2008).

58. Wong, K. K. Y., Ng, S. Y. L., Lee, L. T. O., Ng, H. K. H. & Chow, B. K. C. Orexins and their receptors from fish to mammals: A comparative approach. *Gen. Comp. Endocrinol.* **171**, 124–130 (2011).

59. Prober, D. A., Rihel, J., Onah, A. A., Sung, R. J. & Schier, A. F. Hypocretin/Orexin Overexpression induces an insomnia-like phenotype in zebrafish. *J. Neurosci.* **51**, 13400–13410 (2006).

60. Patel, V. H. *et al.* Functional cardiac orexin receptors: role of orexin-B/orexin 2 receptor in myocardial protection. *Clin. Sci.* **132**, 2547–2564 (2018).

61. Dunn, C. W., Leys, S. P. & Haddock, H. D. The hidden biology of sponges and ctenophores. *Trends Ecol. Evol.* **30**, 282–291 (2015).

62. Meyer, A. & Zardoya, R. Recent Advances in the (Molecular) Phylogeny of Vertebrates. *Annu. Revi. Ecol. Evol. Syst.* **34**, 311–338 (2003).

63. Schoch, R. R. & Sues, H. D. A middle triassic stem-turtle and the evolution of the turtle body plan. *Nature.* **523**, 584–587 (2015).

64. Schoch, R. R. & Sues, H. D. The diapsid origin of turtles. *Zoology.* **119**, 159–161 (2016).

65. Chiari, Y., Cahais, V., Galtier, N. & Delsuc, F. Phylogenomic analyses support the position of turtles as the sister group of birds and crocodiles (Archosauria). *BMC Biol.* **10**, 65 (2012).

66. Crawford, N. G. *et al*. A phylogenomic analysis of turtles. *Mol. Phylogenet. Evol.* **83**, 250–257 (2015).
67. Field, D. J. *et al*. Toward consilience in reptile phylogeny: microRNAs support an archosaur, not a lepidosaur affinity for turtles. *Evol. Dev.* **16**, 189–196 (2014).
68. Kumazawa, Y. & Nishida, M. Complete mitochondrial DNA sequences of the green turtle and blue-tailed mole skink: statistical evidence for archosaurian affinity of turtles. *Mol. Biol. Evol.* **16**, 784–792 (1999).
69. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. JalView version 2 – a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, https://doi.org/10.1093/bioinformatics/btp033 (2009).
70. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol Biol Evol.* **30**, 2725–2729 (2013).
71. Maddison, D. R., Schulz, K. & Maddison, W. P. The tree of life web project. *Zootaxa.* **1668**, 19–40 (2007).

Acknowledgements

Author Contributions

Conceived and designed the analysis: J.R.R. Performed the analyses and analysed the data: J.R.R., M.E.A., M.J.V.S. and C.B. Wrote the paper: J.R.R. critically revised the manuscript: M.E.A., C.B. and M.J.V.S.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-46712-9.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.