

1 **Ambulacrarian insulin-related peptides and their**
2 **putative receptors suggest how insulin and similar**
3 **peptides may have evolved from Insulin-like Growth**
4 **Factor**

5

6 Jan A. Veenstra¹,

7

8 ¹ INCIA UMR 5287 CNRS, Université de Bordeaux, Pessac, France.

9

10 Corresponding Author:

11 Jan A. Veenstra¹

12 INCIA UMR 5287 CNRS, Université de Bordeaux, allée Geoffroy St Hillaire, CS 50023, 33 615

13 Pessac Cedex, France

14 Email address: jan-adrianus.veenstra@u-bordeaux.fr

15 **Abstract**

16 **Background**

17 Insulin is evolutionarily related to the insulin-like growth factors (IGFs) and like the latter
18 stimulates a receptor tyrosine kinase (RTK) that transfers the extracellular hormonal signal into
19 an intracellular response. Other hormones related to insulin, such as relaxin, do not use an RTK,
20 but a G-protein coupled receptor (GPCR). This is unusual since evolutionarily related hormones
21 typically either use the same or paralogous receptors. In arthropods three different IGF-related
22 peptides likely evolved from a gene triplication, as in several species genes coding these three
23 peptides are located next to one another on the same chromosomal fragment. Of these three
24 hormones one, an IGF-like hormone, acts through an RTK, while the other two use a GPCR.
25 This suggests that the ancestral IGF-like peptide may have used both types of receptors. These
26 arthropod insulin-like peptides have homologs in vertebrates, which suggests that the initial gene
27 triplication was perhaps already present in the last common ancestor of deuterostomes and
28 protostomes. It would be interesting to know whether this is indeed so and to establish how
29 insulin and other insulin-like peptides might be related to this trio of IGF-related hormones.

30 **Methodology**

31 Genes coding insulin and related peptides as well as their putative receptors were identified in
32 genomes and transcriptomes from echinoderms and hemichordates.

33 **Results**

34 A similar triplet of genes coding insulin-like peptides is also found in some hemichordates and
35 echinoderms. Two of the three ambulacrarian peptides are orthologs of arthropod IGF and
36 *Drosophila* insulin-like peptide 7 (dilp7), while the third one looks like an ortholog of the
37 arthropod peptide gonadulin. In echinoderms two novel insulin-like peptides emerged, gonad
38 stimulating substance (GSS) and multinsulin, likely from gene duplications of the IGF and dilp7-
39 like genes respectively. However, no novel receptors for insulin-like peptides evolved. If IGF
40 were to act through both a GPCR and an RTK it would suggest that GSS acts on only one of the
41 two receptors, possibly the RTK. The evolution of GSS from IGF may represent a pattern, where
42 IGF gene duplications lead to novel genes coding shorter peptides that have lost their ability to

43 activate a GPCR. It is likely this is how insulin and the insect neuroendocrine insulin-like
44 peptides evolved independently from IGF.

45 **Conclusion**

46 The local gene triplication previously described from arthropods that yielded three genes coding
47 IGF-related peptides was already present in the last common ancestor of protostomes and
48 deuterostomes. It seems plausible that insulin and other insulin-like peptides, such as those
49 produced by neuroendocrine cells in the brain of insects and echinoderm GSS evolved
50 independently from IGF and thus are not true orthologs, but the result of convergent evolution.

51

52

53

54 Key words: insulin; relaxin; receptor tyrosine kinase; G-protein coupled receptor; evolution;
55 gonadulin; octinsulin; multinsulin; dilp7

56 Introduction

57 Many protein hormone and neuropeptide signaling pathways have orthologs in both
58 protostomes and deuterostomes showing that these pathways were already present in their last
59 common bilaterian ancestor. In some cases the orthologs of the ligands show only limited
60 sequence similarity, but their receptors contain protein domains that are sufficiently conserved to
61 establish homology. Virtually all ligands employ either a single receptor or a number of closely
62 related receptors that evolved by gene duplication. Co-evolution of ligands and receptors insures
63 that related protein hormones or neuropeptides use receptors akin to those of their orthologs
64 (Mirabeau & Joly, 2013; Hsueh & Feng, 2020).

65 Insulin-related peptides (irps) are an exception to this rule. Whereas insulin acts through a
66 receptor tyrosine kinase (RTK), relaxin uses a leucine-rich repeat G-protein coupled receptor
67 (LGR). This raises the interesting question as how this apparent jump from one type of receptor
68 to another may have come about. In cockroaches, termites and stick insects three different irp
69 genes, gonadulin, arthropod insulin-like growth factor (aIGF) and arthropod relaxin, are located
70 next to one another in the genome and thus likely originated from a local gene triplication
71 (Veenstra, 2020b). To avoid confusion with the vertebrate relaxins and related peptides, the
72 arthropod relaxins will be referred to as *Drosophila* ilp7 (dilp7) in this manuscript. One of these,
73 aIGF, is known to use an insulin RTK, while gonadulin acts through insect LGR3 (Vallejo et al.,
74 2015; Garelli et al., 2015; Colombani et al., 2015). Bioinformatic evidence suggested that dilp7
75 must be the ligand for insect LGR4 and this has now been confirmed experimentally in
76 *Drosophila* (Veenstra, Rombauts & Grbić, 2012; Imambocus et al., 2020), but dilp7 may also
77 activate an RTK (Linneweber et al., 2014). This suggests that the archtype arthropod IGF-related
78 peptide acted through both an RTK and an LGR and that after a likely gene triplication, some of
79 the ligands may have lost one of the two original receptors. Although it is possible that the gene
80 triplication of the ancestral insulin gene occurred in an early arthropod or protostomian, it may
81 well have occurred in a bilaterian ancestor, as homologs of both aIGF and dilp7 are also present
82 in deuterostomes.

83 Brain neuroendocrine insect irps are more closely related to IGF than either dilp7 or
84 gonadulin and a gene duplication that gave rise to separate genes coding these peptides is
85 therefore likely to have occurred after the triplication that gave rise to the ancestor genes of
86 gonadulin and dilp7. Yet in insect genomes irp genes are not located near the IGF gene, while

87 the IGF gene is in the middle of the triplet. Thus the particular organization of these genes
88 suggests that whereas the gonadulin and dilp7 genes likely originated by two successive local
89 gene duplications, the IGF gene duplication that gave rise to an initial arthropod neuroendocrine
90 brain irp must have materialized in a different fashion. If the earlier mentioned gene triplication
91 was already present in the last common ancestor of the deuterostomes then a similar argument
92 can also be made for the evolution of insulin. Given the importance of insulin as a human
93 hormone and the inherent interest of its evolutionary origin, I explored the evolution of bilaterian
94 insulin-related peptides in more detail and here report on the genes coding such peptides and
95 their receptors in the Ambulacraria that suggest how insulin may have evolved from IGF.

96

97 **Materials and Methods**

98 *Nomenclature*

99 Hormones have often been discovered independently by different groups using different
100 bioassays. The vertebrate insulin-like growth factors are a good example of that. Predicted
101 protostomian peptides and their receptors have sometimes been given names that refer to similar
102 deuterostomian proteins. In some cases this is very confusing, *e.g.* vertebrate LGR-3, -4 and -5
103 are not the orthologs of arthropod receptors that have been given the same names. A similar
104 problem occurs with arthropod relaxin that is not an ortholog of vertebrate relaxin. This peptide
105 will therefore be called dilp7 (*Drosophila* insulin-like peptide 7). I will refer to arthropod LGR3
106 as the gonadulin receptor, arthropod LGR4 as the dilp7 receptor and arthropod LGR5 as
107 GRL101, a GPCR initially identified from the pond snail *Lymnaea stagnalis* (Tensen et al.,
108 1994) that is an ortholog of arthropod LGR5 (Veenstra, 2020b).

109

110 *Sequence analysis*

111 Sequences for insulin related peptides and their likely receptors were identified from a
112 number of Ambulacraria. This was done using using the Artemis program (Rutherford et al.,
113 2000) and the BLAST+ program (<ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast/>) on
114 publicly available genome sequences from the feather star *Anneissia japonica*, the sea urchins
115 *Lytechinus variegatus* (Davidson et al., 2000) and *Strongylocentrus purpuratus* (Sea Urchin
116 Genome Sequencing Consortium, 2006), the sea cucumbers *Apostichopus japonicus* (Jo et al.,
117 2017; Zhang et al., 2017) and *Holothuria glaberrima*, the sea stars *Acanthaster planci* (Hall et

118 al., 2017), *Pisaster ochraceus* (Ruiz-Ramos et al., 2020) and *Patiria miniata*, the brittle star
119 *Ophiothrix spiculata* and the hemichordates *Saccoglossus kowalevskii* and *Ptychodera flava*
120 (Simakov et al., 2015). The genomes were downloaded from
121 <https://www.ncbi.nlm.nih.gov/genome>. For many of these species there are also significant
122 amounts of RNAseq data and these were analyzed using the sratoolkit
123 (<https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=software>) in combination with Trinity
124 (Grabherr et al., 2011) using methods described in detail elsewhere (Veenstra, 2020b). Some
125 protein sequences were found in the NCBI database, but several of them contain errors or are
126 incomplete. Where possible these were corrected and/or completed using the methods described
127 above. As there is only a single crinoid genome assembly available, transcriptome data from
128 *Antedon mediterranea*, *Florometra serratissima* and *Oligometra serripinna* were also included.
129 For the same reason transcriptome data from the brittle star *Amphiura filliformis*, *Ophioderma*
130 *brevispina* and the hemichordate *Schizocardium californicum* were likewise analyzed.
131 Obviously, transcriptome data can only demonstrate the presence of gene but not its absence and
132 their usefulness depends largely on the variety of tissues sampled and the expression levels of the
133 genes of interest. Nevertheless, such data often provide additional sequences that even if they are
134 incomplete increase the robustness of sequence comparisons. Genomic and transcriptomic
135 RNAseq short read archives (SRAs) were downloaded from NCBI
136 (<https://www.ncbi.nlm.nih.gov/sra/>); a list of the SRAs analyzed is provided in the
137 supplementary data.

138 As queries for the insulin-like peptides a number of such peptides from a variety of species
139 was used. Insulin RTKs are easily identified in genome and transcriptome assemblies, as their
140 kinase domains are very well conserved. The LGRs that could function as insulin receptors are
141 more variable. Vertebrate RXFP1 and RXFP2 are LGRs are known receptors for relaxin and
142 Ins3 and *Drosophila* LGR3 and LGR4 for gonadulin, and dilp7 respectively. Other LGRs
143 function as receptors for the various glycoprotein hormones, GPA2/GPB5, bursicon, TSH, FSH
144 and LH. These GPCRs cluster on phylogenetic trees with another protostomian LGR, GRL101.
145 This GPCR was initially identified from the pond snail *Lymnaea stagnalis* and was the first
146 GPCR discovered to have in addition to six leucine-rich repeats also twelve repeats of a
147 sequence that was known to exist in the low density lipoprotein receptor and are now called

148 LDLa repeats (Tensen et al., 1994). I have suggested previously (Veenstra, 2020b) that this
149 receptor might be an IGF receptor.

150 Both the RTK and LGR receptors have large ectodomains. Those of the insulin RTKs are
151 very similar from one receptor to another, while those of the LGRs differ between different
152 types. The latter all contain numerous Leucine-rich repeats (LRRs) and some also have LDL-
153 receptor class A (LDLa) repeats. Both LRRs and LDLa's are present in many other proteins.
154 Initial searches for orthologs receptors were therefore done using the transmembrane regions of
155 various insect and vertebrate LGRs and the protein kinase domain of RTK. Once partial
156 sequences of putative receptors were identified, the coding sequences of these domains were then
157 used to complete the cDNA sequences as best as possible, using either Trinity on RNAseq SRAs
158 or Artemis on genome sequences.

159

160 *Sequence similarity and phylogenetic trees*

161 Peptides having the characteristic insulin signature are notoriously variable in their primary
162 amino acid sequences. Although the various residues allows one to align those sequences, such
163 alignments will not always yield reliable phylogenetic trees as the basic tenet of such analyses is
164 often not met. As an alternative I have proposed to use “sequence similarity trees”. Such trees are
165 constructed using the same methods but do not pretend to illustrate phylogenetic relations, rather
166 similarities between the different proteins. Both phylogenetic and sequence similarity trees use
167 Clustal omega (Sievers et al., 2011) to produce alignments and Fasttree (Price, Dehal & Arkin,
168 2010) to construct trees and estimate probabilities.

169 In order to identify putative receptors for the various irps, LGRs that show homology to
170 various arthropod and vertebrate LGRs were identified and a phylogenetic tree based exclusively
171 on the transmembrane regions of these receptors was constructed.

172

173 *Precursor processing*

174 Precursors of insulin-like peptides contain signal peptides that are removed on entry into
175 endoplasmatic reticulum. Signal P 5.0 (Almagro Armenteros et al., 2019) was used online
176 (<http://www.cbs.dtu.dk/services/SignalP/>) to predict where this cleavage would most likely
177 occur. Some, but not all precursors are further processed by convertases. Of these furin is
178 ubiquitously present in all cell types and can thus potentially cleave any secreted protein. Its

179 consensus cleavage site is K/R-X-K/R-R, the two human IGF precursors are processed at KSAR
180 and KSER respectively (Humbel, 1990). Precursors that are produced in cells with a regulated
181 pathway, such as neuroendocrine and enteroendocrine cells, are also exposed to other
182 convertases like PC1/3 and PC2. Their consensus cleavages site is KR. However, effective
183 proteolytic processing by convertases is strongly influenced by amino acid residues surrounding
184 these consensus cleavage sites. For example bulky residues immediately following the arginine
185 residue, a proline residue before the consensus site or disulfide bridges nearby can cause
186 sufficient steric hindrance to inhibit cleavage. Using rules proposed to predict cleavage by PC1/3
187 and PC2 in both vertebrates and insects (Devi, 1991; Rholam et al., 1995; Veenstra, 2000) I have
188 tried to indicate in Figs. 2-5 where the various precursors might be cleaved. It must be noted
189 though that there is no certainty that these site will be cleaved nor can it be excluded that
190 proteolytic processing occurs at sites that have not been indicated as such.

191

192 *Expression*

193 With a few notable exceptions (*e.g.* Lin et al., 2017), little is known about the expression of
194 the various insulin-like peptides in either echinoderms or hemichordates and except for the GSS
195 our knowledge of their functions is also very limited. Expression data may reveal some
196 preliminary clues as to where and when they are expressed and thus provide a hint as to their
197 function. For this reason the number of reads corresponding to the various insulin-related
198 peptides and their putative receptors was determined in a number of SRAs that might provide
199 evidence as to the time and tissue specific expression of these proteins. The analysis was
200 performed as described previously (Veenstra, 2020b) and the data are supplied in Spreadsheet
201 S2.

202

203

204 **Results**

205 *Peptides related to insulin*

206 Insulin-like peptide precursors are typically characterized as having A, B and C domains that
207 correspond to the A- and B-chains of insulin and the connecting peptide respectively. In IGF D
208 and E domains are also recognized, in which the D domain refers to the extension of the A chain
209 and the E domain to part of the precursor after the D domain that is cleaved from IGF in the

210 Golgi apparatus. For dilp7 orthologs it is appropriate to add an F (front) domain for the sequence
211 in N-terminal of the B-chain that in some peptides is not only larger, but also well conserved.

212 Previous work on insulin-related peptides in in echinoderms have identified two different
213 types of insulin-like peptides, gonad-stimulating substances (GSS) and insulin-like growth
214 factors (Mita et al., 2009; Perillo & Arnone, 2014; Semmens et al., 2016; Smith et al., 2019). The
215 insulin-like growth factors, but not GSS, are also present in hemichordates. While only a single
216 IGF gene was found in the crinoids and hemichordates, other echinoderms have two such genes
217 (Figs. 1, S1, S2; Spreadsheet S1). These proteins have large C-terminal extension that are rich in
218 charged amino acid residues, especially arginine and lysine, but also aspartic and glutamic acid
219 residues. A comparison of the protein sequences and cDNAs from human IGFs identifies the
220 exact separation between the D and E domains in these proteins (Humbel, 1990). However,
221 although the corresponding sequences of the hemichordate and echinoderm IGFs contain
222 numerous arginine and lysine residues (Figs. 1, S1, S2), there are no obvious convertase
223 cleavages sites as many potential arginine residues are succeeded by residues known to inhibit
224 such enzymes in vertebrates. It is thus not impossible that the D domains of these proteins are
225 much larger than in the vertebrate IGFs and if so likely contain numerous positively charged
226 amino acid residues. There are few transcriptome SRAs for specific tissues, the data that is
227 available suggest that the IGFs are expressed by many tissues, with the ovary showing
228 significant expression. *Patiria pectinifera* is the only species with follicle cell specific SRAs and
229 IGF-1 is strongly expressed by these cells and is probably transferred to the oocyte (Spreadsheet
230 S2).

231 The GSS are known to induce oocyte maturation and ovulation in a two step process, where
232 GSS stimulates the follicle cells to produce 1-methyladenine which subsequently induces
233 resumption of meiosis in the oocyte and about 30 minutes later this is followed by ovulation
234 (Chiba, 2020). Interestingly, GSS was not found in either the genome nor the extensive
235 transcriptome data from the feather star *Anneissia japonica* and was similarly not encountered in
236 the transcriptomes of three other crinoids (Suppl data). Transcriptomes may miss expression of
237 some genes and large genome assemblies are never perfect. The short sequence reads in the
238 genomic SRAs from *Anneissia* were therefore also analyzed for the presence of GSS, but again
239 no evidence for such a gene was found. This peptide is thus likely absent from *Anneissia* and
240 perhaps all Crinoidea. In the Holothuroidea and the Asterozoa, but not the Echinoidea, this gene

241 is duplicated with the two paralogous peptides showing significant sequence variability (Figs. 2,
242 S3, S4; Spreadsheet S1). As for all these peptides and their putative receptors expression data is
243 very limited, but in *Apostichopus* the two GSSs are differentially expressed, with GSS-1 being
244 expressed at specific stages during embryonic development as well as by muscle and GSS-2
245 strongly expressed by both the ovary and the testes. Interestingly, it is the ortholog of GSS-1 that
246 in *Holothuria scabra* has been tested for biological activity and induces ovulation (Chieu et al.,
247 2019). This makes one wonder what the effects of GSS-2 on ovulation might be in this species.
248 However, *Apostichopus* was the only species where a significant GSS expression was found in
249 the gonads (Spreadsheet S2).

250 Two other insulin-like peptides are commonly present in both hemichordates and
251 echinoderms, including the Crinoidea. The first is an ortholog dilp7 which has a very
252 characteristic F domain while its A chain is also remarkably well conserved (Figs. 3, S5, S6;
253 Spreadsheet S1). The precursors of this peptide contain typical neuroendocrine KR convertase
254 sites and seems to have its highest expression in the nervous system, although it is also found in
255 other tissues. During embryogenesis its expression occurs relatively late (Spreadsheet S2). The
256 second peptide present in all ambulacrarians has been called octinsulin as it has eight cysteine
257 residues and is thus predicted to have four rather than three disulfide bridges. In echinoderms
258 octinsulin are single copy genes, but hemichordates have several such genes (Fig. 4, S7, S8;
259 Spreadsheet S1). Octinsulin expression levels are the highest in nervous tissue, and significant
260 expression is also found in the gut and stomach of *Strongylocentrotus* and *P. pectinifera*
261 respectively. Although virtually absent from normal gut in *Apostichopus*, it has significant
262 expression during gut regenerating in this species (Spreadsheet S2).

263 The Asterozoa have genes coding for a fifth type of insulin, that is usually present in
264 multiple copies and that are referred to as multinsulins. The predicted peptides share structural
265 similarity with the dilp7 orthologs and their genes have typically four coding exons rather than
266 the two or three of the other irp genes. The sprawl of these peptides is perhaps best illustrated by
267 a phylogenetic tree that suggest independent multiplication of these genes in several species (Fig.
268 S10). Within a single species the various multinsulins thus often seem more closely related to
269 one another than to their putative orthologs of other Asterozoa. Some of the multinsulins, like the
270 octinsulins, have acquired two additional cysteine residues and are thus predicted to have four
271 disulfide bridges, but the location of these additional cysteine residues differs from that in

272 octinsulins (Figs. 5, 6, S9, S10; Spreadsheet S1). Like dilp7 the multinsulins have typical
273 neuroendocrine KR convertase cleavage sites and can thus be expected to be expressed in
274 neuroendocrine and/or enteroendocrine cells but expression data on *P. pectinifera* suggest a
275 relatively ubiquitous expression in several tissues.

276 The genome assemblies of *A. planici* and *Pisaster ochraceus* shows these genes to be
277 clustered in the genome and some RNAseq sequences suggests that at least on occasion coding
278 exons from different genes may be combined (Fig. S10). This and the large numbers of SNPs
279 typically present in animals caught in nature and used for RNAseq preparation make it
280 impossible to reliably determine their exact numbers.

281 Genome assemblies allow identification of the introns in these genes. All insulin genes have
282 a characteristic phase 1 intron somewhere in the conceptual C domain of these molecules. This is
283 the only intron in the coding sequences of the octinsulin and GSS genes. The IGF genes have a
284 phase 0 intron near the end of the coding sequence and at least some of them have another phase
285 1 intron just after the transcription start site. The genes coding the dilp7 orthologs and
286 multinsulins share an additional phase 2 intron and the multinsulin genes have yet another phase
287 1 intron. All these introns appear perfectly conserved (Fig. 6).

288

289 *Synteny of genes producing insulin-like peptides*

290 In the *Strongylocentrotus* genome all five genes are located on the same chromosome, with
291 the two IGF genes and those coding octinsulin and dilp7 orthologs next to one another and GSS
292 at a distance of 6,000,000 bp (base pairs). At least the *Anneissia* octinsulin and IGF genes are
293 likely located next to one another on the same chromosome also, as in the current genome
294 assembly two of the three coding exons of IGF and one of the two octinsulin coding exons are
295 located within about 10,000 bp. The three missing exons of these two genes are all located on
296 minicontigs of less than 2,000 bp, as is one of the coding exons for the dilp7 ortholog. The
297 contigs of the *Lytechinus variegatus* genome assembly are smaller and this may explain why in
298 this species the genes are located on three different scaffolds, with the two ILGF-like peptides
299 and the octinsulin together on a single contig. However in the recently published genome of the
300 closely related *L. pictus* (Warner et al., 2021) the dilp7 ortholog is also closely associated with
301 the other three genes. The GSS gene is on the same chromosome but at a distance of 28,000,000
302 bp. In the *Apostichus japonicus* genome assembly the genes coding the octinsulin and the two

303 IGF genes are located on the same contig, and the other genes each on a different one. In the
304 draft *Holothuria glaberrima* genome assembly only the two IGF genes are located on the same
305 contig, however in a single Oxford nanopore read (SRR9125585.2851.1) from *H. scabra* the
306 octinsulin, dilp7 and two IGF genes are located next to one another as well (Fig. 7).

307 Whereas the various Echinozoa genome assemblies suggest a certain degree of synteny with
308 regard to the various irp genes, the Asterozoa genome show that such synteny is disintegrating.
309 This is most clearly demonstrated in the genome assemblies from *Pisaster ochraceus* and
310 *Acanthaster planci*, where the scaffolds are much larger than from *Patiria miniata*. In these
311 species synteny is largely lost (Fig. 7). Interestingly the various multinsulin genes are present in
312 small clusters on different chromosomes in those species.

313

314 *Sequence similarity tree peptides related to insulin*

315 A sequence similarity tree (Fig. 8) of the various irps from echinoderms and hemichordates
316 show that the structures of the multinsulins are most similar to the dilp7 orthologs. This is not
317 surprising as they look indeed similar (Fig. 5) and they share the position of an intron with the
318 dilp7 orthologs (Fig. 6). The tree also illustrates significant sequence similarity between GSS and
319 the IGF.

320

321 *Orthologs of receptors for insulin-like peptides: Receptor tyrosine kinase*

322 A single insulin RTK gene was found in all species analyzed here. An alternatively spliced
323 form is present in *Acanthaster* and is likely commonly present in echinoderms (Spreadsheet S1).
324 Hundreds of ambulacrarian protein sequences were identified at NCBI using a BLAST search
325 with the *S. kovalevskii* protein kinase domain as a query. After aligning them with Clustal omega
326 the protein kinase domains were used to make a phylogenetic tree. Results revealed no other
327 known or predicted proteins with a similar protein kinase domain (data not shown). The insulin
328 RTK is ubiquitously expressed (Spreadsheet S2).

329

330 *Orthologs of receptors for insulin-like peptides: LGRs*

331 LGR sequences were obtained using the combination of genomic sequences and, where
332 available, transcriptome shotgun sequences and RNAseq SRAs. The latter were used to produce
333 contigs using Trinity (Spreadsheet S1). Short read assemblers are good in combining sequences

334 into larger continuous ones, but they do produce artifacts, which are more easily obtained when
335 very similar sequences are present in multiple copies, such as the multinsulins, or the numerous
336 LDLa and LRR repeats. These repeats are usually individually coded by single exons that are
337 sometimes skipped and when such skipped individual reads enter in the RNAseq SRA, incorrect
338 constructs are obtained. Furthermore, these repeats are present in numerous proteins, and from to
339 time this leads to assembled constructs that are from mRNA species from different genes. It is
340 therefore to be expected that not all assembled transcripts, both those in the databank and those
341 produced here, will be correct. Some errors were corrected by challenging divergent sequences
342 that were discovered on comparing putative orthologs with one another. Other differences could
343 be confirmed as true differences, but it is not impossible that some errors remain, particular for
344 those sequences that are incomplete. LGRs that might function as receptors for the various irps
345 were identified by their homology with such receptors from vertebrates and arthropods. The
346 transmembrane regions of the GPCRs don't have the assembly problems of the LDLa and LRR
347 repeats and are the most characteristic domain of the GPCRs. This makes it easier to construct a
348 phylogenetic trees for these receptors based on their transmembrane regions than that it is to
349 produce complete LGR transcripts.

350 Results show a surprisingly similar distribution of LGRs in the species studied. The tree
351 resolves two major branches, one for the glycoprotein hormone receptors, which itself is divided
352 in two subbranches one for orthologs of the GPA2/GPB5 receptor - containing the receptors for
353 human TSH, FSH and LH - and a second one for the bursicon receptor orthologs. All species
354 studied are represented by one member on each of these two subbranches, except for *Ophiothrix*,
355 where the draft genome reveals two orthologs each for the bursicon and GPA2/GPB5 receptors
356 (Fig. 9). These are likely receptors for the bursicon and GPA2/GPB5 orthologs identified from
357 various echinoderm species (Semmens et al., 2016). It is interesting to see that whereas
358 vertebrates have different receptors for TSH, FSH and LH, most echinoderms have only one
359 GPA2/GPB5-receptor ortholog, even though *A. rubens* has two GPA2 and three GPB5 orthologs
360 (Semmens et al., 2016). The LGRs for the glycoproteins were included in the search for putative
361 receptors for the ambulacrarian irp LGRs in order to be sure that no such receptors would be
362 missed.

363 The second branch of the LGR phylogenetic tree is the one of interest as it contains receptors
364 with irp ligands. It consists of three subbranches, that are characterized by *Drosophila* LGR3 and

365 LGR4 – the receptors for gonadulin and dilp7 respectively - and *Periplaneta* LGR5, an ortholog
366 of *Lymnaea* GRL101. Here in all ambulacrarian species studied only one ortholog was found for
367 each of them, despite extensive attempts to find additional LGRs in the various genomes and
368 transcriptomes.

369 The GRL101 transmembrane regions puts it very close to vertebrate glycoprotein hormone
370 and relaxin LGRs. LRRs are present in many different proteins, but when the LRR part of the
371 *Anneissia* GRL101 (amino acid residues 576-717) are used as query in a protein BLAST against
372 human proteins, it are the glycoprotein hormone and relaxin receptors that are identified as most
373 similar to this ectodomain of GRL101 suggesting that similarity of the GRL101 receptors with
374 vertebrate LGRs is not limited to the transmembrane region of this GPCR.

375 Sequence alignments of these GPCRs show strong sequence similarity (Figs. S11-S13),
376 however the dil7 receptor ortholog varies more between species. A schematic representation of
377 the the ectodomains of the LGRs on this second branch is drawn in Fig. 10. The orthologs of the
378 dilp7 and gonadulin receptors each have a single LDLa repeat, except for the *Patiria* and
379 *Acanthaster* orthologs of the dilp7 receptor which both have two LDLa repeats (Fig. S12). This
380 additional LDLa is likely due to a relatively recent duplication of the LDLa since the two LDLa
381 repeats have very similar amino acid sequences (Spreadsheet S1). All three receptors are
382 expressed in the nervous system and the gonadulin receptor is well expressed in the gonads, both
383 testis and ovary, and strongly so in the follicle cells of *P. pectinifera* (Spreadsheet 2).

384

385 Discussion

386 The genomic and transcriptomic data from both the hemichordates and the echinodermites
387 show that these two groups share three irps, octinsulin, IGF and a dilp7 orthologs, that are
388 present in both echinodermites and hemichordates. IGF and dilp7 are orthologs of the arthropod
389 peptides that together with gonadulin originated from a gene triplication. Evolutionary pressure
390 on gonadulin appears weak, as shown by the variable structure of this peptide and its loss in
391 many arthropod lineages. This may well be the reason that its structure is so poorly maintained,
392 even within insects (Veenstra, 2020b). This may explain why the amino acid sequence of
393 gonadulin looks significantly different from octinsulin. Nevertheless, there are two lines of
394 evidence that suggests that these peptides must be orthologs as well. For one, synteny of the

395 chromosome fragment containing these genes is conserved between the sea urchin
396 *Strongylocentrotus purpuratus*, the hemichordate *Saccoglossus kowalewski* and the cockroach
397 *Blattella germanica*, suggesting that these peptides are likely orthologs. More importantly, all
398 ambulacrarians have an ortholog of the gonadulin receptor and the only plausible ligand for such
399 a receptor encoded by their genomes is octinsulin. Thus the gene triplication previously reported
400 from arthropods must have occurred in a common bilaterian ancestor of the deuterostomes and
401 protostomes.

402 Crinoids have the simplest irp signaling system, one gene each for IGF, octinsulin and the
403 *dilp7* ortholog. Their putative receptors - insulin RTK, GRL101, and the orthologs of the *dilp7*
404 and gonadulin receptors – similarly are also each coded by a single gene. The hemichordates
405 have a very similar repertoire, except that the octinsulin gene is systematically amplified and in
406 some species the *dilp7* ortholog as well. It thus appears likely that the first deuterostome had a
407 single copy of each of these genes.

408 Within the echinoderms the irp genes evolved considerably, as shown both by an increase in
409 the number of irp genes and the loss of synteny. Whereas the feather stars appear to have only a
410 single IGF gene, all other echinoderms have two such genes and two novel insulin-like peptides,
411 GSS and multinsulin, appeared. The GSS sequences are most similar to those of IGF, suggesting
412 that they evolved from a gene duplication event from the IGF gene. Although some GSS genes
413 are located on the same chromosome as the other irps, they are not close to the IGF genes,
414 indicating that the IGF-GSS split was not a local duplication but may have been the result of an
415 incorrectly repaired chromosome break.

416 In the Asterozoa a fifth type of irp gene emerged, those that code for the multinsulins which
417 share significant sequence similarity with the *dilp7* orthologs. The initial multinsulin gene must
418 thus have its origin in a gene duplication of the *dilp7* ortholog gene, with which they furthermore
419 share a characteristic intron. Later the multinsulin gene seems to have undergone several
420 additional gene duplications in this respect the multinsulins resemble the insect neuroendocrine
421 irps.

422 The co-evolution of ligands and receptors allows one to assign the putative receptors for
423 gonadulin, the *dilp7* ortholog and IGF as the orthologs of the receptors of their arthropod
424 orthologs. This allows the identification of the ambulacrarian LGRs that are the orthologs of the

425 gonadulin and dilp7 receptors as likely receptors for octinsulin and the dilp7 respectively, as well
426 as the insulin RTK as a receptor for IGF.

427 The appearance of the multinsulins is not accompanied with the evolution of a novel insulin-
428 receptor. Some animals have multiple insulin RTKs, *e.g.* some arthropods have up to four such
429 genes (Veenstra, 2020a,b), however, in spite of extensive searches for a second insulin RTK in
430 ambulacrarian genomes, none were found. Searches for an additional LGR that might function as
431 a receptor for the GSS and/or multinsulin were neither successful and this raises the question
432 which receptors are activated by these peptides.

433 I have previously argued that the close chromosomal association of the IGF, gonadulin and
434 dilp7 ortholog genes in basal insects suggest that they derived from a gene triplication and that
435 since IGF and gonadulin respectively activate an RTK and an LGR it is likely that the original
436 IGF-like peptide - and possibly its descendants as well - acted on both a GPCR and an RTK
437 (Veenstra, 2020b). Some of the descendant peptides might still do so, while others may have lost
438 the ability to interact with one of the two original receptors. If the ancestral echinoderm dilp7
439 were still to use both a GPCR and an RTK, then the very different multinsulin sequences might
440 be explained by the requirement that dilp7 binds both a GPCR and the RTK while multinsulin is a
441 ligand for only one of those two.

442 The binding of insulin and relaxin to their respective receptors has gotten resolved in much
443 detail in the last couple of years. The effective binding and stimulation of insulin RTK by the
444 small insulin-like peptide from the snail *Conus* to the RTK shows that a small insulin-like
445 peptide can be an effective ligand for this receptor (Menting et al., 2015). On the other hand the
446 complex interaction of relaxin to its LGR makes it more difficult to imagine a smaller peptide as
447 an effective ligand (Hoare et al., 2019). Furthermore, considering the well conserved F-domain
448 of the dilp7 receptor orthologs it is likely that it is necessary for interaction with its LGR
449 receptor. The loss of this structure in multinsulin suggests that it is unlikely to be a dilp7 receptor
450 agonist. On the other hand, the poor sequence conservation in the various *Drosophila* irps that
451 activate a single RTK is reminiscent of the large structural variability of the multinsulins. This
452 seems to suggest that the multinsulins are RTK ligands rather than activate the LGR.

453 The emergence of the GSS is neither accompanied with the evolution of a novel receptor for
454 these irps. This can also be explained by assuming that IGF acts on both the RTK and an LGR

455 and that the GSS have lost their affinity for the LGR. This raises the question whether an IGF
456 LGR might exist.

457 If there were an IGF LGR, one would expect it to be related to the gonadulin and dilp7
458 receptors. GRL101 appears a plausible candidate as its transmembrane regions are closely
459 related to the receptors for gonadulin and dilp7. The ectodomain of GRL101 consists of two
460 parts, a series of LRRs and a second series of LDLa's. In the related GPCRs the LRRs are
461 expected to bind with the insulin core of gonadulin and dilp7 orthologs, just like the human
462 relaxin receptors (Hoare et al., 2019). When the LRR part of the *Anneissia* GRL101, the most
463 basal echinoderm, was used as query for similar human proteins in a BLAST search, it were the
464 glycoprotein hormone and relaxin receptors that were identified as the most similar proteins.
465 This shows that the resemblance of GRL101 to the other LGRs is not limited to the
466 transmembrane regions and reinforces the hypothesis that the ligand of GRL101 has an insulin-
467 like structure. GRL101 has a large number of LDLa's, the ligands of which are typically
468 positively charged surfaces, which in the case of proteins consist of Lys and Arg residues (Daly
469 et al., 1995; Prévost & Raussens, 2004; Fisher, Beglova & Blacklow, 2006; Yasui, Nogi &
470 Takagi, 2010; Dagil et al., 2013). Thus the ligand of GRL101 may consist of two parts, an
471 insulin-like structure and a piece with several positive charges that interact with the LDLa's. The
472 C-terminal tails of the IGFs, whether from arthropods, echinoderms or hemichordates, are all
473 rich in charged amino acid residues. The C-terminal tail of IGF with its numerous positively
474 charged amino acid residue might interact with the LDLa's of GRL101. I therefore posit that in
475 those species that have a GRL101 it functions as the second receptor for IGF. The absence of
476 such a tail in GSS would make it likely that it acts on the RTK rather than an IGF GPCR.

477 The suggestion that GSS activates the RTK goes against the hypothesis that these peptides
478 act through GPCRs. Indeed it has recently been proposed that it is the ortholog of the dilp7
479 receptor that would be activated by the gonad stimulator in *P. miniata* (Mita et al., 2020). Given
480 the clear orthology of both the dilp7 echinoderm orthologs with the *Drosophila* peptide and the
481 similar orthology between the dilp7 receptor and the echinoderm receptor, the conclusion that the
482 two constitute a function ligand receptor combination seems inescapable. It was impossible to
483 find a GSS in either the genome assembly or the individual reads of all the genomic SRAs of
484 *Anneissia japonica*, yet it does have a dilp7 receptor ortholog, thus if the dilp7 receptor were to
485 function as a GSS receptor, it most likely would not be an exclusive receptor. *A priori* this does

486 not exclude the possibility that GSS could function as a ligand for the same receptor. As
487 mentioned above, since the dilp7 orthologs have well conserved F domains, one has to assume
488 that it is important for binding to its receptor. Since this domain is absent from RTK ligands, it is
489 difficult to understand how GSS that similarly lacks this domain would be able to bind the dilp7
490 receptor. It would thus seem unlikely that peptides as different as GSS and dilp7 would be
491 effective ligands of the same LGR. Furthermore, the GSS genes have been duplicated and their
492 structures have evolved considerably. Those duplicate gonad stimulators are present in many
493 species and have not been selected against. Hence they must be physiologically relevant and able
494 to interact with a receptor. Sharing a common evolutionary origin the two gonad stimulators
495 should be expected to act either on the same or paralogous receptors, but the number of putative
496 echinoderm receptors for irps is limited, so it must be the same one. The same arguments that
497 were used to argue that the multinsulins are likely RTK agonists but not LGR ligands, are
498 therefore equally valid here and suggest that GSS is an RTK ligand.

499 Furthermore, the experimental evidence that GSS stimulates the ortholog of the dilp7
500 receptor is not convincing. The reported response to the dilp7 receptor when expressed in Sf9
501 cells is very weak and does not represent a typical response seen in this type of assay. Although
502 the authors have shown high affinity binding of GSS to the follicle cells, such high affinity
503 binding should also have been present in the Sf9 cells expressing the putative GSS receptor, but
504 this was not reported. The follicle cell SRAs from which the putative GSS receptor was
505 identified contains large amounts of RNAseq reads for the gonadulin receptor, a receptor that is
506 more closely related to the vertebrate relaxin receptors than the dilp7 receptor, but surprisingly
507 the authors do not mention this receptor, which they must have found (Mita et al., 2020).

508 I suggest that initially there was an IGF-like hormone that activated both a GPCR and an
509 RTK, after two gene duplications some of the descendant ligands either lost their C-terminal tails
510 or one acquired a larger one and this allowed all three ligands to activate, at least initially, the
511 RTK while each acquired its own LGR. Later, some of the ligands may have lost their affinity
512 for one receptor. Since the primary amino acid sequence of gonadulin is very different from that
513 of the other insulin-like peptides, it likely lost its capacity to activate the RTK. Holometabolous
514 insect species have lost GRL101 and hence in those species IGF can only act on the RTK. Under
515 this hypothesis the arginine-rich C-terminal tail would be useless in such insect species and in
516 higher flies, such as *Drosophila*, it was indeed lost (Veenstra, 2020b). In vertebrates, there is no

517 GRL101 and so IGF can only activate the two RTKs, while the relaxin related peptides are not
518 known to interact with RTK. The presence of a similar arginine-rich E domain of the vertebrate
519 IGF precursors might thus be an evolutionary relict.

520 This scheme raises the question as to how the functions of these two receptors activated by
521 IGF might differ. IGF and the drosophila insulin-like peptides stimulate growth, the echinoderm
522 GSS stimulates oocyte maturation and ovulation (Mita et al., 2009), relaxin and INSL3 affect
523 various developmental and reproductive processes (Ivell et al., 2020; Esteban-Lopez &
524 Agoulnik, 2020), gonadulin is expressed by the gonads as well as the imaginal in flies (Garelli et
525 al., 2012; Liao & Nässel, 2020; Veenstra, 2020b) and dilp7 is expressed in a sex specific manner
526 (Miguel-Aliaga, Thor & Gould, 2008; Yang et al., 2008; Castellanos, Tang & Allan, 2013).
527 These hormones stimulate growth, development and reproduction, processes that are intimately
528 linked; without growth and development reproduction is impossible and growth without
529 reproduction is useless in sexually reproducing species. On the other hand, resources used for
530 growth and development can not be used for reproduction or *vice versa*.

531 Growth is rarely a linear process independent of development; animals are not only getting
532 bigger, but they also mature into adults. Metamorphosis is markedly different between hemi- and
533 holo-metabolous insect species. Every time a cockroach nymph molts, it becomes a little more
534 adult, however during the first molts of a caterpillar the insects mainly become bigger, it is only
535 when it molts into a pupa that it significantly changes its morphology. Cockroaches have
536 GRL101, caterpillars don't. This suggests that the RTK might be more directed toward linear
537 growth, or allowing growth by increasing uptake of resources, such as glucose and amino acids,
538 while the LGRs might be more important for insuring that the animal develops into an adult and
539 becomes sexually competent. Both holometabolous insects and vertebrates have lost GRL101
540 and use steroid hormones to induce sexual maturation. Interestingly, in vertebrates the
541 production of steroid hormones is controlled by glycoprotein hormones, the second group of
542 ligands for LGRs.

543 It is plausible that IGF in an early bilaterian was produced by the tissue that stored energy
544 and perhaps even protein as insects do in the form of storage proteins (Haunerland, 1996).
545 Production and release of IGF might have happened when the animal had sufficient resources to
546 allow for growth and/or reproduction. In arthropods growth has become a discontinuous process
547 in which a new cuticle needs to be made before molting can take place. In those species IGF

548 produced by the fat body may well be the essential growth hormone. However, if the animal is
549 suddenly starved, IGF would no longer be released. If formation of a new cuticle is too advanced
550 to be interrupted, this become problematic. It may have obliged the brain to take at least partial
551 control of growth away from the fat body by releasing one or more of the neuroendocrine
552 insulin-like peptides to force growth and molting to proceed. It is possible that this achieved by
553 simultaneously reducing growth of organs that are needed for (sexual) maturation but not
554 essential for immediate survival, like the gonads. This could be how the neuroendocrine insect
555 irps initially evolved. In echinoderms IGF probably stimulates growth of the follicles and
556 oocytes, but the final growth spurt, the one that permits resumption of meiosis in the oocytes and
557 subsequent ovulation, is delayed until optimal conditions to do so prevail. When the time and
558 place are right the nervous system releases GSS likely in large amounts to finish the maturation
559 process and induce ovulation. In vertebrates, growth and the release of IGF has also been
560 brought under control of the brain but more forcefully by bringing IGF secretion by the liver
561 under control of growth hormone. Whereas in an early ancestor high plasma concentrations of
562 insulin might have led to secretion of IGF, this is no longer the case. Here insulin may have
563 evolved to insure that plasma concentrations of glucose are kept sufficiently low by insuring its
564 absorption by tissues in order to avoid its loss by excretion. In the three cases insulin and insulin-
565 like peptides have very different functions, ovulation in echinoderms, sparing glucose in
566 vertebrates and rescuing interrupted growth in insects. It is plausible then that these hormones
567 each evolved from a non-local IGF gene duplication and that they are thus not proper orthologs
568 but evolved by convergent evolution. This hypothesis would explain, why there is no insulin
569 gene located near the IGF, octinsulin/gonadulin and dilp7 triplet in cockroaches, echinoderms
570 and hemichordates, even though insulin – and other peptides such as the insect neuroendocrine
571 insulin-like peptides and GSS - almost certainly evolved from IGF much later.

572

573 **Conclusions**

574 The gene triplication previously reported from arthropods must have occurred in a common
575 bilaterian ancestor of the deuterostomes and protostomes. The hypothesis that IGF in an ancestral
576 bilaterian used both a GPCR and an RTK may explain the combination of echinoderm irps and
577 putative insulin receptors. This hypothesis implies that insulin is not a hormone that evolved
578 before the split between protostomes and deuterostomes, but that insulin-like peptides evolved

579 independently in different metazoan clades as miniature copies of IGF capable to activate the
580 RTK but unable to stimulate the LGR.

581

582 **Acknowledgements**

583 This manuscript is only made possible by the very large number of those who made their
584 transcriptome and genomic sequences publicly available. I express my sincere gratitude to all of
585 those.

586

587 **References**

588

589 Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Søren Brunak S,
590 von Heijne G, and Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep
591 neural networks. *Nat Biotechnol.* 37:420-423, doi:10.1038/s41587-019-0036-z.

592

593 Castellanos MC, Tang JC, Allan DW. 2013. Female-biased dimorphism underlies a female-
594 specific role for post-embryonic *ilp7* neurons in *Drosophila* fertility. *Development.* 140:3915-
595 3926. doi: 10.1242/dev.094714.

596

597 Chiba K, 2020. Oocyte maturation in starfish. *Cells* 9, 476. doi:10.3390/cells9020476.

598

599 Chieu HD, Turner L, Smith MK, Wang T, Nocillado J, Palma P, Suwansa-Ard S, Elizur A,
600 Cummins SF. 2019. Aquaculture breeding enhancement: Maturation and spawning in sea
601 cucumbers using a recombinant relaxin-like sonad-stimulating peptide. *Front Genet.* 10:77. doi:
602 10.3389/fgene.2019.00077.

603

604 Colombani J, Andersen DS, Boulan L, Boone E, Romero N, Virolle V, Texada M, Léopold P.
605 2015. *Drosophila* *Lgr3* couples organ growth with maturation and ensures developmental
606 stability. *Curr Biol.* 25:2723-2729. doi: 10.1016/j.cub.2015.09.020.

607

608 Dagil R, O'shea C, Nykjær A, Bonvin AM, Kragelund BB. 2013. Gentamicin binds to the
609 megalin receptor as a competitive inhibitor using the common ligand binding motif of
610 complement type repeats insight from the NMR structure of the 10th complement type repeat
611 domain alone and in complex with gentamicin. *J. Biol. Chem.* 288, 4424–4435. doi:
612 10.1074/jbc.M112.434159.

613

614 Daly NL, Scanlon MJ, Djordjevic JT, Kroon PA, Smith R. 1995. Three-dimensional structure of
615 a cysteine-rich repeat from the low-density lipoprotein receptor. *Proc Natl Acad Sci U S A.*
616 92:6334-8. doi: 10.1073/pnas.92.14.6334.

617

618 Davidson PL, Guo H, Wang L, Berrio A, Zhang H, Chang Y, Soborowski AL, McClay DR, Fan
619 G, Wray GA. 2020. Chromosomal-level genome assembly of the sea urchin *Lytechinus*

- 620 *variegatus* substantially improves functional genomic analyses. *Genome Biol Evol.* 12:1080-
621 1086. doi: 10.1093/gbe/evaa101.
- 622
- 623 Devi L. 1991. Consensus sequence for processing of peptide precursors at monobasic sites.
624 *FEBS Lett* 280:189–194. doi: 10.1016/0014-5793(91)80290-j.
- 625
- 626 Esteban-Lopez M, Agoulnik AI, 2020. Diverse functions of insulin-like 3 peptide. *J Endocrinol.*
627 247:R1-R12. doi: 10.1530/JOE-20-0168.
- 628
- 629 Garelli A, Gontijo AM, Miguela V, Caparros E, Dominguez M. 2012. Imaginal discs secrete
630 insulin-like peptide 8 to mediate plasticity of growth and maturation. *Science.* 336:579-582. doi:
631 10.1126/science.1216735.
- 632
- 633 Garelli A, Heredia F, Casimiro AP, Macedo A, Nunes C, Garcez M, Dias ARM, Volonte YA,
634 Uhlmann T, Caparros E, Koyama T, Gontijo AM. 2015. Dilp8 requires the neuronal relaxin
635 receptor *Lgr3* to couple growth to developmental timing. *Nat Commun.* 6:8732. doi:
636 10.1038/ncomms9732.
- 637
- 638 Fisher C, Beglova N, Blacklow SC. 2006. Structure of an LDLR-RAP complex reveals a general
639 mode for ligand recognition by lipoprotein receptors. *Molecular Cell* 22, 277–283. doi:
640 10.1074/jbc.M112.434159.
- 641
- 642 Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L,
643 Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F,
644 Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Full-length
645 transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.*
646 15;29:644-652. doi: 10.1038/nbt.1883.
- 647
- 648 Hall MR, Kocot KM, Baughman KW, Fernandez-Valverde SL, Gauthier MEA, Hatleberg WL,
649 Krishnan A, McDougall C, Motti CA, Shoguchi E, Wang T, Xiang X, Zhao M, Bose U, Shinzato
650 C, Hisata K, Fujie M, Kanda M, Cummins SF, Satoh N, Degnan SM, Degnan BM. 2017. The
651 crown-of-thorns starfish genome as a guide for biocontrol of this coral reef pest. *Nature.*
652 544:231-234. doi: 10.1038/nature22033.
- 653
- 654 Haunerland NH, 1996. Insect storage proteins: gene families and receptors. *Insect Biochem Mol*
655 *Biol.* 26:755-65. doi: 10.1016/s0965-1748(96)00035-5.
- 656
- 657 Hoare BL, Bruell S, Sethi A, Gooley PR, Lew MJ, Hossain MA, Inoue A, Scott DJ, Bathgate
658 RAD.2019. Multi-component mechanism of H2 Relaxin binding to RXFP1 through NanoBRET
659 kinetic analysis. *iScience.* 2019 Jan 25;11:93-113. doi: 10.1016/j.isci.2018.12.004.
- 660
- 661 Hsueh AJW, Feng Y, 2020. Discovery of polypeptide ligand-receptor pairs based on their co-
662 evolution. *FASEB J.* 34:8824-8832. doi: 10.1096/fj.202000779R.
- 663
- 664 Humbel RE, 1990. Insulin-like growth factors I and II. *Eur J Biochem.* 190:445-62. doi:
665 10.1111/j.1432-1033.1990.tb15595.x.

666
667 Imambocus BN, Wittich A, Tenedini F, Zhou F, Hu C, Sauter K, Varela EM, Herédia F,
668 Casimiro AP, Macedo A, Schlegel P, Yang CH, Miguel-Aliaga I, Pankratz MJ, Gontijo AM,
669 Cardona A, Soba P, 2020. Discrete escape responses are generated by neuropeptide-mediated
670 circuit logic. bioRxiv preprint doi:10.1101/2020.09.22.307033.
671
672 Ivell R, Alhujaili W, Kohsaka T, Anand-Ivell R. 2020. Physiology and evolution of the
673 INSL3/RXFP2 hormone/receptor system in higher vertebrates. Gen Comp Endocrinol.
674 299:113583. doi: 10.1016/j.ygcen.2020.113583.
675
676 Jo J, Oh J, Lee HG, Hong HH, Lee SG, Cheon S, Kern EMA, Jin S, Cho SJ, Park JK, Park C.
677 2017. Draft genome of the sea cucumber *Apostichopus japonicus* and genetic polymorphism
678 among color variants. Gigascience. 6:1-6. doi: 10.1093/gigascience/giw006.
679
680 Liao S, Nässel DR. 2020. *Drosophila* insulin-like peptide 8 (DILP8) in ovarian follicle cells
681 regulates ovulation and metabolism. bioRxiv. doi: 10.1101/2020.05.02.073585.
682
683 Lin M, Mita M, Egertová M, Zampronio CG, Jones AM, Elphick MR., 2017. Cellular
684 localization of relaxin-like gonad-stimulating peptide expression in *Asterias rubens*: New
685 insights into neurohormonal control of spawning in starfish. J Comp Neurol. 525:1599-1617.
686 doi: 10.1002/cne.24141.
687
688 Linneweber GA, Jacobson J, Busch KE, Hudry B, Christov CP, Dormann D, Yuan M, Otani T,
689 Knust E, de Bono M, Miguel-Aliaga I., 2014. Neuronal control of metabolism through nutrient-
690 dependent modulation of tracheal branching. Cell 16;156:69-83. doi: 10.1016/j.cell.2013.12.008.
691
692 Menting JG, Gajewiak J, MacRaild CA, Chou DH, Disotuar MM, Smith NA, Miller C, Erchegyi
693 J, Rivier JE, Olivera BM, Forbes BE, Smith BJ, Norton RS, Safavi-Hemami H, Lawrence MC.,
694 2015. A minimized human insulin-receptor-binding motif revealed in a *Conus geographus*
695 venom insulin. Nat Struct Mol Biol. 23:916-920. doi: 10.1038/nsmb.3292.
696
697 Miguel-Aliaga I, Thor S, Gould AP. 2008. Postmitotic specification of *Drosophila* insulinergic
698 neurons from pioneer neurons. PLoS Biol 6: e58. doi:10.1371/journal.pbio.0060058.
699
700 Mirabeau O, Joly JS. 2013. Molecular evolution of peptidergic signaling systems in bilaterians.
701 Proc Natl Acad Sci U S A. 110: E2028-37. doi: 10.1073/pnas.1219956110.
702
703 Mita M, Yoshikuni M, Ohno K, Shibata Y, Paul-Prasanth B, Pitchayawasin S, Isobe M,
704 Nagahama Y. 2009. A relaxin-like peptide purified from radial nerves induces oocyte maturation
705 and ovulation in the starfish, *Asterina pectinifera*. Proc Natl Acad Sci U S A. 106:9507-12. doi:
706 10.1073/pnas.0900243106.
707
708 Mita M, Matsubara S, Osugi T, Shiraishi A, Wada A, Satake H. 2020. A novel G protein-
709 coupled receptor for starfish gonadotropic hormone, relaxin-like gonad-stimulating peptide.
710 PLoS One. 15:e0242877. doi: 10.1371/journal.pone.0242877.
711

- 712 Perillo M, Arnone MI. 2014. Characterization of insulin-like peptides (ILPs) in the sea urchin
713 *Strongylocentrotus purpuratus*: insights on the evolution of the insulin family. Gen Comp
714 Endocrinol. 205:68-79. doi: 10.1016/j.ygcen.2014.06.014.
715
- 716 Prévost M, Raussens V. 2004. Apolipoprotein E-low density lipoprotein receptor binding: study
717 of protein-protein interaction in rationally selected docked complexes. Proteins 55:874-84. doi:
718 10.1002/prot.20080.
719
- 720 Price MN, Dehal PS, Arkin AP. 2010. FastTree 2--approximately maximum-likelihood trees for
721 large alignments. PLoS One. 5:e9490. doi: 10.1371/journal.pone.0009490.
722
- 723 Rholam M, Brakch N, Germain D, Thomas DY, Fahy C, Boussetta H, Boileau G, Cohen P.
724 1995. Role of amino acid sequences flanking dibasic cleavage sites in precursor proteolytic
725 processing. The importance of the first residue C-terminal of the cleavage site. Eur J Biochem.
726 227:707–714. doi: 10.1111/j.1432-1033.1995.tb20192.x.
727
- 728 Ruiz-Ramos DV, Schiebelhut LM, Hoff KJ, Wares JP, Dawson MN. 2020. An initial
729 comparative genomic autopsy of wasting disease in sea stars. Mol Ecol. 29:1087-1102. doi:
730 10.1111/mec.15386.
731
- 732 Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. 2000.
733 Artemis: sequence visualization and annotation. Bioinformatics 16:944-5. doi:
734 10.1093/bioinformatics/16.10.944.
735
- 736 Sea Urchin Genome Sequencing Consortium. 2006. The genome of the sea urchin
737 *Strongylocentrotus purpuratus*. Science 314: 941–952.
738
- 739 Semmens DC, Mirabeau O, Moghul I, Pancholi MR, Wurm Y, Elphick MR. 2016.
740 Transcriptomic identification of starfish neuropeptide precursors yields new insights into
741 neuropeptide evolution. Open Biol. 6:150224. doi: 10.1098/rsob.150224.
742
- 743 Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert
744 M, Söding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein
745 multiple sequence alignments using Clustal Omega. Mol Syst Biol. 7:539. doi:
746 10.1038/msb.2011.75.
747
- 748 Simakov O, Kawashima T, Marlétaz F, Jenkins J, Koyanagi R, Mitros T, Hisata K, Bredeson J,
749 Shoguchi E, Gyoja F, Yue JX, Chen YC, Freeman RM Jr, Sasaki A, Hikosaka-Katayama T, Sato
750 A, Fujie M, Baughman KW, Levine J, Gonzalez P, Cameron C, Fritzenwanker JH, Pani AM,
751 Goto H, Kanda M, Arakaki N, Yamasaki S, Qu J, Cree A, Ding Y, Dinh HH, Dugan S, Holder
752 M, Jhangiani SN, Kovar CL, Lee SL, Lewis LR, Morton D, Nazareth LV, Okwuonu G,
753 Santibanez J, Chen R, Richards S, Muzny DM, Gillis A, Peshkin L, Wu M, Humphreys T, Su
754 YH, Putnam NH, Schmutz J, Fujiyama A, Yu JK, Tagawa K, Worley KC, Gibbs RA, Kirschner
755 MW, Lowe CJ, Satoh N, Rokhsar DS, Gerhart J. 2015. Hemichordate genomes and
756 deuterostome origins. Nature. 527:459-465. doi: 10.1038/nature16150.
757

- 758 Smith MK, Chieu HD, Aizen J, Mos B, Motti CA, Elizur A, Cummins SF. 2019. A Crown-of-
759 Thorns Seastar recombinant relaxin-like gonad-stimulating peptide triggers oocyte maturation
760 and ovulation. *Gen Comp Endocrinol*. 281:41-48. doi: 10.1016/j.ygcen.2019.05.009.
761
- 762 Tensen CP, Van Kesteren ER, Planta RJ, Cox KJ, Burke JF, van Heerikhuizen H, Vreugdenhil E.
763 1994. A G protein-coupled receptor with low density lipoprotein-binding motifs suggests a role
764 for lipoproteins in G-linked signal transduction. *Proc Natl Acad Sci U S A*. 91:4816-4820. doi:
765 10.1073/pnas.91.11.4816.
766
- 767 Vallejo DM, Juarez-Carreño S, Bolivar J, Morante J, Dominguez M. 2015. A brain circuit that
768 synchronizes growth and maturation revealed through dilp8 binding to Lgr3. *Science*.
769 50:aac6767. doi: 10.1126/science.aac6767.
770
- 771 Veenstra JA. 2000. Mono- and dibasic proteolytic cleavage sites in insect neuroendocrine
772 peptide precursors. *Arch Insect Biochem Physiol*. 43:49-63. doi: 10.1002/(SICI)1520-
773 6327(200002)43:2<49::AID-ARCH1>3.0.CO;2-M.
774
- 775 Veenstra JA. 2020a. Gonadulins, the fourth type of insulin-related peptides in decapods. *Gen
776 Comp Endocrinol*. 296:113528. doi: 10.1016/j.ygcen.2020.113528.
777
- 778 Veenstra JA. 2020b. Arthropod IGF, relaxin and gonadulin, putative orthologs of *Drosophila*
779 insulin-like peptides 6, 7 and 8, likely originated from an ancient gene triplication. 10: e9534.
780 doi: 10.7717/peerj.9534.
781
- 782 Veenstra JA, Rombauts S, Grbić M. 2012. *In silico* cloning of genes encoding neuropeptides,
783 neurohormones and their putative G-protein coupled receptors in a spider mite. *Insect Biochem
784 Mol Biol*. 42:277-295. doi: 10.1016/j.ibmb.2011.12.009.
785
- 786 Warner JF, Lord JW, Schreiter SA, Nesbit KT, Hamdoun A, Lyons DC. 2021. Chromosomal-
787 level genome assembly of the painted sea urchin *Lytechinus pictus*, a genetically enabled model
788 system for cell biology and embryonic development. *Genome Biol Evol*:evab061. doi:
789 10.1093/gbe/evab061.
790
- 791 Yang CH, Belawat P, Hafen E, Jan LY, Jan YN. 2008. *Drosophila* egg-laying site selection as a
792 system to study simple decision-making processes. *Science* 319: 1679–1683. doi:
793 10.1126/science.1151842.
794
- 795 Yasui N, Nogi T., Takagi J. 2010. Structural basis for specific recognition of reelin by its
796 receptors. *Structure* 18, 320–331. doi: 10.1016/j.str.2010.01.010.
797
- 798 Zhang X, Sun L, Yuan J, Sun Y, Gao Y, Zhang L, Li S, Dai H, Hamel JF, Liu C, Yu Y, Liu S,
799 Lin W, Guo K, Jin S, Xu P, Storey KB, Huan P, Zhang T, Zhou Y, Zhang J, Lin C, Li X, Xing L,
800 Huo D, Sun M, Wang L, Mercier A, Li F, Yang H, Xiang J. 2017. The sea cucumber genome
801 provides insights into morphological evolution and visceral regeneration. *PLoS Biol*.
802 15:e2003790. doi: 10.1371/journal.pbio.2003790.
803

Figure 1

Sequences of selected ambulacrarian IGF.

Partial IGF sequences from selected ambulacrarians are illustrated to show their sequence similarity. The A-, B- and C-domains of the insulin core are aligned, but not the putative D- and E- domains, as their amino acid sequence is only conserved in closely related species (Fig. S1). Not aligning D- and E- domains allows the visualization the context of putative convertase cleavage sites. None of the arginine or lysine residues conform to a typical arthropod or vertebrate convertase cleavage site. Although the sequence of the latter part of the IGF precursors is not well conserved, all of them are rich in positively charged amino acid residues. Conserved cysteine residues are indicated in red, conserved amino acid residues are highlighted in black and conserved substitutions in grey. The arginine and lysine residues in the D- and E- domains are highlighted in blue.

Anneissia QPNHVRRYCGSELTSELERRCALKGGFNSPKKQESP-----
Strongylocentrotus-1 ---SFPLLCGQELVKAVAAVCNDR-GYYGQPS-----
Strongylocentrotus-2 ---SFRLCGRELADALAVVCKGR-GYYIDDSEI-----
Apostichopus-1 ---SGQRYCGEALLEAMAYVCGDR-GYFGVTSGIRGRS-----
Apostichopus-2 ---AQRYCGTNLADALRIVCADR-GYYTQ---KGAF-----
Acanthaster-1 ---AVIQVCGNDLLDALKSVCGDR-GFYSPPPGYS-----
Acanthaster-2 ---WQRI CGEQLVETVSIVCNTR-GFYSHR-E-----
Ophiothrix-1 ---WQRLCGTFLVDVVSQVCGER-GTYADDSTHDLRKRSLDNFVDRGKMYSAGISKT
Ophiothrix-2 ---SYL CGSQIVQMMAVVCEGR-GYYYTEGSTGQICN-----
Saccoglossus ---WDKLCGRTLVDVLAALICNGR-GYNSGSPKK-----
Ptychodera ---WDRLCGRSLADMLALVCHGR-GYYTDVSRQQ-----
Schizocardium ---WDKLCGRTLADMLSLVCHGR-GYYTDVSKHR-IR-----

Anneissia -----YLWNQEKRIEATS-----TERAKEINIVYECCHSACTEAFIDS
Strongylocentrotus-1 ---KRSAGIELETRAKTEFLKSGISRGETRRSKRGARTGLIVTECCLNRCVSHLES
Strongylocentrotus-2 ---AQKDSPIVPHVASSFLGSSS-ASAHSRQRRRVRTGOIVNECCDKECSNNIMES
Apostichopus-1 ---VSRSPFLTEERANSFLTNE-----RTRKTRRTGRIVTECCDNP CSQRNLES
Apostichopus-2 ---PEVIPRQRASSELTDS---ENHHPRLRRGTGIVTECC EKA CDREVLET
Acanthaster-1 ---KREAELEODERTAKSEFLGTHIG-----SRQRRRTGRIATECC EKV CSYDIVES
Acanthaster-2 SDTNAESNFERGESEAKSEFLGL-----SRQKRRTGRIVHECCNNICNYRIIES
Ophiothrix-1 IVYSFLTDPELLEERANSFLTNE-----RTRTKRRTGRIVTECCDNHCDMQIIIES
Ophiothrix-2 ---KVKRESPEERSGMEANDEFGNIS-SKEKRQRRRSGSGKIVDECC HQA CDYTTLES
Saccoglossus ---KRRRARETVFATQEEANGFEGVG---S---GRTKRRRSGSLIVVECCDKICDYSTIES
Ptychodera ---LPRETIEQTQEDAHKEFGASV-FG---ERTQRRRSGSLIVAECC EKS CDYATIES
Schizocardium

Anneissia FCLSSNKDEDTTVESDVTETTITGKKRKPTRKPKNPKSRKKPKGSSEINSEQSASN
Strongylocentrotus-1 YCNPLPPDAVHDAEVHIRLEKSAEEDADEGRPDQGPSQLDTATGTVPETEMSETRGRV
Strongylocentrotus-2 YCNRRTPPEVPESAISENPFSEITEDSTLRITDGESTEIRITDTNPNATNLEVPSPDANTP
Apostichopus-1 YCNVATTQTTEIPTELTTEGTTTEPAASPRRNSRNIEADGTAAGGGSGQNGRGRGK
Apostichopus-2 YCNPHVPTLALASLVTSIMTKSPTPPSSEPSSSSSSSSSSRNEDKFFPMTDNALGEDY
Acanthaster-1 YCNPPSTSQSQTAAAPRITTTTPDERRANEIVVDETGQTGNTNSQMLRGGNAMGAGSR
Acanthaster-2 YCNPWVVEDRDDPMLAPVAPGRVRODKSADADLLLRPDIAEISEDKSSLLRQAAEKD
Ophiothrix-1 YCNPWPTTTTTATTTQSPEPLPNEQEGGYLTDEIQMKHITRGGQEDNSVDLLSEGSDLR
Ophiothrix-2 YCAPLQEGQVKFTSRNLDVFNVEANNVVEEPSVRVPQQVEENKIVEETVYVYSQAIVGE
Saccoglossus YCAPLPEGVVADSLKRFSLQSFGNDFKDTANEDKLEIVTVVRPSHDEMDGTETRIED
Ptychodera YCAPWPKDIDPAKKIEGFKEGTWEEDYHRYHPESVEQPNANPEEPTPEPTTTDLDK
Schizocardium YCAPWPKMDPALRFAGFKYGSWEDEDYRYKYPHEEFTQPDLTTWIASTDSANHDEHF

Anneissia TELPSQTEPTKDKNGRGDKNKKDKCNKKSRLKNNKPCRRKS RKDSRRKNRKNKKNKPK
Strongylocentrotus-1 RIDAVEKVLSERLIPTSTTGSSPSPSRKKPKRDKSERRNSREAKQARREERRNRER
Strongylocentrotus-2 DATATSDVEQPRSDNTTAVEKPRKKDNGKGNSSLESSTKKNRRTSKGMSKEDRRRIAS
Apostichopus-1 KGRHGKGNRREQDQTVDVTS AETEGNTEPPRPNQENEDNRGTPDQSEERPRDRCRG
Apostichopus-2 EGPTNEGPLTSGEPTPTENRISNGPRGPSSNASSLELPTRTSTATTNSSRIVTEGAHL
Acanthaster-1 ANGTKAPTEVVDGRSDDDDAAGEINTSERVGSLEPDEETGRDVATNRPHKTHSKE
Acanthaster-2 EPIDDLTLENEYADGGNVMQA REGVKEEGAELGKEMEEGGEKMPFPEVVP TKKRRR
Ophiothrix-1 DETPMPTRATGGRSPQLGHDQADVDSGFH RFGDIGAEVEESDSTIDNPSSDR---
Ophiothrix-2 TTSDGFDWSDVSETHNINLGGGDMYINDQPRDDPSENDEEHIPKKEKSLIKQINVTES
Saccoglossus NEHVTPPTKPDVITETSSLLDDINVNKQIISNNTSVEVKS KAGNTKPKREKDRDNS
Ptychodera SLESERDNVDIEESRSKDKSAESESVTENDDLKSEETNDNQDEYSREYLGDRDKGP
Schizocardium TTATPSEETTSSENGEHIKTVNKEDDTQVTGTHELEKDHKFAEYWMNEDKMSRKHKA

Anneissia TRRPKRWDSETTTLSDFLFSQRFQLQRAYSDIDVDEITTEFGVVPPEEDISSSLSEEL
Strongylocentrotus-1 SSGRSTRNGRRKDKNDRASRAKRHGLNLWRNMFSDKFFSDIPGLENQPNLHPVNGRA
Strongylocentrotus-2 DERRASREKKELSERRRKRLKLQQRKDKKKKRLLESAERNRGTDMGLSEDSTLLAR
Apostichopus-1 SKKKGKCRNQDRVEEPTQDRIEEPTTSREVSSTDGLSERERGSGGGRRNRGNGKCK
Apostichopus-2 AESSDQSSLEDTDESEAPT KASHERTRKKQKKQRTTRPKPKKLSREERKERRKKE
Acanthaster-1 RSKNRTSKSERRRRTNRRSSSERMLSSERKREDATRKLRRKEQRLSRKQPHSNKR
Acanthaster-2 VEGRRSRENSRDRNGKSEKSKRKSRSREGGRSFRRRGKNSRRKKGRDGRERSKRWEA
Ophiothrix-1 QTDEILDEDTDSENSESKLSIGATKNSGRTPKPHRPTSRHSRRKNSREKKNKTKTR
Ophiothrix-2 SSKRSHPKPKSRKKQLRQIRKKARGTRKTCLRHVVKSKSTPIQOIETTTTMMKPF
Saccoglossus ERKHKKAPT KLFKEKKLSEDNKKKKSRAKSKNSTKVKPTYVSSMTTSD EETLTPQG
Ptychodera KRLMKERRARKSRTKKGNNTKSKSILGKLVENMETESP TNWQRDGTDERWWTVEDPHF
Schizocardium

Figure 2

Sequences of selected echinoderm GSS.

Sequences of selected echinoderm GSS. Sequence alignment of a few echinoderm GSS showing relatively conserved A- and B- domains of the insulin core sequence and likely KR convertase cleavage sites that can be expected to be cleaved by neuroendocrine convertase as well as a few potential furin sites. Conserved cysteine residues are indicated in red, conserved amino acid residues are highlighted in black and conserved substitutions in grey. The arginine and lysine residues that form likely - or possibly in the case of *Apostichopus* GSS-2 - part of a convertase site are highlighted in blue. For the alignment of a larger number of echinoderm GSS sequences see Fig. S3.

```

Strongylocentrotus  --QQGPRNRYCGLEFARAVETQCSMANKRSDPGAVAESASAARYLA
Apostichopus-1      -----IRLCGPDLSRAVYQICSHG-KRGYIPPTFNS--E-----
Apostichopus-2      ----WSHQRLCGPDLVHALSLVCGERG---YFGGSRLVER---D-V
Acanthaster-1       ----EKFCDNDFHLAVYQTCSTH-KRGDGEPVLSL--K-----
Acanthaster-2       ----DSSSKHCGSAFPQFVWTACSMA-KRS-NRSPRSL--D-----
Ophiothrix-1        --DSARYQPLCGREFTRAVMEICATQVKRTEPLFORFYAN-----
Ophiothrix-2        ----QDSYKSCGREFTRRVMEVCATHVKRTEHF-----

Strongylocentrotus  DTGYEQAEDMPLEWYDVARQGAERLRP-----SL---TDIIF
Apostichopus-1     DDQLNQEFGTDL-----EEYLAETIKEYLKPNSLYDDVERELYPSL
Apostichopus-2     DDGLDEEITTLV-----VGAERT-----SILECLK--AWSPF
Acanthaster-1      D--VLTGS-----RLRG-NIKRSFGSTLE--DEAFF
Acanthaster-2      D--LLETF-----KSAR-----HLDIS--Y--RT
Ophiothrix-1       --LVKRSIDPAFWNNLLEANP-----DLMD
Ophiothrix-2       --MVKRSIDDEFWNDLMESGL-----GL--

Strongylocentrotus  SRFRRSIHNRGQLPMGQLCCVYGCTLVELASVCT-----
Apostichopus-1     RG---FRRVTRTGGIARRCSTGCSSSDIAKLC-----
Apostichopus-2     -----RRRTRGIVEECFRRCTWENLESYCSKTTAYKKADNMI
Acanthaster-1      SR---LVKRSEYDGIASYCCIHGCTPSELAVVC-----
Acanthaster-2      IR---LSKRQDYDGMADYCCIIGCSTNELIASGIC-----
Ophiothrix-1       ----KRQSSAGVGMATHCCQSGCSQEISMVC-----
Ophiothrix-2       ----DKR--SETGMAEHCCQNGCTDQEISMVC-----

```

Figure 3

Sequences of selected dilp7 orthologs.

Sequences of selected dilp7 orthologs. Sequences of *Drosophila* dilp7 and several ambulacrarian orthologs illustrating well conserved sequences, not only in typical insulin core of the peptides, but also in the F-domain (underlined in blue). Note that the sequence conservation of these peptides is stronger than in the IGFs or GSSs (Figs. 1,2). Conserved cysteine residues are indicated in red, conserved amino acid residues are highlighted in black and conserved substitutions in grey. Likely convertase cleavage sites have been highlighted in blue. Sequences are from Spreadsheet S1 and Veenstra 2020b, a comparison of a larger number of sequences is presented in Figs. S5 and S6.

Drosophila	-----LQHTEEGLEMLFRERSQSDWENVVHQETHSR--	C-RDKLVRQLYWAC
Anneissia	-----LRDYSDRSHNDWARVWTVESMRQ--	C-HENLREMVHVS
Strongylocentrotus	-EKFCNCMVLPELTMEDYEDRTPPEEWRESWNMDTLRT--	C-VGPPQLQRVGELAC
Apostichopus	-----TLQELNSRTPQSWEQWLVIVENVPTVD	C-TV-DAVQLHIIAC
Acanthaster	-----APHLPVEQWNSRSKADWVKLWNTERHVNT-	C-NEALLPVWDVAC
Ophiothrix	-----KITDYSSRTKADWQRLWLTEESHQK--	C-NEDLLPLWKIAC
Saccoglossus-1	LTVENEETITFSELNTMYGTRTLTDWQGWKWTSETIHA--	C-GSNLYRISEYV
Ptychodera	----AKDYDTVEDLILKYGTRTEDDWRNVWHTESHNK--	C-RNSLYSYVIFAC
Schizocardium	----ENAYETVESLKAKYSGRTFNDWRAAWNTETHNK--	C-RNSLYSYVIFAC

Drosophila	EKDIYRLTRRNK KRT GNDEAWIKKTTTEPDGSTWLHVNYANMFL RSRR -----
Anneissia	RNDPRK-ISS KRS -----IFIPRNEATGFLSRFL-----
Strongylocentrotus	INDPRKTIVV KRSN -----SDRDLELPAKLAKAFLHYRHRKD--
Apostichopus	SNDVYKDHEGS RRRRF -----SD KRRRS --IELNFSEANNELA-KTKRTHS
Acanthaster	QNDIRK-IT- KRM -----GREFLNEWTAKNFLAGS-----
Ophiothrix	TYDIRK-IN- KRA -----PEFVEDSEAKAFLIGP-----
Saccoglossus-1	YVDIHKSPDR KRTD -----DAFVDSAVAHDFLRGLMEKRTL
Ptychodera	MVDIYATGDK KRAE -----PPEVEHSIANSELAGHRAEKRM
Schizocardium	MKDIHASGDR KRGQ -----FPFLDHEVADKFLSGSL-GKRM

Drosophila	SDGNTPSISNE CC T-KAG CT WEEYA EY - CP SNKRRNH Y -----
Anneissia	RTRRP SELHED CC LDSRG CT WEEVA EIA CINNRRRMHRPGSPVGR-----
Strongylocentrotus	SRRRR VGKDE EE CAEAQ CR WEE LGEY - CTL HTRAYHQSGEQP-----
Apostichopus	RVRRT TTTFSTE CC D--KL CI WEEV GEY - CWH -SRV YH -----
Acanthaster	- KRRKR GLNE EE CH ED LG CV WEEVA EY - CVM HGREKHEDGSPV RGKPGRRR
Ophiothrix	- RRHKR GLNE EE CH ES KG CV WEE IGEY - CRM HSRASHVDGR ISDR -----
Saccoglossus-1	RRYRR TSATSE CC ADDG GV WEE LAEY - CTH QREVRT MDE -----
Ptychodera	TRFRR ASPT EE CC SNA H GN WEE LAEY - CS HQRDD TRRR -----
Schizocardium	ARFRR ANPS EE CC SNR NG CT WEE LAEY - CN HQRDD IDIGR -----

Figure 4

Sequences of selected ambulacrarian octinsulins.

Sequences of selected ambulacrarian octinsulins. Sequence alignment of a number of octinsulin sequences show that these sequences all have typical neuroendocrine convertase KR cleavage sites, suggesting these precursors are processed by enteroendocrine and/or neuroendocrine cells. Conserved cysteine residues are indicated in red, conserved amino acid residues are highlighted in black and conserved substitutions in grey. Likely convertase cleavage sites have been highlighted in blue. Sequences are from Spreadsheet S1, a comparison of a larger number of sequences is presented in Figs. S7 and S8.

```
Anneissia -----ARDWYC-GNAADTLKEFCQS CYASKRAHN--ALSLP
Strongylocentrotus -----QSWHC-GRAAQTIMGMCNS CYASHDKRS-----
Apostichopus -----SWYC-GSAPETVRAICDG CYAGGIHTR--AFKRS
Acanthaster -----DSWYC-SDVYSTVQSLCDS CYAGFDKRT-----
Ophiothrix -----NQWFC-SPVFTMLQSMCGS CYAGVDKRS-----
Saccoglossus-1 -----MSRNWHC-GRPVE TMHEVCQG CYAGHVRPR-----
Ptychodera-1 -----LRREWHC-GRTVETMQGICRG CYAQPSERS-----
Schizocardium-1 -----FTKHWHC-GRIVDTMRAICDG CYASPTARD-----
Saccoglossus-2 RPSGSVD-----DVLTC-KRKL LLDVQICAG CYAPPDI INNVNFDLT
Ptychodera-2 RPRNQGD-----D-VFC-SRTYSMVESVCDG CYATTQDSSPKSESAM
Schizocardium-2 RPNRPSE-----D-VRC-RKTIHMLVLCNG CLAPIESEVENN--TI
Saccoglossus-3 ---LPVDDTTTGVDVDRDRLWLC-GRLVEDLRALCRG CYAG---P-----
Ptychodera-3 -DLLPGNGDY---SNSAKGRRDLC-THLVESMSLICRG CYATDQGV-----
Schizocardium-3 -----QTSADRPKWHCRKSVPEILSGVCRG CYAEP LQPP-----
```

```
Anneissia SIKA-----KKGDFELTKEGASGYLEAKRTRL-----FSS-LHLNHRQHETT
Strongylocentrotus ISKPSY---TPAKPFLHKRNVAHF LRRTTKREIESRPSMGDTAIEVAVERRSTGNR
Apostichopus SSDIIS---LYKDPFLKKSNA LNFLLPRSHTP-----S-----SLIKRGIRRS
Acanthaster NSITRP---I-EEPFVERKNAVDFFKRG TAR-----G-----TRR
Ophiothrix DNSDTLSQKQSLDAEIQKEVAYSFIKRTSVG-----DT---FLRNARNTHH
Saccoglossus-1 NTRS---VDGVQAFISRRDANMFTKGMSPD-----V---KRAIDG
Ptychodera-1 T--N---EAERQAFIGKEEASSETKSV-----TIVKR
Schizocardium-1 V--G---SVKGLPELKKHEASTFTRTRS-----SVQKR
Saccoglossus-2 GK-----GPGY-----EEEIVKN
Ptychodera-2 NE-----SVQI-----QDEDLRE
Schizocardium-2 SN-----ETPP-----DQDVVTA
Saccoglossus-3 -----DISKREASKFMQFNA-----HTIRQSR
Ptychodera-3 -----TVNKRLASSFIPTTTP-----R---NKIQR
Schizocardium-3 G-----KRDILNKQEA SLFLRSP-----NGGDE
```

```
Anneissia NFVTECCYNP CSSFEMIKYCCPTRQIE LHNRNPNSSEDK-----
Strongylocentrotus GFIHECCNKFCDPGEMVLYCC EKROI EWAQFHNLLKA-----
Apostichopus GFIGECC EKNCEIREMVFYCCAEKQREYASEFFEIFRNRIRHT-----
Acanthaster GIVDECCHRQCAVSEMMLYCC EQKQREY YTFVGWLKRR-----
Ophiothrix GLIDECC TQCCDTGEMILYCCQERQREWHIMGLYNK-----
Saccoglossus-1 GLIEECCYSQCSLTHMITYCCAEVQNEFQVEINILGNTDESSENDGDDGEESV
Ptychodera-1 GLLLEDCCYRRCNLQKMMTYCCAERQRELNNEFSLLNQKDNST-----
Schizocardium-1 GIIEHCCNHHCSTFELLIYCCERSEEFYSEFGLLRMDDEDTDASLEKNGDVEA
Saccoglossus-2 QIKEACCKEYCPLPKIIEFC DERQQEFHQFMSSFASTEE-----
Ptychodera-2 KIRDKCCNRRCTIHKMMQFCCEARRNEFHKFLALMGNTDN-----
Schizocardium-2 KVREVCCDNYC SLDKTIEFCCEDLQOEFROEFMSFVSNK-----
Saccoglossus-3 GIIEECCYHTCPTERKIQYCCFEVQAQYRLFMESAI-----
Ptychodera-3 GIIDECCRNRC SVERKIQYCCDEIQKEFAFFFSLFGS-----
Schizocardium-3 LIINECCLRCTVIEKIH YCCREKQIELYII IQSAPWLVDNQR-----
```

```
Anneissia -----
Strongylocentrotus -----
Apostichopus -----
Acanthaster -----
Ophiothrix -----
Saccoglossus-1 HED-----
Ptychodera-1 -----
Schizocardium-1 EDNTDLQGD DDDNTQNNEGGSIDVVVERDVGDVLLKTDKSKP
Saccoglossus-2 -----
Ptychodera-2 -----
Schizocardium-2 -----
Saccoglossus-3 -----
Ptychodera-3 -----
Schizocardium-3 -----
```

Figure 5

Sequence comparison of selected ambulacrarian multinsulins and dilp7 orthologs.

Sequence comparison of selected ambulacrarian multinsulins and dilp7 orthologs. Three different sets of sequences are compared. The top five sequences are dilp7 orthologs, the next five are multinsulins having three disulfide bridges and the last five multinsulins having four disulfide bridges. Note that although the multinsulins and the dilp7 orthologs share some sequences similarity this does not include the F-domain. Like the octinsulins these sequences all have typical neuroendocrine convertase KR cleavage sites, suggesting they are processed by enteroendocrine and/or neuroendocrine cells. Conserved cysteine residues are indicated in red, conserved amino acid residues are highlighted in black and conserved substitutions in grey. Likely convertase cleavage sites have been highlighted in blue. Sequences are from Spreadsheet S1, a comparison of a larger number of sequences is presented in Figs. S9 and S10.


```

Anneissia -----LRDYSDRSHNDWARVWTVE
Antedon -----LQDYSDRSHHDWARVWTVE
Strongylocentrotus -----EKFCNCMVLPELTMEDYEDRTPPEWRESWNMD
Apostichopus -----TLQELNSRTQPSWEQLWIVE
Acanthaster -----APHLPVEQWNSRSKADWVKLWNT
Ophiothrix -----KITDYSRTKADWQRLWLTE

Amphiura-1 -----D-----WVG-
Ophioderma-1 -----DEIGAMKSQNEEPDEHDS----HIGTTQQPVINQRSGMTMEKR----WRG-
Pisaster-2 -----KLNMKPDEVLSVE-SEHCSAVKEFRQVLADNPGVV----KRS-
Asterias-1 -----ELRS-----QDGLMREVRQALARNGRHF----RAA-
Patiria-1 -----APPANDGIEVLTGGKIE----EDDFEGDDFDVKNKNQYYS----RSG-
Acanthaster-1 -----HPDADAKH-GGNL-----A-

Ophioderma-5 NSLTRDADDAPVAMDDKPTSPPRIDIIKRSRHLRCAQFKDMALE-DHFRR-LVGKER-
Asterias-3 -----ELN-----VASCQQLADQVKS-EGKKR----WDG-
Pisaster-1 -----TELD-----EASCQQLADQVKS-EGKKR----WTG-
Acanthaster-3 -----DKTPS-----TASCALAAQIST-EGKKR----WDG-
Patiria-7 -----DKAPS-----AATCEALAAATVKA-EGKKR----WDG-

Anneissia SM--RQCHENIRE--MVHVSERNDPKISS-----KR-SIFIPRNEATGFSLR
Antedon SM--RQCHEDIRE--MVHISCHNDPRKITS-----KR-SIFIPRNEATGFSLR
Strongylocentrotus TLR-TVGPQLQR--VGELACINDPRKTIVVVKR-----SNSDRDLFLPAKLAKAFLHY
Apostichopus NVPTVDCTVDAVQ--LHIIACSNVDYKDHGSRRRRFSDKRRRSIFLNFSEANFLAK
Acanthaster RHV-NTCNEALLP--VWDVACQNDIRKITK-----RMGREFLNEWTAKNFLAG
Ophiothrix SH--QKCNEDILLP--LWKIACTYDIRKIN-----KRA-PEFVEDSEAKAFLIG

Amphiura-1 -FRT-ICDPPFTPSLYIEDFCGVTV--KR-----E-----YEATDPLGFLKM
Ophioderma-1 -YVY-ICEPQLTR--LKNVICNPASVKRS-----DTAGLEFLTEHOAKRFLMQ
Pisaster-2 -PRNYWNTALSQ--RKTCALCGCT---HH-----TLRDDDFMEEKKDATNMLE
Asterias-1 -EDHIYCGVVLLEQ--NRESVCGV--IPG-----SKRNLFVRKEAASEFLE-
Patiria-1 -PRFRVCGTTTHS--WSSFVCHPSGLI-HH-----KRDNDEFLLSAGEANTFLMS
Acanthaster-1 -YRANYCGATTYE--KVRETCCHA-----VR-----GVSNOEFLDSDKASTELEFG

Ophioderma-5 -VFAKYCSPTPQT--VMDNYCQCDVVPRSI-----DDKRAFVDKSSAKSFLNH
Asterias-3 -PSHKFCGETINE--KRYAYCTCGLVPRKR-----EELSEFLNRGKANGFLSA
Pisaster-1 -SWHTFCGETINE--KKYAYCTCGLVPRKR-----ELDSEFLNRGKANGFLSA
Acanthaster-3 -PSHTFCGEDINE--RRNAYCNCQVVPRKR-----ELDSEFLPSGKANAFLSG
Patiria-7 -PAHTFCGETLITE--KANAYCHCQVVPRKR-----EELSEFLTPVKANSFLSG

Anneissia FL---R-T-RRPSELHEDCCLDNRGCTWEEVAEIA CINNRRRMHRPGSPVGR-----
Antedon FL---R-T-RRPSELHEDCCLDNRGCTWEEVAEIA CINNRRRMHRPGSPVGR-----
Strongylocentrotus RHRKDS-R-RRRVGKDEECCAEAQGRWEEELGEYCTLH-TRAYHQSGEQP-----
Apostichopus TKRTHS-RVRRRTTTFSTECDD--KLCIWEEVGEYCWHSR-VYH-----
Acanthaster S-----K-RRKRGLNEECCHEDLGCVWEEVAEYCVMH-GREKHEDGSPVRGKPGRRR
Ophiothrix P-----R-RHKRGLNEECCHEKSGCVWEEI GEYCRMH-SRASHVDGRIDSR-----

Amphiura-1 RGMVKRLNPNWRELMSEECY--ESTSEEIKELC-----
Ophioderma-1 NKR---IWGW-GGLSEECN--EGCSVEEIDEVC-----
Pisaster-2 RGIEKR---SLSEECCH--EGCYEVEVYEV-----
Asterias-1 R--GVQ---GLAECCG--EGCTIEEISES-----
Patiria-1 EGVGKR---GMHEECC--EGCWEEMLLEHC-----
Acanthaster-1 RGIEKR---GLHHECC--EGCSIEEIIYES-----

Ophioderma-5 RAS-TR-----SLDEECN--EGCNLEEIIVELSKTMCSSS-----
Asterias-3 RNVQKR---SLSEECCH--EGCYWEEIEEV-----
Pisaster-1 RNLQKR---SLSEECCH--EGCYWEEIEEV-----
Acanthaster-3 RRITKR---SLSEECCH--EGCYWEEIEEV-----
Patiria-7 RNIAKR---SLTEECCH--EGCYWEEIEEV-----

```

Figure 6

Position of introns in ambulacrarian insulin-like genes.

Schematic representation of the location of the cysteine residues, indicated as purple rectangles, and introns, represented by green T's, in the coding sequences of the various types of ambulacrarian insulin-like genes. Numbers indicate the phase of each intron. All genes share the typical phase 1 intron present in insulin-like genes, whereas dilp7 and multinsulin genes also share a phase 2 intron.

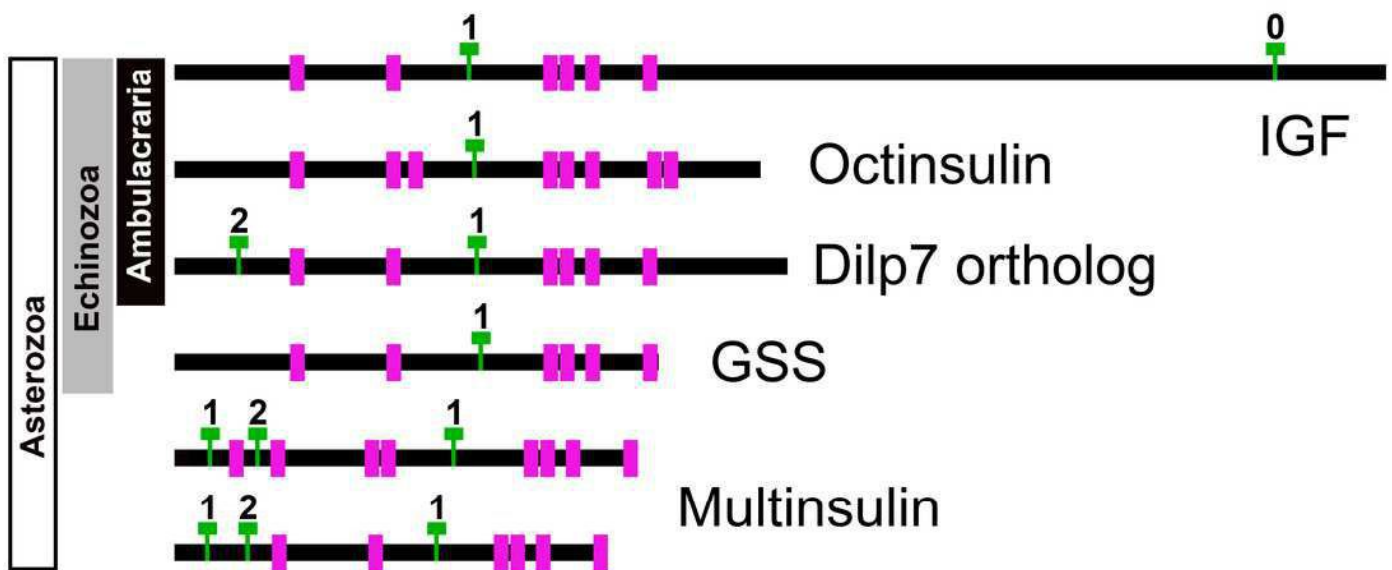


Figure 7

Synteny of ambulacrarian insulin-like genes.

Schematic representation of the relative localization of different IGF and IGF-related genes in several arthropod and ambulacrarian genomes. Arrow heads indicate transcription direction of the various genes, the numbers below the line indicate the number of nucleotides between the coding regions of adjacent genes in kilo base pairs. Note that the relative organization of the two insects - the cockroach *Blattella germanica* and the stick insect *Timema cristinae* - is the same as in the hemichordate *Saccoglossus kowalewskii* and remarkably similar to that of the sea urchin *Strongylocentrotus purpuratus* and the sea cucumber *Holothuria scabra*. In the spider *Pardosa pseudoannulata* and the sea cucumber *Apostichopus japonicus* some of the genes are also next to one another. However, in the sea stars *Acanthaster planci* and *Pisaster ochraceus* synteny has been lost. Arthropod data from Veenstra, 2020b.

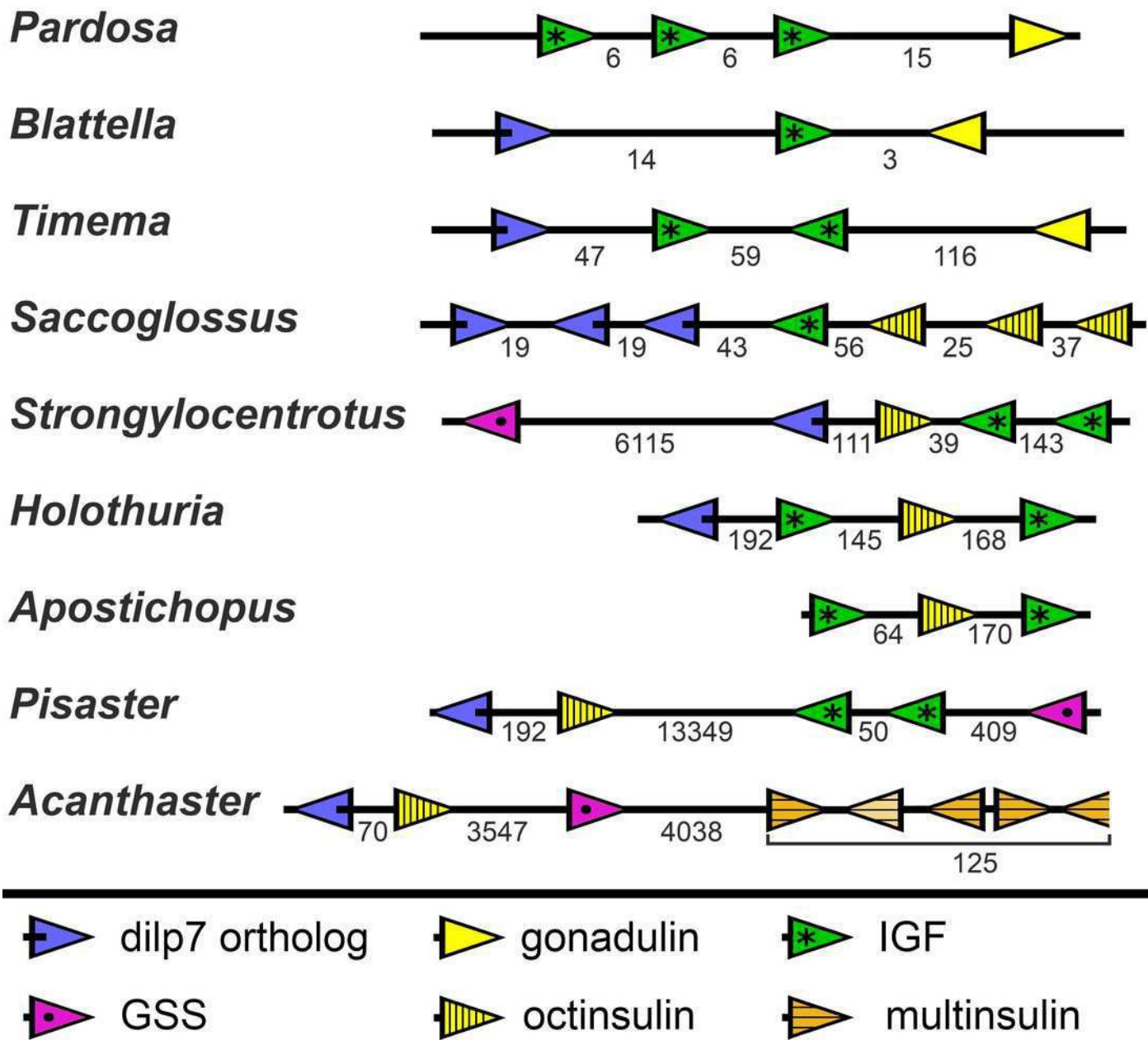


Figure 8

Radial sequence similarity tree of ambulacrarian insulin-like peptides.

The five different types of insulin-related peptides are clearly separated from one another. Note that the GSSs are similar to IGFs and seem to be related to them, while the multinsulins are most similar to the dilp7 orthologs. Echinoderm branches are in black, hemichordate branches in red. More extensive sequence comparisons and sequence trees are the supplementary data (Figs. S1-S10). All sequences are from Spreadsheet S1.

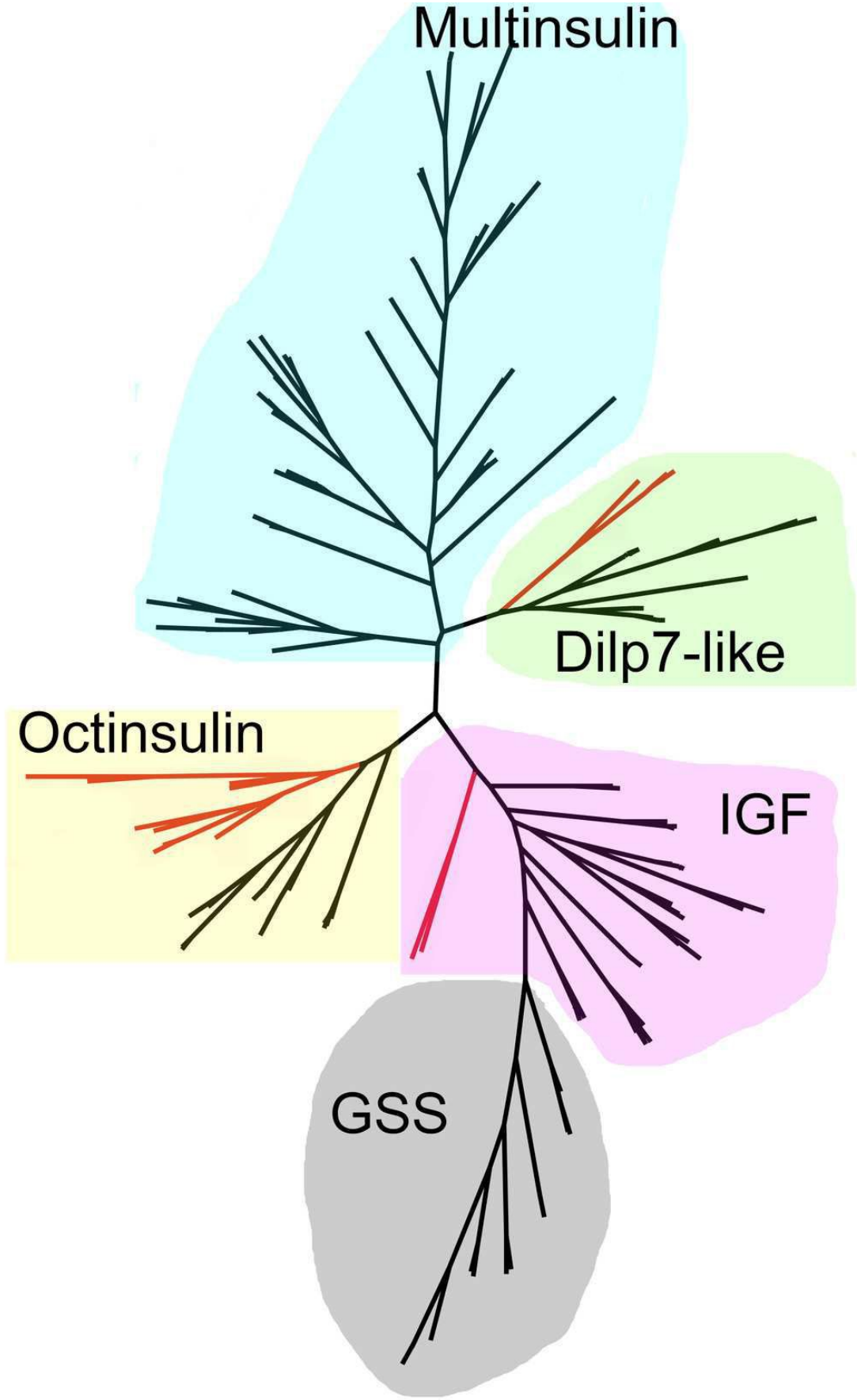


Figure 9

Phylogenetic tree of LGRs.

Phylogenetic tree constructed from the transmembrane regions of ambulacrarian LGRs that are putative receptors for insulin-related peptides. A few human and insect sequences have been added for comparison. The insert at the top shows the same data to which the glycoprotein LGRs have been added and where characteristic ligands for each branch have been identified. Numbers in blue indicate the apparent probabilities as determined by Fasttree.

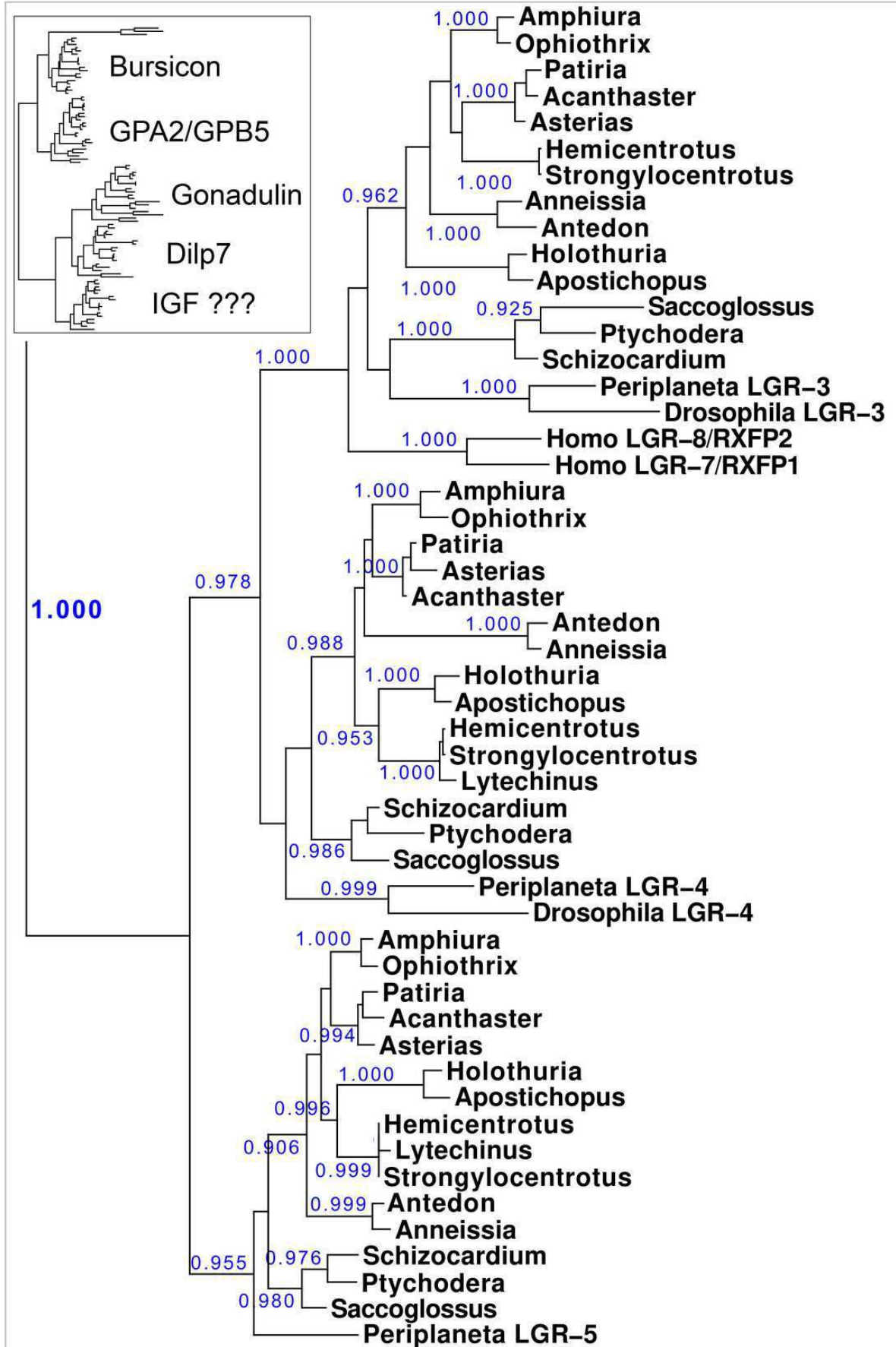


Figure 10

Ectodomains of ambulacrarian LGRs.

Ectodomains of ambulacrarian LGRs. Schematic representation of the various domains of the putative receptors for ambulacrarian insulin-related peptides. Each green circle symbolizes an LDLa repeat and each purple rectangle an LRR repeat, while the yellow oval indicates the seven transmembrane regions. The top representation corresponds to the gonadulin and dilp7 receptors (Figs. S11,S12). Note though, that the latter are somewhat variable, notably in the sea stars of two species of the *Patiria* genus and *Acanthaster planci* those receptors have two LDLa repeats (for details see Fig. S12). The bottom representation corresponds to the GRL101 receptors (Fig. S13).

