bioRxiv preprint doi: https://doi.org/10.1101/2020.03.19.998484; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Gonadulins, the fourth type of Insulin-related peptides in Decapods

Jan A. Veenstra[1],

[1] INCIA UMR 5287 CNRS, University of Bordeaux, Pessac, France.

Corresponding Author:

Jan A. Veenstra[1]

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.19.998484; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

Insulin and related peptides play important roles in the regulation of growth and reproduction. Until recently three different types of insulin-related peptides had been identified from decapod crustaceans. The identification of two novel insulin-related peptides from *Sagmariasus verreauxi* and *Cherax quadricarinatus* by Chandler and colleagues (doi: 10.3390/ijms18091832) suggested that there might a fourth type. Publicly available short read archives show that orthologs of these peptides are commonly present in these animals. Most decapods have two genes coding such peptides, but *Penaeus* species have likely only one and some palaemonids have three. The peptides they encode have been called gonadulins, as they seem consistently expressed by the gonads. Interestingly, expression levels can vary thousand-fold within ovary and testis of *Portunus trituberculatus*, where gonadulin 1 is expressed by the testis and gonadulin 2 by the ovary.

Key words: insulin; relaxin; gonadulin; androgenic insulin-like peptide; decapod; receptor; evolution

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.19.998484; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

Neuropeptides and neurohormones originated early during evolution and are well conserved between crustaceans and insects (*e.g.* Veenstra, 2016). In several cases their functions are also the same or at least similar, this is *e.g.* the case of crustacean cardioactive peptide (CCAP) that plays crucial roles during molting in both decapods and insects (Webster et al., 2013). Other peptides on the other hand likely have very different functions. EFLamide for example is very abundantly expressed and effects many different tissues in decapods (Dickinson et al., 2019), but is almost lacking in insects (Veenstra and Šimo, 2020), while allatotropin is commonly present in insects but absent from decapods (Veenstra, 2016).

Insulin and related peptides seem to be present in all animals and they have been extensively studied in insects where they are important growth hormones. They been particularly well investigated in *Drosophila* (see *e.g*. Brogiolo et al., 2001; Nässel andVanden Broeck, 2016), but also in other species such as the silk worm *Bombyx mori* (Mizoguchi and Okamoto, 2013). The best known crustacean insulin-related peptide is the one made by the androgenic gland. This gland is commonly present in malacostraca and is able to change the sex of animals from female to male as first demonstrated in a amphipod (Charniaux-Cotton, 1954). After the initial identification of its active principle from an isopod as an insulin-related peptide (Martin et al., 1999) it has also been identified from several decapods (*e.g.* Ventura et al., 2009, 2011; Chung et al., 2011). However the physiological relevance of the other insulin-related peptides in crustaceans remains unclear.

Both growth and sex determination are essential for the survival of a species and hence one would expect that the way these processes are regulated is under significant evolutionary pressure and might be conserved between insects and decapods. The insect insulin-like growth hormones and the decapod androgenic insulin-like peptides are evolutionary cousins and I previously suggested that perhaps the androgenic peptides might be orthologs of dilp 8 one of the *Drosophila* insulin-related peptides. Unlike most insulin-related peptides that stimulate a receptor tyrosine kinase (RTK) dilp 8 acts through a leucine-rich repeat G-protein coupled receptor (LGR; Vallejo et al., 2015; Gontijo and Garelli, 2018). Dilp8 is made by the imaginal disks, that are only present in flies, as well as the ovary and testes. The specific LGR (*Drosophila* LGR3) that is the dilp 8 receptor has orthologs in many other arthropods, including decapods (Veenstra, 2016), meaning that dilp 8 orthologs must be common.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.19.998484; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

However, the primary amino acid sequences of dilp 8 and the decapod androgenic insulin-like peptides are very different. The sequence of dilp 8 is also very poorly conserved within insects. When only three different types of insulin-like peptides had been identified in decapods it seemed an attractive hypothesis that dilp 8 might be an ortholog of the androgenic peptides. After all, the other two are clear orthologs of dilp 7, a relaxin-like peptide, and the insect insulin-relaed peptides that stimulate growth. However, soon afterwards two novel decapod insulin-like peptides were identified from *Sagmariasus verreauxi* and *Cherax quadricarinatus* that look more similar to dilp 8 than the androgenic insulin-like peptides (Chandler et al., 2017). This suggests that decapods do not have three but at least four different types of insulin-like peptides and that dilp 8 is probably not an ortholog of the androgenic peptide.

This provided stimulation to exploit the large amount of genomic and especially trancriptomic sequences available for decapods for additional clues as to the presence and distribution of insulin-related peptides in decapods. I here report that the peptides initially described from *S. verreauxi* and *C. quadricarinatus* belong to a novel type of decapod insulin-like peptides and in many species is produced by two and in some species even three paralogous genes, one or two of which seem to be rarely expressed. However, under particular conditions, that remain to be identified, these peptides are occasionally made in very large numbers by the gonads, at least in the crab *Portunus trituberculatus*. As the gonads seem to be an important site of expression of these peptides, they have been baptized gonadulins.

Materials and Methods

I analyzed transcriptome and genome SRAs using the tools from the SRAtoolkit (https://www.ncbi.nlm.nih.gov/sra/docs/toolkitsoft/), Trinity (Grabherr et al., 2011) and Artemis (Rutherford et al., 2000) on the publicly available decapod genomes and draft genomes, as well as a large number of transcriptome (Table S1) and genome SRAs. Details of the methods that were used have been described in more detail elsewhere (Veenstra and Khammassi, 2017; Veenstra, 2019). A list of all SRAs used can be found in the supplementary data. Draft genome assembliess for the following species have been used : *Penaeus vannamei* (Zhang et al., 2019), *P. monodon* (Van Quyen et al., 2020), *Palaemon carinicauda* (Yuan et al., 2017), *Procambarus virginalis* (Gutekunst et al., 2018), *Cherax*

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.19.998484; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

quadricarinatus (unpublished, but available at https://www.ncbi.nlm.nih.gov/assembly/GCA_009761615.1/), and two assemblies each for *Eriocheir sinensis* (Song et al., 2016; Tang et al., 2020b) and *Portunus trituberculataus* (https://www.ncbi.nlm.nih.gov/assembly/GCA_008373055.1/; Tang et al., 2020a).

Decapod genomes tend to be large. A number of draft decapod genomes used illumina technology for the assembly and these have usually relatively short contigs and no or poor scaffolds. For example, the *Cherax quadricarinatus* genome lacks the second coding exon from the gonadulin 1 gene. The sequence of this exon was obtained by using Trinity on individual reads from genomic SRAs. For the genome assembly of *Penaeus vannamei* and the second assemblies for *Portunus trituberculatus* and *Eriocheir sinensis* PacBio sequencing technology was used. This allows the construction of large scaffolds, but it often suffers from significant numbers of indels, which in some genome stretches are numerous and can make analysis difficult. In some cases certain genes did not make it into the final assembly, *e.g.* the *Portunus* RTK4 gene. Nevertheless these genome assemblies are useful for elucidating the exon-intron structures of genes. Other genomic sequences were recovered from several genomic resources created to establish the phylogenetic relations within the decapods (Wolfe et al., 2019). This concerns data from *Coenobita clypeatus, Stenopus hispidus, Fantapenaeus duorarum, Ocypode quadrata, Menippe nodifrons,* and *Procamabarus clarkii*. These genomic SRAs were used in combination with Trinity to deduce the exon-intron structures of several genes. In some cases, such as that of the *Stenopus hispidus* gonadulin gene, this was done after determining its coding sequence from transcriptome data, while in other species the transcripts from related species were used. For example, the partial *Coenobita clypeatus* genes structures were found in the genomic SRA using the gonadulin precursor deduced from transcriptome SRAs of *Paralithodes camtschaticus* (a species from the same infraorder), while the *Palaemon serratus* and *Macrobracium* sequences allowed the identification of the *P. carinicauda* genes. Neither of the two *Eriocheir* genome assemblies contain the gonadulin 1 gene, so the exon-intron structure of this gene was also determined from genomic SRAs.

Phylogenetic trees

The phylogenetic tree for the decapod RTKs (Fig. 5) was made with fasttree (Price et al., 2010) using the conserved parts of all sequences that were available (spreadsheet1 of supplementary data). A second tree was made using exclusively the tyrosine kinase domains (Fig. S1). The

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.19.998484; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

phylogenetic tree of the LGRs (Fig. 6) was made exclusively from the transmembrane regions of these receptors using fasttree.

Expression

Expression data were obtained by using the parts of the transcripts that code for the various peptides and receptors (thus without untranslated 5'- and 3'-sequences) as query for the blastn_vdb command from the SRAtoolkit to look for individual reads in the various transcriptome SRAs. Raw counts and results are expressed as reads per million and reported in spreadsheet2 of supplementary data. In the case of the SRAs from *Portunus,* data for a partial vitellogenin receptor sequence was also included in order to show the likely mix up of the ovary and testis SRAs SRR1920180 and SRR1920182.

Results

This manuscript describes the gonadulins and their putative receptors, however other decapod insulin-like peptides were also identified in an attempt to establish where all the different insulin-related peptides and their putative receptors are expressed. Although gonadulins were identified in many decapod species, identification of their putative receptors and the expression of these genes was limited to eight species. These are *Penaeus vannamei, P. monodon, Macrobrachium nipponense, Cherax quadricarinatus, Procambarus clarkii, Portunus trituberculatus, Carcinus maenas* and *Eriocheir sinensis* (spreadsheet1, spreadsheet2).

Primary amino acid sequences of arthropod insulin-related peptides are notoriously variable, the best conserved sequence features are the the six cysteine residues responsible for the three disulfide bridges. Even that character is not constant, as some peptides have eight cysteine residues and are thus predicted to have four disulfide bridges. This includes some of the decapod gonadulins (Figs 1, 2) as well as some androgenic peptides. It is worth mentioning that, surprisingly, some of the decapod relaxins are predicted to have seven cysteine residues in the propeptide (Table S2). The high primary amino acid sequence variability precludes making reliable phylogenetic trees using these sequences. The primary amino acid residues immediately surrounding the cysteines are often somewhat better conserved, at least in the decapod insulins and relaxins, but hardly or not at all in the androgenic

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.19.998484; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

peptides and gonadulins (Fig. 1). It is thus impossible to predict how the different peptides are related based on their primary amino acid sequences, even though the B-chain sequence logos may suggest that the androgenic peptide is more similar to relaxin than to the other two peptides (Fig. 1). It thus might have its origin in an ancestral gene duplication of the relaxin gene. Deduced amino acid sequences for a number of decapod insulin-related peptides are shown in figures 1 and 2 and more sequences are presented in the supplementary data (Tables S2-S6).

Gonadulins

Interestingly, many decapods may have two different types of gonadulins and the Palaemonids have even three. On the other hand only a single gonadulin was found in three Penaeid species or the single species from the Stenopodidae for which data is available. In the first case there is a large amount of data, including genomes for *Penaeus vannamei* and *P. monodon*, as well as a genomic SRA for *Fantapenaeus duorarum* that covers its genome 27-fold. In none of these data sets evidence for a second gonadulin gene could be found, suggesting that the Penaeids have only one such gene. For *Stenopus hispidus* there is a genomic SRA with 14-fold coverage, suggesting that it too likely has only a single gonadulin gene.

The different gonadulin genes can be distinguished in three different types, with the last type only found in Palaemonidae; they will be referred to as gonadulin 1, gonadulin 2 and gonadulin 3 (Figs 2,3; Tables S3,S4,S5). Of these gonadulin 1 is most widely expressed as detailed below and has also the best conserved primary amino acid sequence. In most species gonadulin 1 has eight cysteine residues rather than the characteristic six typical of insulin (Fig. 2), however gonadulin 2 and 3, which are quite similar to one another, have only six cysteine residues and this is also the case for gonadulin 1 from brachyura. The sequences corresponding to the putative connecting peptides of gonadulin are often poorly conserved, apart from the piece surrounding the two additional cysteine residues in gonadulin 1. The parts of the gonadulin precursors corresponding to the A- and B-chains are much better conserved, but are still quite variable (Figs 2, 3). One structural character that seems to be typical of the gonadulins is the presence of one or more dibasic amino acid residues preceding the first cysteine residue in the A-chain while acidic amino acid residues are absent there. In particular the third residue preceding the first cysteine is always an arginine (Figs 1-3). In contrast, insulin (Table S6),

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.19.998484; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

relaxin and androgenin always have an aspartic residue immediately preceding this cysteine (Fig. 1) and usually lack that arginine residue.

It is not clear at all how these gonadulin precursors are processed. If they are expressed by the ovary or testes, they are likely to be secreted through the constituitive pathway that has furin but is expected to lack the typical endocrine convertases. On the other hand gonadulin 1 is expressed in the eyestalk, where such enzymes are present. Furin, is likely involved in the processing of these precursors, but its precise cleavage preferences are unknown even in vertebrates, although we do know that they differ between insects and vertebrates (Cano-Monreal et al., 2010). This makes predictions as where convertase(s) may cleave the gonadulin precursors impossible.

Gene structure

The availability of several (draft) genome assemblies as well as a number of genome SRAs makes it possible to identify the individual exons of the various gonadulin genes. The gonadulin 1 genes for which a gene structure could be deduced have four, five or six coding exons, while the gonadulin 2 genes have consistently two. The single *Penaeus* gene has three coding exons and represents perhaps an ancestral gene that was duplicated. The single *Stenopus* gene has four coding exons and encodes a gonadulin 1. The difference in number of coding exons is not the only difference between the gonadulin 1 and 2 genes. For those species where this can be determined gonadulin 1 genes stretch out over thousands of base pairs and the intron between the two first coding exons accounts for several thousand nucleotides. Although this latter point is most convincingly established in *Portunus*, due to the use of PacBio sequencing, the genome drafts for *Procambarus virginalis* and *Cherax* also show significant sizes for this intron. The gonadulin 2 gene on the other hand has a small intron between its two coding exons; so small in fact, that Trinity sometimes produces a single contig from genomic DNAseq reads containing the two coding exons. The third gonadulin gene, which has only been detected in Palaemonids, has two coding exons, like the second gene and the predicted peptides are also similar. The gonadulin precursor sequences of the Palaemonid species are sufficiently similar to allow reconstruction of the various coding exons of the three *P. carinicauda* genes (Fig. 4).

Trinity produce neither a complete transcript for the *Eriocheir* nor the *Carcinus* gonadulin 2 transcript, both these sequences, like their *Portunus* ortholog, contain repetitive stretches, making assembly of a transcript difficult. It was also imposssible to find the *Eriocheir* gonadulin 2

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.19.998484; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

gene in either of the two genome assemblies. Genomic SRAs of this species show two very similar, but significantly different genomic sequences for what is likely the first exon of this gene in addition to the repetive sequence. It can therefore not be excluded that this gene is evolving into or has already become a pseudogene. Alternatively, it can neither be excluded that in this species the second gonadulin gene has been duplicated. It is worth noting that all gonadulin gene introns are phase 1, which facilitates the exon shuffling that appears to have occurred with these genes (Fig. 4).

Receptors

Neuropeptides and hormones need receptors to exert their physiological activity. Insulin-related peptides are somewhat unusual as they act on two different types of receptors, receptor tyrosine kinases (RTKs) and leucine-rich repeat G-protein coupled receptors (LGRs). It has previously been reported that there are four different types of decapod RTKs (Herran et al., 2018) and this was confirmed here for a number of species. Although virtually complete RTK sequences could either be recovered from the databases or using Trinity on a collection of transcriptome SRAs, others were only partially identified. These sequences were combined with RTKs previously identified from *Macrobrachium rosenbergii, Sagmariasus verrauxi* and *Fenneropenaeus sinensis* (Sharabi et al., 2016; Aizen et al., 2016; Guo et al., 2018) and four RTKs from the cockroach *Periplaneta americana,* identified here, for the construction of a phylogenetic tree. The results confirms the existence of four different decapod RTKs and shows that decapods and insects independently evolved four paralogs (Fig. 5; Fig. S1).

Two arthropod LGRs have either been shown or suggested to be receptors for insulin-related peptides. LGR3 is the receptor for dilp 8 (Vallejo et al., 2015; Gontijo and Garelli, 2018), while LGR4 has been suggested to be the receptor for relaxins, the dilp 7 orthologs (Veenstra, 2014). Many arthropods, but not holometabolous insect species, have another LGR, orthologs of GRL101 from the pond snail *Lymnaea stagnalis* (Tensen et al., 1994). These LGRs are closely related to LGR3 and LGR4 and generally slightly more similar to LGR4 than to LGR3, suggesting that like those receptors might also have an insulin-like ligand. For this reason these receptors were also included and called here LGR5s. Decapods have typically one LGR4, and two LGR3s. The number of LGR5s are not clear, but there often are at least three such receptors. A phylogenetic tree of the transmembrane regions of

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.19.998484; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

these receptors and their orthologs from *Periplaneta* and *Drosophila* illustrates their occurrence in a number of decapod species (Fig 6) and their phylogenetic relations.

Expression

An attempt has been made to look at expression of the various insulin-related peptides and their putative receptors using the large amount of transcriptomic SRAs available. This approach is obviously limited by a number of factors. Notably, these experiments were done by a large variety of laboratories using different species, different tissues and different physiological conditions. There is a lot of data (spreadsheet2) of which here only some of the more salient results are summarized.

Although the number of samples for the two *Penaeus* species is limited, gonadulin expression in the eyestalk, ovary and the testis of of this species appears to be weak. In other decapods gonadulin 1 RNAseq reads are most abundant in he eyestalk as well as the testis, but not in the ovary. *Macrobrachium* appears to be an exception as in this species gonadulin 1 is also expressed in the ovary. Expression of gonadulin 1 is more widespread than that of gonadulin 2, which seems often more limited to the ovary. Nevertheless, in *Procambarus*, the only species where such tissues have been analyzed, it is also expressed in hematopoetic tissue and the anterior proliferation center of the brain; a very low level of expression is also observed in the brain of *Portunus*. The expression of gonadulin 3, a peptide that has only been found in *Macrobrachium,* was only observed in the ovary. Most intriguing is the very variable level of expression of the gonadulin genes in the gonads. This is prominently illustrated by the number of gonadulin RNAseq reads in a number of *Portunus* SRA experiments (Table 1). This highlights a general problem in using SRAs; there there are very few studies that compare different physiological conditions in the same tissue and hence it is very difficult to know why some SRA experiments yield large read numbers for some peptides and others few or none. That the physiological state is important is nicely illustrated by the expression of relaxin in the ovary of *Penaeus monodon*, where the expression of relaxin seems to be correlated with that of vitellogenin (Table S7).

The RNAseq data suggest expression of androgenin not only in the androgenic gland, but also in the hepatopancreas and/or eyestalk in various species analyzed here. Similar findings have been reported previously (*e.g.* Chung, 2014; Huang et al., 2014; Levy et al., 2017; Shi et al., 2019). Expression levels of both insulin and relaxin are usually low, with insulin mostly expressed in the

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.19.998484; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

eyestalk (and nervous system) and in *Procambarus* and *Portunu*s also in the hepatopancreas. However, in *Penaeus vannamei* insulin is expressed in both the testis and the ovary, while in *P. monodon* the ovary expresses relaxin but the the transcriptome SRAs provide no evidence for the expression of insulin in either the ovary or the testis. There are generally very few relaxin RNAseq reads with most of it in the eyestalk and the nervous system, while in *Penaeus modon,* but not *P. vannamei,* there are large numbers in some ovary samples. One hepatopancreas SRA (SRR10846612) out of six from the same *M. nipponense* experiment has a very large number of relaxin reads, suggesting that this particular sample may have been contaminated with another tissue.

It has been reported that in the penaeid *Fenneropenaeus* the RTK4 ortholog is exclusively expressed in the testis (Guo et al., 2018). In the SRAs analyzed here reads for RTK4 orthologs are found mostly, but not exclusively in the testes and androgenic glands. In *Penaeus monodon* RTK3 is also almost exclusively expressed in the testis, unfortunately there is insufficient data for *P. vanamei*, but in other species, RTK3 seems to be more widely expressed. Interestingly, RTK2 orthologs are especially abundant in ovary SRAs, but it is also expressed in many other tissues, including the androgenic gland where at least in some species this RTK seems as highly expressed as in the ovary (Suppl. Data, spreadsheet2).

Within a single species the various LGR are often expressed more strongly in some tissues, but a pattern that was reproduced between the different species was not found (spreadsheet2).

Discussion

Three different types of insulin-related peptides have previously been shown to be commonly present in decapod crustaceans, the relatively well known androgenic insulin-like peptide, relaxin, and a third peptide, insulin, that shares characteristics with its mammalian homonyme. The similarity of these peptides with their putative orthologs from insects suggested that the androgenic peptide might be an ortholog of dilp 8 (Veenstra, 2016). However, this hypothesis was challenged by the description of yet another type of decapod insulin-like peptide by Chandler et al. (2017). I show here that this novel type of insect-related peptides is ubiquitously present in these crustaceans and propose to call them gonadulins, as they are commonly expressed by the gonads, sometimes in very large quantities.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.19.998484; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Although it remains possible that decapods still have other insulin-like peptides, it is interesting to note that no C-terminally extended insulin was found. Such C-terminally extended insulins are present in several insect species and have sometimes been called insulin-like growth factors as they share characteristics with their homonymes (*e.g.* Defferrari, et al., 2016; Okada et al., 2019).

In several species only incomplete transcripts were obtained for one of the gonadulin genes, but it is nevertheless likely that most decapod species have two such genes. However, several arguments can be advanced to suggest that the penaeids and perhaps all dendrobranchiata, have only one gonadulin gene. Thus evidence for a second gene was not found in the large number of transcriptome SRAs for these species, nor could such a gene be found in the genome assemblies of two *Penaeus* species or the genome SRA for *Fantapenaeus duorarum*. Furthermore, when one compares the gene structures of the various decapod gonadulin genes (Fig. 4), it appears that the penaeid genes differ from the typical decapod gonadulin ones, suggesting that a gene duplication may have occurred after the dendrobranchiata diverged from the other decapods. In a genomic SRA for *Stenopus hispidus* a second gonadulin gene could neither be found, although this is not definitive proof that this species lacks a second gene, given the 13-fold coverage of its genome, it is a distinct possibility. Finally, at least in some Palaemonidae there are three gonadulin genes, as shown by the transcriptome data from two *Macrobrachium* species and the *Palaemon carinicauda* draft genome.

There are thus (at least) four different types of insulin-like peptides in decapods, *i.e.* relaxin, androgenin, gonadulin and insulin. There are also four different types of insulin RTKs, but it is unknown whether each of the four insulin-like peptides has its own RTK or whether one or more insulin-related peptides acts on more than one RTK. Of course, it also remains possible that one or more of these RTKs use neither of these peptides, but yet another ligand. Within this context it is interesting that orthologs of what has been labeled RTK4 here have been shown to interact with androgenin in two different species (Aizen et al., 2016; Guo et al., 2018). In *Fenneropenaeus chinensis* RTK4 is exclusively expressed in males and the expression data analyzed here for the closely related *P. vannamei* show that this receptor is indeed mainly expressed in testes. However, in other species RTK4 orthologs are more widely expressed. Although RTK2 seems to have its highest expression in the ovary, its tissue specificity is clearly not limited to females, as it is also abundantly expressed in other tissues, including the androgenic gland.

There are also several different LGRs that might be activated by these insulin-like peptides. These can be differentiated in three different main types, LGR3, LGR4 and LGR5, and even

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.19.998484; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

more subtypes. The *Drosophila* LGR3 ortholog has been shown to be activated by dilp 8 and one would expect that the decapod LGR3s would have a similar ligand. Although I previously suggested that this might be androgenin, it now seems much more likely that this would be one or more of the gonadulins, that have a primary amino acid sequence that is more similar to that of *Drosophila* ilp 8 (Chandler et al., 2017).

A priori one or more of the RTKs and LGR3 would seem to be the most likely receptors for gonadulins, given its identity as an insulin-like peptide (RTK) and its structure that is somewhat similar to dilp 8 (LGR). One of the intriguing characteristics of insulin-related peptides is that they act both through LGRs and RTKs, but to my knowledge there has so far not been an example where a single peptide acts on receptors of both types.

Decapods have only one LGR4, this receptor is present in the same genomes, from both deuterostomes and protostomes, that also have an ortholog of the relaxin gene, and it is absent from genomes that lack a relaxin gene. This strongly suggests that relaxin may be its ligand (*e.g.* Veenstra, 2014, 2019). Complete relaxin transcripts are so far lacking from several decapods although individual reads containing partial relaxin sequences have been found in several cases (Veenstra, 2016). This suggests that the absence of complete transcripts does not reflect an absence of the gene, but rather the lack of transcript analysis from the tissue where the gene is expressed. If the major site of relaxin expression in decapods would be similar to that in *Drosophila* (Yang et al., 2008; Miguel-Aliaga et al., 2008), one would expect relaxin producing neurons in the abdominal ganglia. As this tissue is rarely, if ever, used for RNAseq analysis in decapods, the scarcity of complete relaxin transcripts is perhaps not surprising. Although relaxin reads are occasionally found in hepatopancreas SRAs, these could be due to the accidental inclusion of (a part of) an abdominal ganglion. The abundant amount of relaxin reads in a single SRA *Macrobrachium nipponense* hepatopancreas might represent such an accident.

The large amount of transcriptome SRAs identify tissues where gonadulins are expressed, it also reveals that expression of the decapod insulin-related peptides is rather variable. Thus, whereas the the ovary of *Penaeus monodon* reveals significant relaxin expression, in *P. vannamei* it makes insulin. It is presently impossible to know whether these apparent differences are authentic, or whether they reflect different physiological states of the animals when they were sacrificied. The extremely variable expression of gonadulins by the gonads of *Portunus trituberculatus,* where expression levels vary more than thousand-fold, is also very intriguing and reinforces the conclusion that it is impossible to draw definitive conclusions as to the expression of these peptides.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.19.998484; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The expression of one of the gonadulins in hematopoetic tissue and the anterior proliferation center of the brain in *Procambarus clarkii* suggest that it, like other arthropod insulin-related peptides, could be a kind of growth hormone, but experiments will be needed to elucidate their function.

Acknowledgements. As always for this type of manuscript I am analyzing data I did not collect myself and using methods developed by others. Without either these tools or data, this work would not be feasible and I express my sincere gratitude to all of those who made this possible. I apologize to all those whose SRA data I used without citing the papers that described them; there are simply too many. Institutional funding from the CNRS is also acknowledged.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.19.998484; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

Aizen, J., Chandler J.C., Fitzgibbon, Q.P., Sagi, A., Battaglene, S.C., Elizur, A., Ventura, T., 2016. Production of recombinant insulin-like androgenic gland hormones from three decapod species: *In vitro* testicular phosphorylation and activation of a newly identified tyrosine kinase receptor from the Eastern spiny lobster, *Sagmariasus verreauxi.* Gen. Comp. Endocrinol. 229, 8-18. doi: 10.1016/j.ygcen.2016.02.013.

Brogiolo, W., Stocker, H., Ikeya, T., Rintelen, F., Fernandez, R., Hafen, E., 2001. An evolutionarily conserved function of the *Drosophila* insulin receptor and insulin-like peptides in growth control. Curr. Biol. 11,213–221. doi: 10.1016/s0960-9822(01)00068-9.

Cano-Monreal, G.L., Williams, J.C., Heidner, H.W., 2010. An arthropod enzyme, Dfurin1, and a vertebrate furin homolog display distinct cleavage site sequence preferences for a shared viral proprotein substrate. J. Insect Sci. 10:29. doi: 10.1673/031.010.2901.

Chandler, J.C., Gandhi, N.S., Mancera, R.L., Smith, G., Elizur, A., Ventura, T., 2017. Understanding insulin endocrinology in decapod crustacea: Molecular modelling characterization of an insulin-binding protein and insulin-like peptides in the eastern spiny lobster, *Sagmariasus verreauxi*. Int. J. Mol. Sci. 18: E1832. doi: 10.3390/ijms18091832.

Charniaux-Cotton, H., 1954. Découverte chez un crustacé amphipode (*Orchestia gammarella*) d'une glande endocrine responsible de la différenciation de charactères sexuels primaires et secondaires mâles. Compt. rend. 239, 780-783.

Chung, J.S., 2014., An insulin-like growth factor found in hepatopancreas implicates carbohydrate metabolism of the blue crab *Callinectes sapidus*. Gen. Comp. Endocrinol. 199, 56–64.

Chung, J.S., Manor, R., Sagi, A., 2011. Cloning of an insulin-like androgenic gland factor. (IAG) from the blue crab, *Callinectes sapidus*: Implications for eyestalk regulation of IAG expression. Gen. Comp. Endocrinol. 173, 4-10.

Defferrari, M.S., Orchard, I., Lange, A.B., 2016. An Insulin-Like Growth Factor in *Rhodnius prolixus* is involved in post-feeding nutrient balance and growth. Front. Neurosci. 201610:566. doi: 10.3389/fnins.2016.00566.

Dickinson, P.S., Dickinson, E.S., Oleisky, E.R., Rivera, C.D., Stanhope, M.E., Stemmler, E.A., Hull, J.J., Christie, A.E., 2019. AMGSEFLamide, a member of a broadly conserved

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.19.998484; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

peptide family, modulates multiple neural networks in *Homarus americanus.* J. Exp. Biol. 222 (Pt 2), jeb194092. pii. doi:10.1242/jeb.194092.

Gontijo, A.M., Garelli, A., 2018. The biology and evolution of the Dilp8-Lgr3 pathway: A relaxin-like pathway coupling tissue growth and developmental timing control. Mech. Dev. 154, 44-50. doi: 10.1016/j.mod.2018.04.005.

Grabherr, M.G., Haas, B.J,, Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644-52. doi: 10.1038/nbt.1883.

Guo, Q., Li, S., Lv, X., Xiang, J., Sagi, A., Manor, R., Li, F., 2018. A putative insulin-like androgenic gland hormone receptor gene specifically expressed in male chinese shrimp. Endocrinology. 159, 2173-2185. doi: 10.1210/en.2017-03253.

Gutekunst, J., Andriantsoa, R., Falckenhayn, C., Hanna, K., Stein, W., Rasamy, J., Lyko, F., 2018. Clonal genome evolution and rapid invasive spread of the marbled crayfish. Nat. Ecol. Evol. 2:567-573. doi: 10.1038/s41559-018-0467-9.

Herran, B., Bertaux, J., Grève, P., 2018. Divergent evolution and clade-specific duplications of the insulin-like receptor in malacostracan crustaceans. Gen. Comp. Endocrinol. 268, 34-39. doi: 10.1016/j.ygcen.2018.07.013.

Huang, X., Ye, H., Huang, H., Yang, Y., Gong, J., 2014. An insulin-like androgenic gland hormone gene in the mud crab, *Scylla paramamosain*, extensively expressed and involved in the processes of growth and female reproduction. Gen. Comp. Endocrinol. 204, 229–238.

Levy, T., Rosen, O., Simons, O., Savaya Alkalay, A., Sagi, A., 2017. The gene encoding the insulin-like androgenic gland hormone in an all-female parthenogenetic crayfish. PLoS One 12:e0189982. doi: 10.1371/journal.pone.0189982.

Martin, G., Sorokine, O., Moniatte, M., Bulet, P., Hetru, C., Van Dorsselaer ,A., 1999. The structure of a glycosylated protein hormone responsible for sex determination in the isopod, Armadillidium vulgare. Eur. J. Biochem. 262,727–736.

Miguel-Aliaga, I., Thor, S., Gould, A.P., 2008. Postmitotic specification of *Drosophila* insulinergic neurons from pioneer neurons. PLoS Biol 6: e58. doi:10.1371/journal.pbio.0060058.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.19.998484; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Mizoguchi, A., Okamoto, N., 2013. Insulin-like and IGF-like peptides in the silkmoth *Bombyx mori*: discovery, structure, secretion, and function. Front. Physiol. 4:217. doi: 10.3389/fphys.2013.00217.

Nässel, D.R., Vanden Broeck, J., 2016. Insulin/IGF signaling in *Drosophila* and other insects: factors that regulate production, release and post-release action of the insulin-like peptides. Cell. Mol. Life Sci. 73, 271-290. doi: 10.1007/s00018-015-2063-3.

Okada, Y., Katsuki, M., Okamoto, N., Fujioka, H., Okada, K., 2019. A specific type of insulin-like peptide regulates the conditional growth of a beetle weapon. PLoS Biol. 17:e3000541. doi:10.1371/journal.pbio.3000541.

Price, M.N., Dehal, P.S., Arkin, A.P. (2010) FastTree 2 -- Approximately maximum-likelihood trees for large alignments. PLoS ONE 5:e9490. doi:10.1371/journal.pone.0009490.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., Barrell, B., 2000. Artemis: Sequence visualization and annotation. Bioinformatics 16, 944–945. doi: 10.1093/bioinformatics/16.10.944.

Sharabi, O., Manor, R., Weil, S., Aflalo, E.D., Lezer, Y., Levy, T., Aizen, J., Ventura, T., Mather, P.B., Khalaila, I., Sagi, A., 2016. Identification and characterization of an insulin-like receptor Involved in crustacean reproduction. Endocrinology 157, 928-941. doi: 10.1210/en.2015-1391.

Shi, L., Han, S., Fei, J., Zhang, L., Ray, J.W., Wang, W., Li, Y., 2019. Molecular characterization and functional study of insulin-like androgenic gland hormone gene in the red swamp crayfish, *Procambarus clarkii.* Genes 10: E645. doi: 10.3390/genes10090645.

Song, L., Bian, C., Luo, Y., Wang, L., You, X., Li, J., Qiu, Y., Ma, X., Zhu, Z., Ma, L., Wang, Z., Lei ,Y., Qiang, J., Li, H., Yu, J., Wong, A., Xu, J., Shi, Q., Xu, P., 2016. Draft genome of the Chinese mitten crab, *Eriocheir sinensis*. Gigascience 5:5. doi: 10.1186/s13742-016-0112-y.

Tang, B., Zhang, D., Li, H., Jiang, S., Zhang, H., Xuan, F., Ge, B., Wang, Z., Liu, Y., Sha, Z., Cheng, Y., Jiang, W., Jiang, H., Wang, Z., Wang, K., Li, C., Sun, Y., She, S., Qiu, Q., Wang, W., Li, X., Li, Y., Liu, Q., Ren, Y., 2020a. Chromosome-level genome assembly reveals the unique genome evolution of the swimming crab (*Portunus trituberculatus*). Gigascience. 9(1). pii: giz161. doi: 10.1093/gigascience/giz161.

Tang, B., Wang, Z., Liu, Q., Zhang, H., Jiang, S., Li., X, Wang, Z., Sun, Y., Sha, Z., Jiang, H., Wu, X., Ren, Y., Li, H., Xuan, F., Ge, B., Jiang, W., She, S., Sun, H., Qiu, Q., Wang, W.,

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.19.998484; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Wang, Q., Qiu, G., Zhang, D., Li, Y., 2020b. High-Quality Genome Assembly of Eriocheir japonica sinensis Reveals Its Unique Genome Evolution. Front Genet. 10:1340. doi: 10.3389/fgene.2019.01340.

Tensen, C.P., Van Kesteren, E.R., Planta, R.J., Cox, K.J., Burke, J.F., van Heerikhuizen, H., Vreugdenhil, E., 1994. A G protein-coupled receptor with low density lipoprotein-binding motifs suggests a role for lipoproteins in G-linked signal transduction. Proc. Natl. Acad. Sci. USA 91, 4816-4820. doi:10.1073/pnas.91.11.4816.

Vallejo, D.M., Juarez-Carreño, S., Bolivar, J., Morante, J., Dominguez, M., 2015. A brain circuit that synchronizes growth and maturation revealed through Dilp8 binding to Lgr3. Science 350, aac6767. doi: 10.1126/science.aac6767.

Van Quyen, D., Gan, H.M., Lee, Y.P., Nguyen, D.D., Nguyen, T.H., Tran, X.T., Nguyen, V.S., Khang, D.D., Austin, C.M., 2020. Improved genomic resources for the black tiger prawn (*Penaeus monodon*). Mar. Genomics. 2020 Feb 4:100751. doi: 10.1016/j.margen.2020.100751.

Veenstra, J.A., 2014. The contribution of the genomes of a termite and a locust to our understanding of insect neuropeptides and neurohormones. Front Physiol. 5:454. doi:10.3389/fphys.2014.00454.

Veenstra, J.A., 2016. Similarities between decapod and insect neuropeptidomes. PeerJ 4:e2043. doi: 10.7717/peerj.2043.

Veenstra, J.A., 2019. Coleoptera genome and transcriptome sequences reveal numerous differences in neuropeptide signaling between species. PeerJ7:e7144. doi: 10.7717/peerj.7144.

Veenstra, J.A., Khammassi, H., 2017. Rudimentary expression of RYamide in *Drosophila melanogaster* relative to other *Drosophila* species points to a functional decline of this neuropeptide gene. Insect Biochem. Mol. Biol. 83, 68-79. doi: 10.1016/j.ibmb.2017.03.001.

Veenstra JA, Šimo L., 2020. The TRH-ortholog EFLamide in the migratory locust. Insect Biochem. Mol. Biol. 116:103281. doi: 10.1016/j.ibmb.2019.103281.

Ventura, T,, Manor, R,, Aflalo, E,D,, Weil, S,, Raviv, S,, Glazer, L,, Sagi, A., 2009. Temporal silencing of an androgenic gland-specific insulin-like gene affecting phenotypical gender differences and spermatogenesis. Endocrinology 150, 1278-1286. doi: 10.1210/en.2008-0906.

Ventura, T., Rosen, O., Sagi, A., 2011. From the discovery of the crustacean androgenic gland to the insulin-like hormone in six decades. Gen. Comp. Endocrinol. 173, 381–388.

Webster, S.G., Wilcockson, D.C., Mrinalini, Sharp, J.H., 2013. Bursicon and neuropeptide cascades during the ecdysis program of the shore crab, *Carcinus maenas*. Gen. Comp. Endocrinol. 82, 54-64. doi: 10.1016/j.ygcen.2012.11.018.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.19.998484; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Wolfe, J.M., Breinholt, J.W., Crandall, K.A., Lemmon, A.R., Lemmon, E.M., Timm, L.E., Siddall, M.E., Bracken-Grissom, H.D., 2019. Proc. Biol. Sci. 286:20190079. doi: 10.1098/rspb.2019.0079.

Yang, C.H., Belawat, P., Hafen, E., Jan, L.Y., Jan, Y.N., (2008) *Drosophila* egg-laying site selection as a system to study simple decision-making processes. Science 319, 1679–1683. doi: 10.1126/science.1151842.

Yuan, J., Gao, Y., Zhang, X., Wei, J., Liu, C., Li, F., Xiang, J,., 2017. Genome sequences of marine shrimp *Exopalaemon carinicauda* Holthuis provide insights into genome size evolution of Caridea. Mar. Drugs 15: E213. doi: 10.3390/md15070213.

Zhang, X., Yuan, J., Sun, Y., Li, S., Gao, Y., Yu, Y., Liu, C., Wang, Q., Lv, X., Zhang, X., Ma, K.Y., Wang, X., Lin, W., Wang, L., Zhu, X., Zhang, C., Zhang, J., Jin, S., Yu, K., Kong, J., Xu, P., Che, J., Zhang, H., Sorgeloos, P., Sagi, A., Alcivar-Warren, A., Liu, Z., Wang, L., Ruan, J., Ch, K.H., Li, B., Li, F., Xiang, J., 2019. Penaeid shrimp genome provides insights into benthic adaptation and frequent molting. Nat. Commun. 10:356. doi: 10.1038/s41467-018-08197-4.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.19.998484; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

SRA	Gonadulin-1	Gonadulin-2	VgR
Testis			
SRR1920182	**2753.7**	**.0**	**.9**
SRR4341948	**.0**	**.0**	**.1**
SRR7160591	**2665.5**	**.0**	**1.1**
Ovary			
SRR1920180	**.2**	**524.7**	**463.9**
SRR2087155	**.2**	**15.1**	**387.1**
SRR4341947	**.0**	**.0**	**.4**
SRR7160590	**.9**	**.0**	**236.9**
SRR8897549	**.0**	**.3**	**165.1**
SRR8897550	**.2**	**.1**	**153.6**
Brain			
SRR8897555	**19.2**	**.7**	**1.2**
SRR8897556	**13.7**	**1.2**	**.8**

Table 1. Expression of gonadulin 1 and -2 in the testis, ovary and brain of *Portunus trituberculatus*. Numbers are the number of RNAseq reads per million spots in the various SRAs. Note that expression of gonadulin 2 by the ovary is extremely variable, varying more than a thousand-fold in these samples. Such variation of expression is also present in the testis SRAs. VgR stands for vitellogenin receptor, as the SRA SRR1920180 and SRR1920182 have been mixed up and the expression of the vitellogenin receptor allows unambiguous identification of the ovary SRA. For more extensive data see spreadsheet2.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.19.998484; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fig. 1. Sequence logos for the A- and B-chains of gonadulin (gon), insulin (ins), relaxin (rel) and androgenin (and) from decapods. Note the conserved arginine located three residues before the first cysteine of the A-chain and the absence of the glutamic acid residue immediately before this cysteine residue that is always present in the other insulin-related peptides. Also note that both insulin and relaxin have the best conserved sequences. Sequences are provided in Tables S2-S6.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.19.998484; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

```
Penaeus        -----MRTLATLVVVALLATASEAGERPSL-------YKRGQLVHVCTPRDVKMMARFVC
Stenopus       -----MRAAGVVAVLVLVVVLVCVVVETRP-------PRIRRGLKICSARDVKFMATFVC
Macrobrachium  -----MRFIGAFALVVLL-VATVAETRPPS-------PREKRGIKICSARDVKFMATFVC
Cherax         MGRCSLRGIGALVF--VLATAALLVETRPP-------YRSRRGMKVCSPRDVKFMATYIC
Procambarus    -----MRGVGALMLVAVLTAATLVVETRPP-------PRNKRGIRLCSARDVKIMATYVC
Paralithodes   -----MRCWVAAAVVVVVCVVGVVVVEPRPQQQHQHYHQEKRGVRLCSARDVKLIATYVC
Portunus       -----MRVVVGVVAVVALVV--AVAEGDSL-------MTRSKKYKLCTARDINMFVTQLC

Penaeus        SLHRRSVRSSSSSSSVAD-------S----VIVSGPLFRIPEFYNRPPSRV-MKCGENG
Stenopus       NLHRRSVRSVVQEEEEEEEDDGDSDDGDMGSPLDFKAIEWLAMMNPSQAYLQ-HWCGHSG
Macrobrachium  NLHKRSVRSVDLDENEDDD-------EDFGLPVTDDDLALLGSLG-MVP-RR-PLCGDDG
Cherax         NLHRRSVRSVDDFEDDFES-------P---------GVSRLSGVN-IPPWRP-SRCGNTG
Procambarus    NLHRRSVRSVHEPDNTFRS-------S---------DAASLVDVS-NRPWRS-SPCGNTG
Paralithodes   NLHRRSVRSVALDENSEGDTY---------------IMPGVGGSSVILPWRSLPPCDGPD
Portunus       SGALTRMRVGRHNSHP----------------------APGILTI------R-IPPRGVRG

Penaeus        DDCLP-----AYDPDYD-----------------------------------------
Stenopus       VPCNP-----QRSFQRQQQAAPE---DQQL--DLRHRITPYVFNDWLSQIGYSTNTSP--
Macrobrachium  RDCGR-----QNKPA------------------NITSSQLNYLQAWLEENGLLLSSTLPE
Cherax         RDCGP-----GLDSSSSLPPLARLHHHSPP--AAATTITAADFNRWLSLNGYHDLS-LQE
Procambarus    WDCIP-----EFDSHPIQSSVVDV--------DDAASITLADIQRWLSANGYHRMF-LPE
Paralithodes   GDCGG-----LNSADSPVQQTQR---A------SPPQAPFNMAKRWYALHRFGTAKGMQA
Portunus       PVVSPAEIGSSYEFPSYRESLLDTQNYSPSELDTVDAPYLRPVSYWES------------

Penaeus        ------------------------------------------SASYFLPPSGDLRTNLIK
Stenopus       -------------NSWRRSEEN--------DEMDDDALANTDTKSPLTVPAQNAFRLKVSK
Macrobrachium  ERSQTRTSDLRHHLLRPLLRN--------IETGSQR-GN--PRIPL-EDLFPSLKGSVTK
Cherax         DS-NGLGNDLVNDPWQAIREN--------GETNQER-ENGLVRVNSATHFLGDIS--LAK
Procambarus    NS-NNLWP----AKIGMINEN--------ADINQDT-ENGFGRVNYAIPFLGDLRTNMAK
Paralithodes   EKENDITEDYRADSNSIDYAPSLSGALPLSELFPEN-VSKFIRVNY-PKYVTSLLPHKGK
Portunus       -------------QSWNNYKESSAE-----TEVPSEYYGNGLTSR--SKKWLNDQ-RNEGE

Penaeus        RDKEVVRA--VGLTIADVRRKCCLNGCLPEDFYGACR
Stenopus       RDKEVDWPSMVPSTLGDIRKSCCVRECSAEDFYGACV
Macrobrachium  RDKEIDWPLVPSMTLGDIRKNCCIRQCRVEDFYGACT
Cherax         RDREVDWPVLVPRSLSDIRRNCCLRECTAEDFYGACS
Procambarus    RDKELDWPVVAPRSLGDVRKNCCLRECSVEDFYGACS
Paralithodes   RDKEMSLPVVVPRSLSAIRQDCCVKECNAEDFFGACS
Portunus       GSSAMIWLRRKRLTFGDIRQKCCTDGCYASELQSLCA
```

Fig. 2. Sequence alignment of a number of gonadulin 1 sequences representing different decapod infraorders: *Penaeus vannemei* (from Trinity), *Stenopus hispidus* (Trinity), *Macrobrachium nipponense* (GHMG01073270.1), *Cherax quadricarinatus* (AIU40993.1), *Procambarus clarkii* (GARH01003670.1), *Paralithodes camtschaticus* (GHJC01030171.1) and *Portunus trituberculatus* (GFFJ01045966.1). Additional gonadulin 1 sequences can be found in table S3. Cysteine resiudes are highlighted in red and show the typical pattern associated with the disulfide bridges in insulin. In most species there is an additional pair of cysteines, but not in *P. trituberculatus.* Sequences are provided in spreadsheet1.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.19.998484; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

```
Macrobrachium  MRTFLLVPLIPTVLVLPLVYAGGPTDD-ANWPRSRRICSPRQVDLRAYEACHLTKRRSVD
Cherax         MGGFETMM--MVIVMVMVVVESHP----SRTKRSGRICSSREVLQAANVVCNLTRRSAPA
Procambarus    MMSFQTAVLLVVVTVMVVVVDSRS----LRTRRTGRICSTRDILQAANAVCNLTRRSTRG
Paralithodes   -----PLQLLLLVLELLLTKEGRAWPALRGRRSTSDLCTPREIRRIANDVCNIARRSIRT
Portunus       MNV--VAVVAVVVVMVVTVVEGDP---------RVTVCSQRDLRNTISNLCSIRRQRSIW

Gonadulin-3    MKGIG-YYLAIAITALLFIRGTESAAQLFRVRRSIRVCTPRDVARIGHFMCKRSVLPTEG

Macrobrachium  GAEEGQEVAGDS--------------------------HLGLTSNQAR-------IPNSPS
Cherax         S-VFSGTLSSASRGVTAV------------A------SGGVTAGDAFGDAYSGVTLSTPS
Procambarus    P-APAASETST---------------------------SGGLSERRGAQE-----------
Paralithodes   Q-LTPSHHETD-WPLLRR------------T------PNGIPAGHEGTNNY---RFSLPS
Portunus       SFLFSSSSPPSS-SSSS--------------------HQSISFPHTF--TYSDSSTSSPS

Gonadulin-3    Q-DGAGNVRPDERSLEHQERSKDSNTFSRAESSGRDAFWEESPEHAFHDIYPFLEFKRVT

Macrobrachium  SVTS------------------TKIFPPTKTHRPTRS----------------------E
Cherax         PRNLD----------WLYGRSYKKPFRVTSPSPETR--------------FILNNGISNR
Procambarus    --HLE----------HLSS---L----------------------------NSFKNDVSNR
Paralithodes   PLTLS----------SDVRPGVHSLIRS---NTKLS--------------QVLVPPLSNT
Portunus       SSYSSSSSSSSSSSLLTHLGIKGKFPRNQGRPRTHSTQEDTTEGLTRTYPLILPSLVGR

Gonadulin-3    --YYE------------GSGQIDGRYFDGSGHTEGS-------------YFGGSGQIEGS

Macrobrachium  DDFG--------------QF-----------KRRMKVLKR-R-------SSPPFKRQVTDV
Cherax         KRAS------------TVV----NLGTTGVVNSWKNLIGTRESDPTGVTGRIRNRQNPVV
Procambarus    KVYS------------RK--------------------VRT-------ANSYPRFSVSSLV
Paralithodes   ------------------I----KARLASQAMVWRLMLA----------GMVRRKRLNRR
Portunus       GDVGLS-MHG--AEGVGRH----RVGAGGRRSRWRVGK----------NKWWVNKDAIKR

Gonadulin-3    YSDGSGLIDGSYFDGSGQIDGSYFDGSGQIDGSYFDGRG-------HIDDSLTERKLQKR

Macrobrachium  PDWRELNFTTYEEVRLFCCRRSCPPEAYYGIC-
Cherax         KRHLNAETLTLRELRIICCVRECTIQDFLGACS
Procambarus    KRNVNSEILTLRDLRVMCCVRECSIQDFMGACT
Paralithodes   QTQHQSTTMTLEELRETCCNAQCSEQDFLAACS
Portunus       HFLTKRKVRTLGDLRRRCCVHGCLESELLSACI

Gonadulin-3    SNKWPRKFETIADIRKYCCHHDCPIELFLAECA
```

Fig. 3. Sequence alignment of a number of gonadulin 2 sequences representing different decapod infraorders together with a gonadulin 3 sequence. Note that gonadulin 2 has the cysteine residues that characterize insulin, but lack the additional disulfide bridge of gonadulin 1. The species are: *Macrobrachium nipponense* (Trinity), *Cherax quadricarinatus* (from genome), *Procambarus clarkii* (from genome SRA), *Paralithodes camtschaticus* (GHJC01008804.1) and *Portunus trituberculatus* (Trinity). These are the same species as in Fig. 1, except for *P. vannamei* and *S. hispidus*, which appear to lack a gonadulin 2 gene. The gonadulin 3 gene is from *M. rosenbergii* and was obtained using Trinity. Sequences are provided in tables S3 and S4.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.19.998484; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

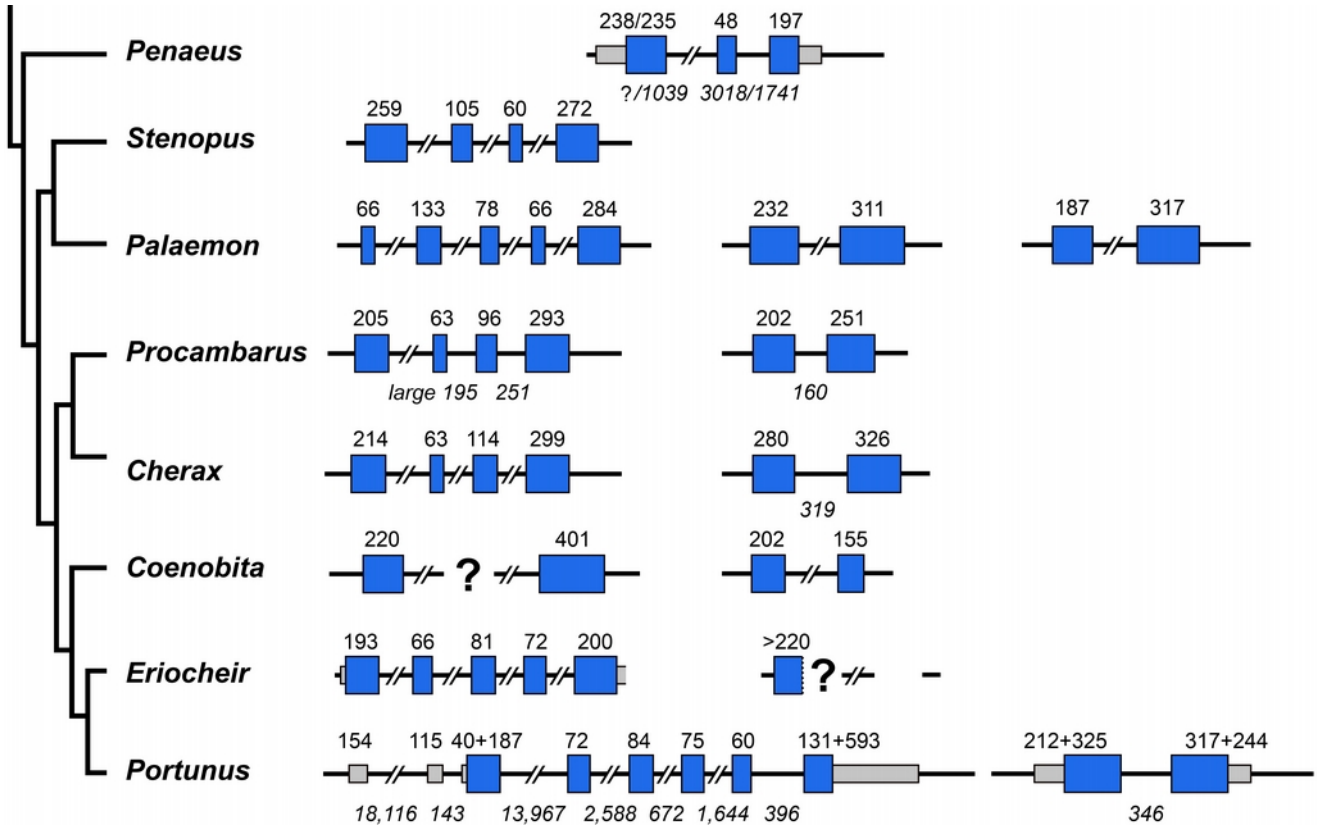

Fig. 4. Configuration of the various decapod gonadulin genes. The peptides coded by the genes from *Penaeus vannamei* and *P. monodon* are most similar to gonadulin 1, as is also evident from the primary sequences of the peptides. Note that the gonadulin 2 genes have much smaller introns than the ones that code gonadulin 1. Lines on the left side of the figure reproduce the most recent consensus regarding their phylogenetic relations (Wolfe et al., 2019). Blue boxes indicate translated and smaller grey ones untranslated sequences while lines represent introns. Numbers above the exons and those in italics below the introns indicate their sizes in nucleotides. Where there are two numbers separated by a slash for the two *Penaeus* sequences, the first refers to *P. vannamei* and the second to *P. monodon.* The sequences for *Penaeus vannamei* and *P. monodon*, *Portunus trituberculatus*, *Procambarus virginalis* and *Palaemon carinicauda* are derived from their genome assemblies, the *Cherax quadricarinatus* genome sequence was corrected with reads from genome SRAs and the exon-intron boundaries of the genes from *Stenopus hispidus*, *Procambarus clarkii*, *Coenobita clypeatus* and *Eriocheir sinensis* were obtained using Trinity on genome SRAs from these species. The exon sizes of *P. clarkii* and *P. virginalis* are identical, the intron sizes indicated are from the *P. viriginalis* assembly.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.19.998484; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

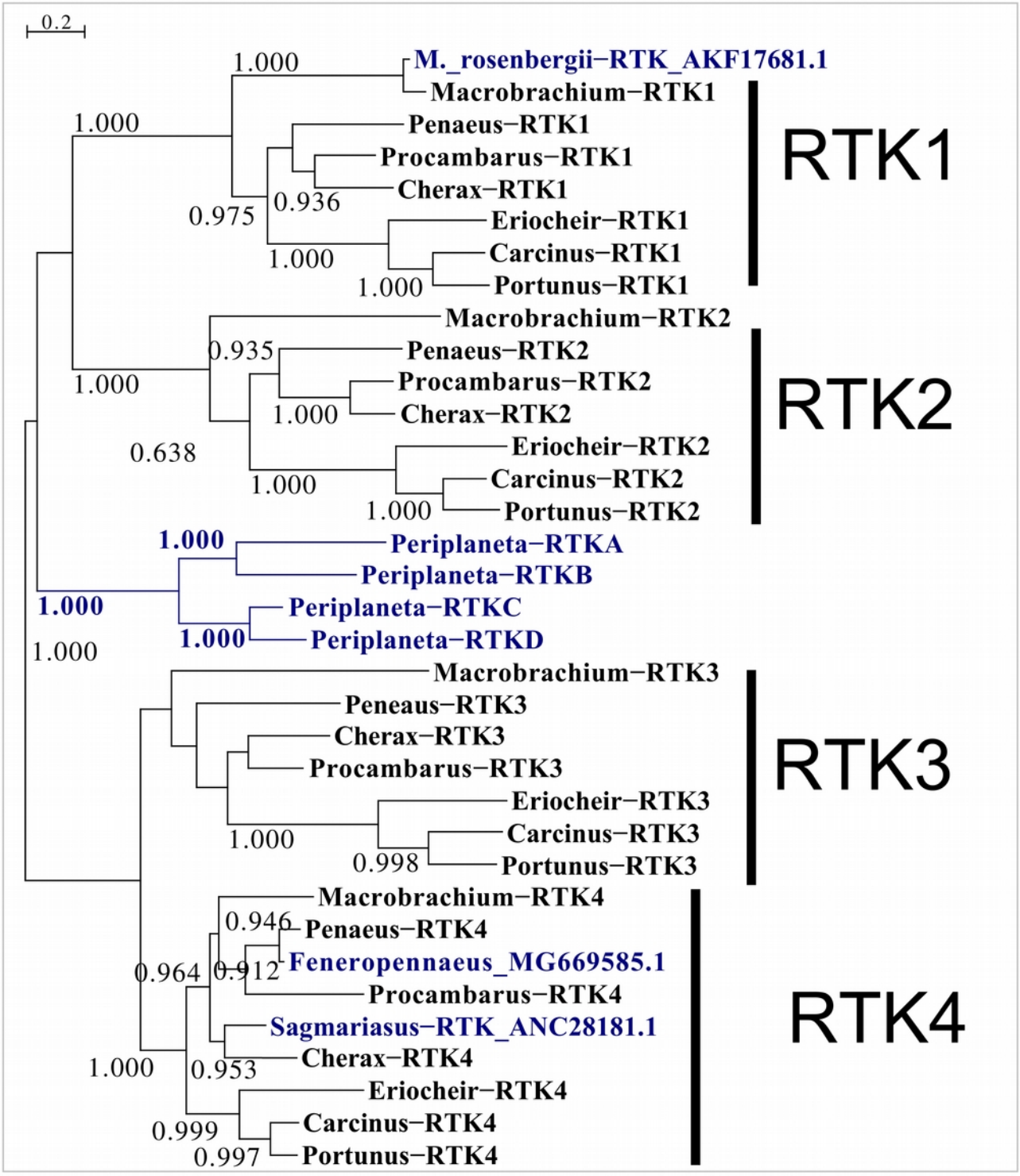

Fig. 5. Phylogenetic tree of various decapod RTKs. For comparison sequences of the four *Periplaneta* RTKs have been added, as well as those of three previously identified decapod RTKs, for two of those, the ones from *Fenneropenaeus* and *Sagmariasus*, there is evidence that they interact with the androgenic insulin-like peptide. Sequences are provided in spreadsheet1.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.19.998484; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

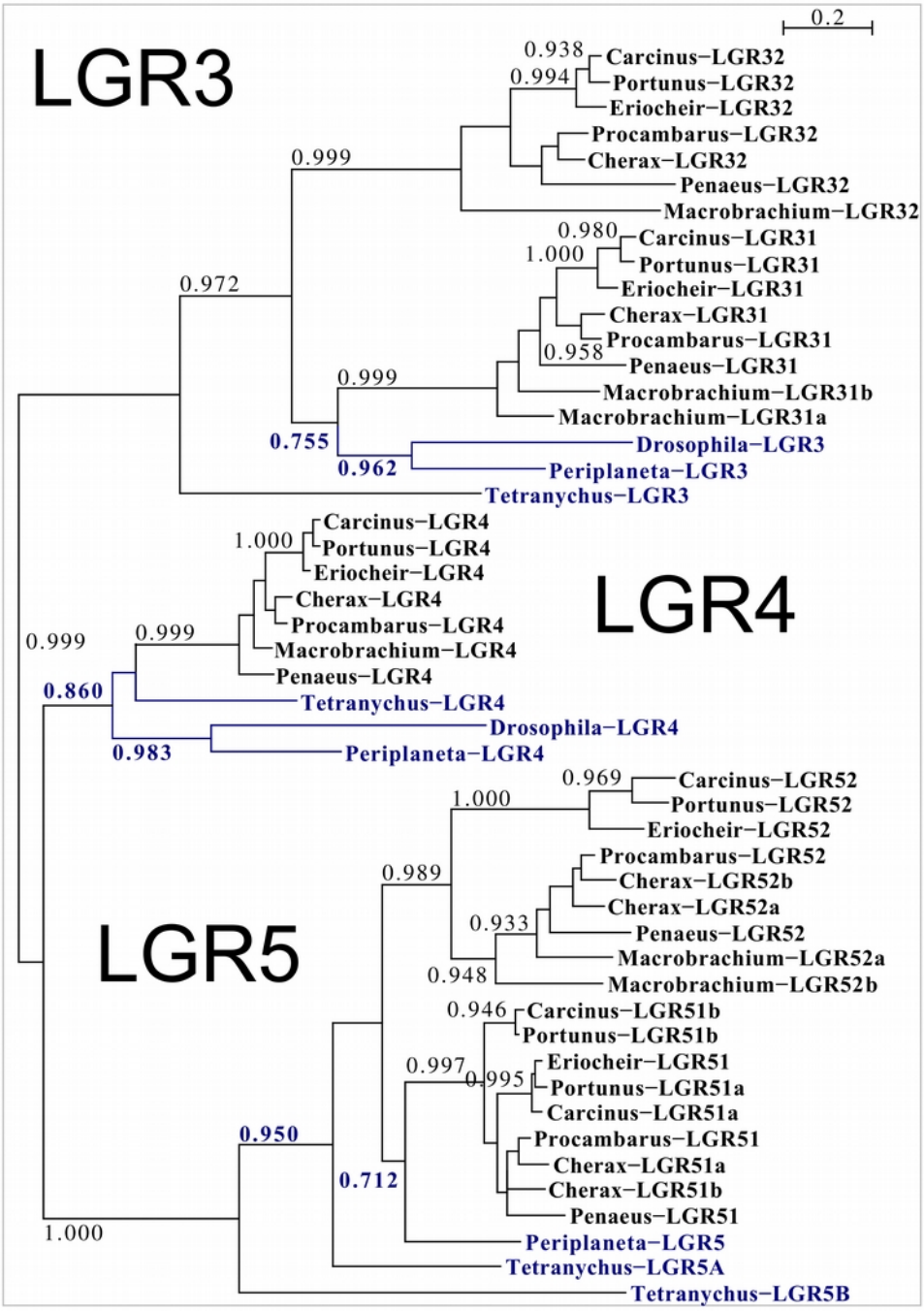

Fig. 6. Phylogenetic tree based exclusively on the transmembrane regions of various decapod LGRs that might function as receptors for insulin-related peptides. For comparison sequences of orthologous receptors from the spider mite *Tetranychus urticae*, the fruit fly *Drosophila melanogaster* and the cockroach *Periplaneta americana* have been added. The only deorphanized receptor is *Drosophila* LGR3 which has *Drosophila* ilp 8 for ligand, while the LGR4 are likely ligands for relaxins. Note that *Drosophila* LGR3 has two decapod orthologs, while *Drosophila* LGR4 has only one. On the other hand decapods have several orthologs of *Periplaneta* LGR5. Sequences are provided in spreadsheet1.