Complete transcriptome assembly and annotation of a critically important amphipod species in freshwater ecotoxicological risk assessment: *Gammarus fossarum*

Domenico R. Caputo\textsuperscript{a,1}, Samuel C. Robson\textsuperscript{b,1}, Inge Werner\textsuperscript{c}, Alex T. Ford\textsuperscript{b,}\textsuperscript{*}

\textsuperscript{a} Institute of Marine Sciences, School of Biological Sciences, University of Portsmouth, Ferry Road, Portsmouth PO4 9LY, UK
\textsuperscript{b} Centre for Enzyme Innovation, St. Michael's Building, University of Portsmouth, White Swan Road, Portsmouth PO1 2DT, UK
\textsuperscript{c} Swiss Centre for Applied Ecotoxicology, Eawag - EPFL, Überlandstrasse 133, 8600 Dübendorf, Switzerland

**A B S T R A C T**

Because of their crucial role in ecotoxicological risk assessment, amphipods (Crustacea) are commonly employed as model species in a wide range of studies. However, despite their ecological importance, their genome has not yet been completely annotated and molecular mechanisms underlying key pathways, such as the serotonin pathway, in development of ecotoxicological biomarkers of exposure to neuroactive pharmaceuticals are still poorly understood. Furthermore, genetic similarities and discrepancies with other model arthropods (e.g., *Drosophila melanogaster*) have not been completely clarified. In this report, we present a new transcriptome assembly of *Gammarus fossarum*, an important amphipod species, widespread in Central Europe. RNA-Seq with Illumina HiSeq technology was used to analyse samples extracted from total internal tissues. We used the Trinity and Trinotate software suites for transcriptome assembly and annotation, respectively. The quality of this assembly and the affiliated targeted homology searches greatly enrich the molecular knowledge on this species. Because of the lack of publicly available molecular information on the serotonin pathway, we also highlighted sequence homologies and divergences of the genes encoding the serotonin pathway components of the well-annotated arthropod *D. melanogaster* and Crustacea with the corresponding genes of our assembly. An inferior number of hits was found when running a BLAST analysis of both *D. melanogaster* and Crustacea mRNA sequences encoding serotonin receptors available in GenBank against the total assembly, compared to other serotonin pathway components. A lack of information on important components for serotonin biosynthesis and vesicle endocytosis (i.e., tryptophan hydroxylase and vesicular monoamine transporter) in Crustacea was also brought to light. Our results will provide an extensive transcriptional resource for this important species in ecotoxicological risk assessment and highlight the need for a more detailed categorization of neuronal pathways components in invertebrates.

**1. Introduction**

Identifying new sentinel species and acquiring molecular data for biomarker analyses are fundamental objectives in today’s ecotoxicological research. Given the number and diversity of chemicals released into natural environments, the need for a deeper understanding of the molecular mechanisms taking place in native species in response to exposure to these substances is increasing more and more. Fortunately, the degree of specificity and above all the number of biomolecules that can be simultaneously analysed in a given species have exponentially increased over the last decade due to the use of multi-omics platforms, such as next generation sequencing (NGS) technologies (Simmons et al., 2015). In research on aquatic species for instance, -omics platforms have allowed the development of new biomarkers of exposure to a considerably higher number of chemicals, allowing an even more precise description of the molecular effects in response to the exposure to a range of ecologically dangerous substances (Hauser-Davis et al., 2012; Jayapal, 2012; Barrera and Ariza, 2017; Poynton et al., 2018). With particular regard to freshwater species, numerous studies have been carried out to evaluate the molecular effects of a large spectrum of substances unremoved by wastewater treatment plants (WWTPs). These include endocrine disruptor...
compounds (Bahamonde et al., 2014; Schneider et al., 2015; Gouveia et al., 2018), pharmaceuticals (Sanchez et al., 2011; Pascoe et al., 2003; Brandão et al., 2013), metals (Atli and Canli, 2007; Gismondi et al., 2017; Lebrun et al., 2017) and polychlorinated biphenyls (Rylander et al., 1998; Leroy et al., 2010) on various species of fish and invertebrates. Other important studies focusing on acquiring genomic sequences or applying multiple omics platforms to elucidate unclear biological mechanisms in species of ecotoxicological interest are also available (Trapp et al., 2016; Macher et al., 2017; Poynton et al., 2018).

Because of their ecologic relevance, invertebrates and in particular amphipods, are commonly employed as test organisms in ecotoxicological assessments. Among amphipods, the genus *Gammarus* represents the greatest number of epigean freshwater species distributed throughout the Northern Hemisphere (Trapp et al., 2014a, 2014b). They are commonly used as model species in aquatic ecotoxicology for several reasons (Kunz et al., 2010). First, they are widespread and found throughout a large habitat range, where they often occur at high densities. Second, they occupy a large trophic repertoire as: herbivores, predators, and detritivores playing a major role in leaf-litter breakdown processes (Dangles and Guéron, 2001). They also constitute a food reserve for macroinvertebrates and fish. Gammarids can be easily maintained in the laboratory or used in field bioassays (Kunz et al., 2010), in which it is possible to assess the impact of pollutants by measuring molecular markers related to diverse modes of action, such as neurotoxicity (Xuerb et al., 2009), as well as by using life-history-trait reproductive features (Geffard et al., 2010). In particular, *Gammarus fossarum* has been shown as notably useful in ecotoxicology studies due to an intrinsic sensitivity to anthropogenic pollutants (Trapp et al., 2016; Wigh et al., 2017) and a spatial distribution across Central Europe (Stráskraba, 1962; Meijering, 1971), making them ideal for freshwater risk assessment in the European geographical area. As a basic premise, a detailed elucidation of the molecular pathways is key to understanding the effects of pollutants on exposed organisms. In the literature, there are several studies that specifically applied sequencing platforms to *G. fossarum*. For example, Weiss et al. (2014) used sequencing of the genes encoding rRNA 16s and cytochrome c oxidase I (CO1) for a taxonomic purpose. Trapp et al. (2014a, 2014b) used a proteogenomics strategy to generate a molecular report specifically on the reproductive tissues of female individuals of *G. fossarum* and Macher et al. (2017) sequenced the complete mitochondrial genome of this amphipod species. However, the lack of publicly available molecular information on amphipod physiological pathways compared to other model species is still clear. This is particularly true when considering neurological pathways, especially the molecular processes at synapses or applying multiple omics platforms to elucidate unclear biological mechanisms in species of ecotoxicological interest are also available (Trapp et al., 2016; Macher et al., 2017; Poynton et al., 2018).

Therefore, a more detailed knowledge of the molecular processes taking place in the serotonergic synapse and a clear elucidation of the genetic differences of the main components (i.e., enzymes for serotonin biosynthesis, synaptic re-uptake and receptors) compared to other well-annotated arthropods such as *D. melanogaster*, will have a highly positive impact on the development of new ecotoxicological biomarkers. This aspect becomes even more important when considering the historical contribution of crustaceans in the synaptic physiology research. For instance, the presence of many parallels found in central synaptic physiology between crustaceans and vertebrates (Wu and Cooper, 2012) and the amount of knowledge about the basic principles underlying the generation, maintenance, and modulation of rhythmically active behaviours (e.g., walking, chewing and breathing) in humans, gained through studies on crustaceans (Selverston and Moulines, 1987; Harris-Warrick et al., 1992; Marder et al., 1995; Selverston et al., 1998; Nusbaum et al., 2001; Skrie, 2001; Cooke, 2002; Fénelon et al., 2003; Selverston, 2005; Selverston and Ayers, 2006; Marder and Bucher, 2007; Stein, 2009; Christie et al., 2010; Blitz and Nusbaum, 2011; Christie, 2011; Dickinson et al., 2016). While much work has focused on examining the behavioural effects of neuroactive drugs active on the serotonergic synapse in invertebrates, commonly released in aquatic environments thorough wastewaters, (De Lange et al., 2006; Guler and Ford, 2010; Bossus et al., 2014), many physiological and molecular details remain largely unknown. Although there are studies that have applied high-throughput omics technologies on various species of amphipods (Short et al., 2014a, 2014b; Gouveia et al., 2017), the genomes of many organisms belonging to this species have not yet been completely annotated. To our knowledge the only study applying a next-generation sequencing approach, in order to give an overall report on the *G. fossarum* transcriptome was performed by Cogne et al., (2019). The authors sequenced the transcriptome of 7 different taxonomic groups of *G. fossarum* and combined the sequencing data with a high-throughput proteomics analysis, in order to provide a broad molecular dataset on this amphipod. Our report will significantly add to the gene discovery work on this critically important species in ecotoxicological risk assessment as well as providing an overview on the publicly available sequence resources on the serotonin pathway in the sub-species A of *G. fossarum*. In this study, we employed a *de novo* transcriptome assembly approach to identify functionally relevant transcripts and explored the serotonin pathway.

2. Materials and methods

2.1. Sampling

The sampling was performed in mid-September 2017 at the Eulach river. The sampling site is located in Elgg, Switzerland at the following geographic coordinates: 47°30′09.40″N – 8°51′09.40″E. *G. fossarum* individuals were collected from beneath stones and leaves at the bottom of the stream, using a standard kick-net method. A net with 1 mm mesh size was used. The amphipods were removed from the net using forceps and sorted, separating the species of interest from leaves and other invertebrates. The animals were placed into 10 L buckets containing stream water and quickly transported to the laboratory where they were placed in controlled conditions. The amphipods were placed in glass tanks filled with continuously aerated stream water (20 cm depth). Incubation conditions were 16 ± 2 °C with a 12/12 light-dark cycle (Blarer and Burkhardt-Holm, 2016). Gammarids were fed *ad libitum* with alder leaves (*Alnus glutinosa*) collected at the sampling site. After a 24 h period, RNA was extracted.

2.2. Total RNA extraction

Total RNA was extracted from *G. fossarum* total internal tissues using RNeasy® Mini Kit (Qiagen, Hombrechtikon, Switzerland), following manufacturer’s instructions. A total of 100 amphipod dissections were conducted, but in order to increase the RNA yield 5 independent amphipods were pooled per replicate, resulting in a total of 20 distinct pools. Each pool was considered an independent biological replicate for downstream analyses. Sampling was conducted independently for both
males and females (10 male pools and 10 female pools). Fresh amphipods were anaesthetised for 10–15 min in a 5% (v/v) clove oil solution prepared in water and washed in DEPC water (Sigma-Aldrich, Shieldsdorf, Germany) prior to dissection, to remove any residual debris. Dissections were performed under a stereo-binocular (x3-4 magnification; SZ2 – ILST, Olympus), using stainless steel forceps. Heads were removed from the body, allowing an easier removal of the internal tissues. Internal tissues were washed in DEPC water and placed in 1.5 mL tubes previously cooled. Tubes were quickly snap-frozen in liquid nitrogen and placed at −80 °C until RNA extractions. For each tube, one pre-treated stainless-steel bead (Qiagen, Hilden, Germany) and 350 μL of lysis buffer plus 10 μL β-mercaptoethanol (Sigma-Aldrich, Buchs, Switzerland) were added and the samples were immediately placed into the adaptors of a Tissue Lyser II© machine (Qiagen, Hilden, Germany). Stainless-steel beads were previously subject to two treatments of 15% (v/v) H2O2 and 70% (v/v) ethanol washings followed by 20 min of UV irradiation, to remove any potential chemical/biological contamination. Three 20 sec mechanical stirring cycles at 30 Hz speed were performed, in order to disrupt the tissues and homogenize the cell suspension. The lysed tissue samples were centrifuged at full speed at room temperature for 3 min to separate the cell debris from the supernatant and the supernatants were transferred to fresh tubes. Total RNA was then extracted from G. fossarum tissue samples. RNA concentrations and purity were assessed measuring the absorbance at 260 and 280 nm using Nanodrop ND-1000 (Witec, Littau, Switzerland). Finally, RNA integrity was checked using an Agilent 2100 Bioanalyzer (Agilent Technologies, Wàhausel, Germany) assay. All RNA samples displayed low background signal with sharp peaks corresponding to intact ribosomal RNA.

2.2.1. RNA sequencing

An Illumina TruSeq Stranded mRNA library kit was used to generate cDNA libraries for each of the 20 RNA samples, which were run on an Illumina HiSeq 2500 sequencer by GATC Biotech (Konstanz, Germany) to generate paired-end 150 bp reads.

2.2.2. Quality control

Quality control of the raw reads was performed using FastQC v0.11.7 (Andrews, 2015). Species-specific sequence purity was assessed by using a multiple genome alignment approach, by mapping reads against a database of different model species taken from the Ensembl database (Aken et al., 2017), a draft transcriptome for Gammarus chevreuxi (Truebano et al., 2016), and all RefSeq entries for Gammarus fossarum using MGA v1.4 (Hadfield and Eldridge, 2014). Read trimming was performed using Trim Galore v0.4.4 (Krueger, 2012) using the following parameters “--illumina -q 20 --stringency 5 -e 0.1 --length 20 --trim-n” to remove Illumina adapter sequence contamination, and to trim reads for ambiguous or low-quality base calls.

2.2.3. Assembly and annotation

Reads were combined across the data set and used to generate a putative transcriptome assembly using Trinity v2.5.1 (Grabherr et al., 2011) with parameters “--seqType fq --max_memory 100G --CPU 24 --min_contig_length 200 --min_kmer_cov 1 --SS_lib_type RF --verbose --full_cleanup”. Raw sequencing data were deposited in the NCBI Sequence Read Archive (SRA) and the complete transcriptome was deposited in the Transcriptome Shotgun Assembly (TSA) database. These data have been collected under BioProject accession code PRJNAS56212. Unique transcript sequences were clustered into potential alternatively spliced isoform groups and paralogous “genes”. TransDecoder v5.0.2 (Haas et al., 2013) was run using default parameters to identify open reading frames (ORF) of 100 amino acids or more within transcripts, and putative protein amino acid sequences were produced. Transcripts were annotated against the Universal Protein Knowledge Base (UniProtKB) SwissProt database (The UniProt Consortium, 2017) using BLAST (Altschul, 1990), either at the protein level by taking the TransDecoder derived peptide sequence (using “blastp”) or from the translated nucleotide sequence directly if no ORF was identified (using blastx). Additional annotation was performed against the Protein family (Pfam) database (Finn et al., 2013) using HMmer (Finn et al., 2011), Clusters of Orthologous Groups of proteins (eggNOG) database (Huerta-Cepas et al., 2015), the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa, 2000), and Gene Ontology (GO) database (Ashburner et al., 2000). Results were collated into a single output table using Trinotate v3.02 (http://trinotate.github.io/). Transcriptome completeness was assessed by comparing the assembly against a database of metazoan universal single copy orthologs using BUSCO v2.0 (Simão et al., 2015).

2.2.4. G. fossarum sub-species assignment

A sequence alignment of the transcript sequences against the complete G. fossarum mitochondrial genome (Macher et al., 2017) was conducted using Blast2GO 5 Basic software (Conesa et al., 2005), in order to identify transcripts corresponding to mitochondrial genes that may allow for taxonomic assignment (i.e., rRNA 16S and CO1) (Müller, 2000; Weiss et al., 2014). The following parameters were used for the analysis:

- E-value: 1.0E−3
- Number of blast hits: 20
- Word size: 11
- HSP length cutoff: 33

BLAST was then used to align putative mitochondrial transcripts against the NCBI non-redundant database (Altschul et al., 1990) for “amphipods – taxid:6821”, to assign both G. fossarum sub-species and CO1 type.

2.3. Serotonin pathway survey

Serotonin-specific nucleotide sequences were identified from GenBank (Benson et al., 2017) and were filtered to identify serotonin-specific mRNAs for enzymes, transporters and receptors, belonging to either D. melanogaster or Crustacea. In cases where multiple transcript variants were identified, all were considered, including in silico predicted transcripts. cDNA and single exon sequences were excluded. Genes encoding serotonin non-specific components (i.e., non-specific ion channels, G proteins, cytochrome p450 enzymes, receptors specific for other neurotransmitters) were not included and their corresponding boxes in Fig. 3 were left blank. A BLAST search for each nucleotide sequence was conducted against our transcriptome using BLAST2GO (Conesa et al., 2005) with parameters as described above to identify G. fossarum-specific serotonin genes. Fig. 1 shows a schematic overview of the workflow used, including both laboratory and bioinformatics procedures.

3. Results

3.1. G. fossarum de novo transcriptome assembly

Sequencing generated a total of 325,393,762 paired end reads across the 20 samples. Base calling quality was excellent across the entire data set, with the vast majority of reads showing average quality scores greater than 30. Mapping of reads against a range of different target genomes identified similar mapping rates for each of the 20 samples, with around 50% of reads mapping to either G. fossarum, G. pulex or E. marinus, whilst the remaining reads importantly did not map to any other model organism indicating no sign of cross-contamination. Trimming of reads to remove Illumina adapters, ambiguous base calls and poor-quality bases resulted in the loss of only 0.1% of reads. In total, 324,958,898 quality-trimmed reads were used for de novo transcriptome assembly. The final assembly consisted of 680,840
transcripts, clustered into 407,060 genes by the Butterfly portion of the Trinity algorithm. The transcriptome consisted of a total of 427,679,404 bp and showed a GC content of 44.68%. Transcripts ranged in size from 100 bp to 31,653 bp, with a median transcript size of 628.16 bp and an N50 of 1,026 bp (Table 1). The Ex90N50 (the N50 score for the transcripts accounting for 90% of the total normalized expression data) was 1922 bp and is accounted for by only 47,563 bp. ORF transcripts (78.7%) were shorter in length than 500 bp (Fig. 2). This suggests that the majority of non-annotated non-coding transcripts are likely a result of fragmented RNA. Annotation for all transcripts identified in this assembly can be seen in Table S1 (Supplementary Information).

### 3.2. *G. fossarum* sub-species assignment

Previous studies have reported strong genetic differences among populations of *G. fossarum* (Siegismund, 1988; Siegismund and Müller, 1991, Müller, 2000). Although this amphipod has been a subject of controversial debate in taxonomy for decades (e.g., Roux, 1971; Goedmakers, 1972; Jadźewski, 1977; Goedmakers, 1980; Scheepmaker and van Dalfsen, 1989; Pinkster and Scheepmaker, 1994; Müller, 1998; Westram et al., 2011), to date at least 3 main sub-species have been identified based on nucleotide differences within the mitochondrial gene encoding rRNA 16S (Müller, 2000). An additional taxonomic system in *G. fossarum* based on the differences in the gene encoding cytochrome oxidase 1 (CO1) was proposed by Weiss et al., (2014). In this study, a BLAST analysis of the transcriptome assembly was conducted against the complete *G. fossarum* mitochondrial genome (Macher et al., 2017). Putative mitochondrial transcript sequences from our assembly were aligned against the amphipod NCBI nucleotide database using BLAST in order to identify transcripts corresponding to rRNA 16S and CO1. Among the 20 putative mitochondrial transcripts, one (TRINITY_DN95725_c7_g2_i2) showed a hit against “*Gammarus fossarum*” type A mitochondrial partial 16S rRNA gene, haplotype A14” suggesting that these data are specific to sub-species assignment (*G. fossarum* A) (Müller, 2000) (Table 2). Additionally, another transcript (TRINITY_DN93353_c3_g3_i1) showed a hit against “*Gammarus fossarum* isolate Glos_47 cytochrome oxidase subunit 1 (CO1) gene, partial cds; mitochondrial” and “*Gammarus fossarum* isolate Glos_45 cytochrome oxidase subunit 1 (CO1) gene, partial cds; mitochondrial” (Table 2). One other transcript (TRINITY_DN93353_c3_g1_i3) also showed hits against CO1, but from the amphipod taxa *Chitoniiidae*.  

### 3.3. Serotonin pathway survey

To further test the depth and accuracy of this dataset, we aimed to identify serotonin-related transcripts by comparing against known serotonin-related genes from the NCBI database GenBank (Benson et al., 2018). We used a BLAST analysis of the transcriptome assembly against the annotated serotonin-related genes from the NCBI database GenBank. This analysis showed hits against known serotonin-related genes for only 80,476 (11.8%) of the transcripts. However, 53.5% of transcripts with an identified open-reading frame (ORF) showed a hit, whilst only 5.6% of transcripts with no ORF showed a hit. The vast majority of non-
Given the well-known effects of exposure to antidepressants (a pharmaceutical class specifically acting on the serotonin pathway) on aquatic organisms (Guler and Ford, 2010; Styrishave et al., 2011; Bossus et al., 2014; Chen et al., 2018), a detailed exploration of these specific transcripts was of interest from an ecotoxicological perspective. *D. melanogaster* is often used as a reference to investigate the metabolic pathways in other arthropods, such as amphipods (Bossus et al., 2013; Trapp et al., 2014a, 2014b; Poynton et al., 2018). However, mRNA
Table 2

<table>
<thead>
<tr>
<th>Transcript ID</th>
<th>Max score</th>
<th>Total score</th>
<th>Query cover</th>
<th>E-value</th>
<th>% identity</th>
<th>Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRINITY_DN95725_c7_g2_i2</td>
<td>725</td>
<td>725</td>
<td>17%</td>
<td>0.0</td>
<td>100%</td>
<td>A296490.1</td>
</tr>
<tr>
<td><strong>Gammarus fossarum</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRINITY_DN93353_c3_g3_i1</td>
<td>1110</td>
<td>1110</td>
<td>73%</td>
<td>0.0</td>
<td>100%</td>
<td>M6986097.1</td>
</tr>
<tr>
<td><strong>Gammarus fossarum</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>M6986097.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>M6986097.1</td>
</tr>
</tbody>
</table>

* Top 2 hits for TRINITY_DN95725_c7_g2_i2 were "Gammarus fossarum mitochondrial, complete genome" and "Gammarus fossarum isolate NID03156 (chromosomal DNA gene, partial sequence, mitochondrial)". The table shows the third hit, which was useful for the sub-species assignment (A), according to Müller (2000) classification.

3.3.1. 5-HT biosynthesis components

Serotonin is synthesized from the amino acid tryptophan in two enzymatic steps: hydroxylation of tryptophan by the neuronal form of tryptophan hydroxylase (TPH), which generates 5-hydroxytryptophan (5-HTP), and conversion to 5-hydroxytryptamine (5-HT) by the enzyme DOPA-decarboxylase (DDC) (Daubert and Condron, 2010). No mRNA sequences encoding TPH were available in GenBank for Crustacea, but 3 out of 4 sequences encoding this enzyme in D. melanogaster showed a putative hit in the G. fossarum transcriptome. No hit was found for the enzyme DDC, represented in GenBank by 3 and 2 mRNA sequences for D. melanogaster and Crustacea, respectively.

3.3.2. 5-HT transport proteins

After synthesis, the vesicular monoamine transporter (VMAT) transports 5-HT into vesicles for storage. Upon vesicle fusion with the plasma membrane, 5-HT is released where it interacts with autoreceptors located on the releasing cell or heteroreceptors, serotonin receptors located on other cell types. The serotonin transporter (SERT) transports 5-HT back into the releasing cell where it is likely to be repackaged for release by VMAT or degraded by monoamine oxidase A (MAO-A) located on the outer mitochondrial membrane (Daubert and Condron, 2010). Similarly to TPH, no mRNA sequences encoding the vesicle monoamine transporter VMAT were found in Crustacea. However, each of the 4 sequences encoding this carrier protein found for D. melanogaster gave a hit with at least one transcript. No mRNA sequences encoding the re-uptake protein SERT were found for both D. melanogaster and Crustacea and all showed hits with at least one transcript (Fig. 3).

3.3.3. 5-HT receptors

To date, at least fourteen different serotonin receptor subtypes have been identified and are grouped into seven families in human (5-HT1–5-HT7) (Daubert and Condron, 2010). Although some of these receptors are present in different isoforms in D. melanogaster and Crustacea (Wu and Cooper, 2012), they mediate the whole range of serotonin’s effects and are distributed throughout the body (Daubert and Condron, 2010). All of the serotonin receptors are G-protein-coupled receptors, except for the 5-HT3 ligand-gated ion channel (Nichols and Nichols, 2008). Compared to other components of the serotonergic synapse, fewer hits were found against D. melanogaster and Crustacea mRNA sequences encoding 5-HT receptors. In fact, of all of the presynaptic receptors, 5-HT1A was the only one that gave at least one hit in Crustacea. In contrast, 5-HT1B, 5-HT2, 5-HT2A, and 5-HT5a gave no hits for Crustacea or for D. melanogaster (Fig. 3). No mRNA sequences were found in GenBank for 5-HT1D, 5-HT4, and 5-HT6 for D. melanogaster nor Crustacea. Interestingly, 5-HT3a, 5-HT4, and 5-HT6 gave at least one hit when using human sequences (Fig. 3).

3.3.4. 5-HT inactivation enzyme

An additional serotonin inactivation mechanism is mediated by the enzyme monoamine oxidase A (MAO-A). This enzyme is located on the outer mitochondrial membrane and is responsible for degradation of 5-HT at the presynaptic terminal (Daubert and Condron, 2010). Surprisingly, no mRNA sequences were found in GenBank for D. melanogaster and none of the 4 sequences found for Crustacea gave hits.
4. Discussion

Much has changed in recent years, and researchers are now more aware of the importance of obtaining detailed molecular information in species of ecotoxicological interest (Simmons et al., 2015). The search for new molecular biomarkers for the evaluation of the status of natural habitats, as well as the study of specific pathways affected as a result of anthropogenic activities, has increased exponentially (Pascoe et al., 2003; Atlí and Canli, 2007; Leroy et al., 2010; Sanchez et al., 2011; Brandão et al., 2013; Bahamonde et al., 2014; Schneider et al., 2015; Gismondi et al., 2017; Lebrun et al., 2017; Gouveia et al., 2018).

In parallel, the use of “-omics” platforms has greatly increased the depth of molecular analyses, allowing the acquisition of data from hundreds of thousands of molecules simultaneously. This provides a valuable resource for ecotoxicological research (Simmons et al., 2015). Particular attention has been given to a sub-group of Crustacea, amphipods, due to their sensitivity to aquatic pollutants (Trapp et al., 2015; Wigh et al., 2017) and their central role in the freshwater food web (Dangles and Guéold, 2001). Furthermore, the potential of amphipods to bioaccumulate toxic substances is likely among the main reasons for their crucial role in ecotoxicological research. Munz et al., (2018) found higher concentrations of frequently detected risk-driving substances (i.e., various classes of pharmaceuticals, corrosion inhibitors, biocides, pesticides and personal care products) in Gammarus fossarum and Gammarus pulex sampled downstream of several Swiss WWTPs compared to upstream sites. In a recent study by Miller et al. (2019), the authors performed a sampling of G. pulex amphipods from several shores of the United Kingdom and assessed the potential presence of various dangerous chemicals within them. Surprisingly, illicit drugs such as cocaine and ketamine, as well as banned pesticides and pharmaceuticals, were found. However, the lack of molecular information on these species still represents a limiting factor and a more detailed genomic annotation is fundamental to highlighting homologies and discrepancies compared with other model organisms. Although some progress has already been made, for instance, the parallels to vertebrate central synaptic physiology of phenomenon described at Crustacea neuromuscular junctions (Wu and Cooper, 2012), there is still a lot of work to be done.

4.1. G. fossarum sub-species assignment

A correct sub-type assignment of the amphipod G. fossarum is essential. In fact, the genetic differences between the sub-types are considered strong enough to prevent crossbreeding in a natural setting (Müller, 1998; Wiese, 2013). Although Pinkster and Scheepmaker (1994) were able to obtain G. fossarum F1 juveniles from ex-situ crossbreeding experiments, Müller, the author who differentiated G. fossarum into 3 distinct cryptic species (A, B, C) (Müller, 1998, 2000), found out that sub-types A and B are to be considered reproductively isolate in a natural setting (Müller, 1998). An additional taxonomic system in G. fossarum, based on the differences in the faster evolving gene encoding cytochrome oxidase 1 (CO1) was proposed by Weiss et al. (2014). In this study, we have attributed the G. fossarum sub-type A, haplotype A14 to our species using a BLAST analysis of our assembly against the complete G. fossarum mitochondrial genome (Macher et al., 2017). Although the BLAST analysis did not allow an unambiguous assignment of the CO1 type, it revealed CO1-45 and CO1-47 as most probable CO1 types of the amphipod subject of study.

4.2. Serotonin pathway survey

Considering that the historical contribution of Crustacea in synaptic physiology is unsurpassed (Atwood, 1976; Wiese, 2013) and much is still unclear concerning the synaptic molecular mechanisms within this taxon (Wu and Cooper, 2012), we also carried out a manual annotation of the genes encoding the serotonin-related molecular components present in the final assembly. We focused on this particular neurotransmitter because of the lack of general molecular information in invertebrates in contrast with the wide range of documented behavioural and transcriptional effects of antidepressants released in aquatic environments (Nentwig, 2007; Guler and Ford, 2010; Bossus et al., 2014; Ford and Fong, 2016; Estévez-Calvar et al., 2017). The acquisition of further molecular details underlying the serotonin pathway in amphipods will also be useful in the field of invertebrate neurobiology. In fact, it has been shown that, similarly to neuroactive drugs, some parasite species (e.g., trematodes) are able to manipulate their invertebrate host’s behaviour through modification of monoamine pathways, such as the serotonin pathway (Guler et al., 2015). We therefore aimed to use BLAST to identify putative candidate genes from our assembly that match serotonin-specific mRNA sequences available in GenBank for D. melanogaster, a widely used arthropod for genomic annotation (Bossus et al., 2013; Trapp et al., 2014a, 2014b; Poynton et al., 2018), and Crustacea.

Interestingly, the 5-HT transporter SERT was the only component of this pathway to show matches from both D. melanogaster and Crustacea within our assembly. Although further investigations will be needed to confirm this, we speculate that the re-uptake carriers at synaptic level might have a higher rate of evolutionary conservation compared to other components of the serotoninergic synapse. In comparison, no mRNA sequences encoding the degradation enzyme MAO-A were found in GenBank for D. melanogaster, and none of the sequences found for Crustacea gave hits against our assembly.

Surprisingly, no mRNA sequences encoding the first enzyme of the serotonin biosynthetic chain (TPH) were found in GenBank for Crustacea, but all splicing variants of this enzyme in D. melanogaster sequences gave at least one hit against our assembly, highlighting the homology of this enzyme in D. melanogaster and amphipods. However, none of the splicing variants found for the second serotonin biosynthesis enzyme in D. melanogaster, DOPA-decarboxylase, gave hits. Although two predicted nucleotide sequences encoding DDC enzyme were found for Crustacea, neither gave hits in the G. fossarum assembly. At present, we are unable to clarify whether this enzyme may have additional isoforms in Crustacea not present in GenBank, or whether sequence differences within this taxon are stronger than expected.

When looking at the serotonin receptors sequences, 5-HT1A was the only receptor that gave at least one hit against Crustacea, but not D. melanogaster sequences. 5-HT1b, 5-HT2a and 5-HT3 gave no hits against either the D. melanogaster or the Crustacea sequences. No mRNA sequences were found in GenBank for 5-HT1c or 5-HT7 for either D. melanogaster and Crustacea. 5-HT7 gave no hits against D. melanogaster and Crustacea sequences.

It has previously been shown that the classification of serotonin receptors in humans is not totally applicable to invertebrates (Wu and Cooper, 2012). For instance, 5-HT4 and 4-HT7 are shown to have different splicing variants (Hoyer, Hannon and Martin, 2002) and 5-HT2 could have different RNA-edited isoforms (Burns et al., 1997; Niswender et al., 1998). This analysis highlights the fact that further studies on a more precise classification of receptors for serotonin in invertebrates is strongly needed, considering the well documented effects of pharmaceutical waste targeting the serotonin pathway can have on aquatic species (Guler and Ford, 2010; Styrisheave et al., 2011; Bossus et al., 2014; Chen et al., 2018). 5-HT1A was the only receptor that gave at least one hit when using Crustacea mRNA sequences found in GenBank. Despite no sequences being found for 5-HT3, we did identify hits in our assembly for human 5-HT3 and 5-HT4 mRNA sequences. Therefore, the lower hit-rate observed for serotonin receptors compared to other molecular components of the pathway might be due to a scarcity of genetic information in Crustacea rather than a strong evolutionary divergence of the 5-HT receptor sequences of this taxon.

Taken together, these findings reveal that molecular comparisons between the serotonin pathway of different organisms, or even within Crustacea taxon, may not be possible for all components. However, the...
amount of genetic information available still represents a limiting factor. Given the importance of the serotonin pathway in ecotoxicological research, further studies on the classification of 5-HT receptors in invertebrates as well as their pharmacological targets will be fundamental.

5. Conclusions

The amphibious species G. fossarum represents an important organism in ecotoxicological risk assessment, and yet there remains a significant lack of genomic information. To further add to our current understanding, we used RNA sequencing and de novo assembly to generate a complete transcriptome of G. fossarum - type A. Our dataset will provide an additional resource of genomic information in this poorly annotated species and will represent a reference source for further and more focused molecular analysis on this and other amphibious species. Given the importance of the serotonin pathway in behavioural and neurological studies on invertebrates (Wu and Cooper, 2012; Guler et al., 2015) an exploration of the publicly available mRNA sequences involved in this pathway in D. melanogaster and Crustacea was also conducted. An inferior number of hits was found when running sequence alignment of our assembly against both D. melanogaster and Crustacea mRNA sequences encoding serotonin receptors available in GenBank, compared to other serotonin pathway components. This may indicate a lack of publicly available sequence information on important components of the serotonin pathway in crustaceans, for instance tryptophan hydroxylase and vesicular monoamine transporter, suggesting the need for a more extensive categorization of neuronal pathway components in key invertebrate species and important model species for both ecotoxicological and neurological studies.

Declaration of Competing Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgements

We thank Dr. Melanie Fischer (aQuaTox - Solutions, Dübendorf, Switzerland), Andrea Schifferli and Thomas Bucher (Swiss Centre for Applied Ecotoxicology, Dübendorf, Switzerland) for their valuable technical support in sampling and RNA extractions. This project was financially supported by a bursary awarded by the Swiss Centre for Education and Research (Portsmouth, UK). Dr Robson was partially funded through an additional resource of genomic information in this poorly annotated freshwater amphipod Gammarus fossarum. Environ. Sci. Pollut. Res. 23 (23), 25322–25332.


Gismondi, E., Thomé, J.P., Uren, N., Uher, E., Biau, D., Mazzucchié, G., De Pauw, E.,...


