Title

The male and female gonad transcriptome of the edible sea urchin, Paracentrotus lividus: identification of sex-related and lipid biosynthesis genes

Authors

André M. Machado^{1*}, Sergio Fernández-Boo^{1*}, Manuel Nande¹, Rui Pinto¹, Benjamin Costas^{1,2} and L. Filipe C. Castro^{1,3**}

Affiliations

1.CIIMAR - Interdisciplinary Centre of Marine and Environmental Research, U. Porto – University of Porto, Terminal de Cruzeiros do Porto de Leixões, Av. General Norton de Matos s/n, 4450-208 Matosinhos, Portugal.

2. Instituto de Ciências Biomédicas Abel Salazar (ICBAS-UP), U. Porto University of Porto, Rua de Jorge Viterbo Ferreira 228, 4050-313 Porto, Portugal.

3. Department of Biology, Faculty of Sciences, U. Porto - University of Porto, 4169-007 Porto, Portugal

*These Authors contributed equally.

**corresponding author(s): Luís Filipe Costa de Castro (filipe.castro@ciimar.up.pt)

Keywords: *Paracentrotus lividus*, gonad, sexual development, lipid biosynthesis genes

31 **Highlights**

32 Assembly of a reference transcriptome of *Paracentrotus lividus* gonads.

33 Differential gene expression between males and female gonads of *Paracentrotus lividus*.

34 Identification and validation of pivotal genes involved in biosynthesis and storage of lipids.

35

36 **Abstract**

37 *Paracentrotus lividus* is the most abundant, distributed and desirable echinoid species in
 38 Europe. Although, economically important, this species has scarce genomic resources
 39 available. Here, we produced and comprehensively characterized the male and female gonad
 40 transcriptome of *P. lividus*. The *P. lividus* transcriptome assembly has 53,865 transcripts, an
 41 N50 transcript length of 1,842 bp and an estimated gene completeness of 97.4% and 95.6% in
 42 Eukaryota and Metazoa BUSCO databases, respectively. Differential gene expression analyses
 43 yielded a total of 3371 and 3351 up regulated genes in *P. lividus* male and female gonad
 44 tissues, respectively. Additionally, we analysed and validated a catalogue of pivotal transcripts
 45 involved in sexual development and determination (206 transcripts) as well as in biosynthesis
 46 and storage of lipids (119 transcripts) in male and female specimens. This study provides a
 47 valuable transcriptomic resource and will contribute for the future conservation of the species as
 48 well as the exploitation in aquaculture settings.

49

50

51

52

53

54 Introduction

55 *Paracentrotus lividus*, commonly known as the purple sea urchin, is the most abundant echinoid
 56 species in Europe, with an overall distribution from the Mediterranean Sea to the eastern
 57 Atlantic coast, from Scotland to Southern Morocco, Canary and Madeira Islands (Ghisaura et al.
 58 2016). In the last few years, *P. lividus* populations have constantly decreased due to intense
 59 harvesting, habitat destruction and climate change (Bertocci et al. 2012, Bertocci et al. 2018).
 60 The significant pressure on this species, results from the high price market of their gonads,
 61 which has led to the populational collapse in some geographic areas (Fernández-Boán et al.
 62 2012, Ouréns et al. 2015). Additional negative factors linked to populational decline, include
 63 environmental changes such as temperature, red tides and ocean acidification (Yeruham et al.
 64 2015, Mos et al. 2016, Ohgaki et al. 2019). Economically, the purple sea urchin is highly
 65 desirable in gourmet markets, not only because their gonads are considered a delicacy but also
 66 due to their high protein and polyunsaturated fatty acid (PUFA) content, especially in omega-3
 67 and omega-6 (Prato et al. 2018, Baião et al. 2019, Rocha et al. 2019). Additionally, the market
 68 price of sea urchin gonads is dependent on taste, colour and firmness. Testis from males have
 69 usually a sweet and milky flavour, while female ovaries taste bitter and sour (Phillips et al. 2009,
 70 Phillips et al. 2010). These characteristics together with the gonadosomatic index of sea urchins
 71 establish the market price of the captures. The reproductive period of *P. lividus* is variable, and
 72 several factors such as diet, availability of food resources, photoperiod and water temperature
 73 can condition their spawn (Ghisaura et al. 2016). Usually, in the Atlantic coast *P. lividus* spawn
 74 is annual (Garmendia et al. 2010, Fernández-Boo et al. 2018), while in the Mediterranean Sea
 75 is bi-annual (Sellem and Guillou 2007). The reproductive apparatus of sea urchin is composed
 76 by five gonads with different colour patterns between sexes, while males present a yellow-
 77 orange pattern, female gonads are red-orange. The gonads are composed by two main types of
 78 cells: germinal cells where the gametes are produced and stored, and somatic cells defined as
 79 nutritive phagocytes which are main storage of nutrients and energetic reserves of the animal
 80 (Garmendia et al. 2010, Ghisaura et al. 2016).

81 Sea urchin aquaculture is based on the enhancement of gonad yield, also known as bulking.
 82 This approach consists in the capture of wild animals with the aim of improving their
 83 gonadosomatic index, gonad sensory attributes or manipulation of the reproductive cycle to
 84 commercialize sea urchin when wild animals are not available (Walker et al. 2015). Despite
 85 multiple attempts to develop intensive aquaculture practices, the poor knowledge of dietary
 86 composition for juveniles and the period of time necessary to reach market size are the main
 87 bottlenecks in *P. lividus* production (James et al. 2015, Liu and Chang 2015). Moreover, sea
 88 urchin aquaculture has been hampered by the lack of genomic resources, with some minor
 89 examples involving the production of sea urchin triploids to increase the roe size by suppression
 90 of gametogenesis and the enhancement of nutritive phagocyte growth (Böttger et al. 2011,
 91 Walker et al. 2015). Importantly, to select parental individuals with valuable genetic traits (e.g.

gonad size, colour, or lipid content), for production purposes, it is crucial to develop *omic* resources of the species.

Decisively, in last few years several strategies have been applied to study and comprehend multiple traits with value in production context (Liu et al. 2005, Wang, Ding et al. 2020). Next generation sequencing approaches, both at genomic and transcriptomic level, are now established to explore general biological species features, whether from the commercial or fundamental point of view (Sea Urchin Genome Sequencing 2006). For instance, transcriptome studies in different sea urchin species have been conducted recently to investigate the genes involved in the synthesis of polyunsaturated fatty acids in *Strongylocentrotus nudus* (Jia et al. 2017, Wei et al. 2019), the sex related genes in *Mesocentrotus nudus* (Sun et al. 2019) or the response to ocean acidification in *S. purpuratus* (Evans et al. 2017). Other species with less economic value, but highly important to local communities, have been also studied and information regarding genes involved in development, fertilization, toxin effects or immune system response against pathogens are now available (Gaitan-Espitia et al. 2016, Laruson et al. 2018). In line with transcriptomic studies, the number of whole genome sequencing projects in sea urchin species has drastically increased after the massive effort in the genome sequencing of *S. purpuratus* (Sea Urchin Genome Sequencing 2006, Janies et al. 2016, Kinjo et al. 2018, Davidson et al. 2020). In *P. lividus*, genomic and transcriptomic resources are scarce in public databases. The few studies available have explored the embryonic development an experimental model for evolutionary and ecotoxicological fields (Gildor et al. 2016, Ruocco et al. 2016, Chassé et al. 2018, Tato et al. 2018, Galasso et al. 2019, Morroni et al. 2019); also, the proteins involved in the attachment by their tube feet were studied for their use in industrial and medical purposes (Pjeta et al. 2020). Regarding the gonadal tissue of adult specimens, only two articles were released, one studying the protein patterns of gonads in males and females at different stages of development (Ghisaura et al. 2016), and second measuring the gene expression level of different pollution biomarkers after metal exposure (Di Natale et al. 2019).

Here, we reported an in-depth transcriptome analysis of the gonads from three adult males and females. These analyses establish the basis for new studies on gonad development, sex differentiation, growth and also lipid and colour traits. Through functional annotation and phylogenetic analyses, multiple pivotal genes involved in sex determination and differentiation, gametogenesis, lipid metabolism and PUFA biosynthesis were scrutinized in this species. This study will be highly valuable not only to improve the information of this keystone species in the Atlantic and Mediterranean Sea, as also could be used to improve and boost the sea urchin aquaculture in Europe.

127

128 **Material and methods**

129 **Experimental setup, RNA extraction and sequencing**

130 Ten adult sea urchins were collected in July 2018 in Vila Chã beach (41.295160 N, 08.737073
131 W) by scuba diving (68.6 ± 4.9 mm diameter; 123.10 ± 22.82 g weight; 13.51 ± 2.29 % gonadal
132 somatic index) (Fig.1a). Initially, four females and six males were identified after gonad
133 extraction by visual identification in a light microscope. A small piece of the gonad was
134 immediately stored in an Eppendorf with 1 ml of RNA later (Sigma) at 4°C for 24 hours and
135 then, frozen at -80°C until RNA extraction. Other piece of the gonad was stored for histological
136 analysis according with Rocha et al. (2019). Sea urchin histological sections were evaluated
137 with light microscopy and all females were in gonadal stage IV and all males in gonadal stage III
138 according to Machado et al. (2019). For RNA extraction, 50-100 mg of gonad tissue was
139 homogenized in 1 ml of trizol (NZY Tech, PT) using a Precellys homogenizer (Bertin Inst.,
140 France). RNA was extracted according manufacturer's instructions. RNA concentration was
141 measured in a DeNovix DS-11 spectrophotometer (DeNovix Inc, USA) and 4 µg of RNA were
142 subjected to DNase treatment (RQ1 – Promega) according manufacturer's instructions. After
143 DNase treatment, RNA was isolated and cleaned using the Total RNA isolation kit (NZY Tech,
144 PT) according manufacturer's protocol and diluted in a final volume of 40 µl in mQ water.
145 Finally, the concentration was measured again in a DeNovix DA-11 spectrophotometer. RNA
146 integrity was observed in a 2% agarose gel and the best 3 males and 3 females were selected
147 for RNA-Sequencing. RNA integrity and quantity were evaluated using the Agilent 2100
148 Bioanalyzer (Agilent technologies, Santa Clara, CA, USA). The RNA integrity number was >6.1
149 in all samples. At the end, six samples (3 females and 3 males) were shipped to Novogene
150 (Honk Kong) Company Limited, and sequenced using Illumina HiSeq-4000 platform (150x2bp,
151 paired-end, 30 million sequencing reads).

152 **Raw data clean up**

153 Initially, the RNA-Seq quality profile, of each sample, was assessed with FastQC (v.0.11.8)
154 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). After that, Trimmomatic (v.0.38)
155 (Bolger et al. 2014) was used to trim and drop reads with quality scores below 5, at leading and
156 trailing ends, with an average quality score below 20 in a 4 bp sliding window and with less than
157 36 bases length. Next, the error correction method, Rcorrector (v.1.0.3) (Song and Florea 2015),
158 was applied to correct random sequencing errors, with the default settings. In the final of the
159 clean-up process, still it was used the Centrifuge (v.1.0.3-beta) (Kim et al. 2016) software
160 (Fig.1b). On this approach, all corrected raw reads were screened against the NCBI-nt
161 database (<ftp://ftp.ccb.jhu.edu/pub/infphilo/centrifuge/data/>) (v.nt_2018_3_3) to obtain a
162 taxonomic classification of each contig, with a minimum hit length of 50. Reads labelled by
163 Centrifuge as non-Echinodermata (NCBI: taxid7586) were considered to be contaminants and
164 excluded from the dataset.

165 **De novo assembly transcriptome, decontamination and quality assessment**

166 After quality filtering control, clean reads were first concatenated and then assembled using
167 Trinity (v. 2.8.4) (Grabherr et al. 2011, Haas et al. 2013) with the specific parameter
168 (SS_lib_type RF). To remove possible sources of contaminations, all transcripts were blasted
169 against NCBI-nt (Download; 20/11/2019) and Univec (Download; 02/04/2019) databases using
170 Blast-n tool (v. 2.9.0) (Fig.1b). Transcripts with a match to *Echinodermata* taxon (NCBI:
171 taxid7586), with an e-value cut-off of 1e-5, identity score of 95 % and minimum alignment length
172 of 100bp, or without matches at all in NCBI-nt database, were retained. Transcripts matching
173 other taxa than *Echinodermata* in NCBI-nt our Univec database were considered exogenous to
174 the *P. lividus* transcriptome (Assembly V0) and exclude from this dataset.

175 **Functional annotation**

176 The Trinotate pipeline (v. 3.0.2) (<https://github.com/Trinotate/Trinotate>) was used to perform the
177 functional annotation of the transcriptome (Bryant et al. 2017). In first step, the TransDecoder
178 software (v.5.3.0) (<https://transdecoder.github.io/>) was applied to predict the open reading
179 frames (ORFs) with at least 100 amino acids. Next, both aminoacid and nucleotide sequences
180 of Assembly V0, were blasted (blast-n tool (v. 2.9.0) (Altschul et al. 1997) and blast-x /p tools of
181 DIAMOND software (v 0.9.24) (Buchfink et al. 2015)), against non-redundant database of NCBI
182 (NCBI-nr) (v. 20/01/2020), NCBI-nt (Download; 30/03/2019), Swiss-Prot (Download;
183 18/02/2020) (The UniProt Consortium 2016), Uniref90 (Download; 04/09/2019) (Suzek et al.
184 2007) databases and searched in Pfam (Download; 18/02/2020) (Punta et al. 2012), eggno
185 (Powell et al. 2011), Kyoto Encyclopedia of Genes and Genomes pathways (Kanehisa and Goto
186 2000). The report generated by the Trinotate pipeline was filtered by an e-value cut-off of 1e-5
187 (Fig.1b).

188 **Transcriptome filtering and redundancy removal**

189 The transcriptome filtering and redundancy removal were performed based on several criteria.
190 First, the functional annotation and ORF prediction of the Assembly V0 version were
191 scrutinized. Thus, all transcripts with at least one blast hit (blast - n, x or p in Uniref90, NCBI-nt
192 or NCBI-nr Databases) in an *Echinodermata* species or codifying to a protein were collected
193 (Annotated Assembly V0) (Fig.1b). Second, the raw reads were mapped onto the Annotated
194 Assembly V0 with Bowtie2 (v. 2.3.4.2) (Settings; --no-mixed --no-discordant --end-to-end --all --
195 score-min L,-0.1,-0.1) and filtered with Corset (v.1.0.9) (Davidson and Oshlack 2014) software.
196 While all the transcripts containing at least 10 reads mapping were kept, the transcripts with few
197 reads mapping were considered spurious and discarded. In addition to the filtering by read
198 coverage, the Corset software clustered of transcripts, based on the ratio of shared reads and
199 expression patterns. Finally, and to remove the redundancy, only the longest transcript per
200 cluster (Unigene) was collected (Assembly V1) (Fig.1b). To evaluate the quality of both
201 transcriptome versions (V0 and V1), several strategies were applied. Completeness and gene
202 content of the transcriptomes were assessed using Metazoa and Eukaryota lineage-specific
203 profile libraries of Benchmarking Universal Single-Copy Orthologs tool (BUSCO v. 3.0.2) (Simão
204 et al. 2015). Accuracy was assessed by the percentage of original clean sequence reads

mapped against the Assembly V0 (RMBT) using Bowtie2 (v. 2.3.4.2) (Langmead and Salzberg 2012) with default settings. Structural integrity and general stats of the transcriptome were examined using TransRate (v. 1.0.3) (Lang-Unnash 1992) with default settings(Fig.1b).

Differential gene expression analyses

Differential gene expression (DGE) analyses were performed using the Unigenes matrix of raw counts, generated by Corset software during the filtration steps, and the Degust (v.4.1.1) platform (Powell 2019) (<http://degust.erc.monash.edu/>). All Unigenes containing less than 1 CPM (count per million mapped reads) in at least three samples were removed from the dataset. In addition, the multidimensional scaling plot was applied to check the variance of the samples. Thereafter, the edgeR (v.3.26.8) package (Robinson et al. 2009) of R (v.3.6.1) was used to perform the DGE analyses between males and females samples of *P. lividus*. During the DGE analyses, the trimmed mean of M-values (TMM) method (Robinson and Oshlack 2010) was applied to perform the normalization of the values across the samples. Finally, the Unigenes were considered differentially expressed if the values of False Discovery Rate – corrected (FDR) p-value < 0.05 and log2|fold change| ≥ 2. The heatmaps were performed in Heatmapper Expression tool (<http://heatmapper.ca/expression/>), within the Average Linkage clustering method and the Euclidean Distance measurement method (Babicki, Arndt et al. 2016)

Gene Ontology analyses and gene cataloguing

To perform the gene ontology (GO) analyses we used the blast-x annotations of the differentially expressed genes (DGEs) against the Uniref90 database. To do that, two datasets were used; firstly, the total number of DGEs; second only the upregulated genes in males and females. Technically, the Uniref90 Id's were subjected to Panther v.15.0 for gene function classification using the sea urchin *Strongylocentrotus purpuratus* as a model organism (Mi et al. 2019). Identification of sex and lipid related genes was done by searching specific GO keywords (Sex related terms: reproductive process GO:0022414; reproduction GO:0000003; developmental process GO:0032502; signalling GO:0023052 and Lipid Related terms: biogenesis GO:0071840; metabolic process GO:0008152; catalytic activity GO:0003824) in panther output and trinitate report, as well as, different genes already described in the bibliography involved in the sex differentiation and lipid related pathways in Echinodermata phylum.

Phylogenetic analysis

Full-length amino acid (aa) sequences were used for phylogenetic analysis, to reduce the impact of nucleotide composition bias. The lipid and sex-related genes as well full-length aa sequences were selected accordingly with their relevance within the main pathways previously identified by Panther and after an exhaustive analysis of the related literature (Suppl. Table. 1). After collecting the target sequences, orthologous proteins were gathered from NCBI database,

242 using the blast-p algorithm with e-value cut-off of 1e-5 and at least 75% identity in echinoderms
 243 (*Strongylocentrotus purpuratus*, *Apostichopus japonicus*, *Asterias rubens*, and *Crassostrea*
 244 *gigas*) and vertebrates model species (*Danio rerio*, *Xenopus tropicalis*, *Gallus gallus*, *Mus*
 245 *musculus*, and *Homo sapiens*) (Suppl. Table. 1). Multiple alignments were performed with all
 246 orthologs sequences using MAFFT v7.402 free software (Katoh et al. 2019) with the L-INS-I
 247 algorithm (Katoh and Standley 2013). Subsequently, the alignments were manually reviewed,
 248 and short sequences and gaps were removed. To remove gaps we used the GapStrip /
 249 Squeeze v2.1.0 (<http://www.hiv.lanl.gov>) software and all columns of the multi alignment with
 250 95% nucleotides fulfilled were kept for the phylogenetic analyses. Phylogenetic trees were
 251 constructed using PhyML 3.0 server (Guindon et al. 2010), using the maximum likelihood
 252 method with the default parameters (specific substitution model details in Suppl. Table. 2). Later
 253 the phylogenetic trees were visualized by Dendroscope (Huson et al. 2007) and rooted with
 254 mollusc sequences from *Crassostrea gigas*.

255

256 **Results and Discussion**

257 **Histological analysis**

258 In the case of females, the gonads presented an advanced stage of development being two of
 259 them at stage III (F1,F2) and one at stage IV (F3) according to scale of Byrne (1990). Stage III
 260 is considered a premature stage where all oocytes are at all stages of development. The
 261 nutritive phagocytes are displaced from the central position to the periphery by the large
 262 oocytes (Fig. 2 a, b, c). The stage IV is a mature stage with ovaries closely –packed and most
 263 of them with a diameter higher than 90 µm. The NP at this stage are in a thin layer around
 264 oocytes (Fig. 2 c). All males were in stage III of maturation. These stages are characterized by
 265 mature testes packed with spermatozoa (S) and nutritive phagocytes (NP) are clearly limited to
 266 the periphery (Fig. 2 d, e, f). The collection of individuals in a similar stage of maturation allowed
 267 to obtain a comparative dataset between males and females.

268 **Sequencing data**

269 From the ten *P. lividus* specimens collected, six (three males (M1, M2, M3) and three females
 270 (F1, F2, F3)) were selected to perform gonadal RNA-Seq sequencing. The RNA-Seq approach
 271 generated a total of 204,865,648 PE reads that after validation with several quality-control and
 272 filtering softwares, were reduced to 203,895,472. Although the conservative approach used to
 273 perform the clean-up, a high percentage of raw reads (99.53% of the initial datasets) showed
 274 phred score $\geq$ Q20 which indicates the high quality of the initial dataset. In the end of this
 275 process, all clean the raw reads were deposited to the NCBI database and can be consulted
 276 under the BioProject accession: PRJNA625933 (Table. 1a).

277 **The *de novo* assembly and filtering**

278 To build the transcriptomic reference of *P. lividus*, we applied *de novo* assembler Trinity. Briefly,
 279 the six samples of *P. lividus* were pooled together and inputted in the Trinity assembler as a
 280 unique dataset. Next, the Trinity software selected about of 44,665,675 (21.91%) reads during
 281 the normalization stage and produced the first draft assembly of the transcriptome. Importantly,
 282 this version of the transcriptome assembly, was submitted to an extensive quality control
 283 against the NCBI-nt and UniVec databases, which allowed to remove sources of contaminations
 284 and to build the first version of the gonadal transcriptome of *P. lividus* (Assembly V0) (Table.
 285 1b). Unexpectedly, this transcriptome version transcriptome showed a huge number of
 286 transcripts (more than 1 M), a N50 transcript length of 720bp and a total length size of
 287 756,496,790bp. Biologically, these values can be explained by several factors. For instance,
 288 features such interspecific variation and heterozygosity can impact the building of *de novo*
 289 assembly references. On the other hand, high repeat content can confuse the *de novo*
 290 assemblers leading to high rates of fragmentation, spliced transcripts in many sub-transcripts,
 291 and production chimeric or artefactual transcripts (Lima et al. 2017). These characteristics were
 292 raised in the genome projects of the sea urchins, *Strongylocentrotus purpuratus*, *Hemicentrotus*

293 *pulcherrimus* and *Lytechinus variegatus*, in which the authors have found high levels of
294 heterozygosity as well as repeat content and several difficulties in the building of the genome
295 references. To remove the redundancy and mitigate the impact of these values in further
296 analyses, we next applied several filtration steps. These filters have used three main
297 approaches, codifying status, functional annotation and read coverage of the transcripts.

298 The codifying status of the transcripts was determined with the TransDecoder software. From
299 1,310,553 initial transcripts, only 138,992 transcripts codify to an ORF with 100 or more
300 aminoacids. The aminoacid and nucleotide sequences were then functional annotated using a
301 set of different databases (Fig 3a). After searches in several databases, the Uniref90, NCBI-nt
302 and NCBI-nr databases were selected to filter the transcriptome and about of 185,644
303 transcripts were found to have at least one blast hit in an echinoderm species. Thus, the
304 Annotated version of the transcriptome (Annotated Assembly V0), was obtained though the sum
305 of coding transcripts and transcripts with blast hit but without an ORF (225,700 transcripts -
306 138,992 protein coding transcripts; 86,708 additional transcripts with blast hit). Subsequently,
307 the 225,700 transcripts were inputted in the Corset software to remove all transcripts with
308 scarce evidences of read mapping and to cluster the remaining transcripts based on multi-
309 mappings. In the end, Corset grouped 141,138 transcripts (>10 reads mapping) in the 53,865
310 clusters / Unigenes- (Assembly V1). Both versions of the transcriptome assembly as well the
311 annotation reports and open read frames can be consulted in figshare digital repository (Link to
312 the reviewers - <https://figshare.com/s/e6d85154080ad008ee7c>).

313 Overall, this strategy has significantly reduced the size and the redundancy of the transcriptome
314 assembly, and is similar to previous analyses with sea urchins transcriptomes (Chen et al. 2015,
315 Gaitan-Espitia et al. 2016, Gaitán-Espitia and Hofmann 2017, Wong et al. 2019, Zhang et al.
316 2019, Zhang et al. 2019, Shi et al. 2020). The final version (Assembly V1) of *P. lividus*
317 transcriptome had only 4,11% of the initial number of transcripts (53,865), an N50 transcript
318 length of 1,842, more than the double of the first version, and 1/10 of the initial transcriptome
319 length (Table.1b). Importantly, both transcriptome versions (V0 and V1) were carefully
320 inspected, assessed and compared with three methods.

321 First, the gene content was analysed using the BUSCO tool. Searching the two libraries profiles,
322 Eukaryota and Metazoa, we observed a high completeness of the transcriptome assemblies.
323 While in V0, 99.4 and 98.1 % of the total gene groups were found in both databases, in the V1
324 we identified 97.4 and 95.6 % respectively. Importantly, all BUSCOs analyses showed a high
325 percentage (>92%) of genes classified as complete, and almost no gene content was lost
326 during the filtration process. On the other hand, the number of missing genes in both versions
327 per database is remarkably low (V0, Eukaryota - 0.6 %, Metazoa – 1.9%; V1, Eukaryota – 2.6
328 %, Metazoa – 4.4 %). Comparing these values with others in literature, we can conclude that *P.*
329 *lividus* gonadal transcriptome is one of the most complete in sea urchins reported to date
330 (Table.1b). For instance, in the sea urchins *Strongylocentrotus intermedius* (Zhan et al. 2019),

331 in *Echinometra* sp. embryos (Uthicke, Deshpande et al. 2019) or in giant red sea urchin
332 *Mesocentrotus franciscanus* (Gaitán-Espitia and Hofmann 2017), *de novo* assemblies achieved
333 82.5%, 91%, 88% of complete genes in eukaryotic database, respectively.

334 Second, we performed a structural comparison between V1 and V0 assemblies. The Fig 3b
335 showed a clear decrease in the percentage of short contigs (< 600 bp) from the V0 (43.7%) to
336 the V1 (22.4%) of the transcriptome. Consistently, in the remaining classes the V1 version
337 showed always higher percentages of transcripts than V0 version, indicating that small contigs
338 were largely affected by the filtration steps, while the percentages of the large contigs remaining
339 almost unmodified.

340 In the final test, the rate of back mapping reads to V0 and V1 transcriptomes was analysed. As
341 expected, the RBMT decrease 10,31% from the V0 to the V1 (Table.1b). Using the structural
342 and RBMT analyses it is possible to conclude that a part of the reads not mapped in the V1
343 version belongs to small and fragmented contigs removed during the filtration set. This results is
344 coherent with the biological features of sea urchins, that with high repetitive regions lead to high
345 percentages of fragmented and spliced transcripts in several sub-transcripts. Lastly, the top 10
346 Echinodermata species with highest number of transcripts matching the NCBI-nr database in
347 both V0 and V1 Assembly versions are shown in Fig.3c. As can be consulted, both versions
348 displayed a very similar distribution.

349 **Functional annotation**

350 The functional annotation was performed to the V0 version of the gonadal transcriptome and
351 thereafter gathered to the V1 version. Overall, the V0 version showed only 15.75 % (206,543) of
352 the total transcripts annotated against at least one database, and 6.29 % annotated in all
353 databases (Fig.3a; Table.1b). On the other hand, the V1 dataset of Unigenes presented a ratio
354 91.05% (49,048) of annotated transcripts (Table. 1b, Suppl. Table. 3). Once the remaining
355 analyses explored the V1 assembly. The following results of functional annotation will be focus
356 on this version. In KEGG, eggNOG, and Pfam databases we found between 33.17 and 39.61 %
357 of the Unigenes. On the other hand, in the general databases of NCBI-nt, NCBI-nr, UNIREF90
358 and Swiss-Prot were found 31,257; 34,584; 35,692 and 22,939 matches. The blast-x results
359 against the NCBI-nr database showed 34,537 match hits in 615 species. Of these, 30,129
360 (87.24%) Unigenes matched against species of Echinodermata, 597 against Cnidaria, 189
361 against Mollusca, 184 against Arthropoda and 3061 against others Phyla. Among the top 10
362 Echinodermata species with highest number of unigenes, the *S. purpuratus* species stand out
363 the remaining species, with 93.9 % of the hits (Fig.3c). Although several genomes and
364 transcriptomes of sea urchins are available (e.g. Sea Urchin Genome Sequencing 2006,
365 Cameron et al. 2015, Cary et al. 2018, Kinjo et al. 2018, Davidson et al. 2020), to the date of
366 this analyses only *S. purpuratus* species has a uniformized genome annotation, CDS and
367 proteins available in NCBI databases. In addition, the *P. lividus* and *S. purpuratus* are close

related species in terms of phylogeny, which makes this result expected (Mongiardino Koch et al. 2018). The remaining eight species of Fig.3c have between 2 and 0.1 % of the hits and comprise one sea star, *Acanthaster planci* (Class: Asteroidea), one sea cucumber, *Apostichopus japonicus* (Class: Holothuroidea) and seven other sea urchins (Class: Echinoidea).

Differential expression analyses

The differential gene expression analyses between the male and female samples were performed in the Degust platform. After submitting the matrix raw of counts to the platform, all Unigenes were filtered by the CPM (1 CPM in at least 3 samples). This filter allowed gathering only Unigenes with strong read support (17,581). Next, it was evaluated the variance across the samples using the multidimensional plot. Importantly, both female and male samples are clearly grouped in two independent clusters. Moreover, the female samples presented higher variance than male samples (Suppl. Fig. 1). The DGE analyses yielded a total of 6722 Unigenes with p-value < 0.05 and log2[fold change] ≥ 2. Overall, were found 3371 Unigenes up regulated in *P. lividus* males and 3351 up regulated in *P. lividus* females (Suppl. Fig. 2, Suppl. Table. 4). Moreover, 163 and 162 Unigenes were found to be female and male specific, respectively (Suppl. Table. 5, 6). Regarding the annotation status of the DGE Unigenes (DGEs), about 5278 DGEs codifying to an ORF, and 6267 have a match hit annotation in UNIPROT or NCBI databases.

Gene ontology analyses

The ontology analyses included all DEG and were mainly focused on the Biological Process category. These analyses showed most of the genes related with cellular process (35%), metabolic process (23.8%) and biological regulation (14.9%), while only a few directly involved in reproduction (0.8%) or reproductive process (0.8%) (Suppl. Fig.3). When the gene ontology annotation was determined per sex, once again most of the genes were related with cellular process (32.6% F vs 37.3% M), metabolic process (24% F vs 23.7% M) and biological regulation (15.8% F vs 14% M), while reproduction has once again a lower expression (0.3% F vs 1.2% M) (Suppl. Fig.4). Remarkably, the reproduction gene ontology term has 4 times more expressed genes in males than females. In terms of developmental process occurs the opposite, with more DGE present in females (4% F vs 1.8% M) (Suppl. Table 7; Suppl. Fig.4).

Next, to identify candidate genes involved in sex differentiation, determination and gonad development of *P. lividus*, the DGE genes were filtered using key GO terms and literature in the Echinodermata phylum. According to Panther classification, only 32 entries were associated to reproduction (GO:0000003) and reproductive process (GO:0022414) gene ontology terms. When the Trinotate report was scrutinized, a higher number of genes linked to gonad development and sex differentiation (GO:0008584, GO:0007548) were founded. The Spermatogenesis category (GO:0007283) was the most abundant category with 115 entries all

of them founded up-regulated in males, corresponding to the 57.21% of the total sex related genes founded in males (Suppl. Table 8). The second category with more entries was Spermatid development (GO:0007286) with 2 up-regulated genes in females and 49 in males. Then, oogenesis (GO:0048477) appeared with 7 entries in females and 10 in males. After that, males gonad development (GO:0008584) with 13 entries 6 in females and 7 in males; fertilization (GO:0009566) with 11 entries all of them in males; oocyte maturation (GO:0001556), 9 entries, 1 in females and 8 in males; males meiosis (GO:0007140), 5 entries, 1 in females and 4 in males; gonad development (GO:0008406), 4 entries, 2 in females and 2 in males; females gonad development (GO:0008585), 3 entries all in males and finally oocyte development (GO:0048599) with 2 entries, one in females and another in males. The highest number of entries were related with lipid metabolic process (GO:0006629) with 30 entries, 15 in females and 15 in males; then, lipid transport (GO:0006869) with 21 hits, 14 in females and 7 in males was the second most abundant group; cholesterol metabolic process (GO:0008203) with 13 hits, 7 in females and 6 in males; lipid glycosylation (GO:0030259) with 12 hits all in females and lipid storage (GO:0019915) with 10 hits, 5 in females and 5 in males were the most abundant groups. Additionally, some genes were involved in fatty acid elongation (GO:0030497) and fatty acid elongase process (GO:0009922) were found (Suppl. Table 9).

Initially, the total number of genes found in female and male gonads was similar. However, after filtering a higher number of genes related with sex and reproduction was detected in males (Fig.4a). Moreover, the DGE present in males was 3.28 times higher than in females (154 vs 47). In contrast, in lipid related genes a total of 119 genes were obtained, 48 hits of genes up-regulated in males and 68 in females. In this case, as expected, the highest number of genes involved in fatty acid production was founded in sea urchin females (Fig.4b).

Catalogue Sex and Lipid-related Unigenes

Unigenes involved in sexual development/determination

As previously described, sea urchins are diploid species with heteromorphic chromosome sex mechanisms of the XY genes (Lipani et al. 1996). A characteristic of *P. lividus* is the lower number of chromosomes ($2n=36$) in comparison with other echinoid species ($2n = 42$ to 44) (Lipani et al. 1996). Usually, sex determination in gonochoric species such as *P. lividus* involves a robust transcriptional regulation (Nef et al. 2005). Interestingly, in the majority of these species the male specimens have a higher number of genes being expressed in the gonads (Nef et al. 2005, Tao et al. 2013, Teaniniuraitemoana et al. 2014, Shen et al. 2020). Notwithstanding, there are some species where the opposite also occurs (González-Castellano et al. 2019, Piprek et al. 2019).

Sex differentiation includes several processes regulated by a high variety of genes and transcription factors. Among these genes, Sox genes are some of the well described in the literature as key factors for sexual determination. In *P. lividus* we found five Sox genes differentially expressed (Suppl. Table 8). While Sox30 (DN1680_c0_g1_i5) was only found up-

regulated in males, other Sox genes such as SoxB1, Sox21 and Sox4 (DN7136_c0_g1_i2; DN7136_c0_g2_i4; DN18728_c0_g1_i1) were uniquely found up-regulated in female gonads. Sox30 also known as Sex determination factor Y has been described as a key regulator of transcription in mouse spermatogenesis (Zhang et al. 2018), suggesting an important role in *P. lividus* testis development. On the other hand, the transcription factor SoxB1, one of the most expressed genes in *P. lividus* female gonads, is described as predominantly expressed in females and can have a function of oocyte protection and regulating the expression of ZP proteins (Yue et al. 2015). Despite, the finding of several Sox genes, other important genes such as Sox9, a major transcription factor for testis development (Zhang et al. 2018), was not found in *P. lividus*. Although unexpected, this result is coherent with the literature, where this and other important genes did not show differential gene expression in testis or ovaries of sea urchins' species (Sun et al. 2019). Moreover, in Echinodea class seems to exist different Sox genes being differentially expressed in gonads. Interestingly, in *Mesocentrotus nudus* species four different Sox genes such as Sox1, 3, 6 and 9 were found (Sun et al. 2019) and none of them was found in *P. lividus* transcriptome. These differences can be explained by the differential developmental stages of the gonads in the different studies (Yue et al. 2015).

The Dmrt1 gene (DN11344_c0_g1_i5), usually involved in testis development, was also found up-regulated in *P. lividus* testis. As in many other species, including sea urchins, Dmrt1 is described as playing a major role in sex determination and differentiation in several vertebrate and invertebrate species (Zhang and Zarkower 2017, Nagasawa et al. 2019, Sun et al. 2019). In accordance with the report of (Sun et al. 2019) in *M. nudus* species, also the Nanos (DN9673_c0_g1_i2) gene in *P. lividus* is up-regulated in female gonads. Although Nanos has a key role in embryonic development in sea urchin (Fujii et al. 2006), in adult ovaries has a pivotal role in proliferation and survival of germline stem cells and cyst development (Forbes and Lehmann 1998, De Keuckelaere et al. 2018). Additionally, several genes related with the Notch and Delta pathway (e.g. Neurogenic locus notch homolog protein 1 and 2 (DN2510_c0_g1_i1; DN17254_c0_g1_i4); notch gene (DN9493_c1_g2_i3), notch ligand (DN44109_c0_g2_i3) and Delta gene (DN44109_c0_g2_i3) were found in female gonads. While in adults this pathway is an essential regulator of cell proliferation during development and oogenesis (Feng et al. 2014, Irles et al. 2016, Sun et al. 2019, Zhang et al. 2019), in sea urchin embryos is related with the development process regulation (Materna and Davidson 2012). Another gene involved in the developing gonads and granulosa cells in adult mice is the GATA-type Zinc finger protein 1 (GLP-1) (Li et al. 2007, Pangas and Rajkovic 2015). This gene, not characterized in *S. purpuratus*, also was expressed in *P. lividus* females (DN31305_c0_g1_i2) and previously described in *M. nudus* (Sun et al. 2019). Importantly, this gene seems to have a crucial role in the regulation of oocyte meiosis and formation of primordial follicles, once the GLP-1 knock-out in both males and females mice results in infertility (Li et al. 2007).

Regarding the spermatogenesis regulation, several genes such as DMC1 (DN19900_c0_g1_i5), Spo11 (DN82908_c0_g1_i1), Hsd3 (DN1732_c1_g1_i2) and two SPATA genes, 24 and 45 (DN632_c1_g1_i2; DN847_c0_g1_i3), were found up-regulated in male

gonads. DMC1 is a recombinase essential for meiosis in male testis, regulating the homologous recombination of chromosomes. Depletion of DMC1 leads to malformation of sperm, or aneuploidy spermatocytes in several species (Chen et al. 2016). DMC1/LIM15 (DN19900_c0_g1_i5) from *S. purpuratus* were founded in testis of *P. lividus* indicating that as in other eukaryotes, the expression of DMC1 is essential for the development of the spermatocytes in sea urchin. Also, Spo11 is essential in meiotic recombination by generating double strand breaks and it is required for meiotic synapsis (Romanienko and Camerini-Otero 2000). Two genes of SPATA family were found in the transcriptome, SPATA24 and SPATA45. The spermatogenesis-associated (SPATA) family consists in several genes with critical roles in spermatogenesis. For instance, the knock-out of these genes in humans led to several problems in sperm motility, sperm production and germ cell development (Sujit et al. 2020). Finally, several testis-specific serine/threonine protein kinases (Tssk) (e.g tssk1, 4 and 5; DN1880_c1_g1_i12; DN20093_c0_g1_i3; DN2835_c2_g1_i6) were detected in the gonadal transcriptome, mainly associated to the male gonad. In general, these genes are associated with the spermatogenesis regulation and they are widespread across the eukaryota domain, including sea urchins (Wang et al. 2016, Sun et al. 2019).

Lipid Unigenes determination

The main components during gametogenesis in *P. lividus* female gonads are proteins, but lipids also play a remarkable role in the viability of gametes (Sanna et al. 2017). Lipids, such as phospholipids (PL) and cholesterol, are structural components of cell membranes and the key to somatic growth (Liu et al. 2007). It increases the synthesis and storage of lipids during gonadal development, reaching maximum values in stage IV and in males in stage III (final stage), mainly for energy reservoirs (Sanna et al. 2017).

Several genes involved in the biosynthesis of long-chain polyunsaturated fatty acid (LC-PUFA) during gametogenesis, such as putative fatty acid elongation protein 3 (up-regulated in males) (DN12877_c0_g1_i2), elongation of very long chain fatty acids protein 6 (DN6887_c0_g2_i2), or fatty acyl desaturases A (Fads A) (DN230613_c0_g1_i7), were identified in the *P. lividus* transcriptome. Fads A mediates the introduction of unsaturation (double bond) into a fatty acyl chain (Guillou et al. 2010), showing Δ^5 -desaturated activity in which it uses 20: 3n-6 and 20: 4n-3 to transform into ARA and EPA (Kabeya et al. 2017). LC-PUFA biosynthesis activity was also reported in other sea urchin's species such as *Strongylocentrotus intermedius* (Han et al. 2019). The gene expression of Fads A in the gonad is congruent with ARA and the EPA as the most abundant LC-PUFAs in the *P. lividus* gonad (Sanna et al. 2017). Different acyl-CoA ligase family members involved in long-chain-fatty-acid-CoA ligase (LC-PUFA-acid CoA ligase) such as LC-PUFA CoA ligase 1 and 6 (DN12809_c0_g1_i8 and DN79126_c1_g1_i1 respectively) were up-regulated in gonads of females while the LC-PUFA CoA ligase 3, 4 and 5 (DN3246_c0_g1_i1, DN3764_c0_g1_i13, and DN12660_c0_g1_i4) were found up-regulated in gonads of males (Suppl. Table. 9).

521 Other gene involved in the LC-PUFA incorporation into the mitochondria, mainly through the
522 hydrolysis of triglycerides (García-Rincón et al. 2016), was carnitine O-palmitoyltransferase 2
523 (DN25985_c0_g2_i1). This Unigene have showed higher expression in female gonads than in
524 males of *P. lividus*. Consistently, this gene showed crucial activity in transcriptome analysis for
525 fatty acid metabolism of other sea urchins such as *Strongylocentrotus intermedius* (Wang et al.
526 2019).

527 Furthermore, PUFAs play a fundamental role as the main components of complex molecules
528 such as phospholipids or triglycerides. An example of intermediate formation products is found
529 in the sn1-specific diacylglycerol lipase beta that catalyzes the hydrolysis of arachidonic acid
530 (AA) esterified diacylglycerol (DAG) to produce the main endocannabinoid, 2-
531 arachidonoylglycerol (2-AG), which can be further cleaved by downstream enzymes to release
532 arachidonic acid (AA) for cyclooxygenase (COX) mediated eicosanoid production. Also, DAG is
533 an intermediate in the glycerolipid and glycerophospholipids metabolism for the biosynthesis of
534 TAG and PL. In male gonads, high expression of sn1-specific diacylglycerol lipase beta
535 (DN17156_c0_g2_i1) was showed.

536 Phospholipids are part of the cell membrane and essential for cell growth (Byrd 1975). Most of
537 the phospholipids are incorporated through the diet, but *P. lividus* has the capacity for its
538 endogenous biosynthesis (Byrd 1975). Thus, phosphatidylcholine (PC) is one of the major
539 phospholipids in the sea urchin (Mita et al. 1994). Importantly, the Kennedy pathway route is
540 used for phospholipids (Gibellini and Smith 2010). Cholinephosphotransferase 1 (chpt1) is the
541 last enzyme in charge of obtaining PC from 1,2-diacyl-sn-glycerol and CDP-choline (EC
542 2.7.8.2). Differential expression analyses showed how cholinephosphotransferase 1
543 (DN72675_c0_g2_i1) was up-regulated in female gonads as an active pathway in the last PC
544 biosynthesis step. In contrats, a significantly lower chpt1 activity was identified in oocytes of
545 *Arbacia punctulata* compared to sperm (Ewing 1973). Furthermore, this study results showed a
546 higher expression of lysophosphatidylcholine acyltransferase 2 (DN14424_c0_g1_i10) in male
547 gonads, as an alternative route for the biosynthesis of PC as a final product. The enzyme
548 encoded by lysophosphatidylcholine acyltransferase 2 plays a role in phospholipid metabolism,
549 specifically in the conversion of lysophosphatidylcholine to phosphatidylcholine in the presence
550 of acyl-CoA (Law et al. 2019). The alternative pathways to obtaining PC between males and
551 females may be related to the composition of PUFA and LC-PUFA in positions sn-2 and sn-1,3.
552 Thus, in females of *P. lividus* gonads, the sn-1,3 position of LC-PUFA such as ARA or EPA
553 determines the use of TAGs in response to acclimatization or involvement in reproduction,
554 respectively (Sanna et al. 2017). Also, the male gametes of the sea urchin use PC and TAG as
555 energy (Mita et al. 1994), while in the gonad, PC is incorporated into the cell structure (Ewing
556 1973).

557 Several genes involved in TAG biosynthesis, such as diacylglycerol kinase zeta
558 (DN39457_c0_g1_i2), nuclear envelope phosphatase-regulatory subunit 1
559 (DN10024_c0_g1_i6), microsomal triglyceride transfer protein large subunit

(DN3378_c1_g1_i6), are presented in gonadal transcriptome. Most of the genes related to triglycerides' metabolism were up-regulated in females gonadal concerning to the males. Thus, TAGs are the highest lipid class in gonads in *P. lividus*, varying seasonally due to lipid turnover after each period (Sanna et al. 2017). Similarly, during the egg development of the sea urchin, *Anthocidaris crassispinu*, TAGs are stored and later used in the early stages of development (YASUMASU, HINO et al. 1984). In contrast, in spermatozoa, the TAGs proportion is low or maybe absent (Mita et al. 1994). This is consistent with the higher activity of triglycerides genes related to gonadal development in *P. lividus* female than male gonads.

Furthermore, several genes related to cholesterol metabolism, transport, and binding such as scavenger receptor class B member 1 (DN1029_c0_g1_i3), NPC intracellular cholesterol transporter 1 (DN7961_c0_g1_i6), low-density lipoprotein (LDL) receptor-related protein 6 (DN74960_c0_g2_i3), and Apolipoprotein D (DN8923_c0_g2) were identified (Suppl. Table 9). A higher expression of these genes was found in female gonads compared with males, with related functions as a membrane receptor, transport, and binding of high-density lipoprotein cholesterol (HDL), low-density lipoprotein (LDL), and cholesterol. Moreover, genes such as sterol O-acyltransferase 1 have increased activity in male gonads. This enzyme plays a role in lipoprotein assembly and dietary cholesterol absorption (Yang et al. 1997). Thus, in female ovary and *P. lividus* spermatozoa, cholesterol was identified as a relevant lipid component (Mita et al. 1994, López-Hernández et al. 1999).

Phylogenetic validation

One of the biggest challenges in gene ortholog identification is related to alterations in gene repertoire among species, as a consequence of loss, duplication, or the appearance of paralogs genes (Altenhoff and Dessimoz 2009). Here, we phylogenetically validated 12 relevant target genes annotated in our transcriptome, with their orthologs full-length aa sequences of other species of echinoderms sea urchin (*Strongylocentrotus intermedius*), starfish (*Asterias rubens*) and, sea cucumbers (*Apostichopus japonicus*), vertebrate model species fish (*Danio rerio*), amphibians (*Xenopus tropicalis*), birds (*Gallus gallus*), mammals (*Mus musculus* and *Homo sapiens*) and mollusks (*Crassostrea gigas*). This is a powerful method to detect annotation and assembly errors caused during the transcriptome assembly for non-model species (Guang et al. 2021). Genes with functional characterization, also validated and described in the literature for echinoderms were included in the analysis such as Fads, PPARs and Gonadotropin-Releasing Hormone-Type (Kabeya et al. 2017, Tian et al. 2017, Capitão et al. 2020). Our results showed that, all target lipid and sex-related full-length aa sequences clustered into a strongly supported clade of echinoderms with the orthologs of sea urchin (*Strongylocentrotus intermedius*), starfish (*Asterias rubens*), and sea cucumbers (*Apostichopus japonicus*) (Suppl. Fig. 5). A similar topology was identified in almost all trees (Suppl. Fig. 5), one group of vertebrates, other from echinoderms, and the last one of molluscs (*Crassostrea gigas*). As expected genes involved in lipid metabolism such as Fads A and chpt1 were clustered in the echinoderm clade with 99% posterior probability support and 100% with the other sea urchin (*Strongylocentrotus*

599 *intermedius*) (Suppl.Table 5). The study of Kabeya et al. (2017) performed the functional
600 characterization of *P. lividus* $\Delta 5$ desaturase (Fads A) and validated by phylogenetic analysis
601 using molluscs, other sea urchins (*Strongylocentrotus purpuratus* and *Lytechinus variegatus*),
602 and other species of marine echinoderms such as lilies (*Oxycomanthus japonicus*), sea
603 cucumbers (*Sclerodactyla briareus*) and starfish (*Patiria miniata*). Supported by the results of
604 Kabeya et al. (2017), we were able to obtain a similar result in our phylogenetic analysis that
605 validates the functional annotation of key target genes in our transcriptome.

606 **Conclusion**

607 To conclude, the gonad transcriptome from both sexes of sea urchins lead to the identification
608 of a larger number of genes involved in sex determination, sex differentiation, gonad
609 development, spermatogenesis and oogenesis and also in several pathways of lipid
610 metabolism, fatty acid elongase process, lipid storage and cholesterol metabolic process. The
611 results obtained in this work will provide useful data on *P. lividus* sea urchin gonad tissue and it
612 will contribute in the future in the conservation of the species and the exploitation for aquaculture
613 purposes.

614 **Acknowledgements**

615 This work was supported by the project CRAGIAMP – PTDC/BIA-BQM/30232/2017, co-
616 financed by COMPETE 2020, Portugal 2020 and the European Union through the ERDF. This
617 research was also supported by national funds “through FCT – Foundation for Science and
618 Technology” within the scope of UIDB/0443/2020 and UIDP/04423/2020.

620 **Author Contributions**

621 AMM: Data curation, Formal analysis, Investigation, Methodology, Visualization and Writing—
622 original draft. SF-B: Data curation, Formal analysis, Investigation, Methodology, Visualization,
623 Resources and Writing—original draft. MN: Formal analysis, Investigation, Methodology,
624 Visualization and Writing - review & editing. RP: Investigation, Methodology, Visualization and
625 Writing - review & editing. BC: Methodology, Validation and Writing—review & editing. LFCC:
626 Conceptualization, Methodology, Validation, Supervision, Project administration, Resources and
627 Writing—review & editing.

628

629 References

- 630 Altenhoff, A. M. and C. Dessimoz (2009). Phylogenetic and Functional Assessment of Orthologs
631 Inference Projects and Methods. *PLOS Computational Biology* **5**(1): e1000262.
- 632 Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman
633 (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search
634 programs. *Nucleic Acids Research* **25**(17): 3389-3402.
- 635 Babicki, S., D. Arndt, A. Marcu, Y. Liang, J. R. Grant, A. Maciejewski and D. S. Wishart (2016).
636 Heatmapper: web-enabled heat mapping for all. *Nucleic acids research* **44**(W1): W147-
637 W153.
- 638 Baião, L. F., F. Rocha, M. Costa, T. Sá, A. Oliveira, M. R. G. Maia, A. J. M. Fonseca, M. Pintado
639 and L. M. P. Valente (2019). Effect of protein and lipid levels in diets for adult sea urchin
640 *Paracentrotus lividus* (Lamarck, 1816). *Aquaculture* **506**: 127-138.
- 641 Bertocci, I., A. Blanco, J. N. Franco, S. Fernandez-Boo and F. Arenas (2018). Short-term
642 variation of abundance of the purple sea urchin, *Paracentrotus lividus* (Lamarck, 1816),
643 subject to harvesting in northern Portugal. *Mar Environ Res* **141**: 247-254.
- 644 Bertocci, I., R. Dominguez, C. Freitas and I. Sousa-Pinto (2012). Patterns of variation of
645 intertidal species of commercial interest in the Parque Litoral Norte (north Portugal) MPA:
646 Comparison with three reference shores. *Marine Environmental Research* **77**: 60-70.
- 647 Bolger, A. M., M. Lohse and B. Usadel (2014). Trimmomatic: a flexible trimmer for Illumina
648 sequence data. *Bioinformatics (Oxford, England)* **30**(15): 2114-2120.
- 649 Böttger, S. A., C. C. Eno and C. W. Walker (2011). Methods for generating triploid green sea
650 urchin embryos: An initial step in producing triploid adults for land-based and near-shore
651 aquaculture. *Aquaculture* **318**(1): 199-206.
- 652 Bryant, D. M., K. Johnson, T. DiTommaso, T. Tickle, M. B. Couger, D. Payzin-Dogru, T. J. Lee,
653 N. D. Leigh, T.-H. Kuo, F. G. Davis, J. Bateman, S. Bryant, A. R. Guzikowski, S. L. Tsai, S.
654 Coyne, W. W. Ye, R. M. Freeman, Jr., L. Peshkin, C. J. Tabin, A. Regev, B. J. Haas and J.
655 L. Whited (2017). A Tissue-Mapped Axolotl De Novo Transcriptome Enables Identification
656 of Limb Regeneration Factors. *Cell reports* **18**(3): 762-776.
- 657 Buchfink, B., C. Xie and D. H. Huson (2015). Fast and sensitive protein alignment using
658 DIAMOND. *Nat Methods* **12**(1): 59-60.
- 659 Byrd, E. W. (1975). Phospholipid metabolism following fertilization in sea urchin eggs and
660 embryos. *Developmental Biology* **46**(2): 309-316.
- 661 Byrne, M. (1990). Annual reproductive cycles of the commercial sea urchin *Paracentrotus*
662 *lividus* from an exposed intertidal and a sheltered subtidal habitat on the west coast of
663 Ireland. *Marine Biology* **104**(2): 275-289.
- 664 Cameron, R. A., P. Kudtarkar, S. M. Gordon, K. C. Worley and R. A. Gibbs (2015). Do
665 echinoderm genomes measure up? *Mar Genomics* **22**: 1-9.
- 666 Capitão, A., M. Lopes-Marques, I. Páscoa, R. Ruivo, N. Mendiratta, E. Fonseca, L. Castro and
667 M. M. J. E. p. Santos (2020). The Echinodermata PPAR: Functional characterization and
668 exploitation by the model lipid homeostasis regulator tributyltin. **263 Pt B**: 114467.
- 669 Cary, G. A., R. A. Cameron and V. F. Hinman (2018). EchinoBase: Tools for Echinoderm
670 Genome Analyses. *Methods Mol Biol* **1757**: 349-369.
- 671 Chassé, H., J. Aubert, S. Boulben, G. Le Corguillé, E. Corre, P. Cormier and J. Morales (2018).
672 Translatome analysis at the egg-to-embryo transition in sea urchin. *Nucleic Acids*
673 *Research* **46**(9): 4607-4621.
- 674 Chen, J., X. Cui, S. Jia, D. Luo, M. Cao, Y. Zhang, H. Hu, K. Huang, Z. Zhu and W. Hu (2016).
675 Disruption of dmc1 Produces Abnormal Sperm in Medaka (*Oryzias latipes*). *Scientific*
676 *Reports* **6**: 30912.
- 677 Chen, Y., Y. Chang, X. Wang, X. Qiu and Y. Liu (2015). De novo assembly and analysis of
678 tissue-specific transcriptomes revealed the tissue-specific genes and profile of immunity
679 from *Strongylocentrotus intermedius*. *Fish and Shellfish Immunology* **46**(2): 723-736.

Davidson, N. M. and A. Oshlack (2014). Corset: enabling differential gene expression analysis for de novo assembled transcriptomes. *Genome Biology* **15**(7): 410.

Davidson, P. L., H. Guo, L. Wang, A. Berrio, H. Zhang, Y. Chang, A. L. Soborowski, D. R. McClay, G. Fan and G. A. Wray (2020). Chromosomal-Level Genome Assembly of the Sea Urchin *Lytechinus variegatus* Substantially Improves Functional Genomic Analyses. *Genome Biology and Evolution* **12**(7): 1080-1086.

De Keuckelaere, E., P. Hulpiau, Y. Saey, G. Berx and F. van Roy (2018). Nanos genes and their role in development and beyond. *Cellular and Molecular Life Sciences* **75**(11): 1929-1946.

Di Natale, M., C. Bennici, G. Biondo, T. Masullo, C. Monastero, M. Tagliavia, M. Torri, S. Costa, M. A. Ragusa, A. Cuttitta and A. Nicosia (2019). Aberrant gene expression profiles in Mediterranean sea urchin reproductive tissues after metal exposures. *Chemosphere* **216**: 48-58.

Evans, T. G., M. H. Pespeni, G. E. Hofmann, S. R. Palumbi and E. Sanford (2017). Transcriptomic responses to seawater acidification among sea urchin populations inhabiting a natural pH mosaic. *Molecular Ecology* **26**(8): 2257-2275.

Ewing, R. D. (1973). Cholinephosphotransferase activity during early development of the sea urchin, *Arbacia punctulata*. *Developmental Biology* **31**(2): 234-241.

Feng, Y.-M., G.-J. Liang, B. Pan, X. Qin, X.-F. Zhang, C.-L. Chen, L. Li, S.-F. Cheng, M. De Felici and W. Shen (2014). Notch pathway regulates female germ cell meiosis progression and early oogenesis events in fetal mouse. *Cell Cycle* **13**(5): 782-791.

Fernández-Boán, M., L. Fernández and J. Freire (2012). History and management strategies of the sea urchin *Paracentrotus lividus* fishery in Galicia (NW Spain). *Ocean & Coastal Management* **69**: 265-272.

Fernández-Boo, S., M. H. Pedrosa-Oliveira, A. Afonso, F. Arenas, F. Rocha, L. M. P. Valente and B. Costas (2018). Annual assessment of the sea urchin (*Paracentrotus lividus*) humoral innate immune status: Tales from the north Portuguese coast. *Marine Environmental Research* **141**: 128-137.

Forbes, A. and R. Lehmann (1998). Nanos and Pumilio have critical roles in the development and function of *Drosophila* germline stem cells. *Development* **125**(4): 679-690.

Fujii, T., K. Mitsunaga-Nakatsubo, I. Saito, H. Iida, N. Sakamoto, K. Akasaka and T. Yamamoto (2006). Developmental expression of HpNanos, the *Hemicentrotus pulcherrimus* homologue of nanos. *Gene Expression Patterns* **6**(5): 572-577.

Gaitán-Espitia, J. D. and G. E. Hofmann (2017). Gene expression profiling during the embryo-to-larva transition in the giant red sea urchin *Mesocentrotus franciscanus*. *Ecology and Evolution* **7**(8): 2798-2811.

Gaitán-Espitia, J. D., R. Sanchez, P. Bruning and L. Cardenas (2016). Functional insights into the testis transcriptome of the edible sea urchin *Loxechinus albus*. *Scientific Reports* **6**.

Galasso, C., S. D'Aniello, C. Sansone, A. Ianora and G. Romano (2019). Identification of Cell Death Genes in Sea Urchin *Paracentrotus lividus* and Their Expression Patterns during Embryonic Development. *Genome Biology and Evolution* **11**(2): 586-596.

García-Rincón, J., A. Darszon and C. Beltrán (2016). Speract, a sea urchin egg peptide that regulates sperm motility, also stimulates sperm mitochondrial metabolism. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1857**(4): 415-426.

Garmendia, J. M., I. Menchaca, M. J. Belzunce, J. Franco and M. Revilla (2010). Seasonal variability in gonad development in the sea urchin (*Paracentrotus lividus*) on the Basque coast (Southeastern Bay of Biscay). *Marine Pollution Bulletin* **61**(4-6): 259-266.

Ghisaura, S., B. Loi, G. Biosa, M. Baroli, D. Pagnozzi, T. Roggio, S. Uzzau, R. Anedda and M. F. Addis (2016). Proteomic changes occurring along gonad maturation in the edible sea urchin *Paracentrotus lividus*. *Journal of Proteomics* **144**: 63-72.

Gibellini, F. and T. K. Smith (2010). The Kennedy pathway-De novo synthesis of phosphatidylethanolamine and phosphatidylcholine. *IUBMB Life* **62**(6): 414-428.

Gildor, T., A. Malik, N. Sher, L. Avraham and S. Ben-Tabou de-Leon (2016). Quantitative developmental transcriptomes of the Mediterranean sea urchin *Paracentrotus lividus*. *Marine Genomics* **25**: 89-94.

González-Castellano, I., C. Manfrin, A. Pallavicini and A. Martínez-Lage (2019). De novo gonad transcriptome analysis of the common littoral shrimp *Palaemon serratus*: novel insights into sex-related genes. *BMC Genomics* **20**(1): 757.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman and A. Regev (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**(7): 644-652.

Guang, A., M. Howison, F. Zapata, C. Lawrence and C. W. Dunn (2021). Revising transcriptome assemblies with phylogenetic information. *PLOS ONE* **16**(1): e0244202.

Guillou, H., D. Zadavec, P. G. Martin and A. Jacobsson (2010). The key roles of elongases and desaturases in mammalian fatty acid metabolism: Insights from transgenic mice. *Progress in Lipid Research* **49**(2): 186-199.

Guindon, S., J.-F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk and O. Gascuel (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology* **59**(3): 307-321.

Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, D. Eccles, B. Li, M. Lieber, M. D. MacManes, M. Ott, J. Orvis, N. Pochet, F. Strozzi, N. Weeks, R. Westerman, T. William, C. N. Dewey, R. Henschel, R. D. LeDuc, N. Friedman and A. Regev (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* **8**(8): 1494-1512.

Han, L., J. Ding, H. Wang, R. Zuo, Z. Quan, Z. Fan, Q. Liu and Y. Chang (2019). Molecular characterization and expression of SiFad1 in the sea urchin (*Strongylocentrotus intermedius*). *Gene* **705**: 133-141.

Huson, D. H., D. C. Richter, C. Rausch, T. DeZulian, M. Franz and R. Rupp (2007). Dendroscope: An interactive viewer for large phylogenetic trees. *BMC Bioinformatics* **8**(1): 460.

Irles, P., N. Elshaer and M.-D. Piulachs (2016). The Notch pathway regulates both the proliferation and differentiation of follicular cells in the panoistic ovary of *Blattella germanica*. *Open Biology* **6**(1): 150197.

James, P., S. I. Siikavuopio and A. Mortensen (2015). Sea Urchin Aquaculture in Norway. *Echinoderm Aquaculture*: 147-173.

Janies, D. A., Z. Witter, G. V. Linchangco, D. W. Foltz, A. K. Miller, A. M. Kerr, J. Jay, R. W. Reid and G. A. Wray (2016). EchinoDB, an application for comparative transcriptomics of deeply-sampled clades of echinoderms. *BMC Bioinformatics* **17**: 48.

Jia, Z., Q. Wang, K. Wu, Z. Wei, Z. Zhou and X. Liu (2017). De novo transcriptome sequencing and comparative analysis to discover genes involved in ovarian maturity in *Strongylocentrotus nudus*. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics* **23**: 27-38.

Kabeya, N., A. Sanz-Jorquera, S. Carboni, A. Davie, A. Oboh and O. Monroig (2017). Biosynthesis of Polyunsaturated Fatty Acids in Sea Urchins: Molecular and Functional Characterisation of Three Fatty Acyl Desaturases from *Paracentrotus lividus* (Lamarck 1816). *PLOS ONE* **12**(1): e0169374.

Kanehisa, M. and S. Goto (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **28**(1): 27-30.

Katoh, K., J. Rozewicki and K. D. Yamada (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinformatics* **20**(4): 1160-1166.

783 Katoh, K. and D. M. Standley (2013). MAFFT multiple sequence alignment software version 7:
784 improvements in performance and usability. *Molecular Biology and Evolution* **30**(4): 772-
785 780.

786 Kim, D., L. Song, F. P. Breitwieser and S. L. Salzberg (2016). Centrifuge: rapid and sensitive
787 classification of metagenomic sequences. *Genome Research*.

788 Kinjo, S., M. Kiyomoto, T. Yamamoto, K. Ikeo and S. Yaguchi (2018). HpBase: A genome
789 database of a sea urchin, *Hemicentrotus pulcherrimus*. *Development, Growth &*
790 *Differentiation* **60**(3): 174-182.

791 Lang-Unnash, N. (1992). Purification and properties of *Plasmodium falciparum* malate
792 dehydrogenase. *Molecular and Biochemical Parasitology* **50**: 17-26.

793 Langmead, B. and S. L. Salzberg (2012). Fast gapped-read alignment with Bowtie 2. *Nature*
794 *Methods* **9**(4): 357-359.

795 Laruson, A. J., S. E. Coppard, M. H. Pespeni and F. A. Reed (2018). Gene expression across
796 tissues, sex, and life stages in the sea urchin *Tripneustes gratilla* [Toxopneustidae,
797 Odontophora, Camarodonta]. *Marine Genomics* **41**: 12-18.

798 Law, S.-H., M.-L. Chan, G. K. Marathe, F. Parveen, C.-H. Chen and L.-Y. Ke (2019). An
799 Updated Review of Lysophosphatidylcholine Metabolism in Human Diseases. *International*
800 *Journal of Molecular Sciences* **20**(5): 1149.

801 Li, S., M. M. Lu, D. Zhou, S. R. Hammes and E. E. Morrissey (2007). GLP-1: A novel zinc finger
802 protein required in somatic cells of the gonad for germ cell development. *Developmental*
803 *Biology* **301**(1): 106-116.

804 Lima, L., B. Sinimeri, G. Sacomoto, H. Lopez-Maestre, C. Marchet, V. Miele, M.-F. Sagot and
805 V. Lacroix (2017). Playing hide and seek with repeats in local and global de novo
806 transcriptome assembly of short RNA-seq reads. *Algorithms for molecular biology : AMB*
807 **12**: 2-2.

808 Lipani, C., R. Vitturi, G. Sconzo and G. Barbata (1996). Karyotype analysis of the sea urchin
809 *Paracentrotus lividus* (Echinodermata): evidence for a heteromorphic chromosome sex
810 mechanism. *Marine Biology* **127**(1): 67-72.

811 Liu, H. and Y.-q. Chang (2015). Sea Urchin Aquaculture in China. *Echinoderm Aquaculture*:
812 127-146.

813 Liu, H., M. S. Kelly, E. J. Cook, K. Black, H. Orr, J. X. Zhu and S. L. Dong (2007). The effect of
814 diet type on growth and fatty-acid composition of sea urchin larvae, I. *Paracentrotus lividus*
815 (Lamarck, 1816) (Echinodermata). *Aquaculture* **264**(1): 247-262.

816 Liu, X.-l., Y.-q. Chang, J.-h. Xiang and X.-b. Cao (2005). Estimates of genetic parameters for
817 growth traits of the sea urchin, *Strongylocentrotus intermedius*. *Aquaculture* **243**(1): 27-32.

818 López-Hernández, J., M. J. González-Castro and M. Piñeiro-Sotelo (1999). Determination of
819 Sterols in Sea Urchin Gonads by High-Performance Liquid Chromatography With
820 Ultraviolet Detection. *Journal of Chromatographic Science* **37**(7): 237-239.

821 Machado, I., P. Moura, F. Pereira, P. Vasconcelos and M. B. Gaspar (2019). Reproductive cycle
822 of the commercially harvested sea urchin (*Paracentrotus lividus*) along the western coast of
823 Portugal. *Invertebrate Biology* **138**(1): 40-54.

824 Materna, S. C. and E. H. Davidson (2012). A comprehensive analysis of Delta signaling in pre-
825 gastrular sea urchin embryos. *Developmental Biology* **364**(1): 77-87.

826 Mi, H., A. Muruganujan, X. Huang, D. Ebert, C. Mills, X. Guo and P. D. Thomas (2019). Protocol
827 Update for large-scale genome and gene function analysis with the PANTHER
828 classification system (v.14.0). *Nature Protocols* **14**(3): 703-721.

829 Mita, M., A. Oguchi, S. Kikuyama, I. Yasumasu, R. De Santis and M. Nakamura (1994).
830 Endogenous Substrates for Energy Metabolism in Spermatozoa of the Sea Urchins
831 *Arbacia lixula* and *Paracentrotus lividus*. *Biological Bulletin* **186**(3): 285-290.

832 Mongiardino Koch, N., S. E. Coppard, H. A. Lessios, D. E. G. Briggs, R. Mooi and G. W. Rouse
833 (2018). A phylogenomic resolution of the sea urchin tree of life. *BMC Evolutionary Biology*
834 **18**(1): 189.

835 Morroni, L., D. Sartori, M. Costantini, L. Genovesi, T. Magliocco, N. Ruocco and I. Buttino
836 (2019). First molecular evidence of the toxicogenetic effects of copper on sea urchin
837 *Paracentrotus lividus* embryo development. *Water Research* **160**: 415-423.

838 Mos, B., M. Byrne and S. A. Dworjanyan (2016). Biogenic acidification reduces sea urchin gonad
839 growth and increases susceptibility of aquaculture to ocean acidification. *Marine*
840 *Environmental Research* **113**: 39-48.

841 Nagasawa, K., T. Thitiphuree and M. Osada (2019). Phenotypic Stability of Sex and Expression
842 of Sex Identification Markers in the Adult Yesso Scallop *Mizuhopecten yessoensis*
843 throughout the Reproductive Cycle (vol 9, 277, 2019). *Animals* **9**(9).

844 Nef, S., O. Schaad, N. R. Stallings, C. R. Cederroth, J. L. Pitetti, G. Schaer, S. Malki, M.
845 Dubois-Dauphin, B. Boizet-Bonhoure, P. Descombes, K. L. Parker and J. D. Vassalli
846 (2005). Gene expression during sex determination reveals a robust female genetic
847 program at the onset of ovarian development. *Dev Biol* **287**(2): 361-377.

848 Ohgaki, S.-I., T. Kato, N. Kobayashi, H. Tanase, N. H. Kumagai, S. Ishida, T. Nakano, Y. Wada
849 and Y. Yusa (2019). Effects of temperature and red tides on sea urchin abundance and
850 species richness over 45 years in southern Japan. *Ecological Indicators* **96**: 684-693.

851 Ouréns, R., I. Naya and J. Freire (2015). Mismatch between biological, exploitation, and
852 governance scales and ineffective management of sea urchin (*Paracentrotus lividus*)
853 fisheries in Galicia. *Marine Policy* **51**: 13-20.

854 Pangas, S. A. and A. Rajkovic (2015). Chapter 21 - Follicular Development: Mouse, Sheep, and
855 Human Models. *Knobil and Neill's Physiology of Reproduction (Fourth Edition)*. T. M. Plant
856 and A. J. Zeleznik. San Diego, Academic Press: 947-995.

857 Phillips, K., P. Bremer, P. Silcock, N. Hamid, C. Delahunty, M. Barker and J. Kissick (2009).
858 Effect of gender, diet and storage time on the physical properties and sensory quality of
859 sea urchin (*Evechinus chloroticus*) gonads. *Aquaculture* **288**(3): 205-215.

860 Phillips, K., J. Niimi, N. Hamid, P. Silcock, C. Delahunty, M. Barker, M. Sewell and P. Bremer
861 (2010). Sensory and volatile analysis of sea urchin roe from different geographical regions
862 in New Zealand. *LWT - Food Science and Technology* **43**(2): 202-213.

863 Pipek, R. P., M. Damulewicz, J.-P. Tassan, M. Kloc and J. Z. Kubiak (2019). Transcriptome
864 profiling reveals male- and female-specific gene expression pattern and novel gene
865 candidates for the control of sex determination and gonad development in *Xenopus laevis*.
866 *Development Genes and Evolution* **229**(2): 53-72.

867 Pjeta, R., H. Lindner, L. Kremser, W. Salvenmoser, D. Sobral, P. Ladurner and R. Santos
868 (2020). Integrative Transcriptome and Proteome Analysis of the Tube Foot and Adhesive
869 Secretions of the Sea Urchin *Paracentrotus lividus*. *International Journal of Molecular*
870 *Sciences* **21**(3): 946.

871 Powell, D. (2019). drpowell/degust 4.1.1. Zenodo.

872 Powell, S., D. Szklarczyk, K. Trachana, A. Roth, M. Kuhn, J. Muller, R. Arnold, T. Rattei, I.
873 Letunic, T. Doerks, L. J. Jensen, C. von Mering and P. Bork (2011). eggNOG v3.0:
874 orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic*
875 *Acids Research* **40**(D1): D284-D289.

876 Prato, E., G. Fanelli, A. Angioni, F. Biandolino, I. Parlapiano, L. Papa, G. Denti, M. Secci, M.
877 Chiantore, M. S. Kelly, M. P. Ferranti and P. Addis (2018). Influence of a prepared diet and
878 a macroalga (*Ulva* sp.) on the growth, nutritional and sensory qualities of gonads of the sea
879 urchin *Paracentrotus lividus*. *Aquaculture* **493**: 240-250.

880 Punta, M., P. C. Coghill, R. Y. Eberhardt, J. Mistry, J. Tate, C. Boursnell, N. Pang, K. Forslund,
881 G. Ceric, J. Clements, A. Heger, L. Holm, E. L. Sonnhammer, S. R. Eddy, A. Bateman and
882 R. D. Finn (2012). The Pfam protein families database. *Nucleic Acids Res* **40**(Database
883 issue): D290-301.

884 Robinson, M. D., D. J. McCarthy and G. K. Smyth (2009). edgeR: a Bioconductor package for
885 differential expression analysis of digital gene expression data. *Bioinformatics* **26**(1): 139-
886 140.

887 Robinson, M. D. and A. Oshlack (2010). A scaling normalization method for differential
888 expression analysis of RNA-seq data. *Genome Biology* **11**(3): R25.

889 Rocha, F., A. C. Rocha, L. F. Baiao, J. Gadelha, C. Camacho, M. L. Carvalho, F. Arenas, A.
890 Oliveira, M. R. G. Maia, A. R. Cabrita, M. Pintado, M. L. Nunes, C. M. R. Almeida and L. M.
891 P. Valente (2019). Seasonal effect in nutritional quality and safety of the wild sea urchin
892 *Paracentrotus lividus* harvested in the European Atlantic shores. *Food Chemistry* **282**: 84-
893 94.

894 Romanienko, P. J. and R. D. Camerini-Otero (2000). The Mouse Spo11 Gene Is Required for
895 Meiotic Chromosome Synapsis. *Molecular Cell* **6**(5): 975-987.

896 Ruocco, N., M. Costantini and L. Santella (2016). New insights into negative effects of lithium
897 on sea urchin *Paracentrotus lividus* embryos. *Scientific Reports* **6**: 32157.

898 Sanna, R., S. Siliani, R. Melis, B. Loi, M. Baroli, T. Roggio, S. Uzzau and R. Anedda (2017).
899 The role of fatty acids and triglycerides in the gonads of *Paracentrotus lividus* from
900 Sardinia: Growth, reproduction and cold acclimatization. *Marine and Environmental*
901 *Research* **130**: 113-121.

902 Sea Urchin Genome Sequencing, C. (2006). The Genome of the Sea Urchin *Strongylocentrotus*
903 *purpuratus*. *Science (New York, N.Y.)* **314**(5801): 941-952.

904 Sellem, F. and M. Guillou (2007). Reproductive biology of *Paracentrotus lividus*
905 (Echinodermata: Echinoidea) in two contrasting habitats of northern Tunisia (south-east
906 Mediterranean). *Journal of the Marine Biological Association of the United Kingdom* **87**(3):
907 763-767.

908 Shen, F., Y. Long, F. Li, G. Ge, G. Song, Q. Li, Z. Qiao and Z. Cui (2020). *De novo*
909 transcriptome assembly and sex-biased gene expression in the gonads of Amur catfish
910 (*Silurus asotus*). *Genomics* **112**(3): 2603-2614.

911 Shi, D., C. Zhao, Y. Chen, J. Ding, L. Zhang and Y. Chang (2020). Transcriptomes shed light on
912 transgenerational and developmental effects of ocean warming on embryos of the sea
913 urchin *Strongylocentrotus intermedius*. *Scientific Reports* **10**(1): 7931.

914

915 Simão, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva and E. M. Zdobnov (2015).
916 BUSCO: assessing genome assembly and annotation completeness with single-copy
917 orthologs. *Bioinformatics* **31**(19): 3210-3212.

918 Song, L. and L. Florea (2015). Rcorrector: efficient and accurate error correction for Illumina
919 RNA-seq reads. *GigaScience* **4**(1): 48.

920 Sujit, K. M., V. Singh, S. Trivedi, K. Singh, G. Gupta and S. Rajender (2020). Increased DNA
921 methylation in the spermatogenesis-associated (SPATA) genes correlates with infertility.
922 *Andrology* **8**(3): 602-609.

923 Sun, Z. H., J. Zhang, W. J. Zhang and Y. Q. Chang (2019). Gonadal transcriptomic analysis and
924 identification of candidate sex-related genes in *Mesocentrotus nudus*. *Gene* **698**: 72-81.

925 Suzek, B. E., H. Huang, P. McGarvey, R. Mazumder and C. H. Wu (2007). UniRef:
926 comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* **23**(10):
927 1282-1288.

928 Tao, W., J. Yuan, L. Zhou, L. Sun, Y. Sun, S. Yang, M. Li, S. Zeng, B. Huang and D. Wang
929 (2013). Characterization of Gonadal Transcriptomes from Nile Tilapia (*Oreochromis*
930 *niloticus*) Reveals Differentially Expressed Genes. *PLOS ONE* **8**(5): e63604.

931 Tato, T., N. Salgueiro-González, V. M. León, S. González and R. Beiras (2018).
932 Ecotoxicological evaluation of the risk posed by bisphenol A, triclosan, and 4-nonylphenol
933 in coastal waters using early life stages of marine organisms (*Isochrysis galbana*, *Mytilus*
934 *galloprovincialis*, *Paracentrotus lividus*, and *Acartia clausi*). *Environmental Pollution* **232**:
935 173-182.

936 Teaniniuraitemoana, V., A. Huvet, P. Levy, C. Klopp, E. Lhuillier, N. Gaertner-Mazouni, Y.
937 Gueguen and G. Le Moullac (2014). Gonad transcriptome analysis of pearl oyster *Pinctada*

938 *margaritifera*: identification of potential sex differentiation and sex determining genes. BMC
939 Genomics **15**(1): 491.

940 The UniProt Consortium (2016). UniProt: the universal protein knowledgebase. Nucleic Acids
941 Research **45**(D1): D158-D169.

942 Tian, S., M. Egertová and M. R. Elphick (2017). Functional Characterization of Paralogous
943 Gonadotropin-Releasing Hormone-Type and Corazonin-Type Neuropeptides in an
944 Echinoderm. Frontiers in Endocrinology **8**: 259-259.

945 Uthicke, S., N. P. Deshpande, M. Liddy, F. Patel, M. Lamare and M. R. Wilkins (2019). Little
946 evidence of adaptation potential to ocean acidification in sea urchins living in Future Ocean
947 conditions at a CO₂ vent. Ecology and Evolution **9**(17): 10004-10016.

948 Walker, C. W., S. A. Böttger, T. Unuma, S. A. Watts, L. G. Harris, A. L. Lawrence and S. D.
949 Eddy (2015). Enhancing the Commercial Quality of Edible Sea Urchin Gonads —
950 Technologies Emphasizing Nutritive Phagocytes. Echinoderm Aquaculture: 263-286.

951 Wang, H., J. Ding, S. Ding and Y. Chang (2019). Transcriptome analysis to characterize the
952 genes related to gonad growth and fatty acid metabolism in the sea urchin
953 *Strongylocentrotus intermedius*. Genes & Genomics **41**(12): 1397-1415.

954 Wang, H., J. Ding, S. Ding and Y. Chang (2020). Integrated metabolomic and transcriptomic
955 analyses identify critical genes in eicosapentaenoic acid biosynthesis and metabolism in
956 the sea urchin *Strongylocentrotus intermedius*. Scientific Reports **10**(1): 1697.

957 Wang, X., H. Li, G. Fu, Y. Wang, S. Du, L. Yu, Y. Wei and S. Chen (2016). Testis-specific
958 serine/threonine protein kinase 4 (Tsk4) phosphorylates Odf2 at Ser-76. Scientific Reports
959 **6**(1): 22861.

960 Wei, Z., X. Liu, Z. Zhou and J. Xu (2019). *De novo* transcriptomic analysis of gonad of
961 *Strongylocentrotus nudus* and gene discovery for biosynthesis of polyunsaturated fatty
962 acids. Genes & Genomics **41**(5): 583-597.

963 Wong, J. M., J. D. Gaitán-Espitia and G. E. Hofmann (2019). Transcriptional profiles of early
964 stage red sea urchins (*Mesocentrotus franciscanus*) reveal differential regulation of gene
965 expression across development. Marine Genomics **48**: 100692.

966 Yang, H., D. Cromley, H. Wang, J. T. Billheimer and S. L. Sturley (1997). Functional Expression
967 of a cDNA to Human Acyl-coenzyme A:Cholesterol Acyltransferase in Yeast: SPECIES-
968 DEPENDENT SUBSTRATE SPECIFICITY AND INHIBITOR SENSITIVITY. Journal of
969 Biological Chemistry **272**(7): 3980-3985.

970 Yasumasu, I., A. Hino, A. Suzuki and M. Mita (1984). Change in the Triglyceride Level in Sea
971 Urchin Eggs and Embryos During Early Development. **26**(6): 525-532.

972 Yeruham, E., G. Rilov, M. Shpigel and A. Abelson (2015). Collapse of the echinoid
973 *Paracentrotus lividus* populations in the Eastern Mediterranean—result of climate change?
974 Scientific Reports **5**: 13479.

975 Yue, H., C. Li, H. Du, S. Zhang and Q. Wei (2015). Sequencing and *De Novo* Assembly of the
976 Gonadal Transcriptome of the Endangered Chinese Sturgeon (*Acipenser sinensis*). PLOS
977 ONE **10**(6): e0127332.

978 Zhan, Y., J. Li, J. Sun, W. Zhang, Y. Li, D. Cui, W. Hu and Y. Chang (2019). The Impact of
979 Chronic Heat Stress on the Growth, Survival, Feeding, and Differential Gene Expression in
980 the Sea Urchin *Strongylocentrotus intermedius*. Frontiers in Genetics **10**: 301.

981 Zhang, D., D. Xie, X. Lin, L. Ma, J. Chen, D. Zhang, Y. Wang, S. Duo, Y. Feng, C. Zheng, B.
982 Jiang, Y. Ning and C. Han (2018). The transcription factor SOX30 is a key regulator of
983 mouse spermiogenesis. Development **145**(11): dev164723.

984 Zhang, J., X. Han, J. Wang, B.-Z. Liu, J.-L. Wei, W.-J. Zhang, Z.-H. Sun and Y.-Q. Chang
985 (2019). Molecular Cloning and Sexually Dimorphic Expression Analysis of nanos2 in the
986 Sea Urchin, *Mesocentrotus nudus*. International Journal of Molecular Sciences **20**(11):
987 2705.

988 Zhang, T. and D. Zarkower (2017). DMRT proteins and coordination of mammalian
989 spermatogenesis. Stem Cell Research **24**: 195-202.

990 Zhang, W., Z. Lv, C. Li, Y. Sun, H. Jiang, M. Zhao, X. Zhao, Y. Shao and Y. Chang (2019).
 991 Transcriptome profiling reveals key roles of phagosome and NOD-like receptor pathway in
 992 spotting diseased *Strongylocentrotus intermedius*. Fish & Shellfish Immunology **84**: 521-
 993 531.
 994 Zhang, W., Z. Wang, X. Leng, H. Jiang, L. Liu, C. Li and Y. Chang (2019). Transcriptome
 995 sequencing reveals phagocytosis as the main immune response in the pathogen-
 996 challenged sea urchin *Strongylocentrotus intermedius*. Fish & Shellfish Immunology **94**:
 997 780-791.

998

999 **Figure 1.** Sampling and bioinformatic workflow of the study. **(a)** Sampling and experimental
 1000 setup used to perform RNA extraction and sequencing. **(b)** Bioinformatics workflow used to
 1001 process the raw datasets, perform the differential gene expression and obtain biological
 1002 insights.

1003 **Figure 2.** Histological sections of gonads used for transcriptome sequencing. Females were in
 1004 stage III (a, b) and IV (c) of maturation. All males were in stage III of maturation. Scale bar (200
 1005 μ m). NP: nutritive phagocyte; O: oocyte; S: spermatid.

1006 **Figure 3.** Functional annotation and clean-up analyses of the V0 and V1 versions of
 1007 transcriptome assembly of *P. lividus*. **(a)** Venn diagram of the transcripts (V0 transcriptome)
 1008 codifying to an open reading frame or matching with NCBI, UNIPROT and PFAM databases. **(b)**
 1009 Length distribution of the transcripts in versions V0 and V1 of the transcriptome assembly. **(c)**
 1010 Blast-x analysis of the V0 and V1 versions of the *P. lividus* transcriptome assembly.

1011 **Figure 4.** Heatmap of up and down regulated genes. **(a)** Sexual development and determination
 1012 genes. **(b)** Lipid related with biosynthesis and storage of lipids.

1013 **Supplementary Figure 1.** Multidimensional scaling analyses. *P. lividus* male samples: M1, M2,
 1014 M3; *P. lividus* male samples: F1, F2, F3.

1015 **Supplementary Figure 2.** Heatmap of all up and down regulated genes.

1016 **Supplementary Figure 3.** GO terms present in all DEG of *P. lividus* gonads according Panther
 1017 v.15 using *Strongylocentrotus purpuratus* as model organism.

1018 **Supplementary Figure 4.** GO terms present in all DEG of *P. lividus* gonads discriminating by
 1019 males and females according Panther v.15 using *Strongylocentrotus purpuratus* as model
 1020 organism.

1021 **Supplementary Figure 5.** Phylogenetic trees were constructed by selected amino acids (aa)
 1022 sequences for lipid metabolism obtained from Illumina-sequenced transcriptomes of sea-urchin
 1023 gonads (*Paracentrotus lividus*) and each orthologs genes of nine invertebrate and vertebrate
 1024 species. The trees were constructed using the Maximum-Likelihood Phylogenies (PhyML 3.0),
 1025 the numbers at each node represent the bootstrap values as percentage and it were rooted in
 1026 the *Crassostrea gigas*. Trees of sex related orthologs genes were showed as: **(a)**, fizzy-related
 1027 protein homolog; **(b)**, amidophosphoribosyltransferase; **(c)**, HORMA domain-containing protein

1028 1; **(d)**, spermatogenesis-associated protein 7 homolog; **(e)**, TBC1 domain family member 20;
 1029 and **(f)**, transforming growth factor beta receptor type 3. Lipid related trees were as **(g)**, fatty
 1030 acids desaturase (Fad); **(h)**, carnitine O-palmitoyltransferase 2; **(i)**, cholinephosphotransferase
 1031 1; **(j)**, peroxisome proliferator-activated receptor alpha (PPAR- α); **(k)**, NPC intracellular
 1032 cholesterol transporter 1; **(l)**, sterol O-acyltransferase 1.

(a) Experimental setup

Sampling
(n Total=10)
(n Selected=6)

(n=3 Males)

(n=3 Females)

Sex identification;
Tissue collection;
Storage in
RNAlater at
-20 °C;
RNA Extraction
(NZY Tech total
RNA Kit);
Storage in mQ
water at -80 °C.

RNA QC ;
Sequencing;
Illumina HiSeq-
4000;
PE150 ;
> 30 million
PE reads /
sample.

RNA-seq Raw Data

(b) Bioinformatic approach

Raw Data Quality-control

Transcriptome processing

Biological insights

Legend :

(a)

(b)

(c)