

SCIENTIFIC REPORTS



OPEN

Functional insights into the testis transcriptome of the edible sea urchin *Loxechinus albus*

Juan Diego Gaitán-Espitia^{1,2}, Roland Sánchez¹, Paulina Bruning¹ & Leyla Cárdenas¹

Received: 08 June 2016
Accepted: 13 October 2016
Published: 02 November 2016

The edible sea urchin *Loxechinus albus* (Molina, 1782) is a keystone species in the littoral benthic systems of the Pacific coast of South America. The international demand for high-quality gonads of this echinoderm has led to an extensive exploitation and decline of its natural populations. Consequently, a more thorough understanding of *L. albus* gonad development and gametogenesis could provide valuable resources for aquaculture applications, management, conservation and studies about the evolution of functional and structural pathways that underlie the reproductive toolkit of marine invertebrates. Using a high-throughput sequencing technology, we explored the male gonad transcriptome of this highly fecund sea urchin. Through a *de novo* assembly approach we obtained 42,530 transcripts of which 15,544 (36.6%) had significant alignments to known proteins in public databases. From these transcripts, approximately 73% were functionally annotated allowing the identification of several candidate genes that are likely to play a central role in developmental processes, nutrient reservoir activity, sexual reproduction, gamete generation, meiosis, sex differentiation, sperm motility, male courtship behavior and fertilization. Additionally, comparisons with the male gonad transcriptomes of other echinoderms revealed several conserved orthologous genes, suggesting that similar functional and structural pathways underlie the reproductive development in this group and other marine invertebrates.

Echinoderms and, in particular, sea urchins, play an important ecological role in coastal ecosystems as a food source for higher animals and as a regulators of seagrass and rocky reef communities¹. These calcifying marine invertebrates are now a focus of intense research as over-fishing of sea urchins for human consumption has reduced their natural populations worldwide². The roe (gonad) is the only edible part of the sea urchin. It is considered a high quality seafood product and its price is greatly influenced by factors such as appearance, colour, texture, and flavour³. Beyond their economic and ecological value, sea urchins are also an important model system for climate change biology due to the critical effects induced by ocean warming and acidification on gonad development and reproduction⁴, and also on the metamorphosis and growth of early ontogenetic stages⁵. Like other marine invertebrates, most sea urchins are broadcast spawners, releasing their gametes into the water column, where fertilization occurs². The quality of the gametes and therefore the successful fertilization in these organisms, depend, in part, on the quality of the gonads and their amino acid and carotenoid composition⁶.

Sea urchins have five gonads attached to the inner wall of their exoskeletons. Each gonad consists of hundreds of acini (i.e., clusters of cells) with populations of germ cells and somatic cells called nutritive phagocytes⁷. The dynamic interactions between the germinal and the somatic cellular populations in the gonads, are fundamental processes during gametogenesis of sea urchins⁸. For instance, the nutritive phagocytes provide a structural and nutritional microenvironment for germinal cells⁹ having similar functions to the Sertoli cells of vertebrates¹⁰. Despite the pivotal role of nutritive phagocytes in the sea urchin gametogenesis, the cellular and molecular mechanisms involved in gonial cell mitosis as well as in the phagocytic and nutritional functions are not fully understood⁸. Some progress has been made in the understanding of the molecular and functional processes underlying gametogenesis in female sea urchins^{8,11,12}. During oogenesis, oocytes express a unique set of genes that appear essential for their growth, for meiotic recombination and division, for storage of nutrients, and for fertilization¹¹. In addition, many transcriptional factors have been identified as responsible for the unique transcriptional activity seen in female gametocytes, such as those of the forkhead family (e.g., FoxL2, FoxX, FoxQ2 and FoxN1/4), the

¹Instituto de Ciencias Ambientales y Evolutivas, Universidad Austral de Chile, Casilla 567 Valdivia, Chile. ²CSIRO Oceans & Atmosphere, GPO Box 1538, Hobart 7001, TAS, Australia. Correspondence and requests for materials should be addressed to L.C. (email: leylacardenas@uach.cl)

ETS factor Sp-Eyg, the doublesex transcription factor *dsx*, and other regulatory proteins that are involved in ovary and oocyte development, and also in sex determination of different metazoans^{11–13}. In contrast, similar genomic and expression analyses are unavailable for sea urchin spermatogenesis. Related processes like formation of the synaptonemal complex and meiosis also occur during spermatogenesis and are probably governed by the same suite of genes observed during oogenesis⁸. However, we are still no closer in understanding the male reproductive toolkit for this highly fecund animal¹⁴.

One of the major obstacles in defining the regulatory and molecular mechanisms underlying reproductive development of sea urchins is the limited genetic and genomic information available for these organisms. In this sense, transcriptome sequencing is an effective way to discover genes participating in specific biological processes when genome sequence is not available¹⁵. In this study, we describe the male gonad transcriptome of the Chilean edible sea urchin *Loxechinus albus* (Molina, 1782). This sea urchin is one of the most economically important species in the littoral benthic systems of the southeast Pacific of South America¹⁶. Harvesting of *L. albus* represents the largest extraction volume among world urchin fisheries¹⁷. However, overfishing of *L. albus* along Chilean coasts is driving the decline or depletion of natural populations of this biological resource¹⁸. A more thorough understanding of sea urchin gametogenesis and gonad development could allow the manipulation of these processes at several stages in order to accomplish particular goals such as: i) To suppress gametogenesis in order to produce high-quality gonads for consumption; and ii) to promote gametogenesis for the increased production of seed stock and natural populations recovery¹⁹.

Results

Sequencing and *de novo* assembly of the *L. albus* transcriptome. Sequencing of the male gonads of *L. albus* produced 1,062,716 raw Roche 454 GS-FLX Titanium reads that were quality and adapter trimmed and size selected to yield 1,016,639 cleaned reads with a mean length of 309.8 bp and 295 Mbp of total sequence data (Fig. 1A). The SRA raw reads have been deposited on GenBank public database under the accession number SRP066399 of the bioproject PRJNA302689. After quality control steps, reads were assembled using the *de novo* and the Reference Annotation Based Transcript (RABT) methods (Fig. 1A). Overall, assembly results obtained from RABT with the genome of *S. purpuratus* gave lower mapped reads and number of contigs compared to the *de novo* approach (Fig. 1A). For RABT, assembly of high quality reads generated 17,557 tentative consensus sequences (non-redundant sequences or unigenes), with an N50 of 956 bp and an L50 of 3,572 (Fig. 1A, Table S1). On the other hand, the *de novo* assembly generated 42,530 consensus sequences with an N50 of 645 bp and an L50 of 10,380 (Fig. 1A, Table 1). Further analyses of these two assembled transcriptomes revealed a lower coverage of homologs in UniRef90 for RABT (11 536) in comparison to the *de novo* assembly (15,072) (Supplementary Table S2). Similarly, the assessment of transcriptome completeness (CEGMA) showed a higher representation of the core eukaryotic genes (CEGs) in the *de novo* transcriptome compared to the RABT (Fig. 1A, Table 1). Based on these results, the *de novo* assembly was chosen for all downstream analyses.

Unigenes in our *de novo* assembled transcriptome of *L. albus*, had a mean length of 539 bp (Fig. 1B) with a sequencing coverage ranging from 1 to 4,459 with a mean of ~7 (highly left-skewed toward low coverage). As expected for a randomly fragmented transcriptome, there was a positive relationship between the length of a given unigene and the number of reads assembled into it (Supplementary Fig. S1). The CEGMA analysis of our *de novo* assembly identified 176 out of the 248 core proteins (71%) as complete (defined as >70% alignment length with core protein) and 205 (83%) as partially present (Fig. 1A, Table 1). The average number of orthologs per CEG was 2.11 with the higher percentages in the more divergent groups (i.e., 1 and 2) whilst the total number of CEGs was higher in the highly conserved groups (i.e., 3 and 4) (Supplementary Table S1).

Functional annotation. Of the 42,530 assembled unigene sequences, 15,544 (36.6%) had significant alignments ($\leq 1e-5$) to known proteins in the public databases UniProt (Swiss-Prot and TrEMBL) and NCBI (nr) (Supplementary Table S2), while 26,986 (63.4%) showed no or poor similarity matches and may represent specific unigenes of *L. albus* with unknown function. Our results showed that matching efficiency (i.e., sequences with hits) increases with the length of the unigene sequences (Supplementary Fig. S2). From the unigenes with significant blast hits, 13,333 (~86%) matched to the sea urchin *S. purpuratus*, followed by the acorn worm *Saccoglossus kowalevskii* (~1%) that is a closely related species to echinoderms, and the sea urchin *Paracentrotus lividus* (<1%) (Fig. 1C). Blast results were used for functional categorization of the assembled unigenes. For this step, Blast2GO annotations (52,725 GO terms for 8,144 sequences) were merged with the InterProScan annotations (18,661 GO terms for 8,022 sequences) to maximize the number of sequences with assigned functions. In total, 58,957 GO terms were identified for 11,400 annotated sequences under the three main ontologies. From these sequences, 8,811 (78%) were assigned to molecular functions (GO:0003674), followed by 7,651 (67%) to biological processes (GO:0008150) and 6,425 (56%) to cellular components (GO:0005575) (Supplementary Fig. S3). Within the molecular function category, binding (GO:0005488) and catalytic activity (GO:0003824) were the most represented GO terms with 74% (6,559 sequences) and 48% (4,302 sequences), respectively. For the biological processes, the three top GO terms were cellular (GO:0009987, 81%), metabolic (GO:008152, 73%) and single-organism (GO:0044699, 66%) processes. Furthermore, in the cellular components category, the predominant GO terms were grouped in cell (GO:0005623, 86%), cell part (GO:0044464, 85%) and organelle (GO:0043226, 69%).

In order to identify functional pathways active in male gonads of *L. albus*, all unigenes were mapped to the reference canonical pathways in KEGG database. In total, 7,067 unigenes were mapped to 446 KEGG pathways within 48 categories (Supporting information). From these, 5,252 unigenes were assigned to the main KO categories (Fig. 2A) with a major representation of Metabolism and Organism System (50.87%), followed by Genetic Information Processing, Environmental Information Processing and Cellular Processes (49.12%) (Fig. 2B). Within these five categories, the highest represented pathways included signal transduction (15.32%),

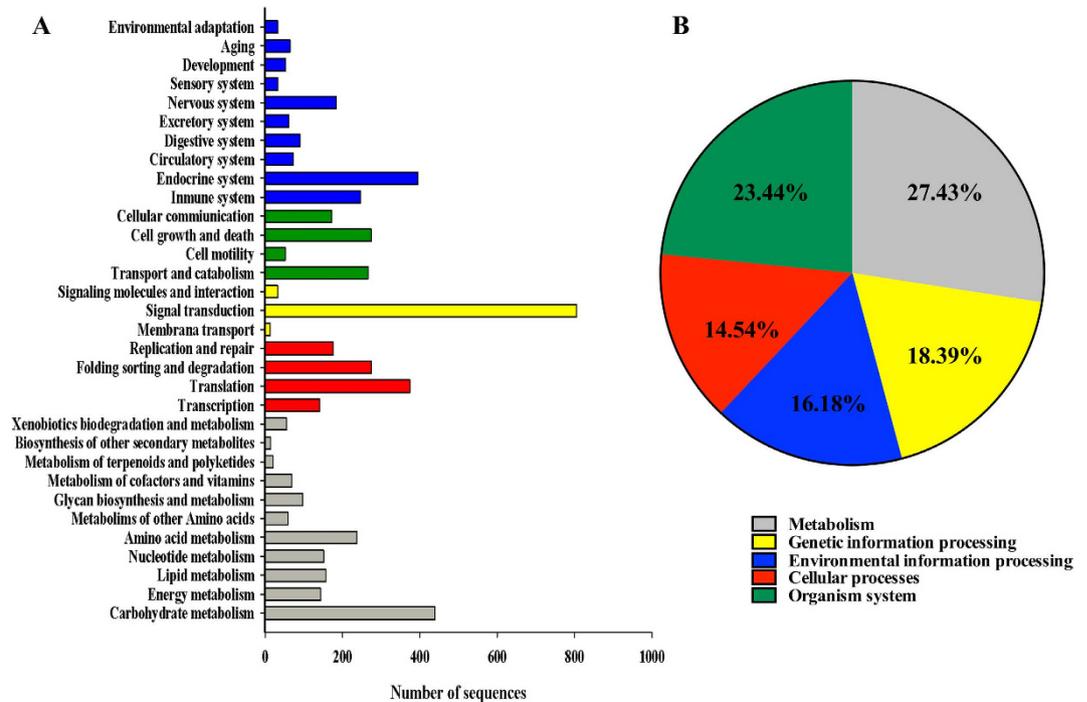


Figure 2. Unigenes homology to KEGG genes obtained from the KAAS server. (A) Number of sequences assigned to each sub-category of the reference hierarchy KOs. **(B)** Percentage distribution of the five top KEGG orthology categories in the male gonad transcriptome of *Loxechinus albus*.

HDT3), mitochondrial (e.g., Cytochrome c oxidase subunit 3) and catalytic (e.g., Isovaleryl-CoA dehydrogenase) activity (Supplementary Table S4 and S5). Posterior analyses of GO terms revealed a significant GO enrichment of the cellular component kinesin complex (GO:0005871), and the molecular functions hydrolase activity (GO:0016787) and microtubule motor activity (GO:0003777). Comparisons between *L. albus* and the testis transcriptome of its closest available phylogenetic relative (i.e., sharing the most common recent ancestor), the black sea urchin *A. lixula*, revealed 190 orthologous clusters (Fig. 4), in which most of the proteins were grouped into the GO terms anterior/posterior axis specification (GO:0009948), establishment of planar polarity (GO:0001736), proton-transporting V-type ATPase (GO:0033180), translation elongation factor activity (GO:0003746), spermatogenesis (GO:0007283), sperm individualization (GO:0007291) and spermatid development (GO:0007286) (Supplementary Tables S4 and S5). On the other hand, transcriptomic comparisons with the most distant available phylogenetic sea urchin, *S. purpuratus*, revealed 729 clusters of orthologous proteins (Fig. 4), with over-representation of metabolic process (GO:0008152) and multicellular organismal development (GO:0007275) (Supplementary Tables S4 and S5).

Discussion

The use of sea urchins as a model system to address major paradigms in development and evolution has been strengthened in recent years by the generation of extensive genomic resources, especially for model species such as *S. purpuratus*, *Lytechinus variegatus*, *Eucidaris tribuloides* and *Paracentrotus lividus*²⁰. Most genomic and transcriptomic studies have been focused on gene regulatory networks during early development e.g. ref. 21, and in a lower proportion on sea urchin gametogenesis, with major progress in oogenesis^{8,11,12}. In order to obtain a more comprehensive understanding of the functional elements and regulatory mechanisms underlying reproductive development and spermatogenesis of sea urchins, we sequenced the male gonad transcriptome of the Chilean edible sea urchin *Loxechinus albus*. In this study, both the *de novo* and the Reference Annotation Based Transcript (RABT) methods were employed to generate the assembled transcriptome of the Chilean sea urchin. Taking advantage of the well-characterized and annotated genome of *S. purpuratus*²², we expected to obtain higher quality and completeness for our assembled transcriptome with the RABT approach compared to the *de novo* method. However, contrary to our expectations, our *de novo* assembly resulted in many more mapped reads, assembled transcripts and conserved eukaryotic core genes than with the RABT approach. The lower performance of the RABT could be explained by the genetic divergence between the reference genome of *S. purpuratus* and our target species^{15,23}, where the estimated divergence time for the split between these two species is dated approximately 55.5 MYA²⁴. In this sense, it has been suggested that the use of a reference genome that is too genetically distant from the target species can yield inaccurate read alignments and substantial data loss, decreasing the ability to detect biological variants, isoforms and exogenous transcripts²⁵. On the other side of the coin, attempting to assemble a transcriptome from short sequences without a reference genome is a computationally challenging task. Some of the issues faced are related to the substantial memory and long computing times required, the existence of large number of reads with artefacts, sequencing errors and different sequencing coverage among different

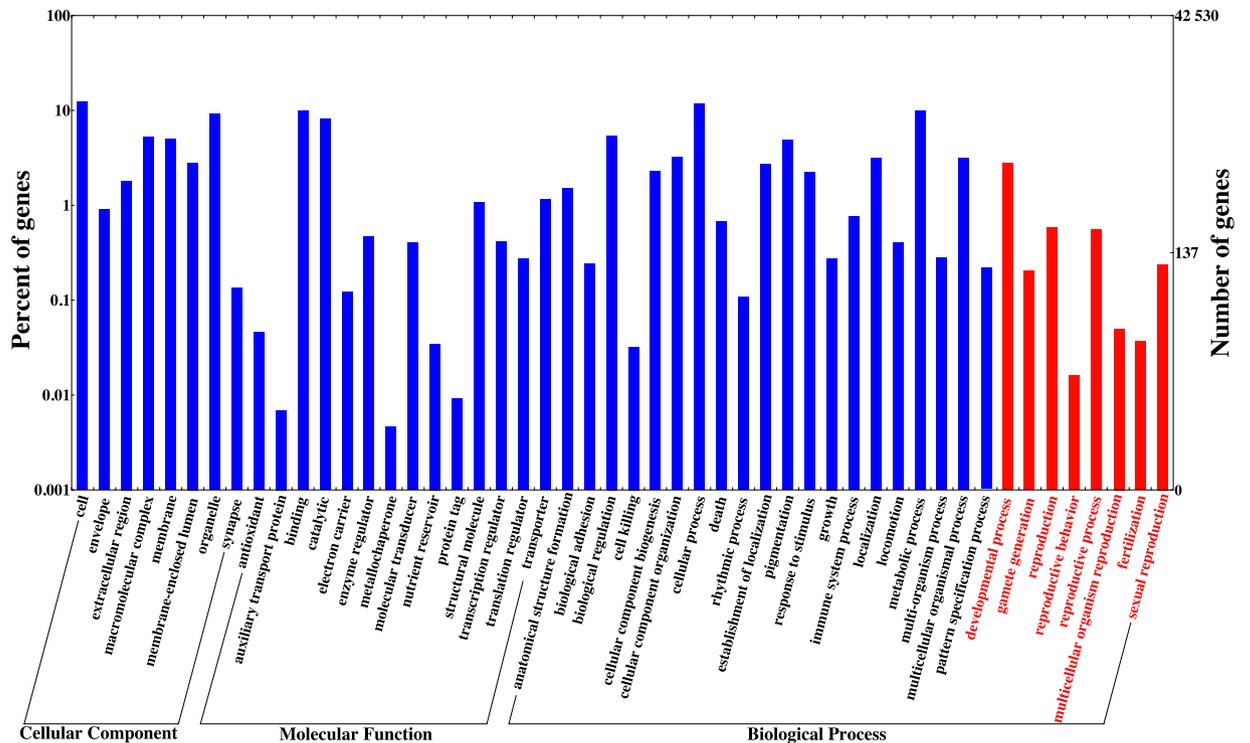


Figure 3. Distribution of Gene Ontology (GO) assignments of assembled unigenes of *L. albus*. GO categories are shown on the x-axis grouped into three main categories: biological processes, cellular components and molecular functions. The y-axis indicates the percentage of total genes in each category.

transcripts²⁶, and also the presence of paralogous genes²⁷. In this study we selected the CLC *de novo* assembler as it has performed well in a number of previous transcriptome assemblies^{28–30} and it has a faster computing pace with comparable or better assembly results than other bioinformatics programs³¹.

Our *de novo* assembled transcriptome showed similar characteristics to the available transcriptomes of other echinoderms. For instance, the GC content of the testis transcriptome of *L. albus* (40.4%) was closer to the testis (40.9%) and ovary (42.1%) transcriptomes of the black sea urchin *Arbacia lixula*³², and showed slightly higher values in relation to the transcriptomes of the Antarctic sea urchin *Sterechinus neumayeri* (38.6%)¹⁵, the purple sea urchin, *S. purpuratus* (36.9%)³³, and the kina sea urchin *Evechinus chloroticus* (39%)³⁴. Moreover, in terms of the number of assembled unigene sequences, our *de novo* assembled transcriptome was within the range reported in other 454 sequencing projects of sea urchins¹⁵, sea cucumbers³⁵, and sea stars³⁶. Although the percentage of unigenes of *L. albus* with a BLAST-hit may appear to be a relatively small number, we found that this value does not differ from other *de novo* transcriptome studies with non-model echinoderms^{14,15,34,37}. In this sense, the lack of homology with known proteins could be explained by biological reasons such as: i) the lack of annotated protein sequences of phylogenetically close species; ii) sequences from untranslated regions (UTR) of mRNA; iii) long noncoding RNA; and iv) orphan proteins³⁸. These results are congruent to those reported for the transcriptome of the sea urchin *S. intermedius*³⁹, the sea star *Acanthaster planci*¹⁴ and the brittle star *Amphiura filiformis*⁴⁰.

Genes involved in the regulation of gonadal development and gamete production. In echinoderms, gametogenesis and gonad development are prolonged processes of interactions between germinal and somatic cells^{8,41}. These processes involve cell growth, new organelle synthesis and storage, cell differentiation, sex determination, meiosis and the activity of a set of genes related to reproduction, growth and metabolism^{12,14,36,41}. By combining results from our different functional analyses we were able to identify several genes that are specifically related to these processes in male sea urchins. For instance, genes such as the DNA helicase MCM8, the Chd5 (chromodomain-helicase-DNA-binding protein Mi-2), the Yy1 (transcriptional repressor protein YY1), the Smbt1 (Scm-like with four MBT domains protein 1) and the Nphp1 (Nephrocystin-1) have been linked to different stages of spermatid differentiation⁴² and male gamete generation^{43–45}. Moreover, genes such as the Dmrt1 (Doublesex- and mab-3-related transcription factor 1), the Rara (Retinoic acid receptor alpha) and the TMF1 (TATA element modulatory factor) have been recognized to play a role during testis development⁴⁶, male germ cell proliferation and differentiation of Sertoli and Leydig cells⁴⁷. These two types of somatic cells are fundamental in spermatophore formation, gamete production and epididymal sperm maturation by regulating the synthesis of testosterone and growth factors⁴⁸ and by providing a structural and nutritional microenvironment for germinal cells⁹. Other key genes involved in spermatogenesis belong to the serine/threonine protein kinase complexes (e.g., TSSK4, VRK1, HASPIN, MAK), which are composed of a regulatory subunit, cyclin and a catalytic subunit, cyclin-dependent kinase (Cdk), that support germ cell maturation¹⁴.

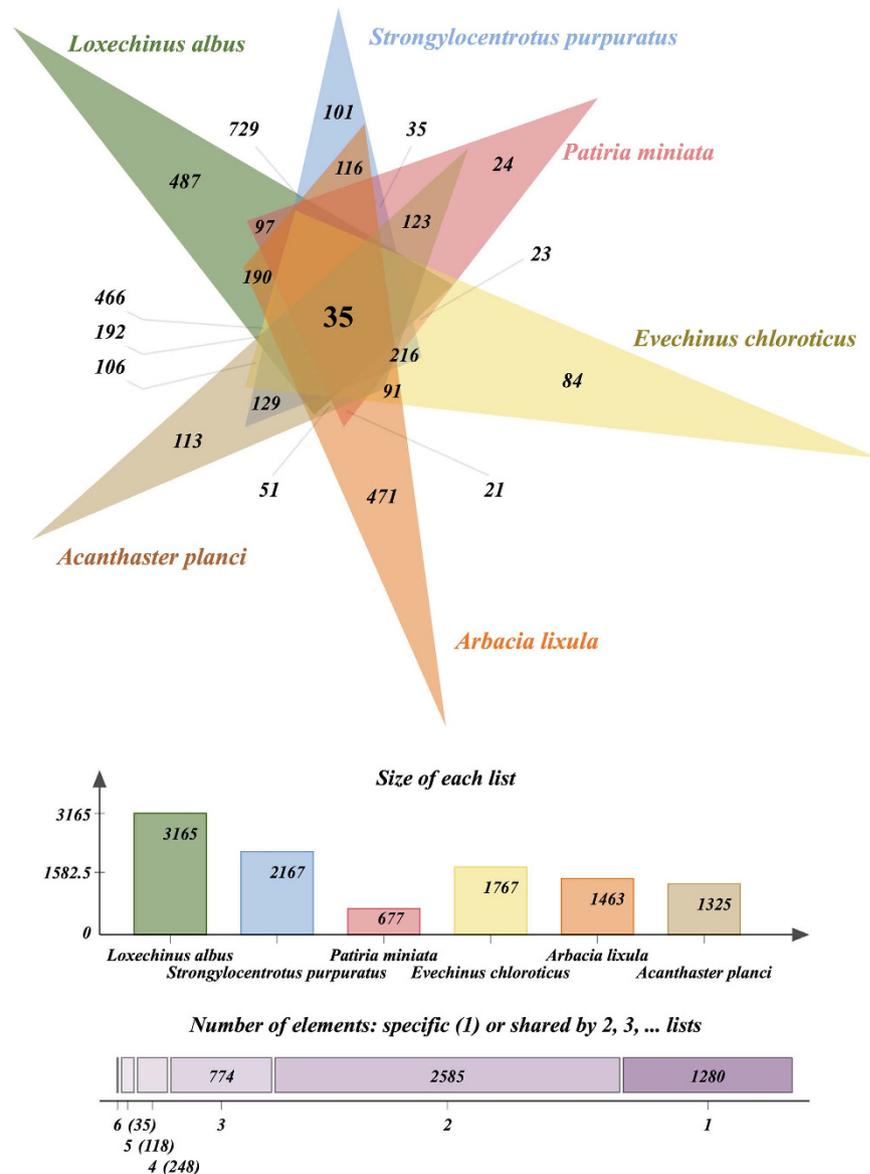


Figure 4. Comparisons of orthologous clusters among male gonad transcriptomes of the sea urchins *Loxechinus albus*, *Arbacia lixula*, *Evechinus chloroticus*, *Strongylocentrotus purpuratus*, and the sea stars *Patiria miniata* and *Acanthaster planci*.

Our functional analyses also found several signalling pathways that have been documented to be essential in spermatozoa and gonadal processes. For instance, signaling pathways of the GTPases Rap1 (KO04015) and Ras (KO04014) have been associated to spermatogenesis and sperm motility of sea urchins and vertebrates^{49,50}. Furthermore, signalling pathways of some protein kinases have been related to capacitation, acrosome reaction and phosphorylation of sperm heads (MAPK: KO04010), energy homeostasis of spermatozoa (AMPK: KO04152) and sperm-oocyte interaction (ErbB: KO04012)⁴⁸. Within these pathways, some kinase receptors (e.g., RTK) bind a variety of growth factors (e.g., TGF- β : KO04350; VEGFR: KO04370), mediating cell-cell interactions during sea urchin development⁵¹. Other signalling pathways that are known to be involved in a variety of biological processes, including development, are the wnt (KO04310), the Hedgehog (KO04340), the Jak-STAT (KO04630), the Notch (KO04330), the Hippo (KO04391), the FoxO (KO04068), the Ca²⁺ (KO04020) the cAMP (KO04024), the cGMP (KO04022), the PI3K-Akt (KO04151) and the Phosphatidylinositol (KO04070) pathway^{12,48,52–55}. Components of these signaling pathways have been highly evolutionarily conserved among metazoans⁵⁶, and are active in the testis transcriptome of *L. albus*.

From a metabolic point of view, and considering the importance of energy allocation in gonad development of sea urchins, our KEGG analysis identified the highest number of sequences in carbohydrate metabolism, followed by amino acid metabolism and lipid metabolism. This is not surprising since gonad tissues have high levels of carbohydrate, protein and lipid reserves to supply energy requirements during gonad growth and gametogenic cycles⁴¹. Proteins and lipids play a dominant role in cellular metabolism because of their dual roles as structural

and functional elements of cell membranes⁵⁷. However, most of the cellular energy is derived from the oxidation of carbohydrates via glycolysis and the Krebs cycle⁵⁸. The synthesis of glycogen from glucose is the most common pathway used by metazoans to store carbohydrates and the anabolic cost of glycogen synthesis is low (0.42 J mg glycogen⁻¹) when compared to the energy costs of protein synthesis (13 J mg protein⁻¹)⁴¹. This may explain the major abundance of transcripts linked to the glycolysis pathway (KO00010), the Citrate cycle (KO00020), the pentose phosphate pathway (KO00030) and pyruvate metabolism (KO000620) in our *de novo* assembled transcriptome. Nonetheless, other carbohydrate metabolic sub-pathways, such as pentose phosphate (KO00030) and fructose and mannose metabolism (KO00051), seem also to play a biochemical role in gonad development and gametogenesis in *L. albus*. These findings support the idea that the sperm of echinoderms can uptake glucose and metabolize mannose and pentose as energy sources¹⁴.

Evolutionary conservation of protein coding genes in testis of echinoderms. The basic processes of spermatogenesis and male gonad development are astonishingly similar in even very different animals, and the genes responsible of these processes are highly conserved⁵⁹. In echinoderms, for example, functional genes involved in protein folding (e.g., heat-shock proteins, endoplasmic), spermatogenesis cell cycle (e.g., serine/threonine protein kinase, cyclin), cell signaling (e.g., calcitonin peptide-receptor), sperm motility (e.g., dyneins, intraflagellar transport proteins) and development (e.g., DEAD-box proteins, Vasa, Nanos) exhibit high evolutionary conservation among the different classes¹⁴. Our comparative analysis identified several of these conserved orthologous genes between sea urchins and sea stars. Shared clusters between these two groups were over-represented by proteins related to the kinesin complex and molecular functions such as microtubule motor activity and hydrolase activity. The kinesin complexes generally possess a force-generating enzymatic activity, or motor, which converts the free energy of the gamma phosphate bond of ATP into mechanical work⁶⁰. These proteins are considered a second type of motor enzyme with a wide range of essential roles in mitotic and meiotic division during spermatogenesis, and modulating the shape and function of Sertoli cells⁶¹. Similarly, proteins related to microtubule motor activity are fundamental components of the dramatic cellular change that takes place during spermatogenesis. These proteins participate in several important events (e.g., nuclear elongation, cytoplasmic redistribution and reduction, and development of the flagellum) through the intricate process of cellular transformation from non-polar spermatogonia into the highly polarized, functional spermatozoa^{8,62}. On the other hand, proteins involved in hydrolase activity such as the Ubiquitin, have been implicated in the control of mammalian gametogenesis, sperm acrosomal function and anti-polyspermy defense⁶³. In sea urchins, these proteins are suggested to be responsible for acrosome reaction and sperm penetration of the vitelline envelope during fertilization⁶⁴.

Finally, our evolutionary comparison among testis transcriptomes of sea urchins and sea stars revealed different degrees of conservation of functional genes depending on the phylogenetic relationships. For example, closely related species such as *L. albus* and *A. lixula*, shared more orthologous proteins associated to spermatogenesis and development than distantly related species (e.g., *L. albus* and *S. purpuratus*), in which most of the clusters were represented by proteins linked to metabolic process and multicellular organismal development. The slow evolutionary rate of many spermatogenesis proteins may explain their greater conservation among closely related species⁶⁵.

Conclusion

The gonads of sea urchins have dual functions, as the main organ of nutrient storage and also as the central core for reproduction. In this work, we identified transcripts related to these functions in male gonads of the Chilean edible sea urchin *Loxechinus albus*. Our *de novo* assembled transcriptome and its functional annotation provide important molecular resources to improve our understanding of the reproductive development and spermatogenesis of sea urchins. The identification of candidate genes related to gonadal growth and maturation, somatic cell differentiation, pigmentation, gamete generation, meiosis, sperm ejaculation, fertilization and sex determination, opens a wide range of possibilities for aquaculture applications (e.g., high-quality gonads for consumption), management and conservation (e.g., natural populations recovery) and studies about the evolution of functional and structural pathways that underlie the reproductive toolkit in marine invertebrates.

Methods

Ethics statement. This study was carried out in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the Comisión Nacional de Investigación Científica y Tecnológica de Chile (CONICYT). All experiments were conducted according to current Chilean law. The protocol was approved by the Committee on the Ethics of Animal Experiments of the Universidad Austral de Chile.

Animal collection, RNA extraction and sequencing. Adult individuals of *L. albus* were collected in Los Molinos (39°40'S-73°12'W), southern Chile before the spawning season of this species (Austral winter-spring) in 2013. Sex was determined by gonad examination under a microscope for the detection of sperm and eggs. Mature animals (i.e., bright gonads filled with fully formed eggs or sperm) with similar gonad indexes (i.e., percentage of the total body weight that is made up by the gonad), were selected for further analyses. Total *L. albus* RNA was extracted from fresh gonads of eleven adult males using a commercial RiboPure™ Kit (AMBION). Quality and quantity of the RNA were analysed by Agilent 2100 Bioanalyzer RNA assays and evaluated by calculating the ratio of the 28S and 18S ribosomal RNA intensity peaks. Total RNA was pooled and submitted to Macrogen, Seoul, Korea (<http://www.macrogen.com>) for cDNA library construction, library normalization, and cDNA pyrosequencing in a 454-GS FLX Titanium System. Data quality was checked at Macrogen and provided in FASTq format.

Assembly and annotation. The sequenced raw data was processed using the CLC Genomics Workbench software v.8.2 (CLC bio, Denmark) in a four-steps pipeline: i) removal of adapter sequences; ii) removal of all reads containing more than 5% of ambiguous nucleotides ('N'), iii) trimming base pairs with a Phred quality score ≤ 20 from the 3'-end of each sequence; and iv) removal of reads shorter than 30 bp after trimming. High quality reads were used in two different assembly strategies, the *de novo* and the Reference Annotation Based Transcript (RABT) methods. In the first case, high quality reads were assembled following the parameters described in Cea *et al.*²³ and default settings (mapping mode = fast, automatic bubble size = yes, minimum contig length = 100 nt, automatic word size = yes, perform scaffolding = yes, auto-detect paired distances = yes) using the CLC *de novo* assembly option. For the RABT approach, the high quality reads were mapped to the *Strongylocentrotus purpuratus* reference genome⁶⁶ using the CLC *assembly to reference* tool. In both cases, reads that were not incorporated into any contig (i.e., singletons) were discarded and excluded from further analyses. The quality and completeness of the new draft assemblies were analysed in three ways: using the software QUASt for assembly statistics⁶⁷, analysing the coverage of homologs in UniRef90⁶⁸ and by mean of the core eukaryotic genes mapping approach (CEGMA)⁶⁹. The last approach is based on a subset of 248 widely conserved eukaryotic core genes that generally lack paralogs in the eukaryotes⁶⁹.

The best-assembled unigenes were used for BLAST searches and annotation against the UniProt (Swiss-Prot and TrEMBL) and NCBI RefSeq (nr) protein databases using the BLASTX algorithm with an e-value cutoff of $1e^{-5}$. Annotated unigenes were further searched for Gene Ontology (GO) terms using the Blast2GO software⁷⁰ according to the main categories of Gene Ontology (GO; molecular functions, biological processes and cellular components)⁷¹. Complementary annotations were done with the InterProScan v.5 software⁷², which provides functional analysis of proteins by classifying them into families and predicting domains and important sites. The annotation results were further fine-tuned with the Annex and GO slim functions of the Blast2GO software in order to improve and summarize the functional information of the transcriptome dataset. After that, the EC (Enzyme Commission) number was assigned and parsed based on the Blast2GO results. The distribution of annotated unigenes among GO categories was mapped using the WEGO software⁷³. Finally, a comparison of overall nucleotide sequence homology between the male gonad transcriptome of *L. albus* and the genome of *S. purpuratus* was completed using the Kyoto Encyclopedia of Genes and Genomes (KEGG) and its automated assignment server (KAAS)⁷⁴.

Comparison among male gonad transcriptomes of echinoderms. In order to explore the level of evolutionary conservation of gene/protein functions in male gonads of echinoderms, we compared the testis transcriptomes of the sea urchins *L. albus*, *Arbacia lixula*³², *Evechinus chloroticus*³⁴, *S. purpuratus*²² and the sea stars *Patiria miniata*³⁶ and *Acanthaster planci*¹⁴. The raw unassembled datasets of these species were downloaded from the NCBI SRA database under the accessions SRP066435, SRR1014624, SRR532121, SRR573710 and SRR1197243 respectively. The same quality control and pipelines applied for *L. albus* were implemented to obtain high-quality contigs of the additional species. Transcripts of the six species were translated and used for comparisons of orthologous clusters with the OrthoVenn software⁷⁵ using default parameters to identify GO categories and any GO enrichment.

Data Availability

Raw reads have been deposited on GenBank public database under the accession number SRP066399 of the bioproject PRJNA302689. In addition, the assembly file has been deposited in Dryad, DOI: 10.5061/dryad.hc7v5.

References

1. Shears, N. T. & Babcock, R. C. Marine reserves demonstrate top-down control of community structure on temperate reefs. *Oecologia* **132**, 131–142 (2002).
2. Lawrence, J. M. *Sea Urchins: Biology and Ecology*. **38**, (Academic Press, 2013).
3. McBride, S. C., Price, R. J., Tom, P. D., Lawrence, J. M. & Lawrence, A. L. Comparison of gonad quality factors: Color, hardness and resilience, of *Strongylocentrotus franciscanus* between sea urchins fed prepared feed or algal diets and sea urchins harvested from the Northern California fishery. *Aquaculture* **233**, 405–422 (2004).
4. Mos, B., Byrne, M. & Dworjanyan, S. A. Biogenic acidification reduces sea urchin gonad growth and increases susceptibility of aquaculture to ocean acidification. *Mar. Environ. Res.* **113**, doi: 10.1016/j.marenvres.2015.11.001 (2015).
5. Byrne, M., Lamare, M., Winter, D., Dworjanyan, S. A. & Uthicke, S. The stunting effect of a high CO₂ ocean on calcification and development in sea urchin larvae, a synthesis from the tropics to the poles. *Philos. Trans. R. Soc. London B Biol. Sci.* **368**, 20120439 (2013).
6. Plank, L., Lawrence, J., Lawrence, A. & Montoya, R. The effect of dietary carotenoids on gonad production and carotenoid profiles in the sea urchin *Lytechinus variegatus*. *J. World Aquac. Soc.* **33**, 127–137 (2002).
7. Unuma, T. In *Sea Urchin From Basic Biol. to Aquac.* (Yukio, Y., Matranga, V. & Smolenicka, Z.) 115–128 (A.A. Balkema Publishers, 2002).
8. Walker, C. W., Lesser, M. P. & Unuma, T. In *Sea Urchins Biol. Ecol.* (Lawrence, J.) **38**, 25–43 (Elsevier, 2013).
9. Walker, C. W., Harrington, L. M., Lesser, M. P. & Fagerberg, W. R. Nutritive phagocyte incubation chambers provide a structural and nutritive microenvironment for germ cells of *Strongylocentrotus droebachiensis*, the green sea urchin. *Biol. Bull.* **209**, 31–48 (2005).
10. Nakanishi, Y. & Shiratsuchi, A. Phagocytic removal of apoptotic spermatogenic cells by Sertoli cells: mechanisms and consequences. *Biol. Pharm. Bull.* **27**, 13–16 (2004).
11. Wessel, G. M., Juliano, C. E., Wong, J. L., Gustafson, E. A. & Song, J. L. Molecular markers of oocyte and primordial germ cell development in the sea urchin. *Echinoderms* 517–528 (2010).
12. Song, J. L., Wong, J. L. & Wessel, G. M. Oogenesis: Single cell development and differentiation. *Dev. Biol.* **300**, 385–405 (2006).
13. Georges, A. *et al.* FOXL2: a central transcription factor of the ovary. *J. Mol. Endocrinol.* **52**, R17–R33 (2014).
14. Stewart, M. J., Stewart, P. & Rivera-Posada, J. De novo assembly of the transcriptome of *Acanthaster planci* testes. *Mol. Ecol. Resour.* **15**, 953–966 (2015).
15. Dilly, G. F., Gaitán-Espitia, J. D. & Hofmann, G. E. Characterization of the Antarctic sea urchin (*Sterechinus neumayeri*) transcriptome and mitogenome: a molecular resource for phylogenetics, ecophysiology and global change biology. *Mol. Ecol. Resour.* **15**, 425–436 (2015).

16. Cárcamo, P. F., Candia, a. I. & Chaparro, O. R. Larval development and metamorphosis in the sea urchin *Loxechinus albus* (Echinodermata: Echinoidea): Effects of diet type and feeding frequency. *Aquaculture* **249**, 375–386 (2005).
17. Vásquez, J. & Donoso, G. In *Sea Urchins Biol. Ecol.* (Lawrence, J. M.) 285–293 (Elsevier, 2013).
18. Olave, S., Bustos, E., Lawrence, J. M. & Carcamo, P. The Effect of Size and Diet on Gonad Production by the Chilean Sea Urchin *Loxechinus albus*. *J. World Aquac. Soc.* **32**, 210–214 (2001).
19. Walker, C. W., Unuma, T. & Lesser, M. P. In *Edible Sea Urchins Biol. Ecol.* (Lawrence, J. M.) 11–33 (Elsevier, 2006).
20. Cameron, R. A., Samanta, M., Yuan, A., He, D. & Davidson, E. SpBase: the sea urchin genome database and web site. *Nucleic Acids Res.* **37**, D750–D754 (2009).
21. Israel, J. W. *et al.* Comparative Developmental Transcriptomics Reveals Rewiring of a Highly Conserved Gene Regulatory Network during a Major Life History Switch in the Sea Urchin Genus *Heliocidaris*. *PLoS Biol.* **14**, e1002391 (2016).
22. Tu, Q., Cameron, R. & Worley, K. Gene structure in the sea urchin *Strongylocentrotus purpuratus* based on transcriptome analysis. *Genome Res.* **22**, 2079–2087 (2012).
23. Cea, G., Gaitán-Espitia, J. D. & Cárdenas, L. Complete mitogenome of the edible sea urchin *Loxechinus albus*: genetic structure and comparative genomics within Echinozoa. *Mol. Biol. Rep.* **42**, 1081–1089 (2015).
24. Hedges, S. B. & Kumar, S. *The timetree of life*. (Oxford University Press, 2009).
25. Martin, J. A. & Wang, Z. Next-generation transcriptome assembly. *Nat. Rev. Genet.* **12**, 671–682 (2011).
26. Martin, J. *et al.* Rnnotator: an automated de novo transcriptome assembly pipeline from stranded RNA-Seq reads. *BMC Genomics* **11**, 663 (2010).
27. Vijay, N., Poelstra, J. W., Künstner, A. & Wolf, J. B. W. Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive *in silico* assessment of RNA-seq experiments. *Mol. Ecol.* **22**, 620–634 (2013).
28. Cárdenas, L. *et al.* Transcriptome analysis in *Concholepas concholepas* (Gastropoda, Muricidae): Mining and characterization of new genomic and molecular markers. *Mar. Genomics* **4**, 197–205 (2011).
29. Kumar, S. & Blaxter, M. L. Comparing de novo assemblers for 454 transcriptome data. *BMC Genomics* **11**, 571 (2010).
30. Ashrafi, H. *et al.* De novo assembly of the pepper transcriptome (*Capsicum annuum*): a benchmark for *in silico* discovery of SNPs, SSRs and candidate genes. *BMC Genomics* **13**, 571 (2012).
31. Brautigam, A., Mullick, T., Schliesky, S. & Weber, A. P. M. Critical assessment of assembly strategies for non-model species mRNA-Seq data and application of next-generation sequencing to the comparison of C3 and C4 species. *J. Exp. Bot.* **62**, 3093–3102 (2011).
32. Pérez-Portela, R., Turon, X. & Riesgo, A. Characterization of the transcriptome and gene expression of four different tissues in the ecologically relevant sea urchin *Arbacia lixula* using RNA-seq. *Mol. Ecol. Resour.* **16**, 794–808 (2016).
33. Kober, K. M. & Pogson, G. H. Genome-wide patterns of codon bias are shaped by natural selection in the purple sea urchin, *Strongylocentrotus purpuratus*. *G3 Genes, Genomes, Genet.* **3**, 1069–1083 (2013).
34. Gillard, G. B., Garama, D. J. & Brown, C. M. The transcriptome of the NZ endemic sea urchin *Kina* (*Evechinus chloroticus*). *BMC Genomics* **15**, 45 (2014).
35. Du, H. *et al.* Transcriptome sequencing and characterization for the sea cucumber *Apostichopus japonicus* (Selenka, 1867). *PLoS One* **7**, e33311 (2012).
36. Hart, M. W. & Foster, A. Highly expressed genes in gonads of the bat star *Patiria miniata*: Gene ontology, expression differences, and gamete recognition loci. *Invertebr. Biol.* **132**, 241–250 (2013).
37. Zhou, Z. C. *et al.* Transcriptome sequencing of sea cucumber (*Apostichopus japonicus*) and the identification of gene-associated markers. *Mol. Ecol. Resour.* **14**, 127–138 (2014).
38. Patel, A., Dettleff, P., Hernandez, E. & Martinez, V. A comprehensive transcriptome of early development in yellowtail kingfish (*Seriola lalandi*). *Mol. Ecol. Resour.* **16**, 364–376 (2016).
39. Chen, Y., Chang, Y., Wang, X., Qiu, X. & Liu, Y. De novo assembly and analysis of tissue-specific transcriptomes revealed the tissue-specific genes and profile of immunity from *Strongylocentrotus intermedius*. *Fish Shellfish Immunol.* **46**, 723–736 (2015).
40. Delroisse, J., Ortega-Martinez, O., Dupont, S., Mallefet, J. & Flammang, P. De novo transcriptome of the European brittle star *Amphiura filiformis* pluteus larvae. *Mar. Genomics* **23**, 109–121 (2015).
41. Marsh, A. G., Powell, M. L. & Watts, S. A. In *Edible Sea Urchins Biol. Ecol.* (Lawrence, J. M.) **37**, 45–57 (Elsevier, 2007).
42. Jiang, S. T. *et al.* Targeted disruption of *Nphp1* causes male infertility due to defects in the later steps of sperm morphogenesis in mice. *Hum. Mol. Genet.* **17**, 3368–3379 (2008).
43. Zhuang, T. *et al.* CHD5 is required for spermiogenesis and chromatin condensation. *Mechanisms of development. Mech. Dev.* **131**, 35–46 (2014).
44. Wu, S., Hu, Y., Liu, H. & Shi, Y. Loss of YY1 impacts the heterochromatic state and meiotic double-strand breaks during mouse spermatogenesis. *Mol. Cell Biol.* **29**, 6245–6256 (2009).
45. Lutzmann, M. *et al.* MCM8- and MCM9-deficient mice reveal gametogenesis defects and genome instability due to impaired homologous recombination. *Mol. Cell* **47**, 523–534 (2012).
46. Raymond, C. S., Murphy, M. W., O'Sullivan, M. G., Bardwell, V. J. & Zarkower, D. *Dmrt1*, a gene related to worm and fly sexual regulators, is required for mammalian testis differentiation. *Genes Dev.* **14**, 2587–2595 (2000).
47. Doyle, T. J. *et al.* Potential functions of retinoic acid receptor A in Sertoli cells and germ cells during spermatogenesis. *Ann. N. Y. Acad. Sci.* **1120**, 114–130 (2007).
48. Gupta, G. S. Proteomics of spermatogenesis. *Proteomics Spermatogenesis*, doi: 10.1007/0-387-27655-6 (2005).
49. Aivatiadou, E., Mattei, E., Ceriani, M., Tilia, L. & Berruti, G. Impaired fertility and spermiogenic disorders with loss of cell adhesion in male mice expressing an interfering *Rap1* mutant. *Mol. Biol. Cell* **18**, 1530–1542 (2007).
50. Cuellar-Mata, P. *et al.* Multiple GTP-binding proteins in sea urchin sperm: Evidence for Gs and small G-proteins. *Dev. Growth Differ.* **37**, 173–181 (1995).
51. Lapraz, F. *et al.* RTK and TGF- β signaling pathways genes in the sea urchin genome. *Dev. Biol.* **300**, 132–152 (2006).
52. Song, J. L. *et al.* Select microRNAs are essential for early development in the sea urchin. *Dev. Biol.* **362**, 104–113 (2012).
53. Peng, J. *et al.* Gonadal transcriptomic analysis and differentially expressed genes in the testis and ovary of the Pacific white shrimp (*Litopenaeus vannamei*). *BMC Genomics* **16**, 1006 (2015).
54. Vacquier, V. D., Loza-Huerta, A., García-Rincón, J., Darszon, A. & Beltrán, C. Soluble adenylyl cyclase of sea urchin spermatozoa. *Biochim. Biophys. Acta* **1842**, 5–12 (2014).
55. Meng, X.-L., Liu, P., Jia, F.-L., Li, J. & Gao, B.-Q. De novo Transcriptome Analysis of *Portunus trituberculatus* Ovary and Testis by RNA-Seq: Identification of Genes Involved in Gonadal Development. *PLoS One* **10**, e0128659 (2015).
56. Hunter, T. Signaling—2000 and beyond. *Cell* **100**, 113–127 (2000).
57. Hochachka, P. & Somero, G. *Biochemical adaptation: mechanism and process in physiological evolution*. (Oxford University Press Inc, 2002).
58. Das, J. The role of mitochondrial respiration in physiological and evolutionary adaptation. *BioEssays* **28**, 890–901 (2006).
59. Bonilla, E. & Xu, E. Y. Identification and characterization of novel mammalian spermatogenic genes conserved from fly to human. *Mol. Hum. Reprod.* **14**, 137–142 (2008).
60. Johnson, C., Buster, D. & Scholey, J. Light chains of sea urchin kinesin identified by immunoadsorption. *Cell Motil. Cytoskeleton* **16**, 204–213 (1990).
61. O'Donnell, L. & O'Bryan, M. K. Microtubules and spermatogenesis. *Semin. Cell Dev. Biol.* **30**, 45–54 (2014).
62. Sperry, A. The dynamic cytoskeleton of the developing male germ cell. *Biol. Cell* **104**, 297–305 (2012).

63. Yi, Y. *et al.* Ubiquitin C-terminal hydrolase-activity is involved in sperm acrosomal function and anti-polyspermy defense during porcine fertilization. *Biol. Reprod.* **77**, 780–793 (2007).
64. Yokota, N. & Sawada, H. Sperm proteasomes are responsible for the acrosome reaction and sperm penetration of the vitelline envelope during fertilization of the sea urchin *Pseudocentrotus depressus*. *Dev. Biol.* **308**, 222–231 (2007).
65. White-Cooper, H. & Bausek, N. Evolution and spermatogenesis. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **365**, 1465–1480 (2010).
66. Sea Urchin, Genome Sequencing Consortium. The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science (80-)*. **314**, 941–952 (2006).
67. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUASt: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).
68. Suzek, B., Huang, H., McGarvey, P., Mazumder, R. & Wu, C. UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* **23**, 1282–1288 (2007).
69. Parra, G., Bradnam, K. & Korf, I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**, 1061–1067 (2007).
70. Conesa, A. *et al.* Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676 (2005).
71. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
72. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
73. Ye, J. *et al.* WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res.* **34**, W293–W297 (2006).
74. Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C. & Kanehisa, M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **35**, W182–W185 (2007).
75. Wang, Y., Coleman-Derr, D., Chen, G. & Gu, Y. Q. OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res.* **43**, W78–W84 (2015).

Acknowledgements

This study was financial supported by Fondo de Fomento al Desarrollo Científico y Tecnológico (FONDEF) with the grant N. D09I1065. JDGE was supported by FONDECYT-Postdoctoral grant no. 3130381. We also acknowledge the support of the FONDAP 15150003 and Millennium Nucleus Initiative NC120086.

Author Contributions

J.D.G.-E. and L.C. conceived and designed the study. R.S. and P.B. sampled the biological material and carried out the DNA extractions. J.D.G.-E. and R.S. performed the bioinformatic analyses. J.D.G.-E. and L.C. drafted the manuscript. All authors read, approved, and contributed to the final manuscript.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Gaitán-Espitia, J. D. *et al.* Functional insights into the testis transcriptome of the edible sea urchin *Loxechinus albus*. *Sci. Rep.* **6**, 36516; doi: 10.1038/srep36516 (2016).

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



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016