MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2016) **16**, 364–376 doi: 10.1111/1755-0998.12451

A comprehensive transcriptome of early development in yellowtail kingfish (*Seriola lalandi*)

A. PATEL, P. DETTLEFF, E. HERNANDEZ and V. MARTINEZ
FAVET-INBIOGEN, Faculty of Veterinary Sciences, University of Chile, Avda. Santa Rosa, 11735 Santiago, Chile

Abstract

***Seriola lalandi* is an ecologically and economically important species that is globally distributed in temperate and subtropical marine waters. The aim of this study was to identify large numbers of genic single nucleotide polymorphisms (SNPs) and differential gene expression (DGE) related to the early development of normal and deformed *S. lalandi* larvae using high-throughput RNA-seq data. A *de novo* assembly of reads generated 40 066 genes ranging from 300 bases to 64 799 bases with an N90 of 788 bases. Homology search and protein signature recognition assigned gene ontology (GO) terms to a total of 15 744 (39.34%) genes. A search against the Kyoto Encyclopedia of Genes and Genomes Pathway database (KEGG) retrieved 6808 KEGG orthology (KO) identifiers for 10 520 genes (26.25%), and mapping of KO identifiers generated 337 KEGG pathways. Comparisons of annotated genes revealed that 1262 genes were downregulated and 1047 genes were upregulated in the deformed larvae group compared to the normal group of larvae. Additionally, we identified 6989 high-quality SNPs from the assembled transcriptome. These putative SNPs contain 4415 transitions and 2574 transversions, which will be useful for further ecological studies of *S. lalandi*. This is the first study to use a global transcriptomic approach in *S. lalandi*, and the resources generated can be used further for investigation of gene expression of marine teleosts to investigate larval developmental biology. The results of the GO enrichment analysis highlight the crucial role of the extracellular matrix in normal skeleton development, which could be important for future studies of skeletal deformities in *S. lalandi* and other marine species.**

Keywords: genetic marker, skeleton deformities, transcriptome, yellowtail kingfish (*Seriola lalandi*)

Received 13 March 2015; revision received 20 July 2015; accepted 27 July 2015

Introduction

Yellowtail kingfish (*Seriola lalandi*), a secondary and tertiary consumer in marine ecosystems, is a species with a worldwide distribution that covers the coast of Australia, New Zealand, Japan, south East China Sea, the Mediterranean Sea, the Pacific coast of America and the northern coast of Chile. Commercial aquaculture production of this species has been successfully established in Japan, Australia, New Zealand and recently in the north of Chile. While more attention has been focused on the commercial exploitation of this species, the assessment and maintenance of the genetic diversity among wild and cultured stocks are critically important and need immediate consideration. Existing genomic resources of *S. lalandi* are insufficient to support sustainable aquaculture activities of this species. In this regard, some initiatives have been undertaken to generate genomic resources of this fish, such as determination of karyotype (Chai *et al.* 2009), construction of an incomplete linkage map (Ohara *et al.* 2005), characterization of a dozens of simple sequence repeat (SSR) markers developed from the *Seriola dumerili* genome (Renshaw *et al.* 2012) and development of 31 genic SSRs from *S. lalandi* (Whatmore *et al.* 2013). There are only 12 mitochondrial genes and 109 protein sequences of yellowtail kingfish deposited in NCBI GenBank.

Although neutral markers have provided new insights into the demographic history and evolution of many populations, the missing element is an understanding of the dynamics of genes that affect fitness (Morin *et al.* 2004). RNA-seq, a transcriptome profiling approach that uses parallel sequencing technology, has emerged as a cost-effective and powerful methodology for gene discovery (Mortazavi *et al.* 2008) and identification of differentially expressed genes, with a greater dynamic range than microarrays (Marioni *et al.* 2008). A benefit of genetic marker discovery with RNA-seq is that the data obtained by sequencing can also be used for

Correspondence: Victor Martinez, E-mail: vmartine@u.uchile.cl, www.genetica-animal.uchile.cl

identifying genes and gene networks that may be seminal in physiology and adaptation (Barshis *et al.* 2013; Schwartz & Bronikowski 2013). Use of genic markers to discern candidate loci for spatial adaptation has been well described in Atlantic salmon (*Salmo salar*; Vasemägi *et al.* 2005) and red abalone (*Haliotis rufescence*; De Wit & Palumbi 2013). Hence, the RNA-seq approach is being increasingly employed to discover genic markers (Davey *et al.* 2011) and to identify the molecular mechanisms involved in the differential phenotypic expression of genes for adaptation (Smith *et al.* 2013).

Teleosts are the most successful vertebrate group, with over 25 000 species, and inhabit virtually every aquatic habitat on the planet. Teleosts are considered to be the first vertebrate group to develop a bony skeleton, along with the molecular machinery necessary for its formation and maintenance. The larval to juvenile transition involves the development of most organs and tissues, the maturation of different physiological functions and the establishment of the immune system (Ferraresso *et al.* 2013). Therefore, the larval phase is crucial in the development of marine teleosts, and the highest mortalities during the entire life cycle occur during this period (Mazurais *et al.* 2011). Teleosts have been recognized as a suitable vertebrate model for understanding skeletogenesis in lower and higher vertebrates for both comparative and evolutionary purposes (Hall 2008; Witten & Huysseune 2009; Sire *et al.* 2009; Fernández *et al.* 2011).

The presence of skeletal deformities in farmed teleosts is a persistent worldwide problem in aquaculture, causing huge economic losses and raising animal welfare issues (Boglione *et al.* 2013a). Understanding the genes and gene networks that control skeletal development is essential for identifying the key genes that are crucial for adaptation (Marsh & Fielman 2005; Streelman *et al.* 2007). There are limited studies on transcriptome-wide expression patterns during the latter stage of larval development of cultured marine fish species, with the exception of a few reports on gilthead sea bream (*Sparus aurata*; Yúfera *et al.* 2012; Calduch-Giner *et al.* 2013), common sole (*Solea solea* L.; Ferraresso *et al.* 2013) and Atlantic bonito (*Sarda sarda*; Sarropoulou *et al.* 2014).

Seriola lalandi have anomalies of various skeletal tissues in a cultured environment mainly during the larval phase. The deformities observed in *S. lalandi* are fused vertebrae, bent lower jaw, shorter jaw, shorter or missing operculum and compacted body and tail (Kolkovski & Sakakura 2004; Cobcroft *et al.* 2004). Such deformities have been reported in other culture marine teleosts such as barramundi (*Lates calcarifer*), gilthead sea bream (*S. aurata*) and convict grouper (*Epinephelus septemfasciatus*; Cobcroft & Battaglene 2013). Several physiological, environmental, genetic, xenobiotic and nutritional factors have been linked to this problem (Lall & Lewis-McCrea 2007). Genetic factors have been reported, such as mutations, hybridization or inbreeding causing skeletal deformities in zebrafish (*Danio rerio*), rainbow trout (*Onchorhynchus mykiss*) and gilthead seabream (Boglione *et al.* 2013b). Limited work has been carried out in *S. lalandi* to understand the effects of nongenetic factors on skeletal development, such as the deleterious effects of food restriction during early development (Chen *et al.* 2007) and the effect of alternative lipids and temperature on growth (Collins *et al.* 2012). However, no global transcriptome-wide study addressing skeletal deformity has been reported to date in this fish species. Martínez-Montaño *et al.* (2014) have characterized FIVE developmental stages of *S. lalandi* (from hatching to juvenile). They identified the flexion stage (15–25 days posthatch), which is an important period of early development, where morphological changes such as upward bending of the notochord, appearance of the first fin rays and changes in swimming behaviour occur.

In this work, we performed large-scale transcriptome sequencing of *S. lalandi* to generate genomic resources for further skeletogenesis studies. We aimed to (i) discover large numbers of gene-associated markers, (ii) perform a functional annotation of the transcriptome on the flexion stage and (iii) detect a set of candidate genes associated with skeletal deformities during larval development. The putative genic markers presented here will be useful in population genetics and genome mapping studies of *S. lalandi*. The annotated transcriptome provides valuable resources for further functional genomic research in skeletogenesis of yellowtail kingfish and other Carangidae species.

Materials and methods

Sampling and library construction

To develop a global overview of the genes associated with the early development of *Seriola lalandi* at a nucleotide level of resolution, six normal and four deformed (which represent bent in lower jaw) larvae were sampled at 23 days posthatch (dph). The RNA was extracted from the body using TRIzol reagent (Life Technologies, Carlsbad, CA, USA) following the manufacturer's protocol. The quality of RNA was determined with an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA; details in the Appendix S1, Supporting information) and sequenced on the Illumina HiSeq™ 2000 platform. Preparation of cDNA libraries and sequencing were carried out at the Genomics and Bioinformatics Facility Center of University of Edinburgh (details in the Appendix S1, Supporting information).

Reads filtering and de novo Assembly

Raw sequence data were processed to remove reads of <101 bases, low-quality reads (average QC value of <30) and adaptors from the data set using fastx toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) commands. We used all reads across all samples of both groups (n = 10, individuals) as inputs for generating a single transcriptome assembly. The transcriptome was assembled using Trinity version_2013-02-25 (http://sourceforge.net/projects/trinityrnaseq/files/trinityrnaseq_r2013-02-25.tgz/download).

To overcome the challenges associated with high computing requirements necessary for *de novo* assembly in Trinity for large numbers of reads, we set –min_kmer_cov parameter to two in the Trinity.pl command, leaving the other parameters with their default settings. This allowed only those kmers occurring at least twice to be assembled by Inchworm. Because error-containing kmers were greatly enriched within the low-abundance kmer counts, setting the –min_kmer_cov value to two instead of the default parameter of one reduced the memory requirements and excluded the error-containing kmers, with minimal loss in the transcript reconstruction sensitivity of Trinity (Haas *et al.* 2013). We clustered the Trinity assembly using CAP3 (Huang & Madan 1999) where the longest contigs were clustered (at least 95% sequence similarity among contigs based on a word size of 8) to remove redundant contigs and isoforms, to generate a single references transcriptome. Additionally, a second *de novo* assembly was built using the same read data with CLC bio's *de novo* assembler (CLC Genomics Workbench, Denmark), with enabled graph parameter settings such as automatic word size, bubble size and autodetect paired distance. This assembly was not used for differential gene analysis.

Annotation and pathway assignments

The BLAST2GO software (Conesa *et al.* 2005) was used to predict gene function and GO terms assignment (Conesa & Götz 2008, p. 2). The transcriptome was queried against the GenBank nonredundant (nr) protein database of NCBI (ftp://ftp.ncbi.nih.gov/blast/db/FASTA/nr.gz downloaded on August, 2013) with BLASTX parameters using an expected value (e-value) threshold <0.001 and 20 blast hits. Mapping and annotation were performed using default parameters (e-value-hit-filter of $1.0E^{-6}$, Annotation cut-off of 55 and GO abstraction was set to five). Genes with enzyme codes (ECs) obtained from BLAST2GO were mapped to the KEGG metabolic pathway database. Afterwards, the InterPro database was interrogated by InterProScan for identifying functional signatures of protein domains, and the GO terms retrieved from homology search were merged with GO terms associated with InterProScan.

While GO offers tremendous advantages, it also has certain limitations. For example, GO terms do not correspond directly to known biological pathways (Mao *et al.* 2005). Therefore, to enrich the pathway annotation and to find the BRITE functional hierarchies, genes were submitted to the KEGG Automatic Annotation Server (KAAS) version 1.68× using a bidirectional best-hit homology (BHH) information method. KAAS annotates sequences with KEGG orthology (KO) identifiers, which represents an orthologous group of genes directly linked to an object in the KEGG pathways and BRITE functional hierarchy, where KEGG BRITE is a collection of manually created hierarchical text files capturing functional hierarchies of various biological objects, especially those represented as KEGG objects. In contrast to KEGG pathways, which are limited to molecular interactions and reactions, KEGG BRITE incorporates many different types of higher order relationships that exist in biological systems (e.g. genetic and environmental information processing, cellular processes and organism systems; Moriya *et al.* 2007).

SNP discovery and primer testing

The two assemblies obtained from Trinity and CLC bio (CLC Genomics Workbench (CLC Genomics Workbench, Denmark)) were used for SNP discovery. Reads of each sample were mapped against the assemblies, and SNP identification was carried out using the quality-based variant detection module of CLC with the following parameters: maximum gap and mismatch count = 2; minimum average quality = 15; minimum central quality = 20; minimum coverage = 10; minimum variant frequency (%) = 35.0; window length = 11. For filtering, we kept SNPs that were uniquely present in the middle of 200-bp regions. These regions were validated *in silico* by aligning one assembly to the other at these specific sites using BLASTN, and those sites with a perfect match were kept for further study.

To validate a subset of these SNPs, we used contigs that had a single SNP for primer design using BATCHPRIMER3 software (You *et al.* 2008). Twenty-three loci were tested by high-resolution melting analysis (HRMA) using all of the individuals sequenced (10) and using an additional 38 *S. lalandi* individuals sampled from two geographical sites (location details and qPCR protocol in the Appendix S1, Supporting information). The observed (H_O), expected heterozygosities (H_E) and Hardy–Weinberg equilibrium (HWE) were calculated using GENEPOP software (Rousset 2008).

Identification of differentially expressed genes

To find differentially expressed genes of the deformed group compared to normal individuals, reads from each sample were mapped to the nonredundant reference transcriptome using CLC Genomics Workbench software (version 7.0.4, CLC bio). During read mapping, penalty for mismatch, insertion and deletion costs were set as default. Paired-end information was reserved, and a maximum of five hits for multimapping with random assignment were allowed. The length fraction was set to 0.7 and similarity was set to 0.9, which meant that at least 70% of the individual reads had at least 90% identity with the reference sequences to be mapped and aligned.

We merged mapping data from all samples and generated a single count table containing gene IDs and their corresponding read count. The original digital expression count table (without any normalization) was used for differential gene expression (DGE) in EDGER (Robinson *et al.* 2010). While executing EDGER, we removed genes that did not have at least one read per million in at least four samples (our smallest group had four samples). We used trimmed mean of *M* values (TMM) normalization of library size to adjust the RNA composition among different samples. We evaluated the relationship between samples in EDGER that compares the relationship between all pairs of samples in terms of biological coefficient of variation (BCV) before proceeding to computations of differential expression. Finally, differential expression between the normal and deformed groups was assessed for each gene using the exactTest () function of EDGER. The Benjamini–Hochberg method was used to control the false discovery rate (FDR) for multiple testing. We considered genes as significantly differentially expressed when the corrected *P*-values (FDR) were <0.01 and expression values differed by more than twofold.

Gene set enrichment analysis

To find overrepresented GO terms in our differentially expressed gene set compared to the full set of the annotated *S. lalandi* transcriptome, we performed enrichment analysis with the statistical analysis package GOSSIP (Blüthgen *et al.* 2004) implemented in BLAST2GO, which uses Fisher's exact test with Benjamini–Hochberg multiple testing correction of FDR. To filter out redundancy, overrepresented GO terms were summarized into a more representative subset of terms using REVIGO (Supek *et al.* 2011) by setting allowed similarity to medium (0.7), GO term database = UniProt and using the clustering algorithm of semantic similarity measure (SimRel).

Results and discussion

After quality assessment and data filtering, approximately 1188.64 million (594.32 M × 2) paired-end reads (mean quality score of 36.2 and mean GC content of 47.32%) were obtained from all libraries. The sequence reads were deposited in the NCBI Sequence Read Archive under Accession no. SRP055812 (SRA, http://www.ncbi.nlm.nih.gov/Traces/sra website). Read assembly resulted in 95 495 and 43 262 contigs with N90 of 641 bases and 725 bases for Trinity and CLC bio, respectively (Table 1). Trinity generated 58 456 components, where each component represented a group of related Trinity transcripts derived from the same *de* Bruijn graph and could be referred to as an individual gene model. CAP3 clustering of Trinity assembly generated 40 066 genes ranging from 300 bases to 64 799 bases with an N50 of 2705 bases and N90 of 788 bases, which is higher than the N50 found in *Trachinotus ovatus* (N50 = 2404; Zhenzhen *et al.* 2014), another member of the Carangidae family. The nonredundant *de novo* transcriptome assembly is available at the Dryad digital repository.

Annotation of genes

BLAST. Of 40 066 genes, 27 333 genes (68.13%) were blasted to known protein sequences in the 'nr' protein database (Table 2), while 12 733 (31.87%) genes did not find any hits with the configured BLASTX setting. There are diverse reasons for the lack of homology with known proteins. Biological reasons include a lack of annotated protein sequences of phylogenetically close species of *Seriola lalandi* in the database, sequences from untranslated regions (UTR) of mRNA, long noncoding RNA or orphan proteins. Additionally, other possible causes might include technical issues such as sequencing and assembly artefact.

Table 1 Assembly summary of *S. lalandi* transcriptome generated on 23 days post- hatch larvae

	Trinity	CLC bio	Genes (after clustering of Trinity assembly by CAP3)
Total length of sequence (bp)	142 422 909	71 208 747	69 721 586
GC %	45.99	45.27	45.26
Total number of contigs (>300 bp)	95 495	43 262	40 066
N50 stats	2356 bases	2581 bases	2705 bases
N90 stats	641 bases	725 bases	788 bases

Table 2 Summary of homology blast search, mapping and annotation results of *S. lalandi* transcriptome expressed on 23 days post-hatch

Type	Number of genes
Without Blast Hits	12 733
Without Mapping	17 517
Failed To Pass The Annotation Threshold	2666
Annotated Sequences from BLASTX	7150
Annotated Sequences from InterProScan	8694
Final Annotations (after merging BLASTX and InterPro)	15 744
Total number of genes	40 066

Various statistics were performed to evaluate BLASTX results to better adjust the annotation rules so that reliable GO terms were retrieved for later steps (Fig. 1). The top-blast hit species distribution showed that the majority of genes matched to *Oreochromis niloticus*, which belongs to the same order as *S. lalandi* (i.e. Perciformes). The other three species that matched with most genes (i.e. *Oryzias latipes*, *Takifugu rubripes* and *Tetraodon nigroviridis*) belong to the superorder Acanthopterygii, which includes the order Perciformes, whereas *Danio rerio* belongs to the superorder Otocephala. The order Cypriniformes placed 6th place on the list of top-hit species. The top-hit species distribution reflects the known phylogenetic affiliations of *S. lalandi*, thus further validating its evolutionary relationship to these species.

Mapping. There were 9816 (35.91%) genes that could be mapped to GO terms, while 17 517 (64.09%) genes (Table 2) could not be mapped to GO terms. This result may demonstrate a limitation of the NCBI nr protein database, which does not have extensive annotation. Evidence from code distribution showed that most GO terms (61.84%) were inferred from electronic annotations; even so, 9286 (8.54%) of the GO terms came from direct assays (Fig. S1-A, Supporting information), and most annotations were obtained from the UniProt Knowledge Base (Fig. S1-B, Supporting information).

Annotation. There were 2666 (27%) genes (Table 2) that had a mapping result but failed to retrieve a GO term. This could be a result of low similarity coverage between query and hits, or evidence codes that were of low reliability and, hence, failed to pass the annotation threshold value (i.e. 55). Homology-based annotation assigned gene functions to 7150 genes, while subsequent mapping between the InterPro domain and GO terms recovered functional information from an additional 8694 genes. Ultimately, the *S. lalandi* transcriptome resulted in 15 744 successfully GO annotated sequences, with a total of 77 592 GO terms at a mean GO level (distance of the GO term to the ontology root term) of 4.38 (Fig. S1-C, Supporting information). To obtain an outline of transcriptome annotation, three categories of GO terms were depicted (Fig. 2) at level two of GO's direct acyclic graph. The annotated transcriptome is available at the Dryad digital repository.

A total of 3490 sequences linked to EC numbers, and they were mapped to six different enzyme classes (Fig. S2, Supporting information) consisting of 124 metabolic KEGG pathways. KAAS retrieved 6808 KO identifiers for 10 520 genes (26.25% of 40 066), and mapping of KO identifiers generated 337 KEGG pathway maps composed of several functional categories (Table 3). We reconstructed some of these pathways to obtain an overview of the completeness of the assembled transcriptome and to ascertain the representation of some important signalling pathways involved in regulation of skeletal development (reviewed by Karsenty 2008; Lefebvre & Bhattaram 2010; Lattanzi & Bernardini 2012) and innate immunity of fishes (reviewed by Rebl *et al.* 2010). For example, the transforming growth factor-beta (TGF-β, Fig. S3-A, Supporting information) superfamily that is composed of over 40 members, including TGF-βs, nodal, activin, and bone morphogenetic proteins (BMPs), plays vital roles in development and tissue homoeostasis by regulating cell proliferation, differentiation, apoptosis and migration. The BMPs are known for their remarkable roles as instructive signals during embryogenesis and in the maintenance and repair of bone and other tissues (Massagué & Chen 2000), while Toll-like receptors (TLRs, Fig. S3-B, Supporting information) are well known for their essential roles in the activation of innate immunity by recognizing specific patterns of microbial components (Takeda & Akira 2004). The nearly completed reconstructed signalling pathway maps of TGF-β and TLRs validated that our transcriptome assembly covers a wide range of nearly complete biological pathways. This information will be an invaluable resource for further characterization of domain specificity in marine teleosts.

SNP discovery and primer testing

A total of 112 254 candidate SNPs were detected initially, but after a quality check, only 6998 SNPs (6.23% of the total) passed the filter criteria. The variants file is available in the Dryad digital repository. We tested 23 SNPs for validation of the procedures; the rest will be used for developing a high-throughput SNP platform (in preparation). Of the 23 tested primer pairs, one was monomorphic, the genotype of

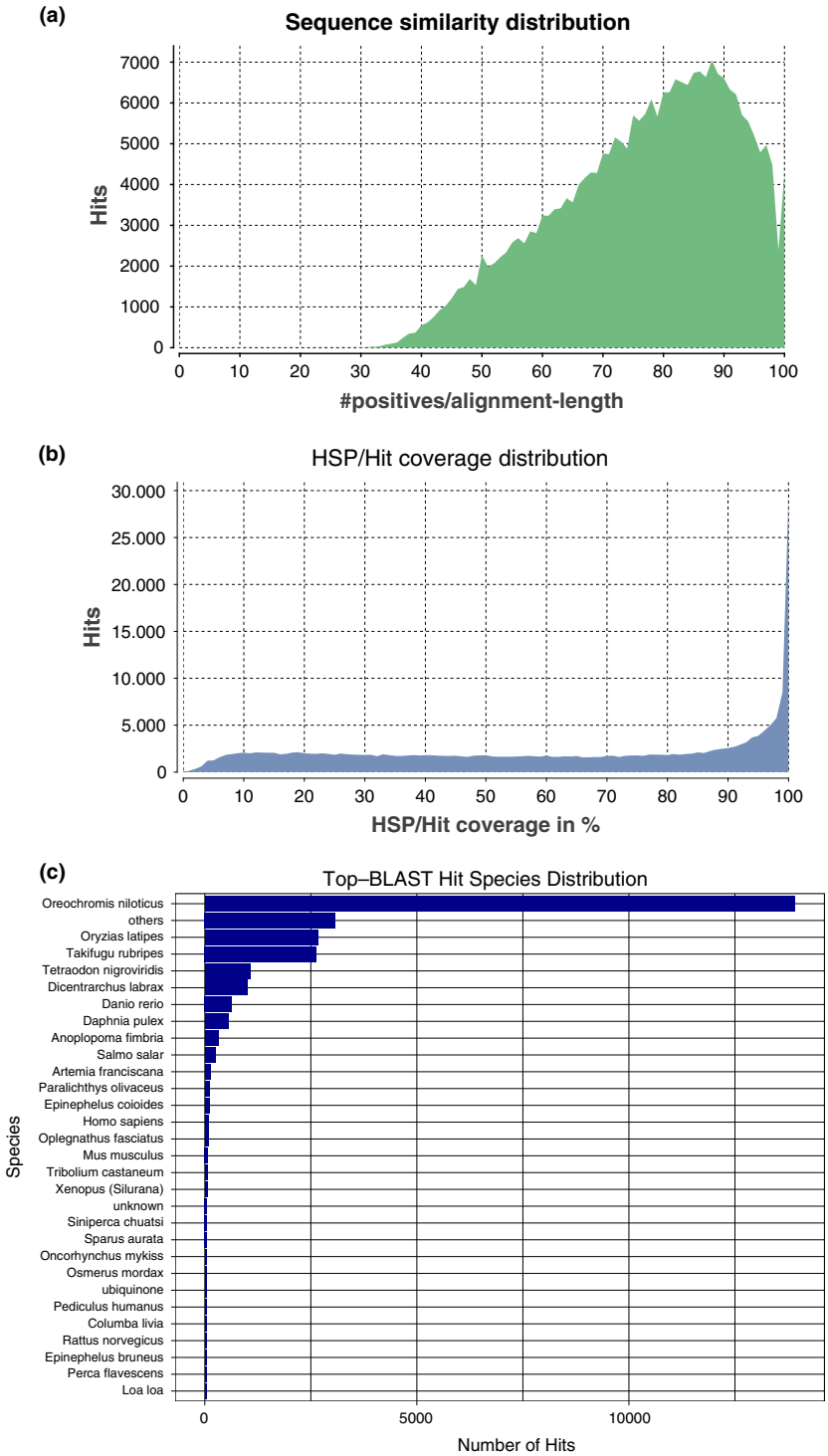

Fig. 1 Overview of homology search of *S. lalandi* genes against the NR database. (a) The sequence similarity distribution chart displays sequence similarities in percentages (Y axis = number of hits, X axis = number of positive aligned base-pairs divided by the alignment length × 100); (b) The HSP (high-scoring segment pairs) per hits distribution (percentages) represent the coverage between the HSPs and its corresponding hits and gives a measure of the accuracy of the sequence alignment; (c) Species distribution of the top BLAST hits for all homologous sequences.

five primers could not be resolved by HRMA, and the remaining 17 loci (74% success) were successfully validated in both groups (Fig. S5, Supporting information). A total of two markers (Sela10532_2771 and Sela11358_281) in Group-1 and two (Sela10664_1141 and Sela10731_3296) in Group-2 deviated from the HWE (Table 4).

Analyses of differentially expressed genes

We obtained quality-filtered 527 456 900 and 447 128 444 reads from normal and deformed individuals, respectively. Nearly 83% of the reads mapped back to the reference transcriptome (Table S1, Supporting information), where only <2.5% of the fragments were

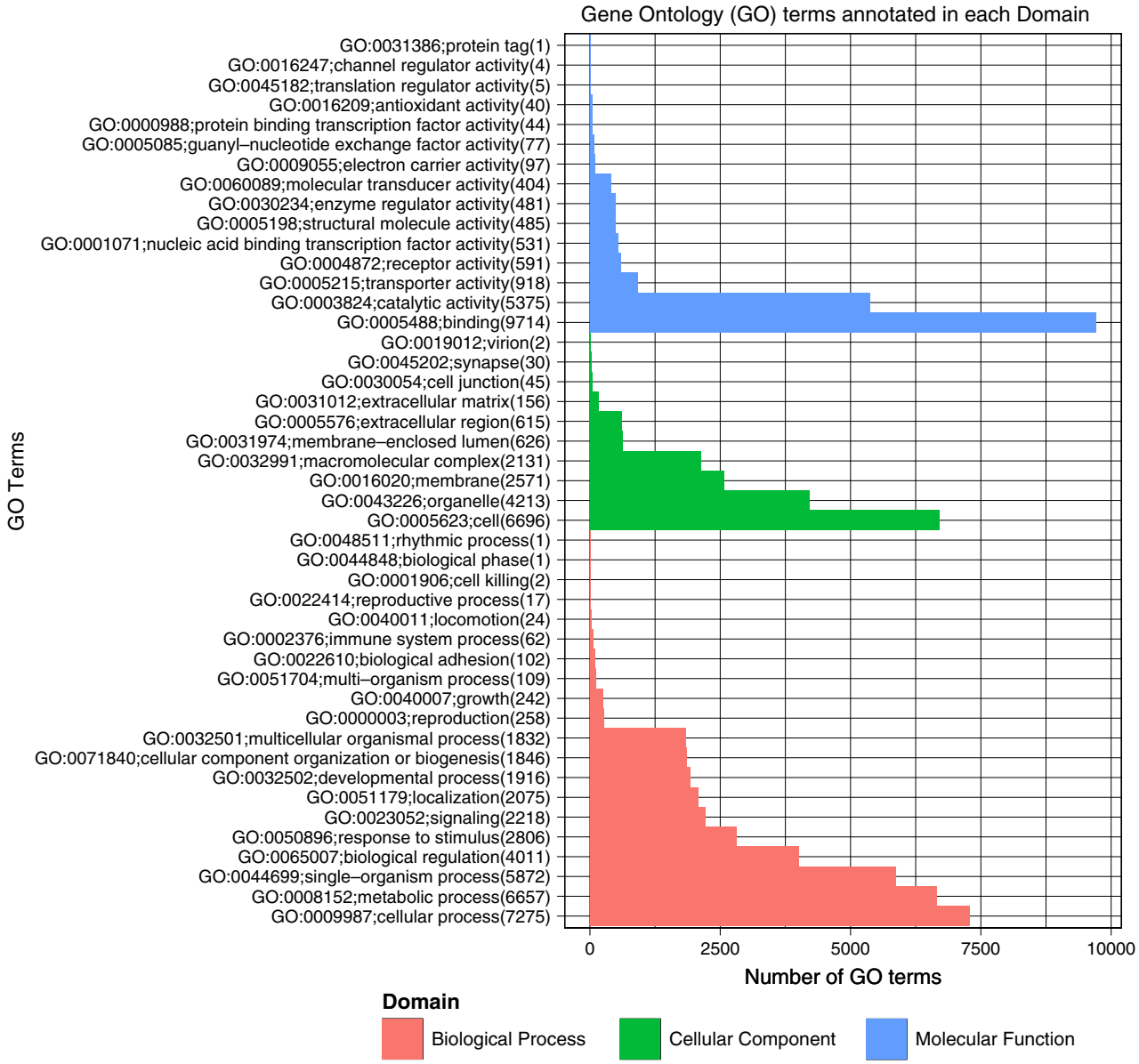

Fig. 2 Histogram of GO classifications (at level-2) of *S. lalandi* genes. Results are summarized for three main GO domains: Cellular Component, Molecular Function and Biological Process; numbers in parentheses are number of genes having respective GO terms.

nonmapped specifically. The high percentages (~80%) of unambiguous mapping and the uniformity of read counts across all samples led to a greater reliability in downstream differential digital gene expression analysis. The complete expression count table is available at the Dryad digital repository.

A total of 2371 noninformative features (genes) that had less than one count per million (CPM) in at least four samples were removed from the count table. The mapped read counts (libraries size) ranged from ~13 to ~18 million reads per sample; therefore, one CPM corresponded to nearly 15 reads. RNA-seq measures the relative abundance of each feature in each sample. When a small number of genes are highly expressed in one sample but not in another, the highly expressed genes represent a major proportion of the total library size, causing the remaining genes to be undersampled in that sample. Therefore, these genes falsely seem to be downregulated in that sample.

EDGER normalizes RNA composition using the trimmed mean of M values (TMM) between each pair of samples as a scaling factor and recomputes the library sizes (effective library size) (Robinson & Oshlack 2010). Six of our libraries had a scaling factor less than one (~0.96) and four had more than one (Table S2, Supporting information). This demonstrated that many reads

Table 3 KEGG biochemical mapping and functional categorization of *Seriola lalandi* transcriptome

	No. of KO
KEGG categories represented	
Carbohydrate metabolism	263
Energy metabolism	148
Lipid metabolism	276
Nucleotide metabolism	182
Amino acid metabolism	292
Metabolism of other amino acid	74
Glycan biosynthesis and metabolism	215
Metabolism of cofactors and vitamins	120
Metabolism of terpenoids and polyketides	27
Biosynthesis of other secondary metabolites	232
Xenobiotics biodegradation and metabolism	67
Genetic information processing	
Transcription	164
Translation	365
Folding, sorting and degradation	373
Replication and repair	194
Environmental information processing	
Membrane transport	32
Signal transduction	1837
Signalling molecules and interaction	349
Cellular processes	
Transport and catabolism	372
Cell motility	123
Cell growth and death	362
Cell communication	388
Organismal systems	
Immune system	745
Endocrine system	699
Circulatory system	183
Digestive system	307
Excretory system	104
Nervous system	530
Sensory system	97
Development	156
Environmental adaptation	92

were assigned to a few highly expressed genes in those six libraries and that less sequencing depth was available for the remaining genes (Chen *et al.* 2014). The degree of variability between the biological replicates considerably affects the calling of differential expression. For example, a single outlying sample could drive increased dispersion estimates and compromise the discovery of differentially expressed features (Anders *et al.* 2013). A multiple dimensional scaling (MDS) plot (Fig. 3A) clearly showed that two biological replicates from the normal group were dissimilar within their group. This could be due to an error in library preparation or biological reasons, such as an infection of specific samples. Therefore, we removed these two samples from further analysis to compute DGE and rechecked sample relationships (Fig. 3B).

EDGER uses negative binomial-based models to calculate the quadratic mean–variance (dispersion) relationship and empirical Bayes methods to moderate the degree of dispersion across features (genes). The common dispersion estimation (where each gene was assigned the same dispersion estimate) of EDGER was 0.103, and it was equal to an overall BCV of 32% (Fig. 4), whereas the gene-wise dispersions varied between 0.02 and 1.52. The BCV was moderately high, but it was not atypical given that our samples were not collected from a specific tissue or a particular cell line. We identified 2309 differentially expressed genes that were significant at a 0.01 FDR and that had at least a twofold change between two groups (Fig. S4). Of these, 1262 genes (54.65%) were downregulated and 1047 genes (45.34%) were upregulated in the group of deformed larva compared to that of phenotypically normal individuals.

Gene ontology analysis

Enrichment of functionally related GO terms among the DE data sets and clustering of significant GO term by REVIGO identified 6 GO terms (Table S3, Supporting information) in downregulated genes in the deformed group. However, no GO terms were enriched in upregulated genes. The REVIGO results are available at the Dryad digital repository.

The terms associated with the proteinaceous extracellular matrix (ECM; GO:0005578) encompassed the highest number of genes (58). These terms included enzymes or enzymatic precursors, such as the 72-kDa type IV collagenase precursor (matrix metallopeptidase-2—MMP-2); members of the disintegrin and metalloproteinase family; structural proteins such as collagen alpha-1, 2, 5 and 10; biglycan; ECM protein fras1; noncollagenous glycoprotein laminin subunit beta-1; SPARC-related modular calcium-binding protein 1; basement membrane-specific heparan sulphate proteoglycan core protein (HSPG2); and cartilage-associated protein (CRTAP) precursor.

Enriched biological processes related to the cell cycle were associated with several genes, including ribonucleoside-diphosphate reductase (RNR), replication factor C subunit (RFC) 3 and RFC4, origin recognition complex (Orc) 4 and Orc5 and members of DNA replication licensing factor (MCM2, 3 and 5). Cell proliferation is a metabolically demanding process, as it requires active reprogramming of cellular bioenergetic pathways (Mauro *et al.* 2011). Consequently, living organisms and cells continuously adapt to changes in their environment. Those changes are particularly sensitive to fluctuations in the availability of energy substrates. The cellular transcriptional machinery and its chromatin-associated proteins integrate environmental inputs to mediate

Table 4 Characterization of seventeen novel SNPs loci of Yellowtail Kingfish (*S. lalandi*)

Marker	Orientation	Tm	Primer	A	Group-1			Group-2		
					H_E	H_O	*P*	H_E	H_O	*P*
Sela10065_274	F	59.8	GCCAGGACTGATCACAGACA	C/T	0.44	0.46	1.00	0.36	0.38	1.00
	R	59.4	CGTGCAGCCAATTACTGAAG							
Sela10162_2695	F	59.6	ACCGGACAAAAAGTTCTGGA	C/T	0.51	0.46	0.692	0.49	0.46	1.00
	R	59.4	ACTGGGAAGCAGCCAAAA							
Sela10532_2771	F	60.0	TGGTTGGGCACTGTGTTAAA	C/T	0.51	0.79	0.011*	0.25	0.29	1.00
	R	59.8	TTCACATTCATTTCCAGGTGAC							
Sela10664_1141	F	57.9	GTCAGATGCAGGAGTTGAAAA	A/T	0.48	0.50	1.00	0.25	0.13	0.047*
	R	57.0	CGTTTCATCATCTCATTCCA							
Sela1068_3136	F	60.1	GAGGTGGAGGACCAACTTCA	A/T	0.47	0.46	1.00	0.36	0.38	1.00
	R	60.0	GGTTGCAGTGGTCTCGGTAT							
Sela10731_3296	F	60.1	CAGATGGAGTTTTGCCCAGT	A/T	0.48	0.58	0.395	0.51	0.87	0.0008*
	R	59.5	CACACCAAGCTGTTCAGGAG							
Sela10851_748	F	59.9	CACAAGTCATCAGCATCAGTCA	A/G	0.48	0.50	1.00	0.49	0.63	0.215
	R	59.9	CCACCAGTCTATTGCCCCTA							
Sela10977_1945	F	59.9	CCATTTAGGCCCAAAGTCAA	G/T	0.50	0.33	0.201	0.51	0.58	0.678
	R	60.0	CCAGTATTTCTGCACCATTTCA							
Sela11050_869	F	59.8	TGTCCATGTGGTGTCTGGAT	C/T	0.47	0.38	0.394	0.50	0.46	0.698
	R	60.4	ACGGAGATTCCTTCCAGGTC							
Sela1098_2386	F	60.2	TGACTTGCTGTCCTGTCTGC	A/G	0.26	0.22	0.411	0.44	0.38	0.634
	R	59.3	AAGAAAACATGGATCCCTTCC							
Sela11153_2022	F	59.7	AGGAGGAGGGCTATTTCCTG	C/T	0.16	0.08	0.124	0.28	0.25	0.499
	R	59.9	CCCTTGTTGGGTAGTCCTCA							
Sela11218_2450	F	59.9	ATGCAACACAACCACCAAAA	A/G	0.41	0.36	0.625	0.49	0.61	0.374
	R	60.6	CTAGCAGCTCAGGCCAACAT							
Sela11304_5489	F	59.7	TGCGCTCACTCTGAGGAATA	G/T	0.19	0.21	1.00	0.12	0.13	1.00
	R	60.5	GCTTTTGGTATTTCTAACCCTTCC							
Sela11358_281	F	60.2	GCGCTAACCTAGCCTCACTG	C/T	0.48	0.25	0.026*	0.45	0.25	0.058
	R	60.2	GCGACAAAAACAGGGTGAGT							
Sela1142_1957	F	59.1	TCAGCTGCAGGATTTGACTC	C/G	0.38	0.25	0.112	0.16	0.17	1.00
	R	60.1	CTGCTGTGACCTGCAGACAT							
Sela1145_419	F	60.3	GGCTAAAACTCGCTGCTCTG	A/G	0.34	0.33	1.00	0.32	0.38	1.00
	R	59.9	TAGCCCTGAGCAGGTCATCT							
Sela10237_1235	F	58.7	GGGGGAACAAAGATAAAAACC	C/T	0.16	0.08	0.127	0.22	0.17	0.297
	R	58.6	CGTTCCAAACTGGGTAATCTC							

Tm, Melting temperature; A, allele variation; Ho, observed heterozygosity; He, expected heterozygosity; P, exact P value for Hardy-Weinberg equilibrium test.
*P = deviated from HWE ($P < 0.05$)

homoeostatic responses through gene regulation (Gut & Verdin 2013).

Completion of a cell cycle carries a high energetic cost; thus, the ability of a cell to sense its energy status and influence cell cycle progression is important for its survival. The AMP-activated protein kinase (AMPK) signalling pathway is an energy-sensing pathway that is conserved throughout eukaryotic evolution. The critical role of AMPK signalling in nutrient-dependent G1 arrest of the cell cycle has been established (Tobin & Saito 2012). The summary of genes found with enriched GO terms supports a logical presumption that at the time of sampling (23 dph), the deformed larvae were feeble and energy deficient due to a decreased food intake ability. Consequently, to maintain energy homoeostasis for survival, deformed larvae had already triggered the arrest of cell proliferation and growth by downregulating genes implicated in cell division.

Skeletal disorders in farmed fishes have been associated with a complex and poorly understood relationship to nutrition, environment and genetic factors (reviewed by Boglione *et al.* 2013b). The process of skeleton formation or skeletogenesis begins in the vertebrate embryo when multipotent mesenchymal cells arise from ectoderm and mesoderm and then migrate to specific locations in the body and commit to a skeletal fate (reviewed by Lefebvre & Bhattaram 2010). The origin of skeletogenic cells such as chondrocytes

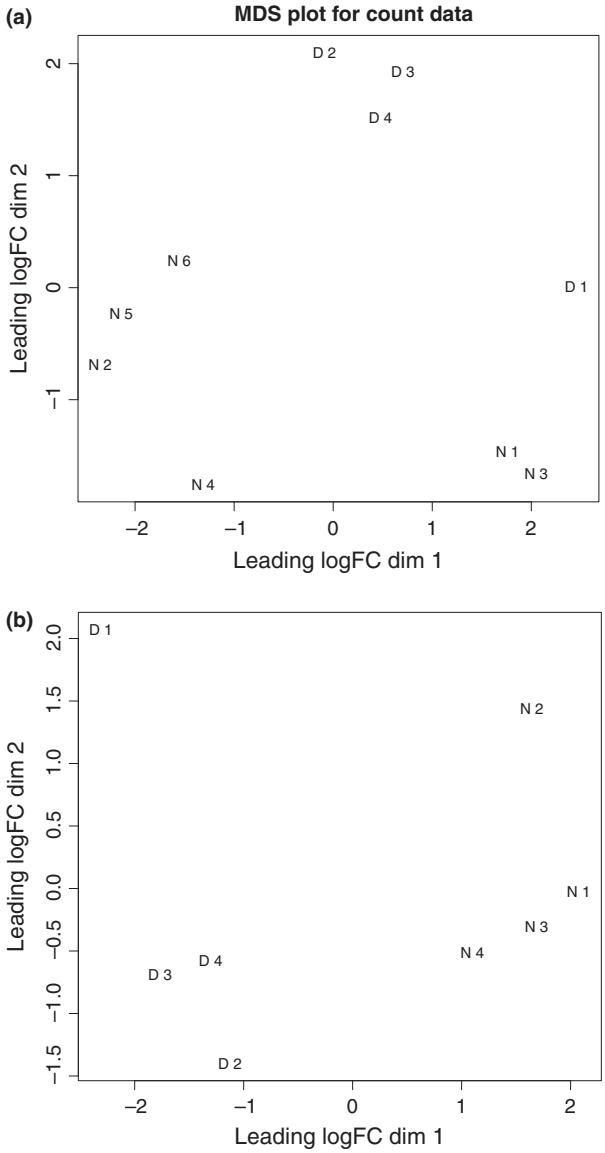

Fig. 3 A multiple dimensional scaling (MDS) plot of *S. lalandi* libraries (a) before removal of dissimilar samples; (b) after removal of samples: The distance between each pair of samples is based on leading fold change, defined as the root-mean-square of the largest 500 log2-fold changes between that pair of samples. Figure (a) shows that biological replicates normal-1 (N-1) and normal-3 (N-3) samples are grouped with deformed group in the first dimension. Figure (b) displays well separated biological replicates between groups in the first dimension after removal of two dissimilar normal samples.

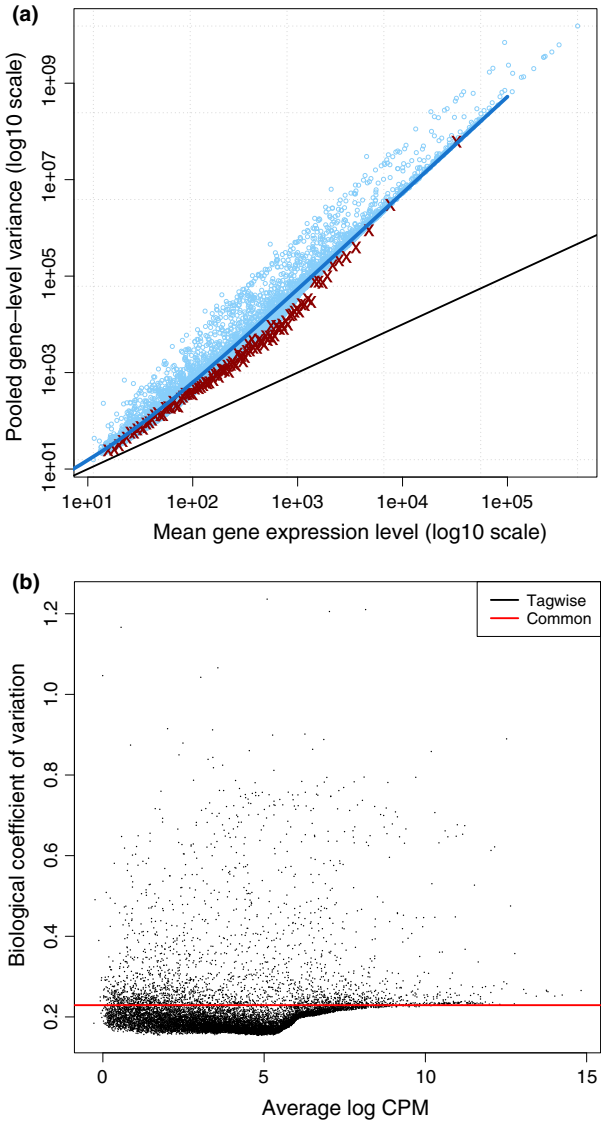

Fig. 4 Mean-variance relationship and dispersion plots. (a) plotMeanVar of edgeR function was used to explore the mean-variance relationship; each dot represents the estimated mean and variance for each gene, with binned variances as well as the trended common dispersion overlaid. (b) plotBCV illustrates the relationship of biological coefficient of variation versus mean log CPM.

and osteoblasts was tracked back to the onset of organogenesis, when the neural crest undergoes epithelial-to-mesenchymal transformation and develops into neuronal cells, melanocytes and skeletogenic cells. Critical roles have been established in early developing embryos for bone morphogenetic protein-2 (BMP-2) in dorsal-ventral patterning (Kishimoto *et al.* 1997), transcription factor Sox9 in cartilage formation (Yan *et al.* 2002), Runx2 and Osx in bone formation (Li *et al.* 2009) and antagonistic effects of noggin in joint formation (Fürthauer *et al.* 1999). In the present study, specific genes that induce cartilage and bone formation (BMP-2, logFC-1.21 and BMP-5, logFC-1.09) and metal ion bonding (collagen type I alpha 1 or COL1A1, logFC-3.43, and collagen type X alpha 1, logFC-2.66, and osteopontin-like protein, logFC-1.06) were down-regulated in deformed individuals.

Bone morphogenetic proteins are multifunctional growth factors within the TGF-β superfamily and are important for the development of bone (Rosen 2009). Among BMPs, BMP-2 plays a key role in bone development, inducing the differentiation of mesenchymal cells into osteoblast precursors and promoting the maturation of osteoblasts through the expression of Runx2/Cbfa1 (Marie *et al.* 2002). In contrast to Fernández *et al.* (2011), research in gilthead sea bream did not find any alternation of COL1A1 and BMP-2 expression in deformed larvae treated with hypervitaminosis Before 60 dph, we found a decrease in expression of COL1A1 and BMP-2 in the deformed group at 23 dph. Downregulation of BMP-2 might have induced downregulation of the SPARC and COL1A1 genes through transcriptional regulation of Runx2/Cbfa1 (Shen *et al.* 2007). This indicates that temporal expression of essential genes related to skeletogenesis may be affected in a different way by environmental and nutritional factors.

A substantial tissue volume in organisms is the extracellular space, which is largely filled by an intricate network of macromolecules constituting the ECM. This network is predominantly composed of collagens, noncollagenous glycoproteins, hyaluronan and proteoglycans. The ECM acts as a scaffold for cells and serves as a reservoir for growth factors and cytokines that modulate cell activation status and turnover (Gentili & Cancedda 2009). Hynes (2009) and Watt & Huck (2013) have highlighted the importance of the ECM in terms of cell and tissue behaviour and skeletal development (Velleman 2000). The interactions between cells and collagen in the ECM are mediated by a variety of widely described receptors (i.e. integrins—ITGs), which, upon collagen binding, may activate other molecules, such as matrikines, matrix metalloproteases (MMPs), tissue inhibitor of MMP (TIMPs), cytokines and growth factors. These molecules are indispensable for remodelling, as well as inflammatory, immune or wound healing responses (Emsley *et al.* 2000; Vogel 2001; Fernández *et al.* 2011; Gao *et al.* 2014).

Some of the most prominent ECM proteins secreted by chondrocytes are the collagens, a single molecule that folds into chains to form a triple helical structure. Mutations in genes that encode these collagens have been associated with the presence of skeletal dysplasias in human (Krakow & Rimoin 2010) and zebrafish (Fisher *et al.* 2003). Therefore, enrichment of the GO term 'extracellular matrix' in our global differential transcriptome analysis highlights the importance of these cellular components in skeletal development processes of *S. lalandi*.

In conclusion, we performed the first global gene expression analysis of early development in yellowtail kingfish. A large number of potential genetic markers were identified from the assembled transcriptome. The SNPs identified will provide a substantial resource for molecular ecology studies and investigations of the effects of selection at the genome level in different yellowtail kingfish populations. Functional classifications of GO and KO terms have identified multiple biological process and signalling pathways, including those for the immune system. Our analysis also revealed DGEs that are essential for normal skeletogenesis and jaw development. Our study is beneficial for future investigations of jaw deformities in yellowtail kingfish, which is an important issue of fitness in natural populations and aquaculture production because of the relatively high incidence of deformities in this species.

Acknowledgements

We thank the Chilean aquaculture diversification Program (Programa de Diversificación de la Acuicultura Chilena, PDACH) CORFO-FIC (09PDAC-7020) for funding this grant and Acuícola del norte S.A for sample collection. Andrew Severin from the Iowa State University is acknowledged for his help in reviewing this manuscript.

References

Anders S, McCarthy DJ, Chen Y *et al.* (2013) Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nature Protocols*, **8**, 1765–1786.

Barshis DJ, Ladner JT, Oliver TA *et al.* (2013) Genomic basis for coral resilience to climate change. *Proceedings of the National Academy of Sciences*, **110**, 1387–1392.

Blüthgen N, Brand K, Cajavec B *et al.* (2004) Biological profiling of gene groups utilizing Gene Ontology. arXiv preprint q-bio/0407034.

Boglione C, Gavaia P, Koumoundouros G *et al.* (2013a) Skeletal anomalies in reared European fish larvae and juveniles. Part 1: normal and anomalous skeletogenic processes. *Reviews in Aquaculture*, **5**, S99–S120.

Boglione C, Gisbert E, Gavaia P *et al.* (2013b) Skeletal anomalies in reared European fish larvae and juveniles. Part 2: main typologies, occurrences and causative factors. *Reviews in Aquaculture*, **5**, S121–S167.

Calduch-Giner JA, Bermejo-Nogales A, Benedito-Palos L *et al.* (2013) Deep sequencing for de novo construction of a marine fish (*Sparus aurata*) transcriptome database with a large coverage of protein-coding transcripts. *BMC Genomics*, **14**, 178.

Chai X, Li X, Lu R, Clarke S (2009) Karyotype analysis of the yellowtail kingfish *Seriola lalandi lalandi* (Perciformes: Carangidae) from South Australia. *Aquaculture Research*, **40**, 1735–1741.

Chen BN, Qin JG, Carragher JF *et al.* (2007) Deleterious effects of food restrictions in yellowtail kingfish Seriola lalandi during early development. *Aquaculture*, **271**, 326–335.

Chen Y, Lun ATL, Smyth GK (2014) Differential expression analysis of complex RNA-seq experiments using edgeR. In: *Statistical Analysis of Next Generation Sequencing Data Frontiers in Probability and the Statistical Sciences* (eds Datta S, Nettleton D), pp. 51–74. Springer International Publishing, Switzerland.

Cobcroft JM, Battaglene SC (2013) Skeletal malformations in Australian marine finfish hatcheries. *Aquaculture*, **396–399**, 51–58.

Cobcroft JM, Pankhurst PM, Poortenaar C, Hickman B, Tait M (2004) Jaw malformation in cultured yellowtail kingfish (*Seriola lalandi*) larvae. *New Zealand Journal of Marine and Freshwater Research*, **38**, 67–71.

Collins GM, Ball AS, Qin JG, Bowyer JN, Stone DA (2012) Effect of alternative lipids and temperature on growth factor gene expression in yellowtail kingfish (*Seriola lalandi*). *Aquaculture Research*, **45**, 1236–1245.

Conesa A, Götz S (2008) Blast2GO: A comprehensive suite for functional analysis in plant genomics. *International Journal of Plant Genomics*, **2008**:619832.

Conesa A, Götz S, García-Gómez JM *et al*. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, **21**, 3674–3676.

Davey JW, Hohenlohe PA, Etter PD *et al*. (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, **12**, 499–510.

De Wit P, Palumbi SR (2013) Transcriptome-wide polymorphisms of red abalone (*Haliotis rufescens*) reveal patterns of gene flow and local adaptation. *Molecular Ecology*, **22**, 2884–2897.

Emsley J, Knight CG, Farndale RW, Barnes MJ, Liddington RC (2000) Structural Basis of Collagen Recognition by Integrin $\alpha 2\beta 1$. *Cell*, **101**, 47–56.

Fernández I, Darias M, Andree KB *et al*. (2011) Coordinated gene expression during gilthead sea bream skeletogenesis and its disruption by nutritional hypervitaminosis A. *BMC Developmental Biology*, **11**, 7.

Ferraresso S, Bonaldo A, Parma L *et al*. (2013) Exploring the larval transcriptome of the common sole (*Solea solea* L.). *BMC Genomics*, **14**, 315.

Fisher S, Jagadeeswaran P, Halpern ME (2003) Radiographic analysis of zebrafish skeletal defects. *Developmental Biology*, **264**, 64–76.

Fürthauer M, Thisse B, Thisse C (1999) Three different noggin genes antagonize the activity of bone morphogenetic proteins in the zebrafish embryo. *Developmental Biology*, **214**, 181–196.

Gao Y, Liu S, Huang J *et al*. (2014) The ECM-cell interaction of cartilage extracellular matrix on chondrocytes. *BioMed Research International*, **2014**, e648459.

Gentili C, Cancedda R (2009) Cartilage and bone extracellular matrix. *Current Pharmaceutical Design*, **15**, 1334–1348.

Gut P, Verdin E (2013) The nexus of chromatin regulation and intermediary metabolism. *Nature*, **502**, 489–498.

Haas BJ, Papanicolaou A, Yassour M *et al*. (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, **8**, 1494–1512.

Hall BK (2008) *Fins into Limbs: Evolution, Development, and Transformation*. University of Chicago Press, Chicago.

Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. *Genome Research*, **9**, 868–877.

Hynes RO (2009) The extracellular matrix: not just pretty fibrils. *Science*, **326**, 1216–1219.

Karsenty G (2008) Transcriptional control of skeletogenesis. *Annual Review of Genomics and Human Genetics*, **9**, 183–196.

Kishimoto Y, Lee KH, Zon L, Hammerschmidt M, Schulte-Merker S (1997) The molecular nature of zebrafish swirl: BMP2 function is essential during early dorso-ventral patterning. *Development (Cambridge, England)*, **124**, 4457–4466.

Kolkovski S, Sakakura Y (2004) Yellowtail kingfish, from larvae to mature fish—problems and opportunities. In: *Avances en Nutrición Acuícola VII. Memorias del VII Simposium Internacional de Nutrición Acuícola* (eds Cruz Suarez LE, Ricque Marie D, Nieto Lopez MG, Villareal D, Scholz U, Gonzalez M), pp. 109–125; 16–19 November 2004, Hermosillo, Sonora, Mexico.

Krakow D, Rimoin DL (2010) The skeletal dysplasias. *Genetics in Medicine*, **12**, 327–341.

Lall SP, Lewis-McCrea LM (2007) Role of nutrients in skeletal metabolism and pathology in fish—an overview. *Aquaculture*, **267**, 3–19.

Lattanzi W, Bernardini C (2012) Genes and molecular pathways of the osteogenic process. In: *Osteogenesis* (ed. Lin Y). InTech, Rijeka, Croatia.

Lefebvre V, Bhattaram P (2010) Vertebrate skeletogenesis. *Current Topics in Developmental Biology*, **90**, 291–317.

Li N, Felber K, Elks P, Croucher P, Roehl HH (2009) Tracking gene expression during zebrafish osteoblast differentiation. *Developmental Dynamics*, **238**, 459–466.

Mao X, Cai T, Olyarchuk JG, Wei L (2005) Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics (Oxford, England)*, **21**, 3787–3793.

Marie PJ, Debiais F, Haÿ E (2002) Regulation of human cranial osteoblast phenotype by FGF-2, FGFR-2 and BMP-2 signaling. *Histology and Histopathology*, **17**, 877–885.

Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y (2008) RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research*, **18**, 1509–1517.

Marsh AG, Fielman KT (2005) Transcriptome profiling of individual larvae of two different developmental modes in the poecilogonous polychaete *Streblospio benedicti* (Spionidae). *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, **304**, 238–249.

Martínez-Montaño E, González-Álvarez K, Lazo JP, Audelo-Naranjo JM, Vélez-Medel A (2014) Morphological development and allometric growth of yellowtail kingfish *Seriola lalandi* V. larvae under culture conditions. *Aquaculture Research*, **442**, 1–10.

Massagué J, Chen Y-G (2000) Controlling TGF-β signaling. *Genes & Development*, **14**, 627–644.

Mauro C, Leow SC, Anso E *et al*. (2011) NF-κB controls energy homeostasis and metabolic adaptation by upregulating mitochondrial respiration. *Nature Cell Biology*, **13**, 1272–1279.

Mazurais D, Darias M, Zambonino-Infante JL, Cahu CL (2011) Transcriptomics for understanding marine fish larval development. *Canadian Journal of Zoology*, **89**, 599–611.

Morin PA, Luikart G, Wayne RK, The SNP Workshop Group (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution*, **19**, 208–216.

Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M (2007) KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research*, **35**, W182–W185.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, **5**, 621–628.

Ohara E, Nishimura T, Nagakura Y *et al*. (2005) Genetic linkage maps of two yellowtails (*Seriola quinqueradiata* and *Seriola lalandi*). *Aquaculture*, **244**, 41–48.

Rebl A, Goldammer T, Seyfert H-M (2010) Toll-like receptor signaling in bony fish. *Veterinary Immunology and Immunopathology*, **134**, 139–150.

Renshaw MA, Buentello A, Gold JR (2012) Characterization of greater amberjack microsatellite markers in lesser amberjacks, yellowtail jacks, almaco jacks, and banded rudderfish. *North American Journal of Aquaculture*, **74**, 522–529.

Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*, **11**, R25.

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.

Rosen V (2009) BMP2 signaling in bone development and repair. *Cytokine & Growth Factor Reviews*, **20**, 475–480.

Rousset F (2008) genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.

Sarropoulou E, Moghadam HK, Papandroulakis N *et al*. (2014) The Atlantic Bonito (*Sarda sarda*, Bloch 1793) transcriptome and detection of differential expression during larvae development. *PLoS One*, **9**, e87744.

Schwartz TS, Bronikowski AM (2013) Dissecting molecular stress networks: identifying nodes of divergence between life-history phenotypes. *Molecular Ecology*, **22**, 739–756.

Shen Z-J, Kook Kim S, Youn Jun D *et al*. (2007) Antisense targeting of TGF-β1 augments BMP-induced upregulation of osteopontin, type I collagen and Cbfa1 in human Saos-2 cells. *Experimental Cell Research*, **313**, 1415–1425.

Sire J-Y, Donoghue PCJ, Vickaryous MK (2009) Origin and evolution of the integumentary skeleton in non-tetrapod vertebrates. *Journal of Anatomy*, **214**, 409–440.

Smith S, Bernatchez L, Beheregaray LB (2013) RNA-seq analysis reveals extensive transcriptional plasticity to temperature stress in a freshwater fish species. *BMC Genomics*, **14**, 375.

Streelman JT, Peichel CL, Parichy DM (2007) Developmental genetics of adaptation in fishes: the case for novelty. *Annual Review of Ecology, Evolution, and Systematics*, **38**, 655–681.

Supek F, Bošnjak M, Škunca N, Šmuc T (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One*, **6**, e21800.

Takeda K, Akira S (2004) TLR signaling pathways. *Seminars in Immunology*, **16**, 3–9.

Tobin DV, Saito RM (2012) Developmental decisions. *Cell Cycle*, **11**, 1666–1671.

Vasemägi A, Nilsson J, Primmer CR (2005) Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in Atlantic salmon (*Salmo salar* L.). *Molecular Biology and Evolution*, **22**, 1067–1076.

Velleman SG (2000) The role of the extracellular matrix in skeletal development. *Poultry Science*, **79**, 985–989.

Vogel WF (2001) Collagen-receptor signaling in health and disease. *European Journal of Dermatology: EJD*, **11**, 506–514.

Watt FM, Huck WTS (2013) Role of the extracellular matrix in regulating stem cell fate. *Nature Reviews Molecular Cell Biology*, **14**, 467–473.

Whatmore P, Nguyen NH, Miller A *et al.* (2013) Genetic parameters for economically important traits in yellowtail kingfish *Seriola lalandi*. *Aquaculture*, **400–401**, 77–84.

Witten PE, Huysseune A (2009) A comparative view on mechanisms and functions of skeletal remodelling in teleost fish, with special emphasis on osteoclasts and their function. *Biological Reviews*, **84**, 315–346.

Yan Y-L, Miller CT, Nissen RM *et al.* (2002) A zebrafish sox9 gene required for cartilage morphogenesis. *Development (Cambridge, England)*, **129**, 5065–5079.

You FM, Huo N, Gu YQ *et al.* (2008) BatchPrimer3: A high throughput web application for PCR and sequencing primer design. *BMC Bioinformatics*, **9**, 253.

Yúfera M, Halm S, Beltran S *et al.* (2012) Transcriptomic characterization of the larval stage in gilthead seabream (*Sparus aurata*) by 454 pyrosequencing. *Marine Biotechnology*, **14**, 423–435.

Zhenzhen X, Ling X, Dengdong W *et al.* (2014) Transcriptome analysis of the *Trachinotus ovatus*: identification of reproduction, growth and immune-related genes and microsatellite markers. *PLoS One*, **9**, e109419.

V.M. conceived the study, A.P. and V.M. conducted the bioinformatics analysis and wrote the first draft of the manuscript. P.D. and E.H. were involved in one or more processes including sample collection, genotyping and manuscript preparation.

Data accessibility

Filtered Illumina reads: NCBI SRA: SRP055812.

Transcriptome assembly, annotated transcriptome, variants (.vcf) files, expression count table, result of EDGER and REVIGO are deposited in Dryad:http://dx.doi.org/10.5061/dryad.sv31k.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Supplementary Materials

Table S1 (A) mapping statistics of *S. lalandi* reads generated by Illumina platform and aligned to the assembled genes by CLC bio aligner and (B) Mapping statistics of fragments (paired reads) of *S. lalandi* generated by Illumina platform and aligned to the assembled unigenes by CLC bio aligner

Table S2 Raw library and effective library size of *S. lalandi* transcriptome

Table S3 Result of GO enrichment test of *S. lalandi* transcriptome

Fig. S1 Mapping and annotation charts of *S. lalandi* genes (A): histogram of number of Gene Ontology terms with a given Evidence Code; (B) histogram of the number of GO obtained form each possible Database source of annotations; (C) GO Annotation Level Distribution: bar chart with the number of sequences having a given annotation.

Fig. S2 *S. lalandi* transcriptome bar chart depicting different categories of enzymes mapped to KEGG

Figure S3 (A) The transforming growth factor-beta (TGF-β) signaling pathway generated by KEGG pathway analysis representing the present and absent gene products in the transcriptome of Seriola lalandi. Gene products appearing in the *S. lalandi* transcriptome are represented in green boxes. (B) Kyoto Encyclopaedia of Genes and Genomes (KEGG) reference map of Toll-like receptor signaling pathway reconstructed in the *Seriola lalandi* transcriptome. Pathway elements in green boxes indicate the components represented by at least, one gene.

Fig. S4 M ('minus') versus A ('add') plots for RNA-seq data. plotSmear function of EDGER display the log-fold change (i.e., the log ratio of normalized expression levels between normal and deformed conditions) against the log counts per million (CPM).

Fig. S5 Normalized melt curves of (A) Sela10065_274 marker (C/T); (B) Sela10664_1141 (A/T); (C) Sela10851_748; (D) Sela10977_1945 (G/T); (E) Sela1142_1957 (C/G).