BMC
Genomics

RESEARCH ARTICLE

Open Access

CrossMark

Development and validation of a mixed-tissue oligonucleotide DNA microarray for Atlantic bluefin tuna, *Thunnus thynnus* (Linnaeus, 1758)

Željka Trumbić[1*], Michaël Bekaert[2], John B. Taggart[2], James E. Bron[2], Karim Gharbi[3] and Ivona Mladineo[4]

Abstract

Background: The largest of the tuna species, Atlantic bluefin tuna (*Thunnus thynnus*), inhabits the North Atlantic Ocean and the Mediterranean Sea and is considered to be an endangered species, largely a consequence of overfishing. *T. thynnus* aquaculture, referred to as fattening or farming, is a capture based activity dependent on yearly renewal from the wild. Thus, the development of aquaculture practices independent of wild resources can provide an important contribution towards ensuring security and sustainability of this species in the longer-term. The development of such practices is today greatly assisted by large scale transcriptomic studies.

Results: We have used pyrosequencing technology to sequence a mixed-tissue normalised cDNA library, derived from adult *T. thynnus*. A total of 976,904 raw sequence reads were assembled into 33,105 unique transcripts having a mean length of 893 bases and an N50 of 870. Of these, 33.4 % showed similarity to known proteins or gene transcripts and 86.6 % of them were matched to the congeneric Pacific bluefin tuna (*Thunnus orientalis*) genome, compared to 70.3 % for the more distantly related Nile tilapia (*Oreochromis niloticus*) genome. Transcript sequences were used to develop a novel 15 K Agilent oligonucleotide DNA microarray for *T. thynnus* and comparative tissue gene expression profiles were inferred for gill, heart, liver, ovaries and testes. Functional contrasts were strongest between gills and ovaries. Gills were particularly associated with immune system, signal transduction and cell communication, while ovaries displayed signatures of glycan biosynthesis, nucleotide metabolism, transcription, translation, replication and repair.

Conclusions: Sequence data generated from a novel mixed-tissue *T. thynnus* cDNA library provide an important transcriptomic resource that can be further employed for study of various aspects of *T. thynnus* ecology and genomics, with strong applications in aquaculture. Tissue-specific gene expression profiles inferred through the use of novel oligo-microarray can serve in the design of new and more focused transcriptomic studies for future research of tuna physiology and assessment of the welfare in a production environment.

Keywords: *Thunnus thynnus*, transcriptome, Microarray, Tissue gene expression, genome mapping

* Correspondence: ztrumbic@unist.hr
[1]University Department of Marine Studies, University of Split, Split, Croatia
Full list of author information is available at the end of the article

Trumbić *et al. BMC Genomics* (2015) 16:1007

Page 2 of 17

Background

Fish represent the oldest and most diverse vertebrate group, inhabiting every available aquatic environment [1]. They display remarkable heterogeneity of life trait characteristics and occupy a historically important place in human culture as a source of nutritionally valuable food. Marked by genome duplication events, the diversification and adaptive radiation of fish represent an attractive evolutionary model to study ecological, behavioural and physiological aspects of biodiversity [2]. Migratory tunas (family *Scombridae*, order *Perciformes*: suborder *Scombroidei*), are pelagic oceanodromous top predators, which display a set of extreme adaptations to life in a prey depauperate pelagic environment. They are obligate ram ventilators with streamlined bodies for rapid locomotion, have increased aerobic scope and metabolic rates that allow them to satisfy the energy needs of multiple high performance functions without compromise, including high digestive rates, increased somatic and gonadal growth and incredibly fast recovery from oxygen debt after exhaustive exercise compared to other teleosts [3, 4]. A key characteristic of tunas is their ability to retain metabolic heat in the brain, red muscle and viscera via vascular specialisations called *rete mirabile*, which comprise numerous juxtaposed arterial and venous vessels acting as a counter-current heat exchange system [4]. This heterothermic physiology shares characteristics of both poikilothermic and homeothermic organisms and is considered to have developed in order to permit thermal niche expansion, especially in correlation with maintenance of central nervous system function [5] or to sustain a rise in aerobic capacity and muscular performance [6].

The largest of the tuna species, Atlantic bluefin tuna, *Thunnus thynnus* (Linnaeus, 1758), inhabits the North Atlantic Ocean and the Mediterranean Sea. Based on the observed homing behaviour to two major spawning grounds, the Gulf of Mexico and the Mediterranean Sea [7], the International Commission for the Conservation of Atlantic Tunas (ICCAT) recognises a smaller Western and larger Eastern population as separate management units divided by the 45°W meridian. Although the differentiation between these populations has been corroborated by genetic data [8, 9], results of electronic tagging reveal a complex ecology, with extensive trans-Atlantic migrations and mixing of individuals at the foraging and spawning grounds [10, 11], complicating *T. thynnus* stock management. In high demand for sushi and sashimi preparation in the Japanese market, both populations of *T. thynnus* are considered to be depleted and overfished [12–14]. While latest assessments report a reduction in fishing mortality and increase in spawning stock biomass, allowing for moderate catch quotas increase after a long time in 2015 [15], these populations remain under threat. Capture-based tuna aquaculture, developed in the Mediterranean Sea over the past two decades, relies on yearly population renewal from the wild, fattening fish to market size in 3 months to 2 years, but is not thought to be sustainable in the long run [16]. Consequently, it is considered that only domestication of *T. thynnus* and development of high welfare aquaculture practices independent of wild resources can preserve the species for future generations. Extensive research efforts have been made regarding the control of *T. thynnus* reproduction in captivity [16, 17], a prerequisite for the development of a sustainable aquaculture industry. As evidenced from the Pacific bluefin tuna (*T. orientalis*) aquaculture development in Japan [18], the first bluefin tuna species successfully grown in captivity, a comprehensive approach will be required to solve many technical problems in order to meet this species' physiological needs.

Large scale transcriptomic studies of important life trait characteristics, such as growth, maturation, environmental tolerance and disease resistance, can provide important knowledge concerning underlying molecular mechanisms and control elements, assisting in the development of welfare-centred and sustainable management strategies for aquaculture species, *e.g.* [19, 20]. With the increasing availability of high throughput sequencing technologies, development of the necessary tools and resources for broad-scale transcriptomics studies can be achieved in a relatively short period of time, even for species with a low level of background information [21]. Until recently, publicly available genomic resources for *T. thynnus* have been relatively scarce, most being derived from a single EST (expressed sequence tag) sequencing project comprising adult liver, ovary and testis mRNA transcripts contributed by Chini et al. [22]. This resource served as the basis for the construction of the first *T. thynnus*-specific oligonucleotide DNA microarray, which was employed for the description of the Gulf of Mexico *T. thynnus* gonad transcriptome [23]. It was also utilised to profile the cardiac transcriptome response to temperature acclimation in congeneric *T. orientalis* [24], highlighting the applicability of such a tool for gene expression analyses in phylogenetically related species. The release of a genome for *T. orientalis* [25] opens new opportunities for tuna comparative genome studies and ongoing transcriptome characterisation projects.

The objectives of the current study were: 1) to contribute to and expand existing transcriptomics resources for studying *T. thynnus*, 2) to use the newly developed resource to design and validate a custom oligo-microarray, and 3) to employ the newly developed microarray to investigate functional differences in transcript expression among tissues. To this end, pyrosequencing technology

Trumbić *et al. BMC Genomics* (2015) 16:1007

Page 3 of 17

was used to sequence adult mixed-tissue normalised cDNA library providing the template for the development of a custom 8 × 15 K Agilent oligo-microarray. The microarray was used to compare: gill, heart, liver, ovaries and testes of adult fish and to infer their respective transcriptional specificities. The derived transcriptome was also compared to *T. orientalis* genome.

Methods

Animal tissue collection and RNA extraction

All fish handling procedures were conducted in accordance with established standards for the care and use of animals of the Ethical committee for animal welfare at the Institute of Oceanography and Fisheries, Croatia. No specific permits were required as samples were collected during standard commercial operations without any form of experimental manipulation.

Atlantic bluefin tuna were reared in mid-Croatian offshore farm cages for two years. During the winter harvest of 2012, adult fish (curved fork length (mean ± SD) = 177 ± 9 cm; weight = 99 ± 14 kg) were sacrificed by pithing and samples of kidney, spleen, liver, gonads, gills, small intestine, heart (*apex ventriculi*), red and white skeletal muscle, skin scrapes and whole blood, were dissected and preserved in RNA*later* stabilisation solution (Qiagen, UK) at –20 °C, after overnight storage at +4 °C. In addition, liver and kidney samples from juvenile moribund fish with signs of septicaemia were available from another study [26] and were included in a normalised cDNA library preparation to increase the diversity of the sequenced transcriptome.

Between 30 – 40 mg of each of the tissues were homogenised separately in 1 mL of TRI Reagent (Sigma-Aldrich, UK) by 2 × 35 s disruption runs on a Mini-Beadbeater-24 (BioSpec Products, Inc, USA) and total RNA was extracted according to manufacturer's instructions. The RNA precipitation step was modified to include a high-salt solution as recommended for polysaccharide- and proteoglycan-rich sources [27]. Integrity of total RNA preparations was assessed by ethidium bromide agarose gel electrophoresis and purity and concentration by spectrophotometry (NanoDrop ND-1000, Thermo Scientific, USA).

Normalised cDNA library preparation and sequencing

A mixed RNA sample comprising equal amounts (5 μg) derived from all sampled tissues, except ovaries, was created from one female, one male and one moribund juvenile specimen and purified using RNeasy columns (Qiagen, UK). Because *T. thynnus* ovaries showed a significantly increased proportion of the small RNA fraction comprising tRNA and 5S RNA, as previously found in amphibians [28, 29] and several other teleosts [30, 31], poly(A) + RNA was first isolated from the

ovarian total RNA population using a Poly(A) Purist kit (Ambion, UK) and subsequently 200 ng of the isolated mRNA were included in the pool. For normalisation, the MINT kit (Evrogen, Russia) was initially used to construct a full-length-enriched double stranded cDNA library from 1.5 μg of the tissue RNA mixture, according to manufacturer's protocol. The TRIMMER kit (Evrogen, Russia) was used to normalise 1 μg of cDNA starting material, applying the duplex-specific nuclease (DSN) method [32]. The final normalised and amplified cDNA library was purified using a MinElute PCR Purification Kit (Qiagen, UK) and ranged in size from 0.25 to 3 kb, as visualised on a 1.5 % agarose/ethidium bromide gel.

GS FLX Titanium library preparation and sequencing were performed by the Edinburgh Genomics facility (University of Edinburgh, UK). Approximately 5 μg of normalised cDNA were used for generating a 454 sequencing library using the GS FLX Titanium Rapid Library Preparation kit (Roche Applied Science, UK), following manufacturer's instructions. Sequencing was performed using The Genome Sequencer™ (GS) Titanium FLX instrument (Roche Applied Science, UK). Bases were called with 454 software and reads trimmed to remove adapter sequences.

Data filtering and assemblies

Low-quality reads (Phred score < 30) were filtered out and PRINSEQ v0.20 [33] was used to remove PCR duplicates and low complexity sequences. Two complementary sequence read assembly methods were chosen for this study: WGS-assembler 7.0 [34] and Mosaik-aligner v2.1 [35] to assemble the contigs based on the Genbank archived *T. thynnus* EST database [36]. Subsequently, to lower the redundancy resulting from *de-novo* assemblies, TRF 4.07b [37] was used to remove all transcripts shorter than 500 bases or exhibiting repetition, with entropy above 1.

Sequence annotation and functional assignments

The longest coding DNA sequences were determined for each transcript using the command-line program getorf from the EMBOSS v6.6.0 package [38]. ESTScan v2 [39] was then used to confirm transcript coding regions and determine sequence orientation. The BLAST algorithm [40] was used for sequence similarity searches against public databases. The coding sequences of the predicted transcripts were annotated using BLASTP searches against the GenBank Reference Proteins database [41], with an expectation value (e-value) cut-off of 10^{-4}, minimum similarity of at least 60 %, and minimum alignment length of 33 amino acids. Additionally, the transcripts were annotated using BLASTN searches against the annotated EST *T. thynnus* database [36]. The inferred annotations were used to retrieve Gene

Trumbić *et al. BMC Genomics* (2015) 16:1007

Page 4 of 17

Ontology (GO) annotation for molecular function, biological process and cellular component [42]. To avoid redundant functional assignments, the best-rated similarity hit with at least one GO annotation was chosen. A custom pipeline converted GO terms to GO Slim terms, using the Protein Information resource and Generic GO Slim files [43]. Kyoto Encyclopedia of Genes and Genomes (KEGG) database [44] was used to infer functional annotation of the sequences through KAAS (KEGG Automatic Annotation Server) [45] using single-directional best hit method for ESTs. Sequences with assigned KO (KEGG Orthology) identifiers were categorised into functional groups according to the KEGG BRITE hierarchy, excluding human diseases.

Genome mapping

Generated sequence data was compared with two perciform fish genomes that have been publicly released: *Thunnus orientalis* (Pacific bluefin tuna; NCBI Assembly GCA_000418415.1;[25]), and *Oreochromis niloticus* (Nile tilapia, NCBI Assembly GCA_000188235.2). The 133,062 contigs (684,497,465 bp) of *T. orientalis* and the 77,755 (927,696,114 bp) of *O. niloticus* genomes were downloaded and BLASTN [40] was used to search for similar transcripts. Default parameters were used for BLASTN searches, with the following exceptions to account for the divergence and short length of the sequences available: minimum alignment size 80 nt, minimum percentage of sequence identity 25 %, maximum e-value 0.001 and low complexity mask on. Results were displayed using a hive plot [46].

T. thynnus microarray design

Transcripts with a significant match to a reference protein were chosen for microarray probe design. In addition, 57 publicly available full mRNA sequences from the *T. thynnus* nucleotide (NCBI) database were also included. Sixty-mer oligonucleotide probes were designed using the eArray (Agilent Technologies, UK) online probe design tool with Base Composition and Best Probe Methodologies, 3'bias and sense orientation. Two probes were designed for each target transcript. After initial screening, unique probes showing no cross-hybridisation potential were selected to produce an 8 × 15 K Agilent custom oligo-DNA microarray design format (Agilent Design ID = 038391), comprising 15,208 user defined features and 536 Agilent positive and negative controls. Target transcripts (N = 6,439) with best annotation were represented with two probes on the array and the remainder (N = 2,190) with a single probe. In all, 6,287 unique proteins were annotated. For calculation of the multiplicative detrending step implemented within the Agilent Feature extraction software 35 probes were replicated five times.

Microarray experiment

For microarray validation, five metabolically distinct tissues were chosen from adult *T. thynnus*: gill, heart, liver, testes and ovaries (4 biological replicates per tissue). A total of 20 arrays (4 replicates x 5 tissues) were assayed in the final common reference pool design, with the pool comprising a contribution from each biological sample. Amplified and fluorescently labelled cRNA (complimentary RNA) for microarray hybridisations was prepared in accordance with the Two-Color Microarray-Based Gene Expression Analysis - Low Input Quick Amp Labelling Protocol v6.6 (Agilent Technologies, USA). The cDNA synthesis was initiated with 150 ng of total RNA from each tissue sample, this being primed with an Oligo dT – T7 RNA polymerase promoter. Subsequently, T7 RNA polymerase incorporated Cyanine 3 (Cy3) or Cyanine 5 (Cy5) – CTPs into a growing chain of antisense cRNA. Amplified cRNA was purified using the GeneJET RNA Purification Kit (Thermo Scientific, UK) following the manufacturer's RNA Cleanup Protocol and was further quality-checked by spectrophotometry (NanoDrop ND-1000, Thermo Scientific, USA) and agarose gel electrophoresis. Following amplification and purification, 300 ng of each Cy3 labelled test and 300 ng of Cy5 labelled reference pool cRNA were mixed together and incubated with Fragmentation and Hybridization buffers according to manufacturer's instructions. Competitive hybridisations were carried out in a rotary oven (Agilent Technologies, UK) over 17 h at 65 °C and 10 rpm. Following Agilent standard protocols, the slides were subsequently washed in Wash Buffer 1, Wash Buffer 2, acetonitrile and Agilent Stabilization and Drying solution, the latter to reduce ozone-induced decay of Cy5 signal. Slides were stored in light proof box and scanned within 1 h of washing.

Microarray data analyses

Scanning was conducted using an Axon Genepix 4200A scanner with Genepix Pro 6.1 image acquisition software (Molecular Devices, UK) with 60 % red and 90 % green laser power at 5 μm resolution. Saturation tolerance was set to 0.05 % and automatic photo-multiplier tube (auto-PMT) gain used to achieve similar mean intensities of Cy3 and Cy5 signals. Acquired data were exported and processed using the Agilent Feature Extraction software (v9.5.3.1) to obtain background-subtracted signals, as well as other spot statistics and quality metrics. Scan data were analysed using GeneSpring GX v12 (Agilent Technologies, UK). Baseline transformation was not employed and data were normalised using a Lowess model. Principal component analysis (PCA) was conducted to visually inspect the distribution of gene expression variance among arrays within and between experimental conditions. Following removal of Agilent control features, stringent quality filtering involved

Trumbić *et al. BMC Genomics* (2015) 16:1007

Page 5 of 17

removal of saturated and non-uniform features, population outliers and features that were not significantly positive with respect to the local background.

One-way ANOVA unequal variance (Welch) ($p < 0.05$) was applied to preselect features showing potential differences in gene expression between tissues. Unsupervised network analysis was conducted on identified features using Biolayout *Express* [3D] [47], using a Pearson coefficient of correlation as a similarity measure and threshold over 0.94. The network graph was clustered into distinct patterns of gene expression using a Markov clustering algorithm (MCL) with inflation value set to 2.0 and smallest cluster allowed being 10, to achieve optimal granularity of clusters and focus only on larger areas of high connectivity [47].

Functional gene set analyses were performed on the entire list of quality filtered entities based on KEGG KO identifiers. During the analyses, gene expression profiles were collapsed into 3,222 unique KOs using the median method. Tests were run to identify significantly up- and down- regulated pathways for each tissue and their pairwise combinations against all other samples using Generally Applicable Gene-set Enrichment (GAGE) analyses [48], implemented as R/Bioconductor package [49], which is robust to small datasets with different sample sizes. A default FDR q-value of 0.1 was used as a cut-off. Pathways were subsequently grouped into KEGG functional categories and the difference between the counts of up- and down- regulated pathways in each category hierarchically clustered and displayed as a heatmap.

RT-qPCR validation

Ten transcripts from different Biolayout derived clusters showing tissue-specific and diverse expression profiles, and spanning a wide range of fold change (FC) values, were chosen for RT-qPCR validation. Three additional genes were tested to serve as internal controls (reference genes): elongation factor-1α (*elf-1α*) [50], beta-actin (*actb*) [26] and a transcript showing a stable expression profile across the experimental conditions, annotated as FtsJ methyltransferase domain containing 2 (*ftsjd2*), from the microarray. Primer3 [51] software was used to design primers for the selected sequences with at least one in the pair overlapping the target hybridisation area of the microarray 60-mer probe. Complementary DNA (cDNA) was synthesised from 2 μg of the same total RNA extractions used for the microarray experiment employing the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, UK). The synthesis was primed with 1.5 μl of random hexamer primers supplied with the kit and 2.5 μM anchored-oligo (dT)$_{20}$ (Eurofins, Germany) per 20 μl reaction. DNase treatment was not performed prior to cDNA synthesis as it may result in RNA degradation [52]. Instead, minus RT (RT-) controls (reverse transcriptase replaced with water) were created to control for the presence of genomic DNA in the extractions.

Real-time PCR assays (20 μl) were performed on a Mastercycler ep realplex2 (Eppendorf, Germany) using Luminaris Color HiGreen qPCR Master Mix (Thermo Scientific, UK), in duplicate with 25 × cDNA template dilutions and 0.3 μM of each primer, according to the following thermal profile: 50 °C for 2 min (Uracil-DNA Glycosylase pre-treatment), 95 °C for 10 min, followed by 40 cycles of denaturation at 95 °C for 15 s, annealing at 60 °C for 30 s and elongation at 72 °C for 30 s. After amplification a melting curve analysis was performed from 55 °C to 95 °C at 0.5 °C increments for 15 s each to verify single product amplification. The presence of the correct size of PCR product was also verified by agarose gel electrophoresis. Six-point standard curves of a 4 × dilution series of a pool of all starting cDNAs were run with every assay to determine primer-specific efficiencies. A standard curve was used to extrapolate non-normalised relative copy numbers for each gene and the BestKeeper application [53] was used to inspect the suitability of reference genes for normalisation.

To compare the fit between qPCR and microarray fold changes (FCs), we calculated the concordance correlation coefficient (CCC) as well as other regression metrics previously described by Miron *et al.* [54]. The statistical significance of individual targets' expression profiles was examined on the combined qPCR and microarray data for each gene, coded by their respective percentile ranks, using a nonparametric Kruskal-Wallis test implemented in R software.

Results and discussion
Pyrosequencing and transcriptome assembly

A total of 976,904 raw sequence reads were generated, with a mean size of 347 bases (b) (Table 1), which is consistent with the sequencing technology used [55]. The reads that passed quality control filtering were trimmed of sequencing primers and adaptors and used for the assembly process. WGS-assembler (Celera) generated a *de-novo* transcriptome assembly of 70,108 contigs and 52,452 singletons (Table 1). Contigs and singletons were scaffolded using Mosaik based on the 10,163 available ESTs. Based on the high quality reads of over 500 b and of higher complexity, 33,105 unique transcripts were assembled, with a mean length of 893 b and an N50 of 870. Of these, 22.0 % of the transcripts had a length more than 1,000 b with 22.8 % of these having a full-length coding region. To evaluate the quality of the assembled transcripts, all the usable sequencing reads were realigned to the transcripts.

Trumbić *et al. BMC Genomics* (2015) 16:1007

Page 6 of 17

Table 1 Statistics of the assembled *T. thynnus* sequences

Raw data	
Number of reads	976,904
Mean read size	347
Total size of reads	338,522,590
After assembly (no singletons)	
Number of transcripts	70,108
Mean size	593
Total size of transcripts	41,564,882
After assembly (singletons only)	
Number of transcripts	52,452
Mean size	431
Total size of transcripts	22,601,482
After filtering (>500 b)	
Number of transcripts	33,105
N50	870
Total size of transcripts	29,512,631
Shortest transcript	500
Longest transcript	4,955
Mean size	893
Median size	799
Mean GC%	43.5 %
N%	0.2 %

Functional inference by searching against public databases

The assembled transcripts were annotated using BLASTP and BLASTN searches against the RefSeq Proteins [41] and EST databases respectively [36]. The results indicated that out of 33,105 transcripts, 11,065 (33.4 %) showed significant similarity to known proteins or gene transcripts in the RefSeq Proteins database. To evaluate the *T. thynnus* coverage of the assembled transcripts, the 10,163 EST available for *T. thynnus* were aligned to the 33,105 transcript sequences generated in this study and vice versa. In total, 4,939 (48.6 %) of all *T. thynnus* ESTs aligned to at least one transcript, whereas 4,394 (13.3 %) of the new transcripts had at least one corresponding *T. thynnus* EST. This demonstrates that the strategy employed captured some of the existing data [22] and added new information to *T. thynnus* database.

GO functional annotation was assigned to the assembled *T. thynnus* transcripts/genes on the basis of RefSeq Proteins annotation. Out of 11,065 transcripts that had similarity to known gene products, 8,384 (25.3 % of total library) were assigned a GO annotation. Figure 1 shows that derived GO descriptions for *T. thynnus* transcripts cover a diverse set of molecular functions, cellular components and biological processes. Similarly, 8,334 (25.2 %) out of the total 33,105 transcripts were assigned
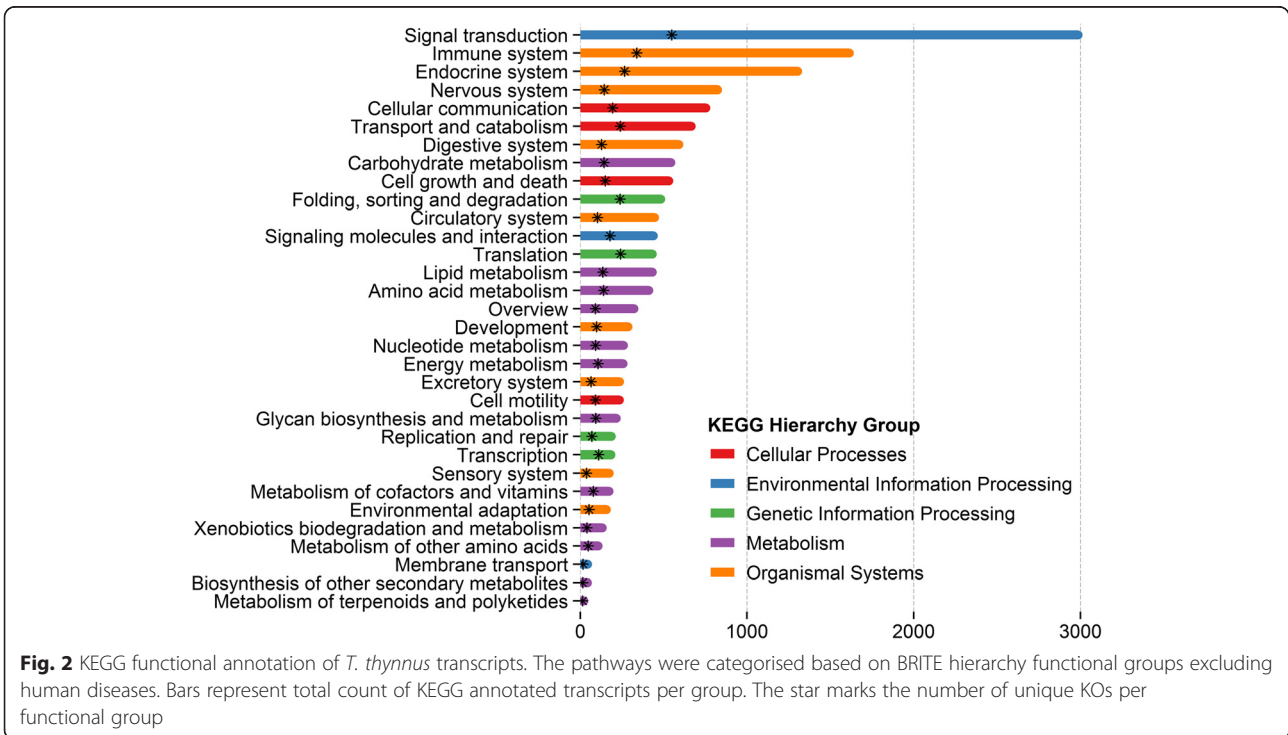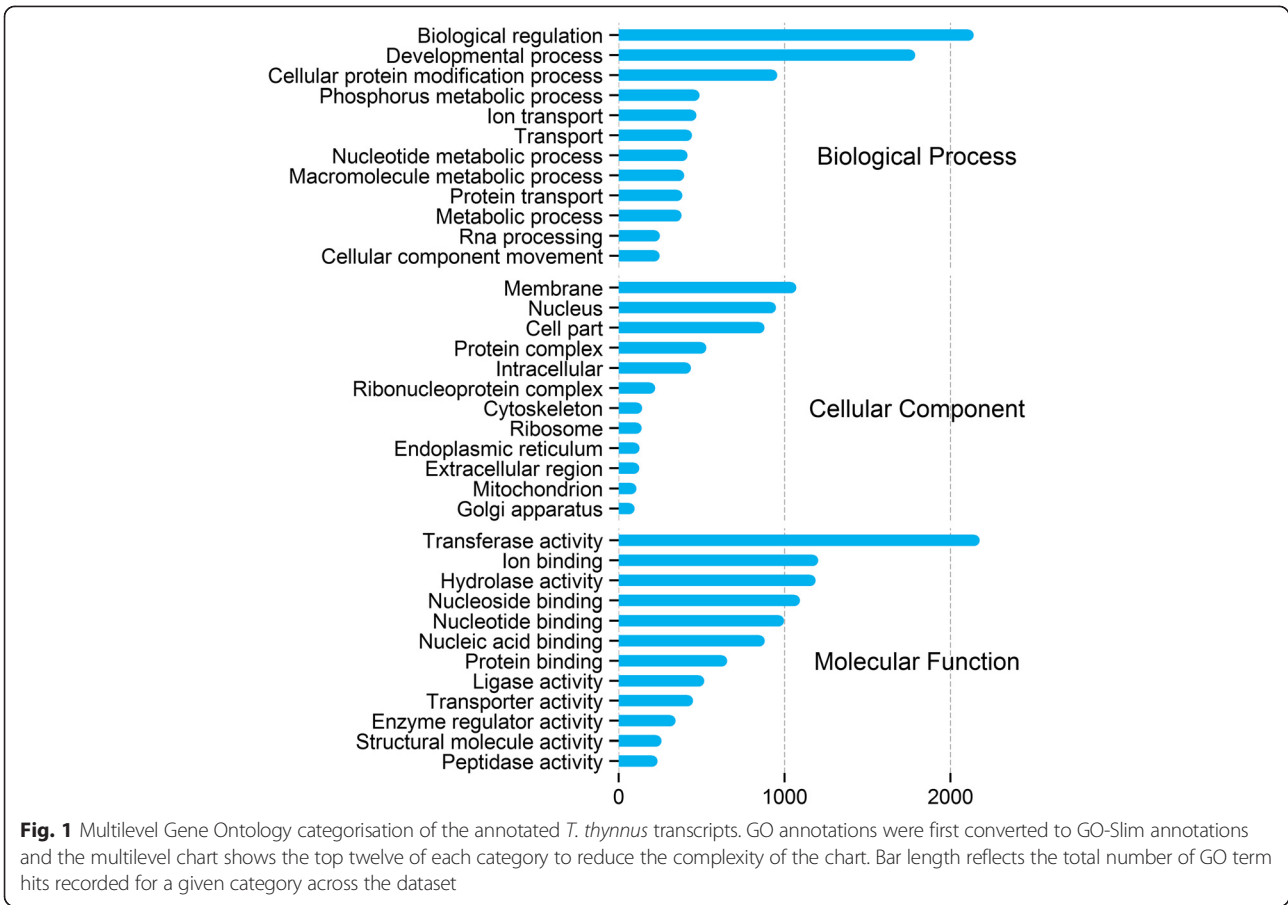
4040 unique KO identifiers and mapped to 262 different biochemical pathways in the KEGG BRITE functional hierarchy, excluding human diseases. Pathways were categorised according to their functional groups and total transcript abundance with unique KO identifiers found in each group is outlined in Fig. 2. The top five best represented categories were signal transduction, immune, endocrine, nervous system and cellular communication. Figure 2 also points to a substantial level of functional redundancy, as evident from lower number of unique KOs found in each group. This is an inherent feature of the KEGG classification as many pathways share the same genes and some KOs are also assigned to different transcripts with similar functional motifs.
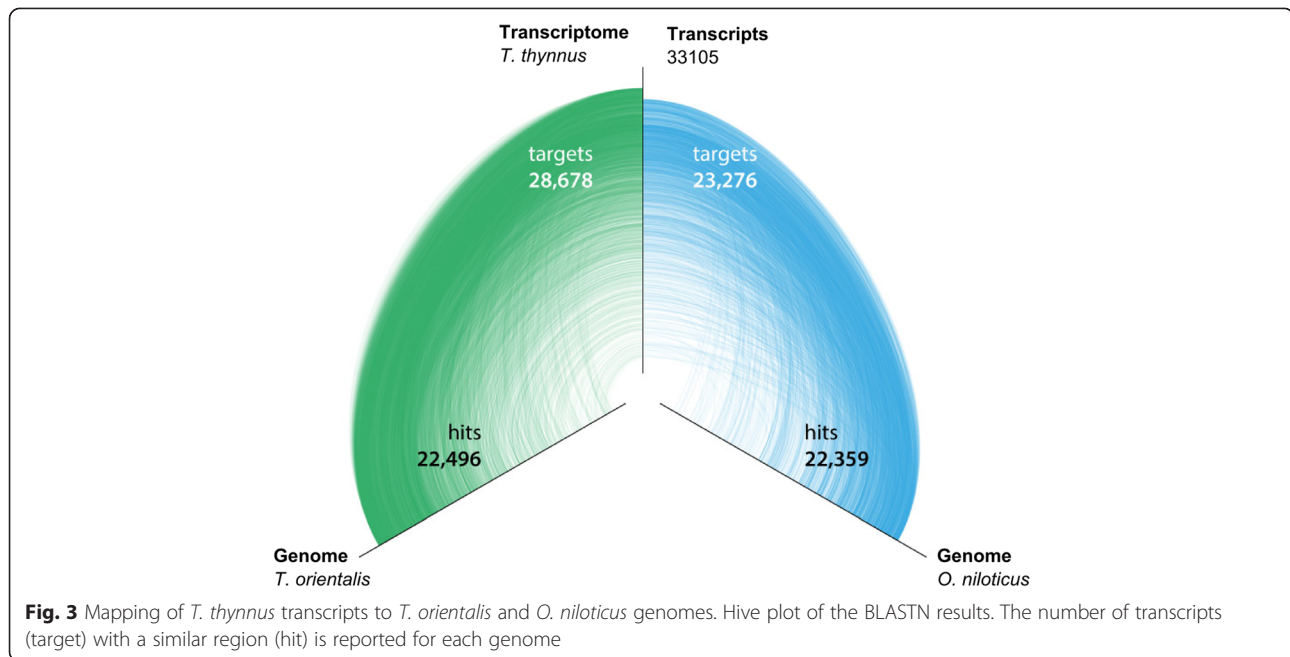
Genome mapping

To assess the similarity of *T. thynnus* transcripts to currently available genomic resources, mapping was performed to two perciform genomic sequences: *T. orientalis* and *O. niloticus* (see Additional files 1 and 2). The genome assembly of *O. niloticus* is in a more advanced state, contigs being organised into linkage groups or chromosomes, while the *T. orientalis* genome is in its initial assembly release with sequences organised into contigs. The results of the mapping to the two genomes were visualised using a hive plot where each genome is placed along a linear axis and hits are drawn as curved links connecting *T. thynnus* transcripts to respective target regions in each genome (Fig. 3). As expected, more targets found a hit with the congeneric *T. orientalis* genome (28,678 or 86.6 %) than the more distantly related *O. niloticus* genome (23,276 or 70.3 %). *T. thynnus* and *T. orientalis* are considered to be very closely related and have been separated as sub-species until recently when molecular data, especially mitochondrial DNA, corroborated their separate species status [4, 56]. Our data also demonstrate their close connection and sequence similarity which makes the *T. orientalis* genome an important resource for *T. thynnus* studies as well, with possible applications for exon-intron gene modelling or reference mapping in future sequencing studies. The fraction of *T. thynnus* targets that did not find a match to *T. orientalis* genome consisted mainly of non-annotated singletons, with this group also showing increased repetitive nucleotide stretches (data not shown). These sequences could be difficult to match, or may represent *T. thynnus*-specific diversification, or both.

Results of microarray transcriptional inference and qPCR validation

Principal component analysis showed distinct separation of arrays into tissue-specific groups (data not shown). The signals from one array deviated considerably from its biological replicates within the heart group, reflecting

Trumbić *et al. BMC Genomics* (2015) 16:1007

Page 7 of 17

Fig. 1 Multilevel Gene Ontology categorisation of the annotated *T. thynnus* transcripts. GO annotations were first converted to GO-Slim annotations and the multilevel chart shows the top twelve of each category to reduce the complexity of the chart. Bar length reflects the total number of GO term hits recorded for a given category across the dataset

Fig. 2 KEGG functional annotation of *T. thynnus* transcripts. The pathways were categorised based on BRITE hierarchy functional groups excluding human diseases. Bars represent total count of KEGG annotated transcripts per group. The star marks the number of unique KOs per functional group

Trumbić *et al. BMC Genomics* (2015) 16:1007

Page 8 of 17

Fig. 3 Mapping of *T. thynnus* transcripts to *T. orientalis* and *O. niloticus* genomes. Hive plot of the BLASTN results. The number of transcripts (target) with a similar region (hit) is reported for each genome

technical difficulties noted during hybridisation, and therefore this array was excluded from further analyses. From 15,068 *T. thynnus*-specific probes, 13,494 passed the quality filtering across the experiment and showed positive and significant signals above background in both channels. Most of these features, 12,306 (91.2 %), were found to be differentially expressed (ANOVA, $p < 0.05$) between at least two tissues, consistent with the expected low number of genes ubiquitously and constitutively expressed in an organism [57]. Network analyses based on these preselected features produced a complex graph with 8,836 nodes, further reduced into 87 clusters of discrete patterns of gene expression (see Additional file 3). Over half of the nodes were grouped within the first six clusters (Table 2), each corresponding to one tissue-specific up-regulated group of features, and an ovary and testis

Table 2 Size and tissue specificity of the first six gene expression clusters of *T. thynnus* obtained by MCL algorithm in Biolayout *Express* [3D]

MCL Clusters	Tissue specificity	N (transcripts)
Cluster 1	Ovary	1,598
Cluster 2	Heart	877
Cluster 3	Liver	654
Cluster 4	Gill	579
Cluster 5	Ovary/Testis	557
Cluster 6	Testis	253
Other	Mixed	2,018
No Class	/	2,300
Grand Total		8,836

combination (Fig. 4). The largest was an ovary-specific cluster while highest intensity values of expression were recorded for liver and heart transcripts. Other clusters not shown in Fig. 4 comprised entities with varying expression profiles across tissues.

Eight transcripts representative of the six largest Biolayout clusters were chosen for the purpose of qPCR validation, taking into account their functional relevance and FC span (see Additional file 4 for details): aquaporin 3b *aqp3b* (Gill), sarcoplasmic reticulum calcium-ATPase 2 *atp2a2a* (Heart), complement C4-like *c4*, coagulation factor V *f5* (Liver), exosome component 2 *exosc2*, pescadillo *pes* (Ovary), somatolactin *sl* (Testis) and mediator complex subunit 6 *med6* (Ovary/Testis). Transcripts specific to particular tissue clusters show high expression in only that tissue, such that examination of expression gives low values for all other tissues. Thus two additional transcripts, biliverdin reductase B *blvrb* and prohibitin 2 *phb2*, were chosen as exemplars showing expression across a range of tissues. This ensured that the microarray validation did not comprise only tissue specific transcripts. However, as functional interpretation of results focuses on tissue-specific signatures, these two transcripts are not specifically discussed. Three reference genes were tested, *actb, elf-1α*, previously used in *T. thynnus* qPCR experiments [23, 26, 50], and a candidate from this microarray study, *ftsjd2*, which subsequently did not show stable expression profiles according to BestKeeper tool, exceeding the recommended quantification cycle (Cq) standard deviation of one [53]. This was not unexpected as we were comparing different tissues of adult fish; however, the variation was
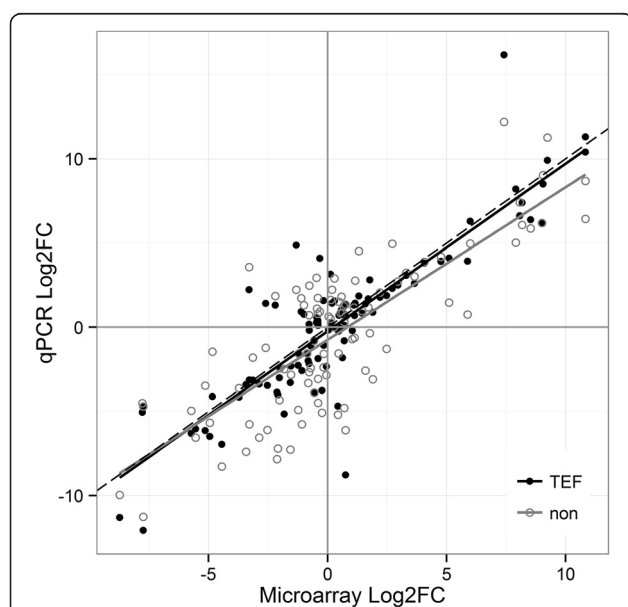
Trumbić *et al. BMC Genomics* (2015) 16:1007

Page 9 of 17
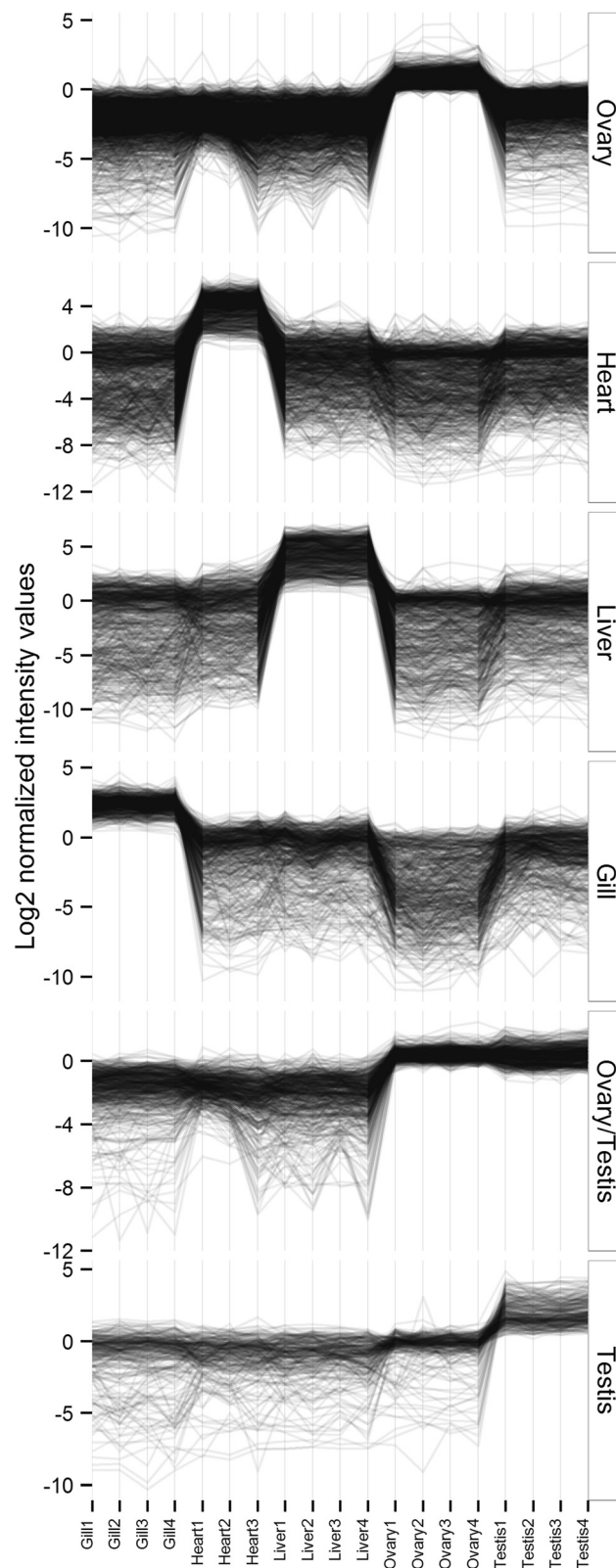
Fig. 4 Clusters of tissue-specific gene expression profiles of adult captive *T. thynnus.* Clusters were derived using network analyses and Markov clustering algorithm in Biolayout *Express* [3D], plotted in descending order according to number of transcripts in cluster

Trumbić *et al. BMC Genomics* (2015) 16:1007

Page 10 of 17

systematically attributed only to ovarian tissue, which was characterised by a very different total RNA population. Hence, non-normalised (normalised to total RNA used for cDNA synthesis) and normalised data (normalised to geometric mean of *actb*, *elf-1α* and *ftsjd2* = TEF) were further compared to microarray results. For all information pertinent to RT-qPCR validation of microarray results see Additional file 4.

Microarray and qPCR fold changes (FCs) among all five tissues for ten target genes were compared in Log2 space using the concordance correlation coefficient (CCC), proposed for global validation of microarray experiments [54]. CCC is a product of a measure of precision (Pearson r), describing linear correlation of the data, and accuracy, how close the least-squares regression line is to the identity line. It takes values from −1 (perfect inverse agreement) to 1 (perfect agreement). We evidenced a good agreement between microarray and qPCR data in this study, both in terms of precision and accuracy (Table 3, Fig. 5). Due to the range enhancement artefact, Miron *et al.* [54] suggested that up- and down-regulated genes should be inspected separately. Gene expression profiles were broken into positive and negative FCs, with former producing better scores (Table 3), which is sometimes observed in microarray and qPCR comparisons [58]. Correlation of microarray and qPCR data is often affected by pitfalls and differences inherent to each technology, *i.e.* probe/primer properties and amplification kinetics, which can differ from one

Table 3 Agreement between qPCR and microarray data

Indice	ALL DATA		UP		DOWN	
	non	TEF	non	TEF	non	TEF
Slope	0.91	0.99	0.88	1.05	0.82	1.05
Y intercept	−0.75	−0.24	−0.54	−0.59	−1.08	0.05
Precision (r)[a]	0.80	0.87	0.77	0.85	0.54	0.73
Accuracy	0.98	0.99	0.96	0.97	0.89	0.94
CCC	0.78	0.86	0.73	0.82	0.48	0.69

Correlation correspondence index (CCC) and other indices describing agreement of microarray and qPCR data (non-normalised (non) and normalised (TEF)) are shown. Log2FCs (Fold Changes) were examined for all data and positively and negatively regulated values separately
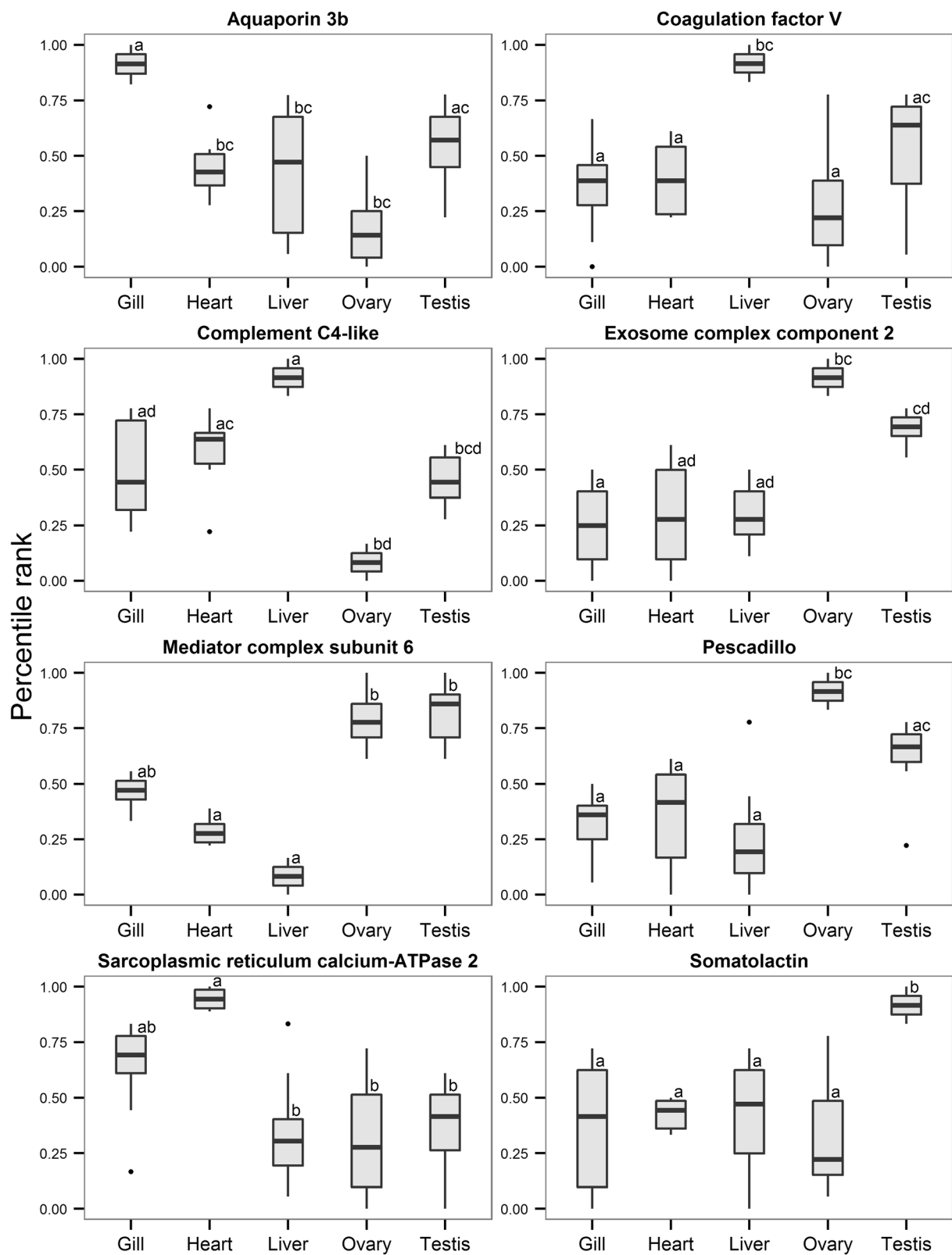[a]Correlation is significant at the 0.01 level (2-tailed)

transcript to another and equally depend on its level of expression [52, 58].
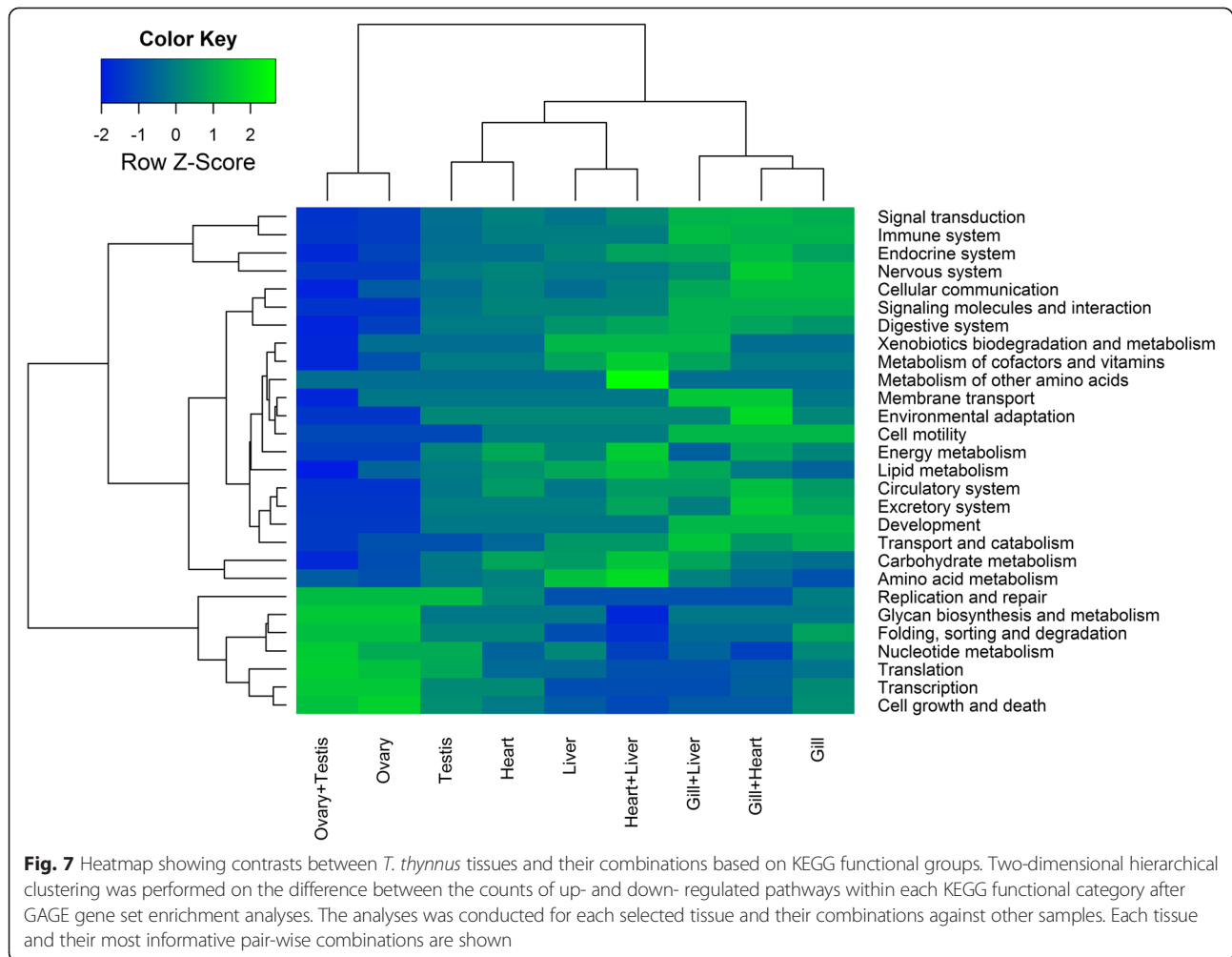
As there was better concurrence between normalised qPCR and microarray data with respect to non-normalised, only the normalised set was included in the joint qPCR and microarray statistical analyses, with data coded by respective percentile ranks. The goal was to maximise power to test differences between tissues by combining two outputs and increasing the number of replicates for each condition. Figure 6 shows derived expression profiles for target transcripts representatives of the six largest Biolayout clusters. Results of post-hoc tests for all transcripts ($p < 0.05$) are found in Additional file 5. Although there was some disagreement between qPCR and microarray data evidenced by increased interquartile ranges in certain instances (Fig. 6), it is evident that selected transcripts exhibit highest expression in the tissue clusters to which they were assigned. However, this was not found to be statistically significant in all comparisons made (Fig. 6).

Functional interpretation of tissue-specific expression profiles

Functional contrasts, inferred through KEGG pathway categorisation into respective functional groups (Fig. 7), were strongest between gill and ovarian tissue, also reflecting the fact that the greatest number of up- and down-regulated pathways was found for these tissues respectively. For the full list of pathways for each tissue and their pairwise combinations see Additional file 6. Gills were characterised by the largest diversity of functional categories, with immune system and signal transduction being the most highly represented, followed by nervous, endocrine, digestive, excretory system and cellular communication *inter alia*. This is consistent with the multi-purpose role of the fish branchial tissue in maintaining systemic homeostasis. The gill provides a key barrier to the outer aquatic environment and serves as the major site of respiratory gas exchange, pH and ion balance, nitrogenous waste excretion and immune

Fig. 5 Least-squares regression fit between microarray and qPCR data. The fit between microarray and qPCR non-normalised (non) and normalised (TEF) Log2FC (Fold Change) is shown in respect to the identity line (dashed) for the selected target genes of *T. thynnus*

Trumbić *et al. BMC Genomics* (2015) 16:1007

Page 11 of 17

Fig. 6 Combined expression profiles of target genes belonging to six largest *T. thynnus* tissue clusters. The profiles were derived by combining qPCR and microarray data for specific genes using their respective percentile ranks. Nonparametric Kruskal-Wallis test was used to determine statistical significance of the differences between tissues. Letter codes denote statistical significance at *p*< 0.05

Trumbić *et al. BMC Genomics* (2015) 16:1007

Page 12 of 17

Fig. 7 Heatmap showing contrasts between *T. thynnus* tissues and their combinations based on KEGG functional groups. Two-dimensional hierarchical clustering was performed on the difference between the counts of up- and down- regulated pathways within each KEGG functional category after GAGE gene set enrichment analyses. The analyses was conducted for each selected tissue and their combinations against other samples. Each tissue and their most informative pair-wise combinations are shown

defence, being controlled and co-ordinated by a complex web of neural and endocrine pathways [59, 60]. Tuna gills in particular have an unusually large surface area and thin blood-water barrier in respect to other active teleosts such as rainbow trout [61] permitting extremely efficient oxygen extraction. It is considered that this entails a higher metabolic cost with respect to the maintenance of water and ion balance, explaining in part the increased metabolic rates of tunas [3, 62]. As a representative of gill-specific processes, aquaporin 3b (*aqp3b*), a member of major intrinsic channel-forming proteins involved in teleost osmoregulation, water balance and nitrogen excretion [63], was confirmed in gill-specific Cluster 4 (Fig. 6).

In contrast to the gills, the ovaries showed overrepresentation of glycan biosynthesis and metabolism, folding, sorting and degradation, as well as cell growth and death, nucleotide metabolism, transcription, translation, replication and repair (Fig. 7). Primary oocyte growth is characterised by intense transcriptional activity and the presence of abundant ribosomes in highly

granulated ooplasm [64, 65]. Maternal factors, protein and RNA species, are produced at this time and serve to support embryo development until the zygotic genome is activated [64]. Our analyses also suggest these processes are coupled with RNA degradation and mRNA surveillance pathways, crucial for maintenance of the integrity of transcriptional activity. Expression of a pescadillo (*pes*) homologue, important for early embryonic development in zebrafish [66], cell cycle progression and 60S ribosomal subunit synthesis [67, 68], and a transcript corresponding to exosome component 2 (*exosc2*) of the exosomal complex, comprising several 3′-5′ exoribonucleases involved in mRNA, 5.8S rRNA and small RNA processing [69], were confirmed as ovary-specific *cf.* other tested tissues (Fig. 6). These processes could also reflect the fact that the specific total RNA population of *T. thynnus* ovary seems to be dominated by the 5S RNA fraction, thus resembling the oocyte RNA complement of African clawed frog *Xenopus laevis* [28], Iberian ribbed newt *Pleurodeles waltl* [29] and a number of teleost species [30, 31]. It is believed that the 5S

Trumbić *et al. BMC Genomics* (2015) 16:1007

Page 13 of 17

fraction is stored with aminoacyl-tRNAs in 7S and/or 42S ribonucleoprotein particles in the ooplasm and is progressively transferred, starting with vitellogenesis, into ribosomes by the end of oocyte differentiation [28]. This observation concurs well with our samples being taken during winter when the most advanced oocyte stage found in *T. thynnus* ovaries is perinucleolar [70]. Other typical ovarian transcripts were also present in ovary-specific Cluster 1 (see Additional file 3), including the aquaporin 1a homologue, previously detected in *T. thynnus* ovaries [22, 23] and involved in the process of oocyte hydration prior to ovulation, which is necessary for providing buoyancy in pelagic fish eggs [64, 71]. A major driving force of this water influx is the osmotic pressure produced by oocyte yolk protein cleavage by lysosomal proteases termed cathepsins. Cathepsins B, D and L have been implicated in this process in fish [64, 72]. Gardner *et al.* [23] found transcripts similar to cathepsin S, L and cathepsin Z precursor differentially expressed in ovarian relative to testis tissue in *T. thynnus* and postulated that the species likely uses different cathepsin proteases for this purpose. In this study, expression profiles of different cathepsin-like transcripts (A, L, F, S, Z) were detected across tissue clusters (see Additional file 3) with transcripts similar to cathepsin A and S placed in the ovary-specific Cluster 1, thus supporting this hypothesis.

Testes displayed a number of transcripts in common with ovaries leading to a separate, shared "gonad-specific" cluster using Biolayout *Express* ³ᴰ (Cluster 5). This cluster was characterised by the presence of transcripts generally important for germline development. These included a subunit of mediator complex (*med6*), which is a transcriptional co-activator of RNA polymerase II that is particularly important for expression of developmentally regulated genes [73]. Fanconi anaemia complementation group 1 (*fanci*) was also present, part of the DNA repair mechanism employed during homologous recombination in meiosis [74], as well as Piwi-like 2 (*piwil2*) homologue, essential for the germline integrity via transposable element repression and gonad development [75]. Analysis of ovary- and testis-associated transcripts as a single amalgamated group (ovary + testis) did not reveal any new additional pathways relative to separate analysis of the two tissues (see Additional file 6). A further Biolayout cluster comprised largely testis-specific transcripts (Cluster 6). Testis-associated pathways (Fig. 7) were more similar to those of other inspected tissues than to ovaries, however there was an overlap. Shared pathways were associated with nucleotide metabolism, replication, repair and translation. Some of the transcriptional differences between ovaries and testes have been identified for *T. thynnus* [23]. A similar transcript to one found differentially expressed in this previous study, was

present here in the testis-specific Cluster 6; a component of t-complex *tcte1*, important for molecular interaction between sperm and egg *zona pellucida* [76]. This was also identified as testis-specific in rainbow trout by Rolland *et al.*, when compared to ovary, liver, muscle, gill and brain transcripts [77]. Intestinal fatty acid binding protein 2 (*fabp2*), suggested to play a role in sex dependent energy balance in *T. thynnus* [23], was found in heart-specific Cluster 2 (see Additional file 3) with its expression markedly elevated in testes compared to ovaries. The presence of several typically pituitary hormone transcripts, growth hormone (*gh*) and somatolactin (*sl*), were also apparent in the testis-specific cluster. The expression profile for somatolactin was re-investigated with qPCR and joint analysis of both datasets statistically confirmed ($p < 0.05$) this observation (Fig. 6). Both hormones are members of the growth hormone/prolactin family and are implicated in pleiotropic functions such as somatic growth, osmoregulation, lipid metabolism and reproduction [78, 79]. Extra-pituitary expression of growth hormone has been demonstrated in mammals and chicken testes [80], as well as fish [81] being localised to spermatogonia and primary spermatocytes in chicken and Sertoli cells surrounding the germ cells in Japanese eel. It has a paracrine role during spermatogenesis, stimulating proliferation of spermatogonia. This mitotic phase of spermatogenesis occurs in Mediterranean *T. thynnus* between October and January [82], corresponding to the collection period for our samples. Accordingly, a germ cell line marker vasa homologue (*vasa*), a putative RNA helicase found to be expressed in Pacific bluefin tuna spermatogonia [83], was also found to be up-regulated in testes of *T. thynnus* examined here. Primarily expressed in the pituitary gland, somatolactin mRNA has been detected across various tissues, including gonads in fish [84, 85]. These findings suggest *T. thynnus* testes actively transcribe *gh* and *sl*, warranting further investigation, especially in the context of reproductive cycle in both gonads.

During their spawning migrations, tuna utilise perigonadal mesenteric fat deposits as a source of metabolic energy for reproductive maturation and active swimming, which are primarily processed by the liver [82, 86]. In this study the *T. thynnus* liver was functionally characterised by lipid, amino acid and carbohydrate metabolism, fat digestion and absorption, bile secretion, metabolism of cofactors and vitamins, xenobiotic biodegradation and immune system (Fig. 7), consistent with the basic metabolic role of this organ [87] and agreeing with previous findings for the liver of *T. thynnus* [22]. Among the transcripts showing predominant expression in the liver, members of a haem-containing superfamily of cytochrome P450 monooxygenases, involved in metabolic biotransformation of xenobiotics and lipids often associated with liver [88],

Trumbić *et al. BMC Genomics* (2015) 16:1007

Page 14 of 17

were found (see Additional file 3). The liver is also a key production site for coagulation and complement proteolytic cascade components, mediators of blood clotting and elements of the innate immune system. Two representative transcripts, homologues of blood coagulation factor V (*f5*), pivotal proteins in maintaining haemostasis with dual pro-coagulant and anti-coagulant activities in humans [89], and complement component C4 (*c4*), which plays a crucial role in the classical and lectin pathways of complement activation as a subunit of the C3 convertase [90], were found to be significantly up-regulated in the liver compared to other tissues (Fig. 6).

Tuna cardiac function has been the subject of extensive investigation, as the heart operates at ambient temperature while supporting high metabolic rates and endothermic physiology [4]. Tuna have large hearts that function at close to maximum stroke volumes and depend, more than other teleosts, on oxygen consuming fatty acid oxidation as a fuel source, similar to mammals [62, 91]. Mitochondria isolated from skipjack tuna ventricles have a higher capacity to oxidise lactate than other metabolites, suggesting that lactate is also an alternative fuel source, especially during high intensity swimming when blood lactate is elevated [92]. These processes were well represented in the pathways and functional categories assigned to *T. thynnus* heart (*apex ventriculi*) compared to gill, liver and gonads in this study. Lipid (fatty acid degradation), energy (oxidative phosphorylation), carbohydrate metabolism (pyruvate metabolism, glycolysis/gluconeogenesis) and circulatory system (cardiac muscle contraction) were most prominent categories in functional contrasts displayed for the heart in Fig. 7. The ability to maintain cardiac performance during temperature acclimation, especially to cold, has been linked at the transcriptomic and functional level to Ca^{2+}-dependent excitation-contraction coupling (EC) in *T. orientalis* [24, 93]. Up-regulation of sarcoplasmic reticulum calcium-ATPase 2 (*atp2a2a* or *serca2*) homologue, vital for Ca^{2+} removal from cytosol into sarcoplasmic reticulum in vertebrates [94], has been identified in *T. thynnus* heart-specific Cluster 2 (Fig. 6 and see Additional file 3).

Amalgamating transcripts by pairs of tissues (*e.g.* gill + liver, gill + heart, gill + testis *etc.*), in order to resolve shared and unique pathways, showed pairings which included gill, heart and liver to be most informative and revealed functional signatures shared between these tissues. For instance, the most pronounced feature in Fig. 7 is the up-regulation of metabolism of other amino acids, β-alanine metabolism specifically (see Additional file 6), found exclusively for the heart + liver combination. β-Alanine is produced in vertebrate liver and is used as a rate-limiting precursor for synthesis of histidine-containing dipeptides such as carnosine and anserine in

skeletal muscle, heart and brain [95, 96]. These compounds are involved in intracellular pH buffering by tissues exhibiting high dependence on anaerobic glycolytic metabolism in which lactic acid is generated, such as fast-twich white skeletal muscle in fish [95]. High levels of histidine-containing dipeptides have been recorded for skipjack tuna white muscle, with anserine largely exceeding carnosine [97]. Aside from strong buffering capacity, carnosine has also been found to exhibit antioxidant, pro-contractile, chelating and anti-glycating properties and it is considered that β-alanine diet supplementation can have protective cardiac effects due to increased levels of carnosine [96]. Other important pathways mediating environmental and physiological adaptation were also identified. HIF-1 (hypoxia inducible factor) signalling pathway was found to be up-regulated in gill + heart and heart + liver combinations. Hif is a transcriptional factor governing the physiological response to hypoxia that induces a wide range of changes in energy metabolism, red blood cell formation, vascularisation, oxygen transport and apoptosis [98]. Regulation of *hif-1α* has been reported in haemaotopoietic and gas-exchange organs in *T. orientalis* as a result of cold temperature acclimation [99]. ABC transporter systems were found to be up-regulated in gill + heart and gill + liver combinations (see Additional file 6). The ATP-binding cassette (ABC-transporters) are trans-membrane spanning proteins involved in substrate translocation across membranes and form the basis of multi-xenobiotic resistance mechanism in aquatic animals, with high detoxification relevance. They also participate in transport of lipids, ions and other metabolites and have been attributed various tissue-specific roles [100]. Finally, insulin signalling pathway was found common to all pairwise combinations of gill, heart and liver (see Additional file 6). Exerting its effect through insulin receptors and associated effector signalling cascade actively transcribed in various peripheral organs, insulin has been implicated in wide range of processes in fish, including vital carbohydrate and lipid metabolism, appetite regulation, growth and development [101].

Conclusions

Sequence data generated from a novel mixed-tissue *T. thynnus* cDNA library provides an important transcriptomic resource that can be further employed for study of various aspects of *T. thynnus* ecology and genomics, with strong applications in aquaculture. We have developed and validated a new and flexible gene expression profiling tool, in the form of a 15 K oligonucleotide DNA microarray, which produced biologically relevant information agreeing with prior observations made in the literature, providing external validation for our new resource. The resolved expression profiles of gill, heart, liver and gonads provide a useful starting point for exploring gene expression patterns in *T.*

Trumbić *et al. BMC Genomics* (2015) 16:1007

Page 15 of 17

thynnus and for planning new transcriptomic studies. The oligo-microarray can also be further expanded to include un-annotated transcripts from the generated library with the aim of extending their functional characterisation through correlation with annotated entities. It is important to keep in mind that even tissue specific profiles can be population dependent [102], warranting further research on captive and wild tuna populations.

Availability of supporting data

The raw sequence data have been submitted to the EBI Sequence Read Archive (SRA) study PRJEB7253 (http://www.ebi.ac.uk/ena). The microarray design and data are available in the EBI ArrayExpress database (www.ebi.ac.uk/arrayexpress) under accession numbers A-MTAB-553 and E-MTAB-3412, respectively.

Additional files

Additional file 1: Mapping of *T. thynnus* transcripts to *O. niloticus* genome. (TXT 2853 kb)

Additional file 2: Mapping of *T. thynnus* transcripts to *T. orientalis* genome. (TXT 5528 kb)

Additional file 3: Results of network analyses and clustering using Markov algorithm for *T. thynnus* transcripts. Log2 normalised averaged signals are provided for each probe. (XLSX 904 kb)

Additional file 4: RT-qPCR related data. Table sheets containing primer sequences for genes included in qPCR analyses, BestKeeper stability for reference genes and Log2 transformed data for all targets. (XLSX 25 kb)

Additional file 5: Kruskal-Wallis test statistic for combined expression profiles of target genes based on microarray and qPCR data. (XLSX 14 kb)

Additional file 6: KEGG pathway analyses of *T. thynnus* tissue expression profiles. The analyses were performed by choosing each tissue as experimental state and all other as reference. The same was repeated by denoting combinations of two tissues together as a test sample. Data are given only for statistically significant pathways (q < 0.1) and the sign of stat. mean value indicates the direction of regulation. (XLSX 51 kb)

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

ŽT carried out the laboratory work, all components of microarray and RT-qPCR data analyses and drafted the manuscript. MB advised on experimental design, assembled and annotated the sequences and performed genome mappings. JBT participated in the laboratory work, experimental design and planning of the study. JEB contributed to experimental design and data analyses. KG supervised the sequencing at the Edinburgh Genomics Facility. IM participated in the experimental design, sample collection, planning and supervision of the study. All authors read, improved and approved the final manuscript.

Author details

[1]University Department of Marine Studies, University of Split, Split, Croatia. [2]Institute of Aquaculture, School of Natural Sciences, University of Stirling, Stirling, Scotland, UK. [3]Edinburgh Genomics, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3FLScotland, UK. [4]Institute of Oceanography and Fisheries, Split, Croatia.

References

1. FishBase. World Wide Web electronic publication. www.fishbase.org. Accessed Aug 2014.
2. Volff J-N. Genome evolution and biodiversity in teleost fish. Heredity. 2005; 94:280–94.
3. Brill RW. Selective advantages conferred by the high performance physiology of tunas, billfishes, and dolphin fish. Comp Biochem Physiol Part A Physiol. 1996;113:3–15.
4. Block BA, Stevens ED, editors. Tuna: Physiology, Ecology, and Evolution. San Diego: Academic; 2001.
5. Block BA, Finnerty JR, Stewart AF, Kidd J. Evolution of endothermy in fish: mapping physiological traits on a molecular phylogeny. Science. 1993;260: 210–4.
6. Block BA, Finnerty JR. Endothermy in fishes: a phylogenetic analysis of constraints, predispositions, and selection pressures. Environ Biol Fishes. 1994;40:283–302.
7. Rooker JR, Secor DH, De Metrio G, Schloesser R, Block BA, Neilson JD. Natal homing and connectivity in Atlantic bluefin tuna populations. Science. 2008; 322:742–4.
8. Carlsson J, McDowell JR, Carlsson JEL, Graves JE. Genetic identity of YOY bluefin tuna from the eastern and western Atlantic spawning areas. J Hered. 2007;98:23–8.
9. Boustany AM, Reeb CA, Block BA. Mitochondrial DNA and electronic tracking reveal population structure of Atlantic bluefin tuna (*Thunnus thynnus*). Mar Biol. 2008;156:13–24.
10. Block BA, Teo SL, Walli A, Boustany A, Stokesbury MJ, Farwell CJ, et al. Electronic tagging and population structure of Atlantic bluefin tuna. Nature. 2005;434:1121–7.
11. Walli A, Teo SLH, Boustany A, Farwell CJ, Williams T, Dewar H, et al. Seasonal movements, aggregations and diving behavior of Atlantic bluefin tuna (*Thunnus thynnus*) revealed with archival tags. PLoS One. 2009;4, e6151.
12. Fromentin J-M, Powers JE. Atlantic bluefin tuna: population dynamics, ecology, fisheries and management. Fish Fish. 2005;6:281–306.
13. MacKenzie BR, Mosegaard H, Rosenberg AA. Impending collapse of bluefin tuna in the northeast Atlantic and Mediterranean. Conserv Lett. 2009;2:25–34.
14. Taylor NG, McAllister MK, Lawson GL, Carruthers T, Block BA. Atlantic bluefin tuna: a novel multistock spatial model for assessing population biomass. PLoS One. 2011;6, e27693.
15. ICCAT. Report of The Standing Committee on Research and Statistics (SCRS). ICCAT: Madrid, Spain; 2014.
16. Mylonas CC, De La Gándara F, Corriero A, Ríos AB. Atlantic bluefin tuna (*Thunnus Thynnus*) farming and fattening in the Mediterranean Sea. Rev Fish Sci. 2010;18:266–80.
17. Rosenfeld H, Mylonas CC, Bridges CR, Heinisch G, Corriero A, Vassallo-Aguis R, et al. GnRHa-mediated stimulation of the reproductive endocrine axis in captive Atlantic bluefin tuna, *Thunnus thynnus*. Gen Comp Endocrinol. 2012; 175:55–64.
18. Sawada Y, Okada T, Miyashita S, Murata O, Kumai H. Completion of the Pacific bluefin tuna *Thunnus orientalis* (Temminck et Schlegel) life cycle. Aquac Res. 2005;36:413–21.

Trumbić *et al. BMC Genomics* (2015) 16:1007

Page 16 of 17

19. Taggart JB, Bron JE, Martin SAM, Seear PJ, Høyheim B, Talbot R, et al. A description of the origins, design and performance of the TRAITS-SGP Atlantic salmon *Salmo salar* L. cDNA microarray. J Fish Biol. 2008;72:2071–94.

20. Nielsen J, Pavey S. Perspectives: gene expression in fisheries management. Curr Zool. 2010;56:157–74.

21. Ekblom R, Galindo J. Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity. 2011;107:1–15.

22. Chini V, Cattaneo AG, Rossi F, Bernardini G, Terova G, Saroglia M, et al. Genes expressed in Blue Fin Tuna (*Thunnus thynnus*) liver and gonads. Gene. 2008;410:207–13.

23. Gardner LD, Jayasundara N, Castilho PC, Block B. Microarray gene expression profiles from mature gonad tissues of Atlantic bluefin tuna. Thunnus thynnus in the Gulf of Mexico BMC Genomics. 2012;13:530.

24. Jayasundara N, Gardner LD, Block BA. Effects of temperature acclimation on Pacific bluefin tuna (*Thunnus orientalis*) cardiac transcriptome. Am J Physiol Regul Integr Comp Physiol. 2013;305:R1010–20.

25. Nakamura Y, Mori K, Saitoh K, Oshima K, Mekuchi M, Sugaya T, et al. Evolutionary changes of multiple visual pigment genes in the complete genome of Pacific bluefin tuna. Proc Natl Acad Sci U S A. 2013;110:11061–6.

26. Lepen Pleić I, Secombes CJ, Bird S, Mladineo I. Characterization of three pro-inflammatory cytokines, TNFα1, TNFα2 and IL-1β, in cage-reared Atlantic bluefin tuna *Thunnus thynnus*. Fish Shellfish Immunol. 2013;36:98–112.

27. Chomczynski P, Mackey K. Short technical reports. Modification of the TRI reagent procedure for isolation of RNA from polysaccharide- and proteoglycan-rich sources. Biotechniques. 1995;19:942–5.

28. Picard B, Wegnez M. Isolation of a 7S particle from *Xenopus laevis* oocytes: a 5S RNA-protein complex. Proc Natl Acad Sci U S A. 1979;76:241–5.

29. Van den Eynde H, Mazabraud A, Denis H. Biochemical research on oogenesis. RNA accumulation in the oocytes of the newt *Pleurodeles waltl*. Development. 1989;106:11–6.

30. Mazabraud A, Wegnez M, Denis H. Biochemical research on oogenesis. RNA accumulation in the oocytes of the newt *Pleurodeles waltl*. Dev Biol. 1975;44: 326–32.

31. Cerio O, Rojo-Bartolomé I, Bizarro C, Ortiz-Zarragoitia M, Cancio I. 5S rRNA and accompanying proteins in gonads: powerful markers to identify sex and reproductive endocrine disruption in fish. Environ Sci Technol. 2012;46:7763–71.

32. Zhulidov PA, Bogdanova EA, Shcheglov AS, Vagner LL, Khaspekov GL, Kozhemyako VB, et al. Simple cDNA normalization using kamchatka crab duplex-specific nuclease. Nucleic Acids Res. 2004;32, e37.

33. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinformatics. 2011;27:863–4.

34. Myers E, Sutton G, Delcher A, Dew I, Fasulo D, Flanigan M, et al. A whole-genome assembly of *Drosophila*. Science. 2000;287:2196–204.

35. Lee W-P, Stromberg MP, Ward A, Stewart C, Garrison EP, Marth GT. MOSAIK: a hash-based algorithm for accurate next-generation sequencing short-read mapping. PLoS One. 2014;9, e90581.

36. NCBI *Thunnus thynnus* EST database. http://www.ncbi.nlm.nih.gov/nucest/?term=thunnus. Accessed 6 Oct 2013.

37. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27:573–80.

38. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 2000;16:276–7.

39. Iseli C, Jongeneel C, Bucher P. ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. Proc / Int Conf Intell Syst Mol Biol ISMB. 1999;138–148.

40. Altschul S, Gish W, Miller W, Myers E, Lipman D. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.

41. RefSeq: NCBI Reference Sequence Database. http://www.ncbi.nlm.nih.gov/refseq. Accessed 6 Oct 2013.

42. Ashburner M, Ball C, Blake J, Botstein D, Butler H, Cherry J, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium Nat Genet. 2000;25:25–9.

43. Bekaert M. Slimlt. p. Creative Commons Attribution 3.0 License. http://www.isalmon.eu/slimit/. Accessed 6 Oct 2013.

44. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 2000;28:27–30.

45. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 2007;35(Web Server issue):W182–5.

46. Krzywinski M, Birol I, Jones SJ, Marra MA. Hive plots–rational approach to visualizing networks. Brief Bioinform. 2012;13:627–44.

47. Theocharidis A, van Dongen S, Enright AJ, Freeman TC. Network visualization and analysis of gene expression data using BioLayout *Express* [3D]. Nat Protoc. 2009;4:1535–50.

48. Luo W, Friedman MS, Shedden K, Hankenson KD, Woolf PJ. GAGE: generally applicable gene set enrichment for pathway analysis. BMC Bioinformatics. 2009;10:161.

49. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 2004;5:R80.

50. Morais S, Mourente G, Ortega A, Tocher JA, Tocher DR. Expression of fatty acyl desaturase and elongase genes, and evolution of DHA:EPA ratio during development of unfed larvae of Atlantic bluefin tuna (*Thunnus thynnus* L.). Aquaculture. 2011;313:129–39.

51. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3–new capabilities and interfaces. Nucleic Acids Res. 2012;40, e115.

52. Pfaffl MW. Quantification Strategies in Real-Time PCR. In A-Z of Quantitative PCR. Edited by Bustin SA. La Jolla, CA, USA: International University Line; 2004:87–112.

53. Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper -Excel-based tool using pair-wise correlations. Biotechnol Lett. 2004;26:509–15.

54. Miron M, Woody OZ, Marcil A, Murie C, Sladek R, Nadon R. A methodology for global validation of microarray experiments. BMC Bioinformatics. 2006;7: 333.

55. Gilles A, Meglécz E, Pech N, Ferreira S, Malausa T, Martin J-F. Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. BMC Genomics. 2011;12:245.

56. Chow S, Nakagawa T, Suzuki N, Takeyama H, Matsunaga T. Phylogenetic relationships among *Thunnus* species inferred from rDNA ITS1 sequence. J Fish Biol. 2006;68:24–35.

57. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, et al. A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci U S A. 2004;101:6062–7.

58. Morey JS, Ryan JC, Van Dolah FM. Microarray validation: factors influencing correlation between oligonucleotide microarrays and real-time PCR. Biol Proced Online. 2006;8:175–93.

59. Evans DH, Piermarini PM, Choe KP. The multifunctional fish gill: dominant site of gas exchange, osmoregulation, acid–base regulation, and excretion of nitrogenous waste. Physiol Rev. 2005;85:97–177.

60. Uribe C, Folch H, Enriquez R, Moran G. Innate and adaptive immunity in teleost fish: a review. Vet Med. 2011;56:486–503.

61. Brill RW, Bushnell PG. The cardiovascular system of tunas. In: Block BA, Stevens ED, editors. Tuna: Physiology, Ecology, and Evolution, vol. 19. San Diego: Academic; 2001. p. 79–120.

62. Bushnell PG, Brill RW. Oxygen transport and cardiovascular responses in skipjack tuna (*Katsuwonus pelamis*) and yellowfin tuna (*Thunnus albacares*) exposed to acute hypoxia. J Comp Physiol B. 1992;162:131–43.

63. Cerdà J, Finn RN. Piscine aquaporins: an overview of recent advances. J Exp Zool. 2010;313A:623–50.

64. Babin PJ, Cerdà J, Lubzens E, editors. The Fish Oocyte From Basic Studies to Biotechnological Applications. Dordrecht, The Netherlands: Springer; 2007.

65. Abascal FJ, Medina A. Ultrastructure of oogenesis in the bluefin tuna, *Thunnus thynnus*. J Morphol. 2005;264:149–60.

66. Allende ML, Amsterdam A, Becker T, Kawakami K, Gaiano N, Hopkins N. Insertional mutagenesis in zebrafish identifies two novel genes, pescadillo and dead eye, essential for embryonic development. Genes Dev. 1996;10: 3141–55.

67. Kinoshita Y, Jarell AD, Flaman JM, Foltz G, Schuster J, Sopher BL, et al. Pescadillo, a novel cell cycle regulatory protein abnormally expressed in malignant cells. J Biol Chem. 2001;276:6656–65.

68. Oeffinger M, Leung A, Lamond A, Tollervey D. Yeast Pescadillo is required for multiple activities during 60S ribosomal subunit synthesis. RNA. 2002;8: 626–36.

69. Allmang C, Kufel J, Chanfreau G, Mitchell P, Petfalski E, Tollervey D. Functions of the exosome in rRNA, snoRNA and snRNA synthesis. EMBO J. 1999;18:5399–410.

70. Corriero A, Desantis S, Deflorio M, Acone F, Bridges CR, De la Serna JM, et al. Histological investigation on the ovarian cycle of the bluefin tuna in the western and central Mediterranean. J Fish Biol. 2003;63:108–19.

Trumbić *et al. BMC Genomics* (2015) 16:1007

Page 17 of 17

71. Fabra M, Raldúa D, Bozzo MG, Deen PMT, Lubzens E, Cerdà J. Yolk proteolysis and aquaporin-1o play essential roles to regulate fish oocyte hydration during meiosis resumption. Dev Biol. 2006;295:250–62.

72. Carnevali O, Carletta R, Cambi A, Vita A, Bromage N. Yolk formation and degradation during oocyte maturation in seabream *Sparus aurata*: involvement of two lysosomal proteinases. Biol Reprod. 1999;60:140–6.

73. Kwon JY, Park JM, Gim BS, Han SJ, Lee J, Kim YJ. *Caenorhabditis elegans* mediator complexes are required for developmental-specific transcriptional activation. Proc Natl Acad Sci U S A. 1999;96:14990–5.

74. Titus TA, Yan Y-L, Wilson C, Starks AM, Frohnmayer JD, Canestro C, et al. The Fanconi anemia/BRCA gene network in zebrafish: Embryonic expression and comparative genomics. Mutat Res. 2009;668:117–32.

75. Zhou Y, Wang F, Liu S, Zhong H, Liu Z, Tao M, et al. Human chorionic gonadotropin suppresses expression of Piwis in common carp (*Cyprinus carpio*) ovaries. Gen Comp Endocrinol. 2012;176:126–31.

76. Juneja R, Agulnik SI, Silver LM. Sequence divergence within the sperm-specific polypeptide TCTE1 is correlated with species-specific differences in sperm binding to zona-intact eggs. J Androl. 1998;19:183–8.

77. Rolland AD, Lareyre J-J, Goupil A-S, Montfort J, Ricordel M-J, Esquerré D, et al. Expression profiling of rainbow trout testis development identifies evolutionary conserved genes involved in spermatogenesis. BMC Genomics. 2009;10:546.

78. Kaneko T, Hirano T. Role of prolactin and somatolactin in calcium regulation in fish. J Exp Biol. 1993;184:31–45.

79. Pérez-Sánchez J. The involvement of growth hormone in growth regulation, energy homeostasis and immune function in the gilthead sea bream (*Sparus aurata*): a short review. Fish Physiol Biochem. 2000;22:135–44.

80. Harvey S, Baudet M-L, Murphy A, Luna M, Hull KL, Aramburo C. Testicular growth hormone (GH): GH expression in spermatogonia and primary spermatocytes. Gen Comp Endocrinol. 2004;139:158–67.

81. Miura C, Shimizu Y, Uehara M, Ozaki Y, Young G, Miura T. Gh is produced by the testis of Japanese eel and stimulates proliferation of spermatogonia. Reproduction. 2011;142:869–77.

82. Abascal F, Megina C, Medina A. Testicular development in migrant and spawning bluefin tuna (*Thunnus thynnus* (L.)) from the eastern Atlantic and Mediterranean. Fish Bull. 2004;102:407–17.

83. Nagasawa K, Takeuchi Y, Miwa M, Higuchi K, Morita T, Mitsuboshi T, et al. cDNA cloning and expression analysis of a *vasa*-like gene in Pacific bluefin tuna *Thunnus orientalis*. Fish Sci. 2009;75:71–9.

84. Yang BY, Arab M, Chen TT. Cloning and characterization of rainbow trout (*Oncorhynchus mykiss*) somatolactin cDNA and its expression in pituitary and nonpituitary tissues. Gen Comp Endocrinol. 1997;106:271–80.

85. Lynn SG, Shepherd BS. Molecular characterization and sex-specific tissue expression of prolactin, somatolactin and insulin-like growth factor-I in yellow perch (*Perca flavescens*). Comp Biochem Physiol Part B Biochem Mol Biol. 2007;147:412–27.

86. Mourente G, Megina C, Díaz-Salvago E. Lipids in female northern bluefin tuna (*Thunnus thynnus thynnus* L.) during sexual maturation. Fish Physiol Biochem. 2002;24:351–63.

87. Wolf JC, Wolfe MJ. A brief overview of nonneoplastic hepatic toxicity in fish. Toxicol Pathol. 2005;33:75–85.

88. Andersson T, Förlin L. Regulation of the cytochrome P450 enzyme system in fish. Aquat Toxicol. 1992;24:1–20.

89. Segers K, Dahlbäck B, Nicolaes GAF. Coagulation factor V and thrombophilia: Background and mechanisms. Thromb Haemost. 2007;98:530–42.

90. Boshra H, Li J, Sunyer JO. Recent advances on the complement system of teleost fish. Fish Shellfish Immunol. 2006;20:239–62.

91. Moyes CD. Cardiac metabolism in high performance fish. Comp Biochem Physiol Part A Physiol. 1996;113:69–75.

92. Moyes C, Mathieu-Costello OA, Brill RW, Hochachka PW. Mitochondrial metabolism of cardiac and skeletal muscles from a fast (*Katsuwonus pelamis*) and a slow (*Cyprinus carpio*) fish. Can J Zool. 1992;70:1246–53.

93. Shiels H, Di Maio A, Thompson S, Block B. Warm fish with cold hearts: thermal plasticity of excitation–contraction coupling in bluefin tuna. Proc R Soc B Biol Sci. 2011;278:18–27.

94. Landeira-Fernandez AM, Castilho PC, Block BA. Thermal dependence of cardiac SR Ca^{2+}-ATPase from fish and mammals. J Therm Biol. 2012;37:217–23.

95. Abe H. Role of histidine-related compounds as intracellular proton buffering constituents in vertebrate muscle. Biochem. 2000;65:757–65.

96. McCarty MF, DiNicolantonio JJ. β-Alanine and orotate as supplements for cardiac protection. Open Hear. 2014;1, e000119.

97. Abe H, Brill R, Hochachka P. Metabolism of L-histidine, carnosine, and anserine in skipjack tuna. Physiol Zool. 1986;59:439–50.

98. Nikinmaa M, Rees BB. Oxygen-dependent gene expression in fishes. Am J Physiol Regul Integr Comp Physiol. 2005;288:R1079–90.

99. Mladineo I, Block BA. Expression of Hsp70, Na$^+$/K$^+$ ATP-ase, HIF-1α, IL-1β and TNF-α in captive Pacific bluefin tuna (*Thunnus orientalis*) after chronic warm and cold exposure. J Exp Mar Bio Ecol. 2009;374:51–7.

100. Luckenbach T, Fischer S, Sturm A. Current advances on ABC drug transporters in fish. Comp Biochem Physiol Part C Toxicol Pharmacol. 2014;165:28–52.

101. Caruso MA, Sheridan MA. New insights into the signaling system and function of insulin in fish. Gen Comp Endocrinol. 2011;173:227–47.

102. Whitehead A, Crawford DL. Variation in tissue-specific gene expression among natural populations. Genome Biol. 2005;6:R13.