Marine Genomics 29 (2016) 27–29

Contents lists available at ScienceDirect

Marine Genomics

journal homepage: www.elsevier.com/locate/margen

Genomics/technical resources

Transcriptome sequencing of maraena whitefish (*Coregonus maraena*)

Andreas Brietzke [a,1], Andreas Borchel [a,b,1], Simone Altmann [a], Mareen Nipkow [a], Alexander Rebl [a], Ronald M. Brunner [a], Tom Goldammer [a,*]

[a] Leibniz Institute for Farm Animal Biology, Institute for Genome Biology, Fish Genetics Unit, Wilhelm-Stahl-Allee 2, 18196, Dummerstorf, Germany
[b] SLRC — Sea Lice Research Centre, Department of Biology, University of Bergen, Thormøhlensgata 55, 5008 Bergen, Norway

ARTICLE INFO

Article history:
Received 1 April 2016
Received in revised form 30 May 2016
Accepted 30 May 2016
Available online 4 June 2016

Keywords:
Maraena whitefish
Salmonid
454 FLX titanium sequencing
454-Newbler assembly

ABSTRACT

Maraena whitefish (*Coregonus maraena*, Bloch, 1779) is a high-quality food fish belonging to the family Salmonidae with considerable economic relevance in the Baltic area. Aquaculture of this species is fundamental for its successful conservation and thus sustainable fisheries. Robust fishes obtained from breeding lines build the basis for effective aquaculture. Doubtless, the utilization of transcriptome sequencing and identification of genetic markers contribute to this aim. 454 FLX Titanium Sequencing provided 1.31 million sequence reads representing a first insight into the *C. maraena* transcriptome. The 454 Newbler Assembly arranged 29,094 contigs with an average length of 798 bp. We found a whole series of transcripts highly probably resulting from ancient genome duplication and annotated 2887 different transcripts with an average length of 812 bp. Functional annotation obtained a transcript composition predominantly comprising enzyme-coding genes.

1. Introduction

Maraena whitefish (*Coregonus maraena*) is a white-fleshed food fish in the riparian states of the Baltic Sea, which is in neither flavor nor nutrition inferior to its famous salmonid relatives, Atlantic salmon (*Salmo salar*) and rainbow trout (*Oncorhynchus mykiss*). A combination of extensive fishing, eutrophication of its estuary spawning areas, and habitat fragmentation through anthropic obstruction have brought the population of *C. maraena* to the edge of extinction, especially in the southern Baltic Sea area (Olsson et al., 2012). Intensive restocking operations in the 1990s led to short-term stabilization of the yields (Jennerich and Schulz, 2009; Landesamt für Landwirtschaft, L. und F.M., 2015; Schulz, 2000). To relieve the native population and guarantee a continuous supply of whitefish as an interregional delicacy, farming of *C. maraena* was successfully launched in Mecklenburg-Western Pomerania, Germany (Arndt and Jansen, 2008). The establishment of *C. maraena* breeding lines has encountered aquaculture-related challenges such as slow growth rates or susceptibility towards husbandry stress and pathogens (Altmann et al., 2015). Marker-assisted breeding strategies require profound knowledge of the genetic constitution of *C. maraena*, but only few relevant sequences are publicly available to date (Altmann et al., 2015; Brietzke et al., 2014; Rebl et al., 2011; Verleih et al., 2012). Therefore, we sequenced the transcriptome of *C. maraena* across tissues and ages as a starting point for future investigations.

2. Methods and analysis

2.1. Sample collection and sequencing

To obtain the broadest picture of the transcript composition possible, we collected various tissues from differently aged individuals (Suppl. Fig. 1). Tissues were homogenized with TriZol Reagent (Life Technologies). Total RNA was extracted with the RNeasy Mini Kit (Qiagen); DNA contamination was removed using the RNase-free DNase Set (Qiagen). Messenger RNA was purified from 10 μg total RNA using the mRNA-only Eukaryotic mRNA Isolation Kit (Epicentre). Synthesis of double-stranded cDNA was conducted with the Mint-Universal cDNA Synthesis Kit (Evrogen) and 1 μg mRNA as a template. The cDNA was normalized with the Trimmer Kit (Evrogen), thus reducing the number of highly abundant transcripts. Size selection and amplification of normalized cDNA are listed in Suppl. Table 1. Library generation for 454 FLX sequencing was carried out according to standard protocols (Roche/454 Life Sciences). In brief, the concatenated normalized cDNA inserts were sheared randomly by nebulization into fragments ranging in size from 400 bp to 900 bp, end-polished and ligated with 454 A and B adaptors (containing identifier tags). The entire constructed fragment library was sequenced on a picotiterplate (PTP) on the GS FLX system using Roche/454 FLX + chemistry. Table 1 gives

Abbreviations: CKMT2, creatine kinase, mitochondrial 2; HNRPDL, heterogeneous nuclear ribonucleoprotein D like; TGFBI, transforming growth factor beta induced; TUBA1C, tubulin alpha 1c.

* Corresponding author.
E-mail address: tomgoldammer@fbn-dummerstorf.de (T. Goldammer).

[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.margen.2016.05.006
1874-7787/

Table 1
Characteristics of the whitefish transcriptome sequencing project, compliant with the MIGS standard.

Item	Description
Investigation_type	Eukaryote
Project_name	Transcriptome sequencing of maraena whitefish (*Coregonus maraena*)
Collected_by	S. Altmann and M. Nipkow
Collection_date	From 2013-04-26 to 2013-10-8 (detailed in Suppl. Fig. 1)
Lat_lon	54.380 N 12.523E
Country	Germany (Mecklenburg-Western-Pomerania)
Environment	Bodden chain
Ref_biomaterial	Altmann et al. (2015)
Biotic_relationship	Farmed fish
Rel_to_oxygen	Aerobe
Isol_growth_condt	Altmann et al. (2015)
Sequencing_meth	Pyrosequencing
Num_replicons	NA
Assembly	Newbler 2.8
Finishing_strategy	Draft
Estimated_size	1.87E9 (Berthelot et al., 2014)
Biome	ENVO:00002030 (aquatic biome)
Feature	ENVO:00000294 (fish farm)
Material	ENVO:00002019 (brackish water)
Geo_loc_name	Darss-Zingst, Germany
Sample-material	See Suppl. Fig. 1
Motility	Yes
Assembly method	Newbler 2.8
Sequencing technology	Pyrosequencing
Ploidy	Pseudo-tetraploid

a brief overview of the rearing conditions of the sampled whitefish and summarizes the key characteristics of the transcriptome sequencing project, compliant with the MIGS standard.

2.2. Transcriptome assembly

We obtained altogether 1.31 m reads from the sequencing procedure. Prior to assembly, we screened these sequencing reads to clip out the Sfi-linker used for library preparation. The processed reads were then assembled into individual transcripts using the Roche/454 Newbler software at default settings (454 Life Sciences Corporation, Software Release: 2.8 — 20120726_1306). Assembly shifted the average read length of 477 bp to an average contig length of 798 bp (Suppl. Fig. 2). Moreover, assembly yielded a total number of 26,880 (≥200 bp) contigs that were utilized for functional annotation. Suppl. Table 2 provides a summary of the raw and processed data.

Table 2
Summary of the ingenuity pathway analysis.

Top canonical pathways	Overlap	P-value
Mitochondrial dysfunction	32/171 (18.7%)	6.37E−11
Oxidative phosphorylation	25/109 (22.9%)	8.50E−11
Protein ubiquitination pathway	32/255 (12.5%)	1.23E−06
Epithelial adherens junction signaling	22/146 (15.1%)	2.89E−06
Actin nucleation by ARP-WASP complex	12/56 (21.4%)	1.46E−05
Top biofunctions	**Molecules**	**P-value range**
A) Molecular and cellular functions		
Cell death and survival	412	4.36E−04–2.13E−28
Cellular growth and proliferation	431	4.62E−04–1.63E−27
Gene expression	286	9.78E−05–4.32E−18
Cellular development	371	4.62E−04–7.71E−17
Protein synthesis	184	3.65E−04–257E−14
B) Physiological system development and function		
Connective tissue development and function	187	3.70E−04–1.53E−12
Organismal survival	249	2.44E−12–2.44E−12
Tissue development	273	4.62E−04–9.34E−11
Cardiovascular system development and function	151	4.01E−04–3.17E−09
Embryonic development	230	4.62E−04–3.85E−08

2.3. Functional annotation

To establish a preferably large scale of full transcript sequences with a maximum reliable annotation, we first performed the blastn function using exclusively the sequence database from salmonid species (≥85% identity, ≥85% query coverage). To obtain standardized human gene symbols (Gray et al., 2015), the sequence of the best blastn hit was used for blastp analysis against Swissprot-DB. Conditions for an annotation were ≥50% protein identity and ≥85% query coverage with the human amino acid sequence. Large read lengths, as obtained in this experiment, allow for identifying highly similar duplicated genes differing in only few nucleotide exchanges. The presence of multiple paralogues is characteristic of salmonid genomes having undergone a further round of duplication in comparison to other piscine taxa (Allendorf and Thorgaard, 1984; Berthelot et al., 2014). We found multiple copies of the transcripts encoding transforming growth factor beta-induced protein (TGFBI, Suppl. Fig. 3), heterogeneous nuclear ribonucleoprotein D-like protein (HNRPDL), mitochondrial creatine kinase 2 (CKMT2) or tubulin-alpha 1c (TUBA1C), to name but a few.

Gene annotation with official gene symbols was required to obtain gene ontology terms. The genes and gene products search tool (http://geneontology.org/; Ashburner et al., 2000) assisted the correction of

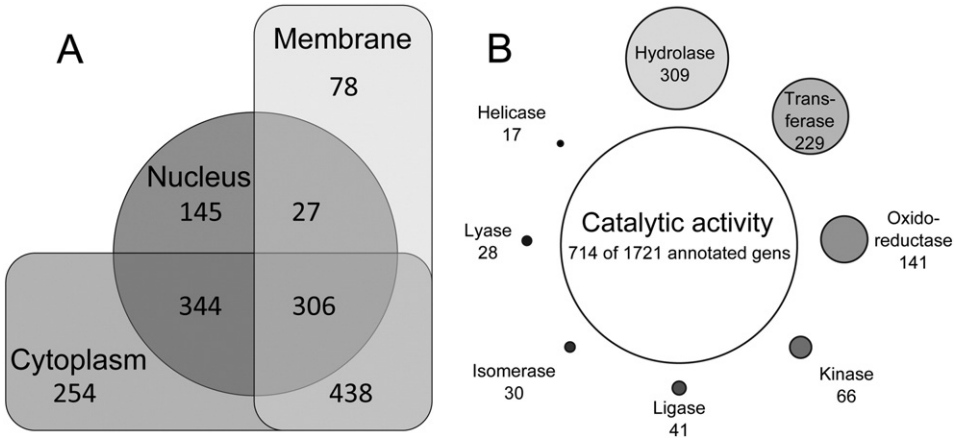

Fig. 1. Localization and functional aspects of the annotated genes. 1721 GO terms were included in the analysis of functional aspects with the GO term mapper. Expression distribution on the compartments of the annotated genes is presented in A, diversity and appearance of catalytic activities in B.

obsolete, false, or missing nomenclature. A total number of 1721 annotated transcripts were analyzed with the GO term mapper (http://go.princeton.edu/cgi-bin/GOTermMapper).

Regarding the cellular spatial distribution of the encoded factors, we found 254, 145, and 78 factors exclusively associated with the cytoplasm, nucleus, and membrane, respectively; 306 factors were ubiquitously present (Fig. 1A).

From a functional point of view, we found 714 transcripts coding for factors with catalytic activity, of which 309 possessed hydrolase and 229 transferase activity (Fig. 1B). In addition, IPA ingenuity pathway analysis identified representative pathways and biofunctions. The top canonical pathway found was "Mitochondrial Dysfunction" (Table 2, "Top Canonical Pathways"). Within the category "Molecular and Cellular Functions" cellular growth- and proliferation-related factors were the most abundant molecules (Table 2, "Top Biofunctions", A). Factors with relevance in tissue development were most abundant in the category "Physiological System Development and Function" (Table 2, "Top Biofunctions", B).

3. Data deposition

All data belongs to BioProject PRJNA302355. Raw reads were deposited in the NCBI Sequence Read Archive (SRA accession: SRP066290). Annotated sequence files of this Transcriptome Shotgun Assembly project have been deposited at DDBJ/EMBL/GenBank under the accession GEDA00000000. The version described in this paper is the first version, GEDA01000000.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.margen.2016.05.006.

Acknowledgments

We thank Ingrid Hennings, Brigitte Schöpel and Luisa Falkenthal for their excellent technical assistance. This work is funded by the European Fisheries Fund (EFF) and the Ministry of Agriculture, Environment, and Consumer Protection Mecklenburg-Western Pomerania, Germany (Pilot project VI-560/730-32614).

References

Allendorf, R.W., Thorgaard, G.H., 1984. Evolutionary Genetics of Fishes. Plenum Press, New York.

Altmann, S., Rebl, A., Kühn, C., Goldammer, T., 2015. Identification and de novo sequencing of housekeeping genes appropriate for gene expression analyses in farmed maraena whitefish (*Coregonus maraena*) during crowding stress identification and de novo sequencing of housekeeping genes appropriate maraena). Fish Physiol. Biochem. 41, 397–412.

Arndt, G.-M., Jansen, W., 2008. Erste Ergebnisse Zur Aufzucht von Ostseeschnäpeln (*Coregonus lavaretus*) in Einer Teichwirtschaft in Mecklenburg-Vorpommern 5–13.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G., 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25 (1), 25–29.

Berthelot, C., Brunet, F., Chalopin, D., Juanchich, A., Bernard, M., Noël, B., Bento, P., Da Silva, C., Labadie, K., Alberti, A., Aury, J.-M., Louis, A., Dehais, P., Bardou, P., Montfort, J., Klopp, C., Cabau, C., Gaspin, C., Thorgaard, G.H., Boussaha, M., Quillet, E., Guyomard, R., Galiana, D., Bobe, J., Volff, J.-N., Genêt, C., Wincker, P., Jaillon, O., Roest Crollius, H., Guiguen, Y., 2014. The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. Nat. Commun. 5, 3657.

Brietzke, A., Goldammer, T., Rebl, H., Korytář, T., Köllner, B., Yang, W., Rebl, A., Seyfert, H.-M., 2014. Characterization of the interleukin 1 receptor-associated kinase 4 (IRAK4)-encoding gene in salmonid fish: the functional copy is rearranged in *Oncorhynchus mykiss* and that factor can impair TLR signaling in mammalian cells. Fish Shellfish Immunol. 36, 206–214.

Gray, K.A., Yates, B., Seal, R.L., Wright, M.W., Bruford, E.A., 2015. Genenames.org: the HGNC resources in 2015. Nucleic Acids Res. 43 (Database issue), D1079–D1085.

Jennerich, H.-J., Schulz, N., 2009. Zur situation des ostseeschnäpels (*Coregonus lavaretus* Balticus, Thienemann, 1922) in Mecklenburg-Vorpommern. Beiträge Zur Fischerei, Mitteilungen Der Landesforschungsanstalt für Landwirtschaft Und Fischerei 45, pp. 12–20.

Landesamt für Landwirtschaft, L. und F.M, 2015. Fangstatistik der Kl. Hochsee- Und Küstenfischerei M-V 2005–2014 (http://www.lallf.de/fileadmin/media/PDF/fischer/5_Statistik/Fangstatistik_2014Gebiete.pdf).

Olsson, J., Florin, A.B., Mo, K., Aho, T., Ryman, N., 2012. Genetic structure of whitefish (*Coregonus maraena*) in the Baltic Sea. Estuar. Coast. Shelf Sci. 97, 104–113.

Rebl, A., Köbis, J.M., Fischer, U., Takizawa, F., Verleih, M., Wimmers, K., Goldammer, T., 2011. MARCH5 gene is duplicated in rainbow trout, but only fish-specific gene copy is up-regulated after VHSV infection. Fish Shellfish Immunol. 31, 1041–1050.

Schulz, N., 2000. Das Wiedereinbürgerungs- und Besatzprogramm des Ostseeschnäpels *Coregonus lavaretus* Balticus (Thienemann) in der vorpommerschen Boddenlandschaft, Rückblick und Ausblick. Fisch und umwelt M-V e.V. Jahresheft 2000, 45–59.

Verleih, M., Rebl, A., Köllner, B., Korytář, T., Anders, E., Wimmers, K., Goldammer, T., 2012. Comparative molecular characterization of the regucalcin (RGN) gene in rainbow trout (*Oncorhynchus mykiss*) and maraena whitefish (*Coregonus marena*). Mol. Biol. Rep. 39, 4291–4300.