SCIENTIFIC REPORTS

OPEN

Received: 15 March 2016
Accepted: 28 July 2016
Published: 15 September 2016

The *Tetraodon nigroviridis* reference transcriptome: developmental transition, length retention and microsynteny of long non-coding RNAs in a compact vertebrate genome

Swaraj Basu[1,†], Yavor Hadzhiev[2], Giuseppe Petrosino[1], Chirag Nepal[3], Jochen Gehrig[2,§], Olivier Armant[4], Marco Ferg[4], Uwe Strahle[4], Remo Sanges[1] & Ferenc Müller[2]

Pufferfish such as fugu and tetraodon carry the smallest genomes among all vertebrates and are ideal for studying genome evolution. However, comparative genomics using these species is hindered by the poor annotation of their genomes. We performed RNA sequencing during key stages of maternal to zygotic transition of *Tetraodon nigroviridis* and report its first developmental transcriptome. We assembled 61,033 transcripts (23,837 loci) representing 80% of the annotated gene models and 3816 novel coding transcripts from 2667 loci. We demonstrate the similarities of gene expression profiles between pufferfish and zebrafish during maternal to zygotic transition and annotated 1120 long non-coding RNAs (lncRNAs) many of which differentially expressed during development. The promoters for 60% of the assembled transcripts result validated by CAGE-seq. Despite the extreme compaction of the tetraodon genome and the dramatic loss of transposons, the length of lncRNA exons remain comparable to that of other vertebrates and a small set of lncRNAs appears enriched for transposable elements suggesting a selective pressure acting on lncRNAs length and composition. Finally, a set of lncRNAs are microsyntenic between teleost and vertebrates, which indicates potential regulatory interactions between lncRNAs and their flanking coding genes. Our work provides a fundamental molecular resource for vertebrate comparative genomics and embryogenesis studies.

Pufferfish species of the Tetraodontidae family such as fugu[1] and tetraodon[2] carry the smallest genomes among all vertebrates (350–400 Mb) whose size is about 1/8th of the human genome. Their genomes are thought to be enriched for functional elements because, with respect to other vertebrates, they are characterized by lower percentages of repetitive sequences, shorter introns and intergenic regions, higher gene density and chromosomal stability[2]. Therefore, they were suggested to be ideal models for comparative analysis[3]. The compact nature of these genomes could also aid in the characterization of syntenic relationships and potentially highlight regulatory interactions between *cis* regulatory elements which are distributed over megabases in larger vertebrate genomes. The sequencing of the tetraodon genome permitted the first accurate prediction of the number of human protein

[1]Biology and Evolution of Marine Organisms, Stazione Zoologica Anton Dohrn, Villa Comunale, Naples - 80121, Italy. [2]Institute of Cancer and Genomic Sciences, College of Medical and Dental Sciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK. [3]BRIC - Biotech Research & Innovation Centre, University of Copenhagen, Ole Maaløes Vej 5, DK-2200 Copenhagen N, Denmark. [4]Karlsruhe Institute of Technology (KIT), Campus North, Institute of Toxicology and Genetics (ITG), Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany. †Present address: Department of Medical Biochemistry and Cell Biology, Institute of Biomedicine, Sahlgrenska Academy, University of Gothenburg, 405 30 Gothenburg, Sweden. §Present address: Acquifer AG, Sophienstraße 136, 76135 Karlsruhe, Germany. Correspondence and requests for materials should be addressed to R.S. (email: remo.sanges@gmail.com) or F.M. (email: f.mueller@bham.ac.uk)