

BRIEF REPORT

Open Access



Comparative transcriptome reveal the potential adaptive evolutionary genes in *Andrias davidianus*

Qiaomu Hu^{1*}, Quanhe Wang^{1,2}, Yan Meng¹, Haifeng Tian¹ and Hanbing Xiao^{1*}

Abstract

To search the evidence of molecular evolution mechanism for aquatic and cave habitat in *Andrias davidianus*, the evolution analysis was carried out among several species transcriptome data. The transcriptome data of *Notophthalmus viridescens*, *Xenopus tropicalis*, *Cynops pyrrhogaster*, *Hynobius chinensis* and *A. davidianus* were obtained from the Genbank and reassembled except *Xenopus tropicalis*. The BLAST search of transcriptome data obtained 1244 single-copy orthologous genes among five species. A phylogenetic tree showed *A. davidianus* to have the closest relationship to *H. chinensis*. Fourteen positively selected genes were detected in *A. davidianus* and *N. viridescens* group and fifteen in *A. davidianus* and *H. chinensis* group. Five genes were shared in the both groups which involved in the immune system, suggesting that *A. davidianus* adaptation to an aquatic and cave environment required rapid evolution of the immune system compared to *N. viridescens* and *H. chinensis*.

Keywords: *Andrias davidianus*, Comparative transcriptome, Evolution analysis, Adaptive genes

Background

Amphibians played an important role as a transitional group linking aquatic to terrestrial in the evolution of vertebrates [1]. To elucidate evolutionary history, the genome and mitochondrial DNA are traditionally used to estimate divergence time [2]. Transcriptome sequencing has become a viable alternative to provide rapid developing genomic resources in non-model organisms [3, 4]. Comparative transcriptome analysis is used to estimate the non-synonymous substitution (K_a) and synonymous substitution (K_s) rates to calculate the evolutionary rate [5, 6] and hence, to identify genes involved in environmental adaptation. Distribution of synonymous substitutions can be used to calculate the divergent time based on the coding sequence [2, 7].

The Chinese giant salamander *Andrias davidianus* is a typical urodele, and an important species both as a biological resource and with respect to its value as a living fossil [8]. The species was historically widespread in China, but environmental degradation and human killing have led to its severe decline in the wild. From 1980s, it is classified as

endangered by the International Union for Conservation of Nature and Nature Resources. Because of its irreplaceable protection status and good taste, artificial propagation technology was studied and succeeded at the end of 1990s. Success of artificial propagation technology provided a value way to protect the wild resources. In wild, it is aquatic in all life stages and typically inhabits rocky crevices in banks of streams and lakes, as well as subterranean rivers. To identify genes possibly related to *A. davidianus* adaptation to its aquatic life history and to a cave habitat, transcriptome data of other amphibian species were obtained from GenBank, and comparative transcriptome analysis was carried out to detect genes positively selected for in evolution.

Methods

RNA extraction and sequencing

Total RNA was extracted from five ovaries and testes using Trizol reagent (Invitrogen, USA) according manufacturer's instructions and treated with RNase-free DNase I (Takara, China) to remove the genomic DNA, respectively. After RNA quality and quantity test, RNA was broken into short fragment, and first-strand cDNA was synthesized, and then the sequencing adapter was added. The cDNA libraries were constructed and sequenced on the Illumina

* Correspondence: hqmu0806@163.com; xhb@yfi.ac.cn

¹Yangtze River Fisheries Research Institute, Chinese Academy of Fishery Sciences, Wuhan, Hubei 430223, China

Full list of author information is available at the end of the article

sequencing platform (Illumina HiSeq™ 2500). All raw reads, low quality sequences, and reads containing adaptor sequences were removed, and the clean reads were obtained.

Identification of orthologues genes and phylogenetic analysis

Two gonad transcriptome data (SRR3308418 and SRR3308420) of *A. davidianus* were provided by my lab. To expand data of *A. davidianus*, transcriptome data of skin (SRX729810) and spleen (SRX729743) were obtained from NCBI database and reassembled with the gonad transcriptome data. Transcriptome data of *N. viridescens* from heart, lens, brain, eye, liver, lung, spleen, kidney, testis, and ovary (ERR108189), *C. pyrrhogaster* lens and neural retina (SRR1051839), *H. chinensis* whole body (SRR1042328) and *X. tropicalis* from genome sequencing (GCA_000004195) were also obtained. The unigenes were reassembled from the downloaded raw reads, except for *X. tropicalis*. The numbers of unigenes for each species is given in Table 1. BLASTN software was used to align sequences, with the cutoff E-value set at $1e^{-7}$ [9]. OrthoMCL software was applied to classify the gene family [10]. Orthologous genes were obtained, and Venn diagrams were used to obtain the gene number [11]. The orthologous genes were used to construct the phylogenetic tree by the NJ method with 1000 bootstrap replications.

Estimate of substitution rates among species

Form the orthologous gene, only one orthologous gene in other species was classed as single-copy orthologous by PERL package [12, 13]. The single-copy orthologous genes were identified to calculate the synonymous substitution rates (Ks) and non-synonymous rate (Ka). The amino acid sequences were aligned by Muscle software [14]. The aligned sequences were converted to corresponding nucleotide sequences. Synonymous substitution rates (Ks) and non-synonymous rates (Ka) were estimated between species pairs by sit model under Codeml program in PAML package [15]. The best threshold was set at 0.5 based on the Ka/Ks value according to previous reports [5, 6]. Value of two fold log-likelihood difference was used to perform a Chi-squared test and the difference of the parameter number was set as the degree in the Chi-squared. Positively selected sites were allowed when P was < 0.05 and posterior probability was > 0.95 [16]. A Ka/Ks

value > 1 indicated strong positive selection, from 0.5 to 1 indicated weak positive selection, and a value < 0.1 indicated negative selection.

Results

Orthologue identification and phylogenetic analysis

To identify the phylogenetic relationship among the species, large-scale transcriptome characterizations were carried out for *N. viridescens*, *X. tropicalis*, *C. pyrrhogaster*, *H. chinensis*, and *A. davidianus*, and transcriptome data were downloaded and reassembled (Table 1). Comparative analysis yielded 4279 gene families and 34,246 putative orthologous genes (Fig. 1). To construct the phylogenetic tree with *X. tropicalis* as out-group, 1244 single-copy orthologous genes were identified. The phylogenetic tree showed *A. davidianus* to have the closest relationship to *H. chinensis*, with *N. viridescens* and *C. pyrrhogaster* clustered on one separate branch (Additional file 1: Figure S1).

Evolutionary profile of *Andrias davidianus* genes

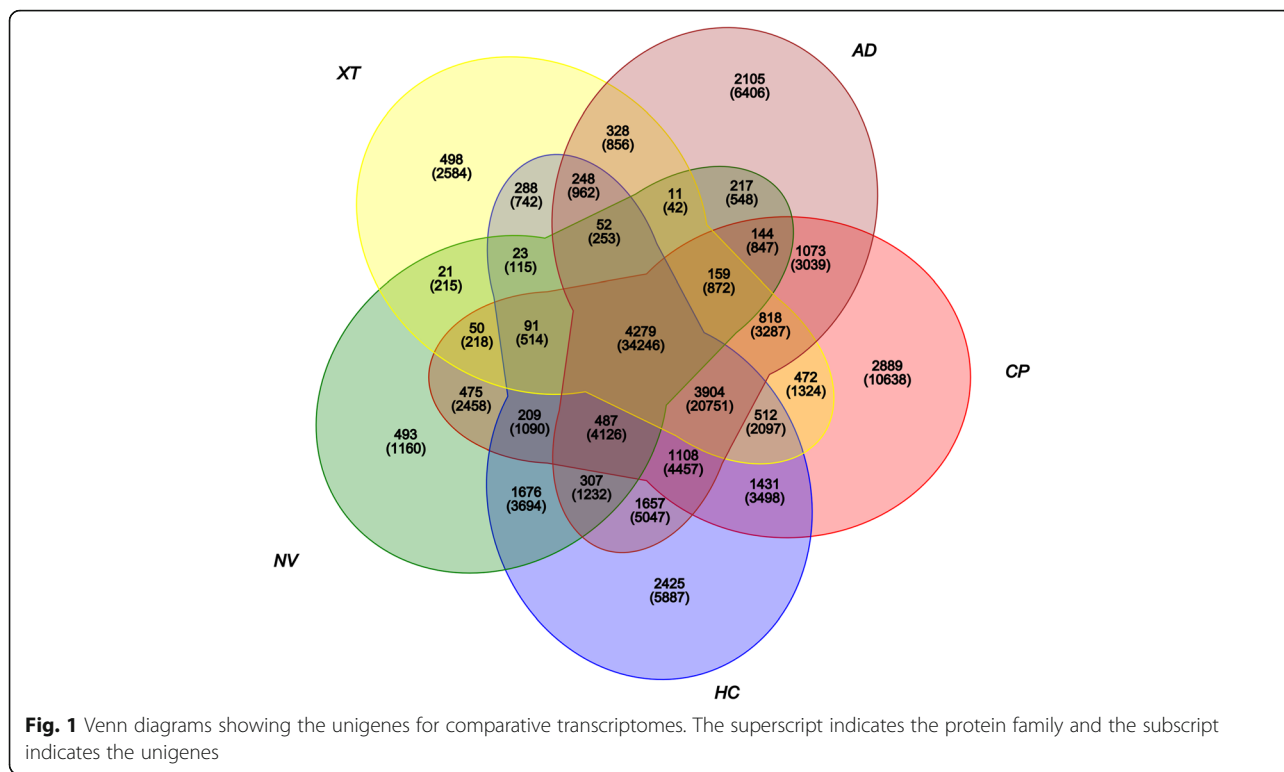
We analyzed the evolutionary pattern of 1244 single-copy orthologous genes in *A. davidianus*, *H. chinensis*, and *N. viridescens*. Synonymous (Ks) and non-synonymous (Ka) substitutions per site were observed (Fig. 2). A majority of sequence pairs showed a $Ka/Ks < 0.5$, implying that these genes involved negative selection. Fifteen rapidly evolving sequences were identified with $Ka/Ks > 0.5$ between *A. davidianus* and *H. chinensis*, and 14 such sites were observed between *A. davidianus* and *N. viridescens* (Additional file 2: Table S1).

Discussion

Next-generation sequencing technology yielded a large number of sequences at the low cost and provides more sequences compared to traditional sequencing methods [17, 18]. Due to the cost and the throughput, genome-wide detection of the adaptive evolution gene was performed in many species by next-generation sequencing [17, 19]. Comparative phylogenetic analysis at the genome level improved the precision of evolutionary inference compared to single gene [20]. However, because of the large genome of the *A. davidianus*, evolutionary analysis by comparative genome was hard to carry out. Transcriptome sequencing was a valuable way to obtain large-scale sequences without reference

Table 1 Results of the assembly for each study species

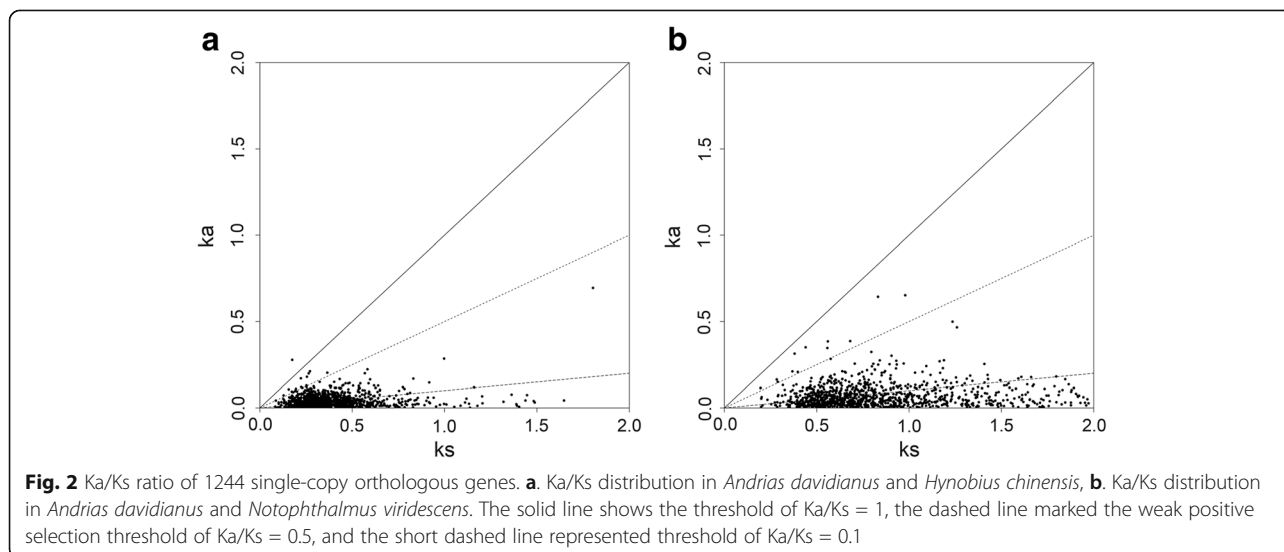
Species	Transcriptome	N50	GC %	Max Length bp	Min Length bp	CDS
<i>Xenopus tropicalis</i> (XT)	22,855	2418	45.21	94,440	114	22,718
<i>Cynops pyrrhogaste</i> (CP)	122,913	1596	44.61	18,379	201	49,986
<i>Notophthalmus viridescens</i> (NV)	31,998	392	44.48	9697	201	17,943
<i>Hynobius chinensis</i> (HC)	103,800	426	47.54	15,293	201	51,362
<i>Andrias davidianus</i> (AD)	85,868	1492	48.61	17,741	201	43,402



genome [21, 22]. Phylogenetic analysis of transcriptome sequence data exhibited high supported tree topologies in many species [23, 24].

To elucidate the phylogenetic evolution of *A. davidianus*, comparative transcriptome analysis was conducted to construct the phylogenetic tree with *X. tropicalis* as out-group. To search adaptive gene for aquatic and cave life, molecular evolution was analyzed among the related species. Synonymous substitution rates (Ks) and non-synonymous substitution rates (Ka) were calculated according to the

phylogenetic tree by PAML software [15, 25], with the optimal threshold for selecting the positively expressed sequence tag (EST) of 0.5 based on previous study [25]. Several positively selected genes were detected. Similar results were found in topmouth culter *Erythroculter ilishaeformis* and zebrafish *Danio rerio*, in which 38 candidate genes exhibited signs of positive selection with dN/dS ratios > 0.5 [6]. Five genes related to the immune system [26–29] [cystatin-like, oncostatin-M-specific receptor subunit beta isoform X1(OSMF), exonuclease, cell death regulator Aven,



and centromere protein H] were detected in the *A. davidianus*/*H. chinensis* and *A. davidianus*/*N. viridescens* groups. *Andrias davidianus* is aquatic and inhabiting subterranean rivers and caves while *N. viridescens* and *H. chinensis* are mainly terrestrial and only special stage in water.

Aquatic and cave dwelling organisms generally encounter more bacteria than do terrestrial animals. Thus, the *A. davidianus* immune system should show more rapid mutations, as was confirmed in our investigation. Due to lack of full-length according to the transcriptome sequencing, many gene relevant to positive selection was omitted and Ka/Ks ratio was decreased from normal level [6]. Further study will be carried out to identify the genes under positive selection.

Additional files

Additional file 1: Figure S1. Phylogenetic tree of selected species based on 1244 single-copy orthologous genes. (TIFF 212 kb)

Additional file 2: Table S1. Orthologs gene under positive selection among species. (DOCX 18 kb)

Abbreviations

EST: Expressed sequence tag; Ka: Non-synonymous substitution rates; Ks: Synonymous substitution rates; OSMF: Oncostatin-M-specific receptor subunit beta isoform X1

Acknowledgments

The authors would like to thank the Linen lab group for comments on data analysis.

Funding

This work was supported by National Nature Science Foundation of China (31502155). This work was supported by National Nonprofit Institute Research Grant (2017JBF0205). Key Laboratory of Freshwater Aquatic Biotechnology and Breeding, Ministry of Agriculture (KF-2017-06).

Availability of data and materials

The variation data reported in the paper was deposited in the GenBank.

Authors' contributions

QH analyze the data and write the manuscript. QW collected the data. YM and HT participated in the data analyzed and revised the manuscript. The study was conceived by HX. All the authors participated during the discussion and approved of its final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Yangtze River Fisheries Research Institute, Chinese Academy of Fishery Sciences, Wuhan, Hubei 430223, China. ²College of Life Science, Yangtze University, Jingzhou 434025, China.

Received: 22 October 2017 Accepted: 29 January 2018

/ Published online: 20 February 2018

References

- Lowcock LA, Green DM, Sessions SK. Amphibian cytogenetics and evolution. *Copeia*. 1991;1992(2):588.
- Zhang P, Chen YQ, Zhou H, Liu YF, Wang XL, Papenfuss TJ, et al. Phylogeny, evolution, and biogeography of Asiatic salamanders (Hynobiidae). *Proc Natl Acad Sci U S A*. 2006;103(19):7360–5.
- Hiremath PJ, Farmer A, Cannon SB, Woodward J, Kudapa H, Tuteja R, et al. Large-scale transcriptome analysis in chickpea (*Cicer arietinum* L.), an orphan legume crop of the semi-arid tropics of Asia and Africa. *Plant Biotechnol J*. 2011;9(8):922–31.
- Fraser BA, Weadick CJ, Janowitz I, Rodd FH, Hughes KA. Sequencing and characterization of the guppy (*Poecilia reticulata*) transcriptome. *BMC Genomics*. 2011;12(1):202.
- Zhang L, Yan HF, Wu W, Yu H, Ge XJ. Comparative transcriptome analysis and marker development of two closely related primrose species (*Primula poissonii* and *Primula wilsonii*). *BMC Genomics*. 2013;14(1):329.
- Ren L, Tan XJ, Xiong YF, Xu K, Zhou Y, Zhong H, et al. Transcriptome analysis reveals positive selection on the divergent between topmouth culter and zebrafish. *Gene*. 2014;552(2):265–71.
- Niu SH, Li ZX, Yuan HW, Chen XY, Li Y, Li W. Transcriptome characterisation of *Pinus Tabuliformis* and evolution of genes in the *Pinus* phylogeny. *BMC Genomics*. 2013;14(1):263.
- Gao KQ, Shubin NH. Earliest known crown-group salamanders. *Nature*. 2003;422(6930):424–8.
- Zhang Z, Schwartz S, Wagner L, Miller W. A greedy algorithm for aligning DNA sequences. *J Comput Biol*. 2000;7(1-2):203–14.
- Li L, Jr SC, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res*. 2003;13:2178–89.
- Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics*. 2011;12(1):35.
- Gabriel V, Francesc R, Gabriel C. A perl package and an alignment tool for phylogenetic networks. *Bmc Bioinformatics*. 2008;9(1):1–5.
- Wu F, Mueller LA, Crouzillat D, Pétiard V, Tanksley SD. Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative, evolutionary and systematic studies: a test case in the euasterid plant clade. *Genetics*. 2006;174(3):1407.
- Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32(5):1792–7.
- Yang Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 2007;24(8):1586–91.
- Yang Z, Bielawski JP. Statistical methods for detecting molecular adaptation. *Trends Ecol Evol*. 2000;15(12):496.
- Tine M, Kuhl H, Gagnaire PA, Louro B, Desmarais E, Martins RS, et al. European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nat Commun*. 2014;5:5770.
- Liu H, Chen CL, Gao ZX, Min JM, Gu YM, Jian JB, et al. The draft genome of blunt snout bream (*Megalobrama amblycephala*) reveals the development of intermuscular bone and adaptation to herbivorous diet. *Gigascience*. 2017;6:1–13.
- Shao CW, Bao BL, Xie ZY, Chen XY, Li B, Jia XD, et al. The genome and transcriptome of Japanese flounder provide insights into flatfish asymmetry. *Nat Genet*. 2017;49(1):119–24.
- Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, et al. Evolution of genes and genomes on the drosophila phylogeny. *Nature*. 2007;450(7167):203–18.
- Che RB, Sun YN, Wang RX, Xu TJ. Transcriptomic analysis of endangered Chinese salamander: identification of immune, sex and reproduction-related genes and genetic markers. *PLoS One*. 2014;9:e87940.
- Cai JF, Yang W, Chen D, Zhang YZ, He Z, Zhang WM, et al. Transcriptomic analysis of the differentiating ovary of the protogynous ricefield eel *Monopterus albus*. *BMC Genomics*. 2017;18(1):573.
- Roeding F, Borner J, Kube M, Klages S, Reinhardt R, Burmester T. A 454 sequencing approach for large scale phylogenomic analysis of the common emperor scorpion (*Pandinus imperator*). *Mol Phylogenet Evol*. 2009;53(3):826–34.
- Meusemann K, von Reumont BM, Simon S, Roeding F, Strauss S, Kuck P, et al. A phylogenomic approach to resolve the arthropod tree of life. *Mol Biol Evol*. 2010;27(11):2451–64.

25. Swanson WJ, Wong A, Wolfner MF, Aquadro CF. Evolutionary expressed sequence tag analysis of drosophila female reproductive tracts identifies genes subjected to positive selection. *Genetics*. 2004;168(3):1457–65.
26. Li FM, Gai XM, Wang LL, Song LS, Zhang H, Qiu LM, et al. Identification and characterization of a cystatin gene from Chinese mitten crab *Eriocheir sinensis*. *Fish Shellfish Immunol*. 2010;29(3):521–9.
27. Kutuk O, Temel SG, Tolunay S, Basaga H. Aven blocks DNA damage-induced apoptosis by stabilising Bcl-xL. *Eur J Cancer*. 2010;46(13):2494–505.
28. He WL, Li YH, Yang DJ, Song W, Chen XL, Liu FK, et al. Combined evaluation of centromere protein H and Ki-67 as prognostic biomarker for patients with gastric carcinoma. *Eur J Surg Oncol*. 2013;39(2):141–9.
29. Larrea E, Echeverria I, Riezu-Boj JI, Aldabe R, Guembe L, Sola I, et al. Characterization of the CD40L/Oncostatin M/Oncostatin M receptor axis as an antiviral and immunostimulatory system disrupted in chronic HCV infection. *J Hepatol*. 2014;60(3):482–9.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

