Data in Brief 19 (2018) 501–505

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Transcriptome of *Xenopus andrei*, an octoploid frog, during embryonic development

Mark E. Pownall, Ronald R. Cutler, Margaret S. Saha *

College of William & Mary, Biology Department, Williamsburg, VA 23185, United States

ARTICLE INFO

Article history:
Received 16 March 2018
Accepted 4 May 2018
Available online 16 May 2018

Keywords:
Development
Octoploid
Polyploidy
Transcriptome
RNA-Seq
Xenopus

ABSTRACT

Although polyploidy occurs throughout the fish and amphibian lineages, the *Xenopus* genus exhibits a high incidence of polyploidy, with 25 out of the 26 known species being polyploid. However, transcriptomic information is currently available for only one of these species, the tetraploid *Xenopus laevis*. *Xenopus andrei*, an octoploid species within the *Xenopus* genus, offers an opportunity for assessing a novel polyploid transcriptome during vertebrate development. RNA-Seq data was generated at nine different developmental stages ranging from unfertilized eggs through swimming tadpole stages and raw FASTQ files were deposited in the NCBI SRA database (accession number SRP134281). Additionally, RNA-seq data from all nine stages were pooled to create a *de novo* assembly of the transcriptome using Trinity and has been deposited in the NCBI GEO database (accession number GSE111639). To our knowledge, this represents the first published assembly of an octoploid vertebrate transcriptome. In total, 849 Mb were assembled, which led to the identification of 1,650,048 transcripts in the assembly with a contig N50 of 630 bases. This RNA-Seq and transcriptome data will be valuable for comparing polyploid transcriptomes across *Xenopus* species, as well as understanding evolutionary implications of whole-genome duplication and polyploidy in vertebrates.

* Correspondence to: College of William & Mary, Biology Department, Integrated Science Center, 540 Landrum Dr., Williamsburg, VA 23185, United States.
E-mail address: mssaha@wm.edu (M.S. Saha).

https://doi.org/10.1016/j.dib.2018.05.017

Specifications Table

Subject area	Biology
More specific subject area	Developmental Biology and Transcriptomics
Type of data	Transcriptome assembly, raw sequences
How data was acquired	High-throughput sequencing carried out by Oklahoma Medical Research Facility (OMRF) Clinical Genomics Center using an Illumina HiSeq. 3000.
Data format	Analyzed, raw
Experimental factors	One pair of adult frogs was mated to produce embryos sampled in groups of five individual embryos at nine different developmental time points.
Experimental features	Total RNA was extracted from samples of five pooled embryos. Library preparation and sequencing was performed by OMRF. Analysis was carried out in-house.
Data source location	Adult *X. andrei* were purchased from Xenopus Express Inc. (Brooksville, FL, USA).
Data accessibility	Raw FASTQ files were deposited on NCBI SRA database with accession number SRP134281. Transcriptome fasta file was deposited on NCBI GEO database with accession number GSE111639. All data can be accessed at the following link https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE111639.

Value of the data

- To our knowledge, this is the first published *de novo* transcriptome assembly of an octoploid vertebrate species.
- *Xenopus andrei* presents a novel model system for studying transcriptional regulation of an octoploid genome during vertebrate development.
- Comparisons between *X. tropicalis*, *X. laevis*, and *X. andrei* transcriptomes will provide insight into how multiple genome duplications have affected gene expression.

1. Data

Although comparatively less prevalent than in plants, polyploidy in animals occurs throughout the amphibian and fish lineages as well as in mammalian tissues including heart, placenta, and in pathogenic conditions such as wound healing and cancer [1,2]. The publication of the completed tetraploid *Xenopus laevis* genome [3] as well as associated transcriptome data has presented the possibility of studying comparative gene expression in polyploid vertebrate animals. *Xenopus andrei* is an octoploid ($2n = 72$) species [4] and provides the unique opportunity to compare transcriptomes among a range of ploidy levels in related frogs. We generated RNA-seq data at nine developmental time points and produced a draft assembly of the transcriptome. The data presented here can be accessed at SRA: SRP134281 for raw FASTQ files of all sequencing, and GEO: GSE111639 for assembly of the transcriptome.

Using Trinity [5], a *de novo* assembly of the transcriptome was produced by combining all of the sequence data generated from the nine samples in this experiment. Statistics of the assembly are summarized in Table 1.

A filtered transcriptome assembly was also generated using a minimum transcript per million (TPM) threshold of 2 to filter out lowly expressed transcripts in the initial *de novo* assembly. This is

Table 1
Statistics of *X. andrei* transcriptome assembly.

Number of reads used in assembly	341,149,792
Number assembled bases	849,005,380
Number of assembled genes	1,023,069
Number of assembled transcripts	1,650,048
Assembly GC percent	41.15
Contig N50	631
Contig Ex90N50	1215
Average contig length	514.53

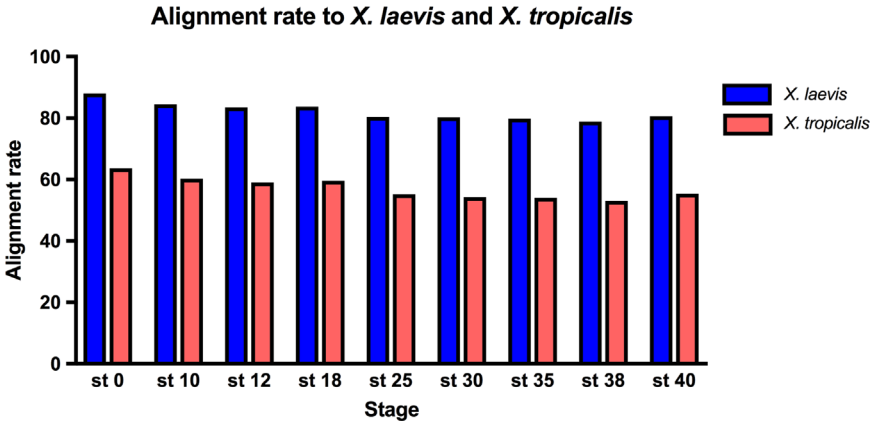

Fig. 1. Overall alignment rate to *X. laevis* and *X. tropicalis*. *X. andrei*. RNA-Seq reads at each of the indicated stages were aligned to *X. laevis* and *X. tropicalis* reference genomes.

because many assembled transcripts lacked read support, meaning that paired-end reads did not concordantly align to the transcriptome. This has been observed by others [6,7] when using Trinity to assemble polyploid transcriptomes. The filtered assembly consists of 149,471 transcripts and 100,936 genes, representing 9.06% and 9.86% of the transcripts and genes in the initial *de novo* assembly, respectively. Data including TPM abundance measures, genes from the transcriptome assembly, and the filtered transcriptome assembly are available as supplementary files. While the tetraploid *Xenopus laevis* genome has approximately 46,000 genes [3], our filtered assembly contains approximately twice as many genes in octoploid *Xenopus andrei*.

To provide additional comparative data, *X. laevis* v9.2 and *X. tropicalis* v9.1 reference genomes were downloaded from Xenbase.org [8]. At each developmental stage, RNA-seq reads were aligned to both reference genomes. *X. andrei* reads align to the *X. laevis* reference on average at 82% and to the *X. tropicalis* reference on average at 57% (Fig. 1).

2. Experimental design, materials and methods

2.1. Animal usage

All animal care and use was approved by the College of William and Mary Institutional Animal Care and Use Committee (IACUC). Mating was induced by subcutaneous injection of human chorionic gonadotropin (hCG) as described by Sive et al. [9], with the female receiving 150 units and the male 100 units. Embryos were collected and dejelled in basic 2% cysteine for five minutes. After dejelling, embryos were washed three times in 0.1 × Marc's Modified Ringers (MMR) with 50 μg/ml gentamicin. Embryos were then reared in 0.1 × MMR + gentamicin at room temperature (22 °C) until the

desired stage was reached. All staging was based on equivalent Nieuwkoop and Faber [10] *Xenopus laevis* stages.

2.2. Sample preparation and RNA extraction

At stages 0 (unfertilized egg), 10 (early gastrula), 12 (gastrula), 18 (neurula), 25 (tailbud), 30 (hatching), 35 (swimming tadpole), 38 (swimming tadpole), and 40 (swimming tadpole), groups of five embryos were flash frozen in liquid nitrogen and stored at − 80 °C. RNA was extracted by homogenizing embryos in 600 μl TRI reagent (Ambion) then dividing the homogenate equally between two tubes. Extraction was then carried out using 60 μl of chloroform. This partially separated mixture was spun down in a Phase Lock Gel Heavy tube (QuantaBio) and RNA was then purified using the MagMAX™-96 for Microarrays (Thermo Fisher Scientific) kit according to the manufacturer's instructions. RNA quality and yield was assessed by spectrophotometry (NanoDrop®) and agarose gel electrophoresis.

2.3. RNA-seq

Total RNA samples were shipped to OMRF where libraries were generated using the Illumina TruSeq Stranded rRNA depletion Library Prep kit. Library quality was confirmed with Kapa qPCR and Agilent TapeStation prior to paired-end sequencing with an Illumina HiSeq 3000.

2.4. Transcriptome analysis and de novo assembly

Raw reads were first analyzed by FastQC [11] for per base sequence quality. Reads were then aligned using Hisat2 with relaxed parameters [12] to the *Xenopus laevis* reference genome, and alignment files were used to infer read strandedness using RSeQC (Version 2.6.4) [13]. Trimmomatic (Version 0.36) [14] was used to trim Illumina Truseq paired-end adaptors and bases with phred quality scores less than 5. *De novo* transcriptome assembly was performed using Trinity (Version 2.4.0) [5] with all reads from the nine samples run on a high performance computing cluster with settings to normalize read coverage to 50 × and minimum of 2 k-mers to be assembled by the Inchworm module. Raw reads were then mapped back to the *de novo* assembled transcriptome using Bowtie2 (Version 2.3.4.1) [15] with default parameters to assess assembly quality and RNA-Seq read representation (Supplementary Fig. 1). Transcript isoforms were collapsed into 'genes' to construct a genome-like reference using the Trinity SuperTranscripts module. Transcript abundance quantification was performed with Salmon (Version 0.8.2) [16] and transcripts were then filtered using a minimum TPM threshold of 2 using the Trinity transcript filtering module. Detailed commands, scripts, and parameters used in this pipeline are available in the supplementary files.

Acknowledgements

This work was funded by the National Institutes of Health (1R15HD077624-01) to MSS and the National Science Foundation (IOS-1257895) to MSS. The authors declare no competing interests.

Transparency document. Supplementary material

Transparency data associated with this article can be found in the online version at https://doi.org/10.1016/j.dib.2018.05.017.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.dib.2018.05.017.

References

[1] K.P. Schoenfelder, D.T. Fox, The expanding implications of polyploidy, J. Cell Biol. 209 (2015) 485–491. http://dx.doi.org/10.1083/jcb.201502016.

[2] T.L. Orr-Weaver, When bigger is better: the role of polyploidy in organogenesis, Trends Genet. 31 (2015) 307–315. http://dx.doi.org/10.1016/j.tig.2015.03.011.

[3] A.M. Session, Y. Uno, T. Kwon, J.A. Chapman, A. Toyoda, S. Takahashi, A. Fukui, A. Hikosaka, A. Suzuki, M. Kondo, S.J. van Heeringen, I. Quigley, S. Heinz, H. Ogino, H. Ochi, U. Hellsten, J.B. Lyons, O. Simakov, N. Putnam, J. Stites, Y. Kuroki, T. Tanaka, T. Michiue, M. Watanabe, O. Bogdanovic, R. Lister, G. Georgiou, S.S. Paranjpe, I. van Kruijsbergen, S. Shu, J. Carlson, T. Kinoshita, Y. Ohta, S. Mawaribuchi, J. Jenkins, J. Grimwood, J. Schmutz, T. Mitros, S.V. Mozaffari, Y. Suzuki, Y. Haramoto, T.S. Yamamoto, C. Takagi, R. Heald, K. Miller, C. Haudenschild, J. Kitzman, T. Nakayama, Y. Izutsu, J. Robert, J. Fortriede, K. Burns, V. Lotay, K. Karimi, Y. Yasuoka, D.S. Dichmann, M.F. Flajnik, D.W. Houston, J. Shendure, L. DuPasquier, P.D. Vize, A.M. Zorn, M. Ito, E.M. Marcotte, J.B. Wallingford, Y. Ito, M. Asashima, N. Ueno, Y. Matsuda, G.J.C. Veenstra, A. Fujiyama, R.M. Harland, M. Taira, D.S. Rokhsar, Genome evolution in the allotetraploid frog Xenopus laevis, Nature 538 (2016) 336–343. http://dx.doi.org/10.1038/nature19840.

[4] B.J. Evans, T.F. Carter, E. Greenbaum, V. Gvodík, D.B. Kelley, P.J. McLaughlin, O.S.G. Pauwels, D.M. Portik, E.L. Stanley, R. C. Tinsley, M.L. Tobias, D.C. Blackburn, Genetics, morphology, advertisement calls, and historical records distinguish six new polyploid species of African clawed frog (Xenopus, Pipidae) from west and central Africa, PLoS One 10 (2015) 1–51. http://dx.doi.org/10.1371/journal.pone.0142823.

[5] M.G. Grabherr, B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B.W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, Full-length transcriptome assembly from RNA-Seq data without a reference genome, Nat. Biotechnol. 29 (2011) 644–652. http://dx.doi.org/10.1038/nbt.1883.

[6] R. Chopra, G. Burow, A. Farmer, J. Mudge, C.E. Simpson, M.D. Burow, Comparisons of de novo transcriptome assemblers in diploid and polyploid species using peanut (Arachis spp.) RNA-Seq data, PLoS One 9 (2014) 1–16. http://dx.doi.org/10.1371/journal.pone.0115055.

[7] J.F. Sánchez-Sevilla, J.G. Vallarino, S. Osorio, A. Bombarely, D. Posé, C. Merchante, M.A. Botella, I. Amaya, V. Valpuesta, Gene expression atlas of fruit ripening and transcriptome assembly from RNA-seq data in octoploid strawberry (Fragaria × ananassa), Sci. Rep. 7 (2017) 1–13. http://dx.doi.org/10.1038/s41598-017-14239-6.

[8] K. Karimi, J.D. Fortriede, V.S. Lotay, K.A. Burns, D.Z. Wang, M.E. Fisher, T.J. Pells, C. James-Zorn, Y. Wang, V.G. Ponferrada, S. Chu, P. Chaturvedi, A.M. Zorn, P.D. Vize, Xenbase: a genomic, epigenomic and transcriptomic model organism database, Nucleic Acids Res. 46 (2018) D861–D868. http://dx.doi.org/10.1093/nar/gkx936.

[9] H. Sive, R.M. Grainger, R.M. Harland, Early Development of *Xenopus Laevis*: A Laboratory Manual, first ed., Cold Spring Harbor Laboratory Press, New York, 2000.

[10] P.D. Nieuwkoop, J. Faber, Normal Table of *Xenopus laevis* (Daudin), Garland Publishing Inc., New York, 1994.

[11] S. Andrews, FastQC: a quality control tool for high throughput sequence data, 2010. ⟨Http://Www.Bioinformatics.Babraham.Ac.Uk/Projects/Fastqc/⟩. ⟨http://dx.doi.org/citeulike-article-id:11583827⟩.

[12] D. Kim, B. Langmead, S.L. Salzberg, HISAT: a fast spliced aligner with low memory requirements, Nat. Methods 12 (2015) 357–360. http://dx.doi.org/10.1038/nmeth.3317.

[13] L. Wang, S. Wang, W. Li, RSeQC: quality control of RNA-seq experiments, Bioinformatics 28 (2012) 2184–2185. http://dx.doi.org/10.1093/bioinformatics/bts356.

[14] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics 30 (2014) 2114–2120. http://dx.doi.org/10.1093/bioinformatics/btu170.

[15] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, Nat. Methods 9 (4) (2012) 357–359. http://dx.doi.org/10.1038/nmeth.1923.

[16] R. Patro, G. Duggal, M.I. Love, R.A. Irizarry, C. Kingsford, Salmon provides fast and bias-aware quantification of transcript expression, Nat. Methods 14 (2017) 417–419. http://dx.doi.org/10.1038/nmeth.4197.