Data in Brief 000 (2020) 106126

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Transcriptome dataset for RNA-seq analysis of axolotl embryonic oropharyngeal endoderm explants

Lauren Marazzi[a], Priya Kohli[b], Deborah Eastman[a,*]

[a] Department of Biology, Connecticut College, United States
[b] Department of Math and Statistics, Connecticut College, United States

ARTICLE INFO

Article history:
Received 23 July 2020
Revised 29 July 2020
Accepted 29 July 2020
Available online 04 August 2020

Keywords:
Ambystoma mexicanum
Axolotl
deNovo Assembly
RNAseq
Transcriptome

ABSTRACT

Animal nutrition and toxin deterrence rely on the ability to taste, which occurs through columnar taste cells clustered within taste buds. Taste buds in mammals are located within specialized tissues, called papillae. However, taste buds in fish and amphibians, such as axolotls (*Ambystoma mexicanum*), are not housed in papillae, rather they are embedded within the pharyngeal epithelium. This simplified tissue level organization, along with the ability of cultured oropharyngeal explants from early embryos to produce taste buds on the same time-line as embryos, make the axolotl an excellent model to identify molecules specifically involved in taste bud cell differentiation. We performed de novo transcriptomic analysis on RNA sequences from three different stages of oropharyngeal explants: stages 37/38, 39, and 41. RNA-seq data from 17 total samples representing these stages were pooled to generate a de novo assembly of the transcriptome using a Trinity pipeline. From 27.9Gb of raw sequences, we identified 21,244 transcripts. To our knowledge, this is the first published assembly of axolotl oropharyngeal endoderm explants. This data and transcriptome assembly relate to the research article "Transcriptome Analysis of Axolotl Oropharyngeal Explants During Taste Bud Differentiation Stages" (Kohli et al. 2020). This RNA-seq data and transcriptome assembly provide information on genes expressed

DOI of original article: 10.1016/j.mod.2020.103597
* Corresponding author.
E-mail address: daeas@conncoll.edu (D. Eastman).

https://doi.org/10.1016/j.dib.2020.106126

in the oropharyngeal endoderm and will be valuable in the identification of taste bud development genes.

Specifications Table

Subject	Biology
Specific subject area	Amphibian transcriptome
Type of data	Transcriptomic data, table
How data were acquired	Single-end read RNAseq of RNA isolated from axolotl embryonic oropharyngeal endoderm explants. High-throughput sequencing performed with Illumina NextSeq 500 System.
Data format	Raw FASTQ and processed FASTA sequence files, including assembled transcriptome FASTA files Analyzed Filtered
Parameters for data collection	Oropharyngeal endoderm explants from stage 15 axolotl embryos that were cultured to stage 37/38, stage 39 or stage 41 were considered for data collection.
Description of data collection	All axolotl embryos used were from the Ambystoma Genetic Stock Center at the University of Kentucky. All experiments with these embryos was in compliance with the National Institutes of Health Guidelines for the care and use of animals in research. Descriptions in Bordzilovskaya et al. [2] was used to determine embryonic stages. Isolation and culturing of oropharyngeal endodermal explants (OPEs) from stage 15 embryos was performed as described by Parker et al. [3] These cultured explants were incubated alongside control embryos from the same clutch to confirm staging. Explants from stages 37/38, 39, and 41 were harvested for further analysis. OPE replicates from stages 37/38, 39, and 41 were stored in RNALater until further processed. Samples were homogenized with a Dounce homogenizer in RLT buffer containing ß-mercaptoethanol (Qiagen). The homogenate was then processed using an RNAEasy kit (Qiagen, Hilden, Germany). Replicate samples for each stage were analyzed for RNA quantity and quality on a Nanodrop 8000 spectrophotometer. Seventeen high quality RNA samples were used to prepare libraries. Single-end sequencing was then performed on the Illumina platform. De novo transcriptome assembly was performed on trimmed sequencing reads using the Trinity Pipeline [4] and annotation of the assembly was done using EnTAP (Eukaryotic Non-Model Transcriptome Annotation Pipeline) [5].
7	Institution: Connecticut College City/Town/Region:New London, CT Country:USA
Data accessibility	Repository name: NCBI GEO Data identification number: GSE15153 Direct URL to data: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE151535
Related research article	Kohli P, Marazzi L, Eastman D. Transcriptome analysis of axolotl oropharyngeal explants during taste bud differentiation stages. *Mech Dev*. 2020;161:103597. doi:10.1016/j.mod.2020.103597.

Value of the data

- These data are important as they provide the first molecular-level data on axolotl oropharyngeal explants.
- This data is a resource for other researchers examining taste bud differentiation and axolotl development. They will be able to identify and study genes expressed in these tissues.
- This data provides candidate molecules that may be involved in taste bud development and sequence information can be used to design probes for experiments such as in situ hybridizations, RTqPCR and CRISPR.

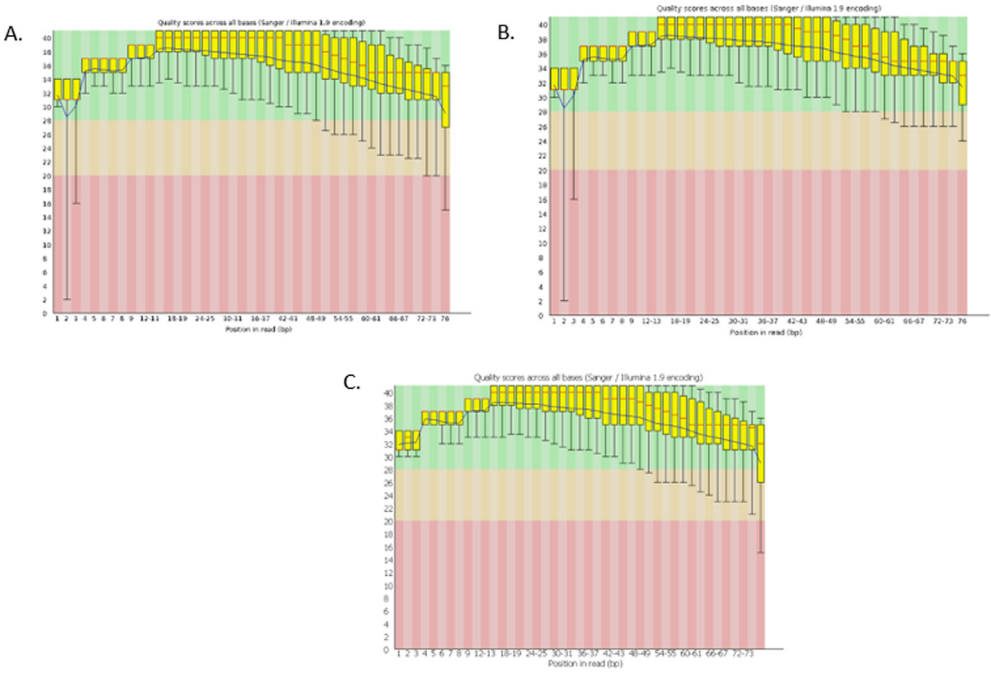

Fig. 1. Comparison of FASTQC Per base sequence quality for a PHRED quality cutoff of (A) 5, (B) 10, (C) 15. PHRED 15 cutoff scores displayed higher quality scores and was thus chosen for further assembly.

1. Data description

We performed de novo transcriptomic analysis on RNA sequences from three different stages (37/38, 39, and 41) of oropharyngeal explants. RNA-seq data from 17 total samples representing these stages were pooled to generate de novo assembly of the transcriptome using a Trinity pipeline.

The raw reads are available at NCBI Biorepository: GEO accession ID GSE15153. Supplementary Files 1 contains FASTQC [6] output reports of quality trimming cutoffs for the of PHRED score 5, PHRED score 10, and PHRED score 15 raw data. Supplementary File 2 is the assembled transcriptome of 21244 transcripts produced by the Trinity de novo transcriptome assembly pipeline [4]. Supplementary File3 contains the gene count matrix of normalized produced by RSEM [7]and bowtie-2 [8] across the studied samples. Supplementary File 4 includes assembly annotation the Eukaryotic Non-Model Transcriptome Annotation Pipeline (EnTAP) [5], using the DIAMOND sequence aligner [9] against the NCBI non-redundant protein database, and Swissprot database [10,11]. Gene Ontology (GO) and KEGG pathway mapping of the transcriptome was performed using EnTAP with the eggnogg-mapper [12]. Supplementary File 5 describes the assembly annotation against the axolotl genome using NCBI blastx. Fig. 1 displays a comparison of the quality scores of all bases for quality score trimming cutoffs of PHRED5, PHRED10, and PHRED15. Fig. 2 shows the top 20 GO terms and KEGG terms represented in the transcriptome assembly.

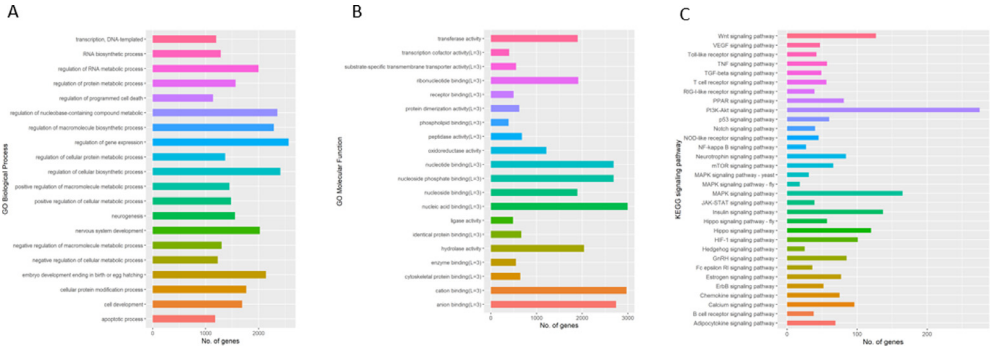

Fig. 2. Top 20 (A) GO Biological Process (level 5), (B) Molecular Function (level 3) terms in the transcriptome assembly. (C) KEGG signaling pathways represented in the assembly.

2. Experimental design, materials, and methods

2.1. Experimental design

Axolotl embryos were from the Ambystoma Genetic Stock Center at the University of Kentucky. National Institutes of Health Guidelines for the care and use of research animals were used in all experiments involving axolotl embryos. Embryos were incubated in 20% Holtfreter's solution at 22°C and staged according to Bordzilovskaya et al. [1]. Isolation and culturing of oropharyngeal endodermal explants (OPEs) from stage 15 embryos was performed as described by Parker et al. [2]. Specifically, jelly coats were removed from stage 15 embryos, which were then treated with 0.1% formalin for 10 minutes. Using sterile technique, embryos were pinned, ventral side up, to silicone coated dishes. Tungsten needles were used to remove the epidermis and the lateral plate mesoderm surrounding the oropharyngeal endoderm. The OPE was then lifted out and washed in 65% L15 media (Sigma). OPEs were incubated in individual agar-coated wells containing 65% L15 media. Whole control embryos were developed in 20% Holtfreter's solution alongside the explants for staging purposes.

OPE replicates from each stage (37/38, 39, 41) were stored in RNALater. Samples were homogenized with a Dounce homogenizer in RLT buffer containing ß-mercaptoethanol (Qiagen). The homogenate was then processed using an RNAEasy kit (Qiagen, Hilden, Germany). Replicate samples for each stage were analyzed for RNA quantity and quality on a Nanodrop 8000 spectrophotometer.

RNA-seq libraries from high quality RNA samples were prepared as described by Hunt [13]. Specifically, polyA RNA was enriched using oligo(dT) beads (New England Biolabs) and was then fragmented by incubating at 95°C for 2 minutes. cDNA synthesized using SMARTSCRIBE (TAKARA) was then PCR amplified using Phire DNA polymerase (Thermofisher). Single-end sequencing was performed on the Illumina platform (NextSeq 500 System) at the Genomics Core Laboratory at the University of Kentucky.

2.2. de novo assembly

Raw RNA-Seq reads across samples were demultiplexed and processed using Trimmomatic [14] for quality trimming. To trim the single end reads data, the first five bases at the 5' end, the last five bases at the 3' end, and reads with less than 25 base pairs were removed. While it is recommended under *de novo* settings to trim values with a quality score below a PHRED 5 value [15], we used FASTQC [6] to compare reads trimmed with PHRED 5, PHRED 10, and PHRED 15 cutoff scores (supplemental file 1). The PHRED15 score cutoff reads passed more FASTQC

quality thresholds than PHRED5 and PHRED10 (Fig. 1), and thus was selected for transcriptome assembly. As described by the Trinity *de novo* assembly pipeline [4], all 17 samples were pooled for transcriptome assembly.

De novo transcriptome assembly was performed using Trinity (version 2.4.0). The minimum assembled contig left was set to 200 bps, and *in silico* read normalization was set to 50 due to read sequencing depth (Supplemental File2). After contig clustering, de Brujin graph construction, and full length transcript reporting, Bowtie2 was used to map raw reads back to the *de novo* assembly and RSEM was implemented to estimate transcript abundance in the assembly across the samples the using the "align_and_estimate_abundance.pl" command in Trinity [8]. Gene isoform counts were collapsed to trinity gene counts for quantification across the samples using Trinity's "abundance_estimates_to_matrix.pl" command (Supplemental File3).

3. Assembly annotation

Annotation of the assembly was done using EnTAP (Eukaryotic Non-Model Transcriptome Annotation Pipeline) using human reference databases [5]. TransDecoder in EnTAP was used to identify long open reading frames within the transcriptome. Next, a similarity search was performed using DIAMOND sequence aligner [9] against NCBI non-redundant protein and Swissport databases for human proteins with expect values (E) set at a maximum of 0.001 (Supplemental File 4). Contaminants were filtered by EnTAP to retain only high-quality transcripts for annotation. Within EnTAP, eggNOG mapper was used for protein domain mapping, Gene Ontology (GO) and KEGG pathway mapping of the transcriptome [12]. Overall, 12,135 transcripts were annotated with gene families, and 10,745 genes had at least one gene ontology (GO) assignment (Fig. 2(A) and (B)). 5698 genes are found in KEGG pathways and 32 KEGG signaling pathways are represented in the transcriptome (Fig. 2(C)). Furthermore, we annotated the transcriptome assembly against the axolotl genome [GCA_002915645.2] [16] using NCBI BLASTx with an E value of 0.001. We found 2119 genes aligned to the axolotl genome (Supplemental File 5).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have, or could be perceived to have, influenced the work reported in this article.

Ethics statement

Animals in all experiments involving axolotl embryos were cared for and used according to the National Institutes of Health Guidelines for the care and use of Laboratory Animals (NIH Publications No. 8023, revised 1978).

Acknowledgments

We acknowledge design and technical support from Arthur Hunt and Marc Peterson. This work was funded by a Research Matters award from Connecticut College and an NSF RCNUBE Award to Juniata College, 1248096.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.dib.2020.106126.

References

[1] P. Kohli, L. Marazzi, D. Eastman, Transcriptome analysis of axolotl oropharyngeal explants during taste bud differentiation stages, Mech. Dev. 161 (2020) 103597.
[2] N.P. Bordzilovskaya, T.A. Dettlaff, S.T. Duhon, G.M. Malacinski, in: Developmental-Stage Series of Axolotl Embryos, Oxford University Press, New York, 1989, pp. 201–219.
[3] M.A. Parker, M.L. Bell, L.A. Barlow, Cell contact-dependent mechanisms specify taste bud pattern during a critical period early in embryonic development, Dev. Dyn. 230 (2004) 630–642.
[4] B.J. Haas, A. Papanicolaou, M. Yassour, M. Grabherr, P.D. Blood, J. Bowden, M.B. Couger, D. Eccles, B. Li, M. Lieber, et al., De novo transcript sequence reconstruction from RNA-Seq using the trinity platform for reference generation and analysis, Nat. Protoc. 8 (2013) 1494–1512.
[5] A.J. Hart, S. Ginzburg, M.S. Xu, C.R. Fisher, N. Rahmatpour, J.B. Mitton, R. Paul, J.L. Wegrzyn, EnTAP: bringing faster and smarter functional annotation to non-model eukaryotic transcriptomes, Mol. Ecol. Resour. (2019).
[6] Andrews, S. FastQC: a quality control tool for high throughput sequence data. 2010.
[7] B. Li, C.N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome, BMC Bioinform. 12 (2011) 323.
[8] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with bowtie 2, Nat. Methods 9 (2012) 357–359.
[9] B. Buchfink, C. Xie, D.H. Huson, Fast and sensitive protein alignment using DIAMOND, Nat. Methods 12 (2015) 59–60.
[10] Uniprot Consortium, the Uniprot Database (2002–2019) .
[11] The UniProt Consortium, UniProt: the universal protein knowledgebase, Nucl. Acids Res. 45 (2017) D158–D169.
[12] J. Huerta-Cepas, D. Szklarczyk, D. Heller, A. Hernandez-Plaza, S.K. Forslund, H. Cook, D.R. Mende, I. Letunic, T. Rattei, L.J. Jensen, et al., eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses, Nucl. Acids Res. 47 (2019) D309–D314.
[13] A.G. Hunt, A rapid, simple, and inexpensive method for the preparation of strand-specific RNA-Seq libraries, in: A. Hunt, Q. Li (Eds.), Polyadenylation in Plants. Methods in Molecular Biology, 1255, Humana Press, New York, NY, 2015, pp. 195–207.
[14] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for illumina sequence data, Bioinformatics 30 (2014) 2114–2120.
[15] M.D. Macmanes, On the optimal trimming of high-throughput mRNA sequence data, Front. Genet. 5 (2014) 13.
[16] J.J. Smith, N. Timoshevskaya, V.A. Timoshevskiy, M.C. Keinath, D. Hardy, S.R. Voss, A chromosome-scale assembly of the axolotl genome, Genome Res. 29 (2019) 317–324.