



ELSEVIER

Contents lists available at ScienceDirect

Data in Brief

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



Data Article

Genome-scale embryonic developmental profile of gene expression in the common house spider *Parasteatoda tepidariorum*



Sawa Iwasaki-Yokozawa^a, Yasuko Akiyama-Oda^{a,b},
Hiroki Oda^{a,c,*}

^a Laboratory of Evolutionary Cell and Developmental Biology, JT Biohistory Research Hall, 1-1 Murasaki-cho, Takatsuki, Osaka 569-1125, Japan

^b Microbiology and Infection Control, Osaka Medical College, Takatsuki, Osaka, Japan

^c Laboratory of Biohistory, Department of Biological Sciences, Graduate School of Science, Osaka University, Japan

ARTICLE INFO

Article history:

Received 19 April 2018

Received in revised form

16 May 2018

Accepted 18 May 2018

Available online 24 May 2018

ABSTRACT

We performed RNA sequencing (RNA-Seq) at ten successive developmental stages in embryos of the common house spider *Parasteatoda tepidariorum*. Two independent datasets from two pairs of parents represent the normalized coverage of mapped RNA-Seq reads along scaffolds of the *P. tepidariorum* genome assembly. Transcript abundance was calculated against existing AUGUSTUS gene models. The datasets have been deposited in the Gene Expression Omnibus (GEO) Database at the National Center for Biotechnology Information (NCBI) under the accession number GSE112712.

© 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>).

* Corresponding author at: Laboratory of Evolutionary Cell and Developmental Biology, JT Biohistory Research Hall, 1-1 Murasaki-cho, Takatsuki, Osaka 569-1125, Japan.

E-mail address: hoda@brh.co.jp (H. Oda).

<https://doi.org/10.1016/j.dib.2018.05.106>

2352-3409/© 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Specifications Table

Subject area	Biology
More specific subject area	Developmental Biology
Type of data	wiggle format, tab-delimited text
How data was acquired	RNA-Seq by Illumina MiSeq
Data format	Processed
Experimental factors	Successive developmental stages, two replicates
Experimental features	mRNA from whole embryos at each stage was sequenced and the sequence reads were processed.
Data source location	Osaka, Japan
Data accessibility	The data have been deposited in the Gene Expression Omnibus database at NCBI under the accession number GSE112712. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE112712

Value of the data

- The datasets are useful for identifying and characterizing transcripts from the regions of interest in the *P. tepidariorum* genome.
- The datasets are useful for examining developmental changes in expression of a gene of interest.
- The datasets may be used to characterize genes based on their expression levels and profiles.
- The datasets can be visualized in a genome browser where the *P. tepidariorum* genome assembly is embedded.

1. Data

The genome of *Parasteatoda tepidariorum* has been sequenced and annotated using gene models [1]. Embryonic development of *P. tepidariorum* is divided into more than 10 stages [2–4]. The datasets described here represent the normalized coverage of mapped RNA sequencing (RNA-Seq) reads along scaffolds of the *P. tepidariorum* genome assembly for ten successive developmental stages (stage (st) 1, st2, st3, st4, st5 early, st5 late, st6, st7, st8, and st10). The normalized abundance of the transcripts was calculated against the annotated gene models and is shown in a data table, together with the features of the gene models.

2. Experimental design, materials and methods

Two developmental series of mRNA were independently obtained from embryos produced by two pairs of parents and analyzed by RNA-Seq. Embryos were incubated at 25 °C until they reached the appropriate stages. Poly(A) mRNA was extracted using the Dynabeads mRNA DIRECT Kit (Ambion) from 10–100 embryos of each developmental stage. mRNA was fragmented using the NEBNext RNase III RNA Fragmentation Module (New England BioLabs). Sequencing libraries were constructed from the fragmented RNA using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England BioLabs) and NEBNext Multiplex Oligos for Illumina (Index Primers Set 1, New England BioLabs). The libraries were sequenced by Illumina MiSeq. Sequence reads were trimmed to remove the adaptor and primer sequences and low-quality sequences, using the CLC Genomics Workbench 7.0.3 (Qiagen). The parameter settings for the quality trimming were as follows: trim using quality scores, limit=0.05; trim ambiguous nucleotides, maximum number of ambiguities=2; and filter on length, discard reads below length=30. Trimmed reads were then aligned to Ptep_1.0 genome assembly (Ptep_1.0, GCA_000365465.1) using the BLAT algorithm version 34 in the DDBJ Read Annotation Pipeline (<https://p.ddbj.nig.ac.jp/pipeline/>) with default settings. The output alignments

were filtered based on quality, coverage, and uniqueness using a PERL script filterPSL.pl available from the AUGUSTUS 3.0.1 scripts folder (<http://bioinf.uni-greifswald.de/augustus/>). The parameter settings were as follows: minimum coverage, 60%; minimum percent identity, 90%; uniqueness threshold, 0.96. The filtered alignments were converted into SAM files with a custom PERL script. Using the SAM files, the abundance of the mRNAs fitting AUGUSTUS gene models [1] was calculated with htseq-count v. 0.6.1p1 in -s reverse, -m union settings and was normalized in reads per kilobase of exon per million reads (RPKM). The alignments were converted into wig files using the aln2wig script available from the AUGUSTUS 3.0.1 scripts folder. Scores in the wig files represent the normalized read coverage (per 10 million reads).

Acknowledgements

We thank Akiko Noda for technical assistance. This work was supported in part by the Japan Society for the Promotion of Science (JSPS) Grants-in-Aid for Scientific Research (KAKENHI) awards to HO (15K07139) and YA (26440130).

Transparency document. Supporting information

Transparency data associated with this article can be found in the online version at <https://doi.org/10.1016/j.dib.2018.05.106>.

References

- [1] E.E. Schwager, P.P. Sharma, T. Clarke, D.J. Leite, T. Wierschin, M. Pechmann, Y. Akiyama-Oda, L. Esposito, J. Bechsgaard, T. Bilde, A.D. Buffry, H. Chao, H. Dinh, H. Doddapaneni, S. Dugan, C. Eibner, C.G. Extavour, P. Funch, J. Garb, L.B. Gonzalez, V.L. Gonzalez, S. Griffiths-Jones, Y. Han, C. Hayashi, M. Hilbrant, D.S.T. Hughes, R. Janssen, S.L. Lee, I. Maeso, S.C. Murali, D.M. Muzny, R. Nunes da Fonseca, C.L.B. Paese, J. Qu, M. Ronshaugen, C. Schomburg, A. Schönauer, A. Stollewerk, M. Torres-Oliva, N. Turetzek, B. Vanthournout, J.H. Werren, C. Wolff, K.C. Worley, G. Bucher, R.A. Gibbs, J. Coddington, H. Oda, M. Stanke, N.A. Ayoub, N.M. Prpic, J.F. Flot, N. Posnien, S. Richards, A.P. McGregor, The house spider genome reveals an ancient whole-genome duplication during arachnid evolution, *BMC Biol.* 15 (2017) 62.
- [2] Y. Akiyama-Oda, H. Oda, Early patterning of the spider embryo: a cluster of mesenchymal cells at the cumulus produces Dpp signals received by germ disc epithelial cells, *Development* 130 (2003) 1735–1747.
- [3] K. Yamazaki, Y. Akiyama-Oda, H. Oda, Expression patterns of a twist-related gene in embryos of the spider *Achaearanea tepidariorum* reveal divergent aspects of mesoderm development in the fly and spider, *Zool. Sci.* 22 (2005) 177–185.
- [4] B. Mittmann, C. Wolff, Embryonic development and staging of the cobweb spider *Parasteatoda tepidariorum* C.L. Koch, 1841 (syn.: *Achaearanea tepidariorum*; Araneomorphae; Theridiidae), *Dev. Genes Evol.* 222 (2012) 189–216.