Supplemental Materials for

De novo assembly and comparative transcriptome characterization of *Poecilobdella javanica* provide insight into blood feeding of medicinal leech

Zichao Liu^{1,*,a}, Yanjie Wang^{2,a}, Xiangrong Tong^{1,a}, Yuan Su¹, Lijiang Yang¹, Debin Wang^{1,*}, Yuqi Zhao^{3,*}

1. Department of Life Science & Technology, Kunming University, Kunming Key Laboratory of Hydroecology Restoration of Dianchi Lake, Key Laboratory of Special Biological Resource Development & Utilization of Universities in Yunnan Province, Kunming, 650214, China.

2. Jules Stein Eye Institute, Department of Ophthalmology, University of California, Los Angeles, CA 90095, USA.

3. Department of Integrative Biology and Physiology, University of California, Los Angeles, Los Angeles, CA 90095, USA.

^a These authors contributed equally to this work.

^{*} Author for Correspondence:

Zichao Liu:

Department of Life Science & Technology, Kunming University, Kunming 650214, China, Email: abclzc@aliyun.com

Debin Wang:

Department of Life Science & Technology, Kunming University, Kunming 650214, China, Email: wangdebin1202@sina.com

Yuqi Zhao:

Department of Integrative Biology and Physiology, University of California, Los Angeles, Los Angeles, CA 90095, USA, E-mail: zhaoyuqi616@gmail.com

Running title: De Novo Assembly of Leech Transcriptome



Supplement Figure S1. Clean reads in leech RNA-seq. (A) Reads counts in all the RNA-seq libraries. (B) An example of per base sequence quality of clean reads in 20day RNA-seq library. Clean reads were generated by filtering out low-quality reads (containing Ns and with more than 30% of bases with Phred quality score of \leq 30) and adaptor sequences.



Supplement Figure S2. The abundance histogram of K-mer in leech RNA-seq. The KmerGenie was used to estimate the best k-mer length for the de novo assembly [1]. The best k-mer lengths range from 19 to 31.



Supplement Figure S3. The gene expression correlations among time points. The gene expression levels were estimated using RSEM (v1.3.0) [2].



Supplement Figure S4. Heatmap of differentially expressed BUSCO single-copy orthologs in assembled transcriptomes.

References

- 1. Chikhi R, Medvedev P (2014) Informed and automated k-mer size selection for genome assembly. Bioinformatics 30: 31-37.
- 2. Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. Bmc Bioinformatics 12.