

***De novo* sequencing, assembly and analysis of eight different transcriptomes from the Malayan pangolin**

The International Pangolin Research Consortium (IPaRC)

Aini Mohamed Yusoff^{1,4}, Tze King Tan^{1,4}, Ranjeev Hari^{1,4}, Klaus-Peter Koepfli⁷, Wei Yee Wee^{1,4}, Agostinho Antunes^{8,9}, Frankie Thomas Sitam⁵, Jeffrine Rovie Ryan Japning⁵, Kayal Vizi Karuppanan⁵, Guat Jah Wong¹, Leonard Lipovich^{10,11}, Wesley C. Warren³, Stephen J. O'Brien^{2,6}, Siew Woh Choo^{1,4}

Affiliation addresses

¹Genome Informatics Research Laboratory, High Impact Research (HIR) Building, University of Malaya, 50603 Kuala Lumpur, Malaysia.

²Theodosius Dobzhansky Center for Genome Bioinformatics St. Petersburg State University St. Petersburg, Russia 199004

³McDonnell Genome Institute, Washington University, St Louis, MO 63108, USA.

⁴Department of Oral Biology and Biomedical Sciences, Faculty of Dentistry, University of Malaya, 50603 Kuala Lumpur, Malaysia

⁵Ex-Situ Conservation Division, Department of Wildlife and National Parks, 10 Jalan Cheras, 56100 Kuala Lumpur, Malaysia.

⁶Oceanographic Center, 8000 N. Ocean Drive, Nova Southeastern University, Ft Lauderdale, Florida 33004, USA.

⁷National Zoological Park, Smithsonian Conservation Biology Institute, Washington, DC 20008, USA

⁸CIIMAR/CIMAR, Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Rua dos Bragas, 177, 4050-123 Porto, Portugal.

⁹Department of Biology, Faculty of Sciences, University of Porto, Rua do Campo Alegre, 4169-007 Porto, Portugal.

¹⁰Center for Molecular Medicine and Genetics, Wayne State University, Detroit, MI 48201, USA.

¹¹Department of Neurology, School of Medicine, Wayne State University, Detroit, MI 48201, USA.

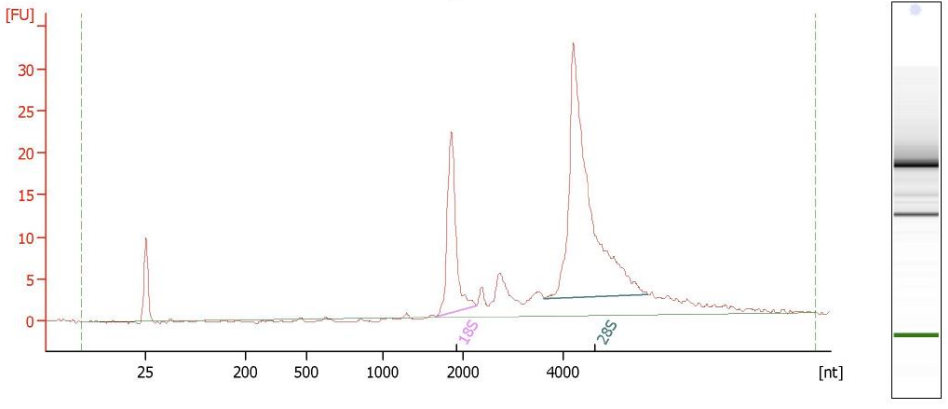
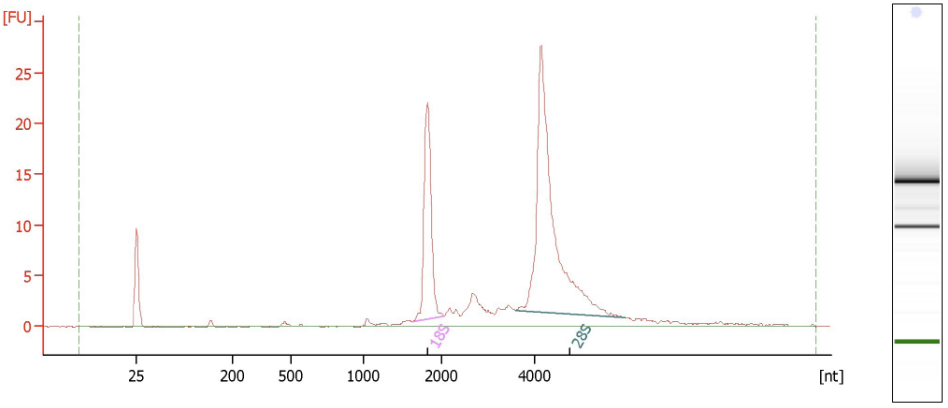
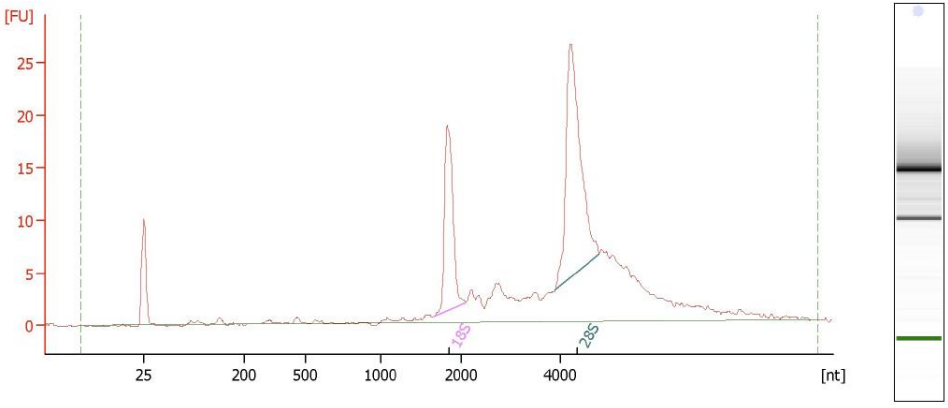
Corresponding author:

Siew Woh Choo

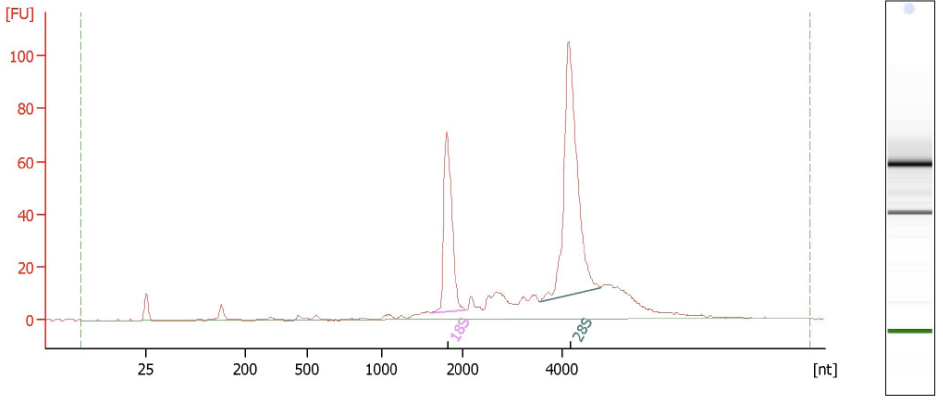
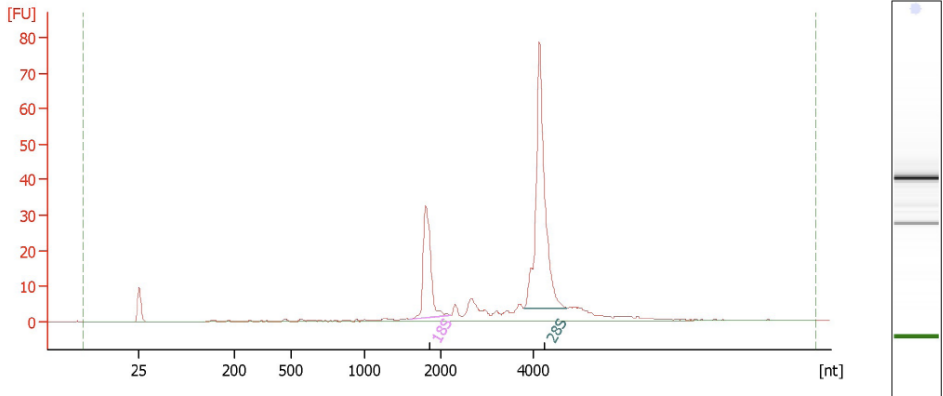
lchoo@um.edu.my or lchoo@genomesolutions.com.my

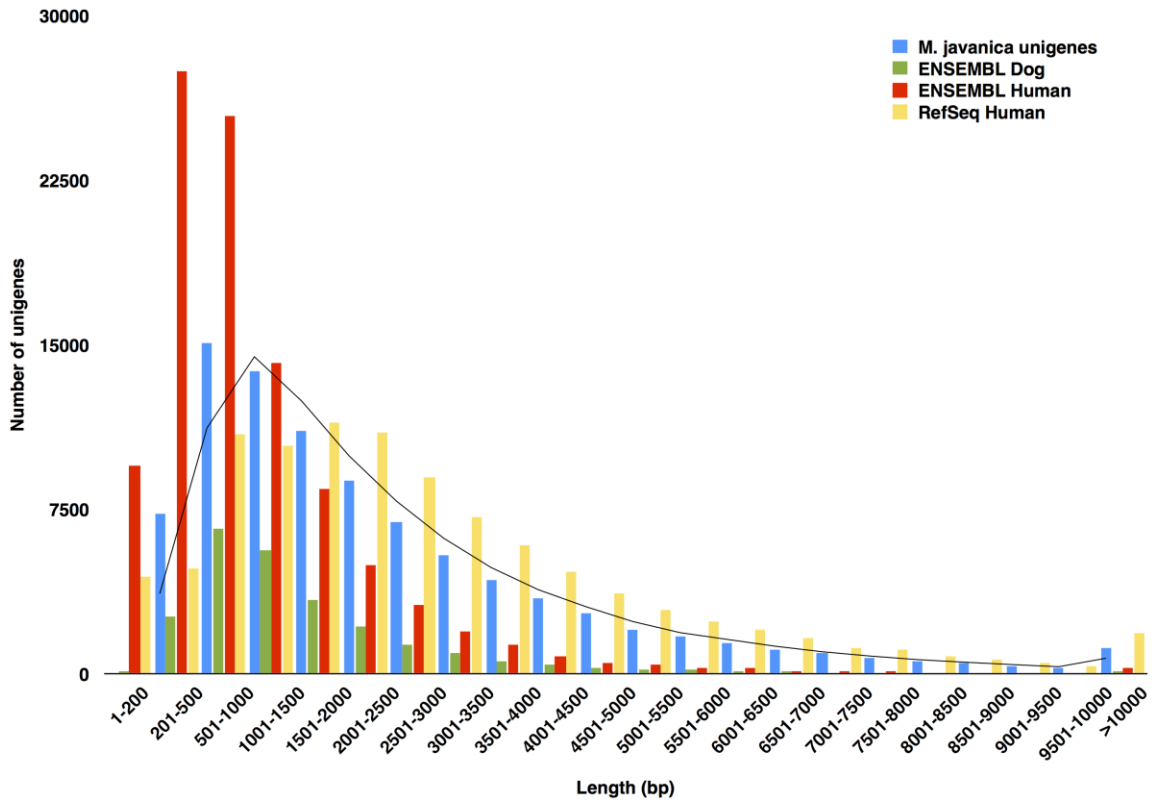
Supplementary Figures

Supplementary Figure 1: Table of Bioanalyzer results indicating the quality of RNA extracted from respective tissue.

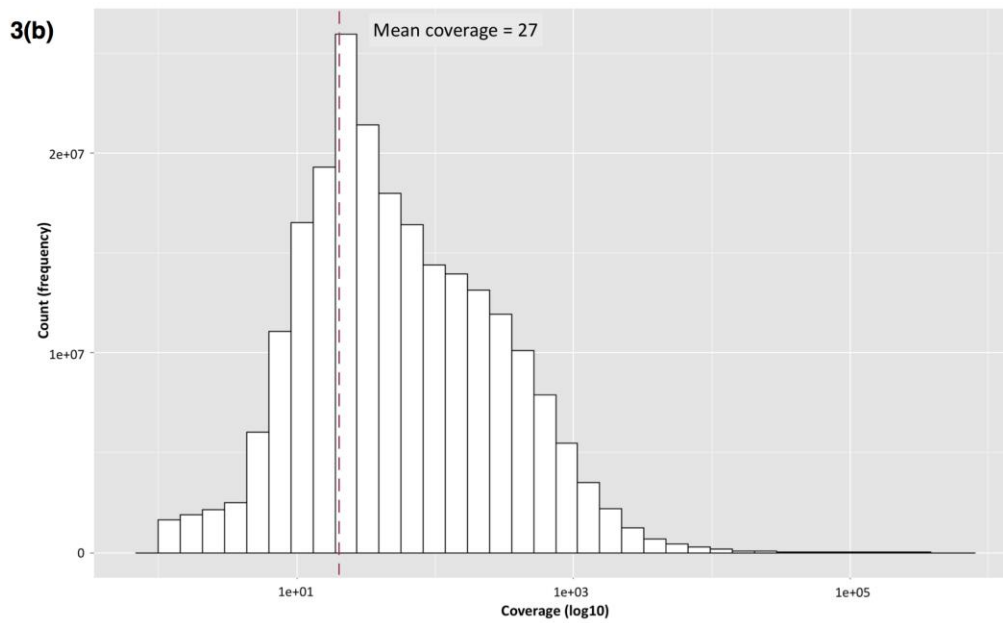
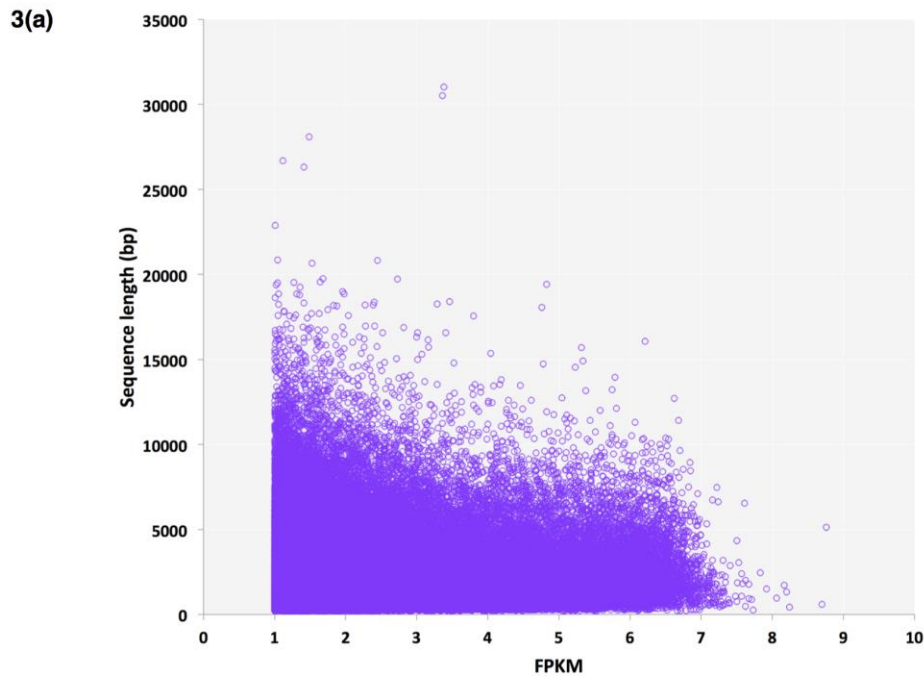
Organ/ details	Electropherogram
<p>Cerebellum</p> <p>RIN: 10.0</p> <p>RNA area:196.6</p> <p>RNA concentration: 113 ng/μl</p>	
<p>Cerebrum</p> <p>RIN: 10.0</p> <p>RNA area: 140.1</p> <p>RNA concentration: 120 ng/μl</p>	
<p>Heart</p> <p>RIN: 9.30</p> <p>RNA area:187.3</p> <p>RNA concentration: 107 ng/μl</p>	

Organ/ details	Electropherogram	
<p>Kidney</p> <p>RIN: 9.90</p> <p>RNA area:639.6</p> <p>RNA concentration: 550 ng/μl</p>		
<p>Liver</p> <p>RIN: 9.10</p> <p>RNA area:441.7</p> <p>RNA concentration: 253 ng/μl</p>		
<p>Lungs</p> <p>RIN: 10.0</p> <p>RNA area:324.7</p> <p>RNA concentration: 279 ng/μl</p>		

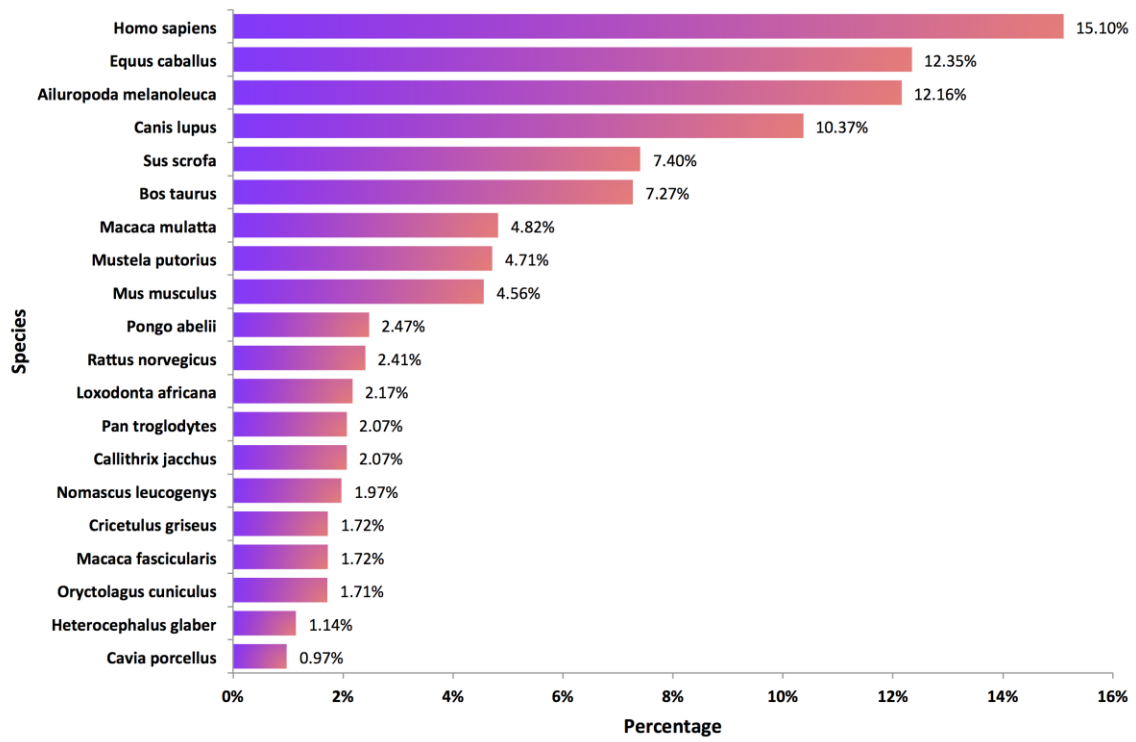
Organ/ details	Electropherogram
<p>Spleen</p> <p>RIN: 9.80</p> <p>RNA area:544.0</p> <p>RNA concentration: 312 ng/μl</p>	
<p>Thymus</p> <p>RIN: 9.60</p> <p>RNA area:236.9</p> <p>RNA concentration: 204 ng/μl</p>	



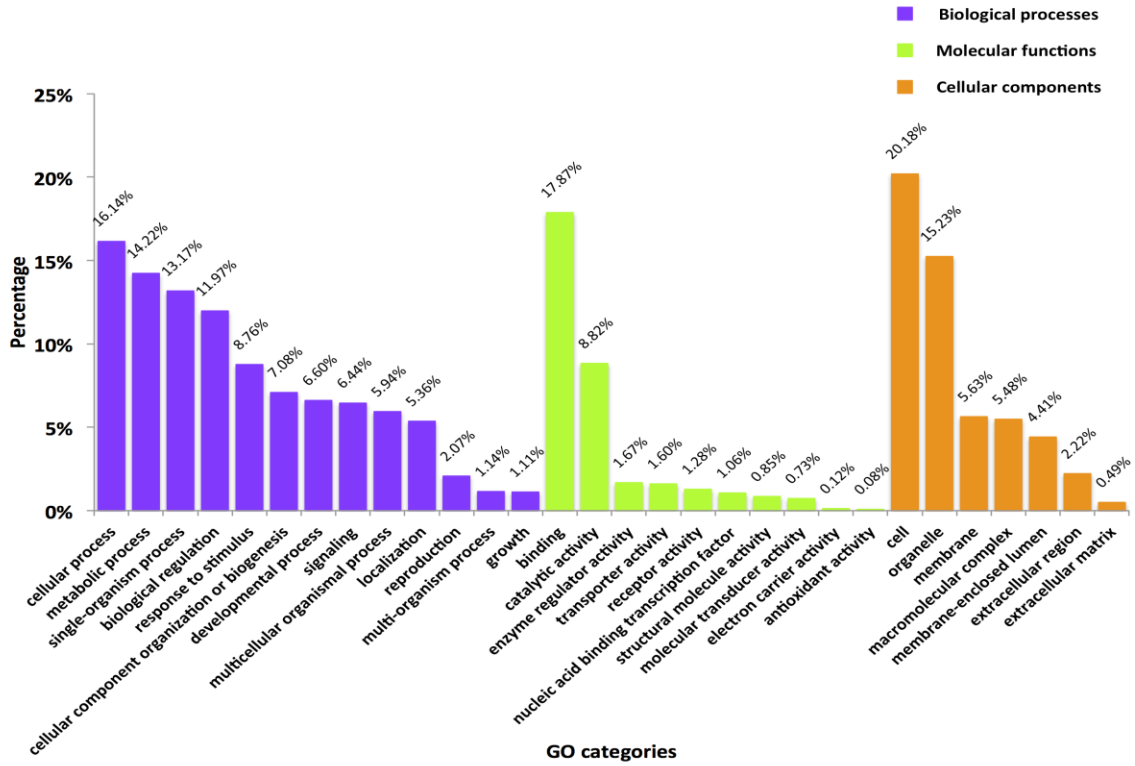
Supplementary Figure 2: *M. javanica* unigene length distribution in comparison with ENSEMBL dog and human genes, and RefSeq human genes.



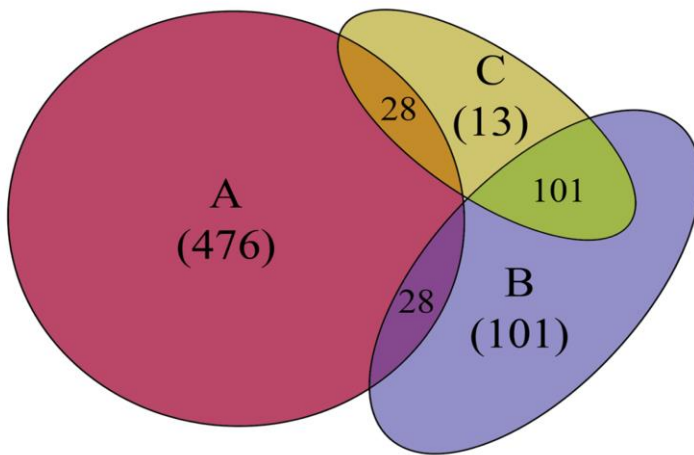
Supplementary Figure 3: *M. javanica* transcript validation and quality assessment. Figure 3a) describes the relationship between FPKM values and unigene length. The minimum supported FPKM is 1.00. Figure 3b) shows the mapping coverage distribution of *M. javanica* unigenes. The red dotted line indicates the mean mapping coverage, denoting the quality of mapping.



Supplementary Figure 4: Top 20 BLAST hits distribution to known species.



Supplementary Figure 5: Functional classification of assembled *M. javanica* unigenes.



Supplementary Figure 6: Mapping of unigenes to metabolic pathways. A=Glycerolipid metabolism; B=Fatty acid degradation; and C=Linoleic and alpha-linolenic metabolism.