



# Transcriptomic analysis identifies genes and pathways related to myrmecophagy in the Malayan pangolin (*Manis javanica*)

Jing-E Ma<sup>1</sup>, Lin-Miao Li<sup>1</sup>, Hai-Ying Jiang<sup>1</sup>, Xiu-Juan Zhang<sup>1</sup>, Juan Li<sup>1</sup>, Guan-Yu Li<sup>1</sup>, Li-Hong Yuan<sup>1</sup>, Jun Wu<sup>2</sup> and Jin-Ping Chen<sup>1</sup>

<sup>1</sup>Guangdong Key Laboratory of Animal Conservation and Resource, Guangdong Public Laboratory of Wild Animal Conservation and Utilization, Guangdong Institute of Applied Biological Resources, Guangzhou, Guangdong, China

<sup>2</sup>Wildlife Disease Surveillance and Molecular Ecology Research Center, Nanjing Institute of Environmental Sciences under Ministry of Environmental Protection, Nanjing, Jiangsu, China

## ABSTRACT

The Malayan pangolin (*Manis javanica*) is an unusual, scale-covered, toothless mammal that specializes in myrmecophagy. Due to their threatened status and continuing decline in the wild, concerted efforts have been made to conserve and rescue this species in captivity in China. Maintaining this species in captivity is a significant challenge, partly because little is known of the molecular mechanisms of its digestive system. Here, the first large-scale sequencing analyses of the salivary gland, liver and small intestine transcriptomes of an adult *M. javanica* genome were performed, and the results were compared with published liver transcriptome profiles for a pregnant *M. javanica* female. A total of 24,452 transcripts were obtained, among which 22,538 were annotated on the basis of seven databases. In addition, 3,373 new genes were predicted, of which 1,459 were annotated. Several pathways were found to be involved in myrmecophagy, including olfactory transduction, amino sugar and nucleotide sugar metabolism, lipid metabolism, and terpenoid and polyketide metabolism pathways. Many of the annotated transcripts were involved in digestive functions: 997 transcripts were related to sensory perception, 129 were related to digestive enzyme gene families, and 199 were related to molecular transporters. One transcript for an acidic mammalian chitinase was found in the annotated data, and this might be closely related to the unique digestive function of pangolins. These pathways and transcripts are involved in specialization processes related to myrmecophagy (a form of insectivory) and carbohydrate, protein and lipid digestive pathways, probably reflecting adaptations to myrmecophagy. Our study is the first to investigate the molecular mechanisms underlying myrmecophagy in *M. javanica*, and we hope that our results may play a role in the conservation of this species.

Submitted 31 July 2017

Accepted 15 November 2017

Published 22 December 2017

Corresponding author

Jin-Ping Chen, chenjp@gdei.gd.cn,

chenjp@giabr.gd.cn

Academic editor

Edward Braun

Additional Information and  
Declarations can be found on  
page 17

DOI 10.7717/peerj.4140

© Copyright  
2017 Ma et al.

Distributed under  
Creative Commons CC-BY 4.0

OPEN ACCESS

**Subjects** Conservation Biology, Genomics, Zoology

**Keywords** Digestion, Myrmecophagy, Conservation, Pangolin

## INTRODUCTION

Pangolins, also known as scaly anteaters, are eutherians and placental mammals. Eight pangolin species are recognized: four from Asia, *Manis javanica*, *M. pentadactyla*, *M. crassicaudata*, and *M. culionensis*, and four from Africa, *Phataginus tricuspis*, *P. tetradactyla*, *Smutsia gigantea*, and *S. temminckii*, the genus *Manis* to the four Asian species while assigning the two African tree pangolin species to the genus *Phataginus* and the two terrestrial African pangolin species to the genus *Smutsia* (Du Toit Z et al., 2014; Gaudin, Emry & Wible, 2009). *M. javanica* is found mainly in Southeast Asia (Trageser et al., 2017). In pangolins, unlike other placental mammals, the skin is covered by large and overlapping keratinized scales (Meyer et al., 2013). Furthermore, pangolins are edentulous or toothless and are thus specialized within an already unusual mammalian dietary niche (Pietersen et al., 2016; Yang et al., 2007). Pangolins feed mainly on ants and termites (Ashokkumar et al., 2017; Pietersen et al., 2016; Mahmood et al., 2013; Richer, Coulson & Heath, 1997; Swart, Richardson & Ferguson, 1999). Pangolins also have a well-developed muscular system for fossorial or arboreal behavior and a remarkable olfactory system (Choo et al., 2016). As predators of ants and termites, pangolins have a specialized diet and perform an important ecological role in regulating insect populations (Hua et al., 2015). Individual adult pangolins have been estimated to consume more than 70 million insects annually and pangolins provide an important regulatory function in controlling both ants and termites (Hua et al., 2015; Pietersen et al., 2016; Yang et al., 2007). In addition to their ecological value, pangolins are extremely economically important animals. Pangolins are the most poached and trafficked mammal in the world because of the high demand for their meat, which is considered a delicacy, and their scales are used in traditional medicine (Trageser et al., 2017).

*Manis javanica* is classified as critically endangered according to the International Union for Conservation of Nature (IUCN) Red List of Threatened Species® (Challender et al., 2014) and has been included in Appendix I of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES I). The numbers of *M. javanica* in the wild have been dramatically declining for several reasons. A major threat is the rapid loss and deterioration of their natural habitat as a result of deforestation; illegal hunting, which might reflect the pangolin's economic value; and human agricultural expansion (Hua et al., 2015). Therefore, artificial breeding may be the best choice for ensuring pangolin survival. A suitable artificial diet is one of the critical limiting factors in raising pangolins in captivity because they can have many digestive problems (Hua et al., 2015). Pangolins have adapted to a highly specialized diet of ants and termites; thus it is difficult to replace their natural food completely by artificial food (Hua et al., 2015; Yang et al., 2007). Pangolins favor high-protein, high-fat, and high-calorie food, and they have the notable ability to digest and absorb chitin in the digestive system (Hua et al., 2015; Yang et al., 2007). Chitin is a linear polymer of N-acetyl-beta-glucosaminidase subunits linked by a beta-1, 4 glycosidic bond and is one of the constituents of insect exoskeletons and the peritrophic membranes of ants (Kawashima et al., 2016). The internal degradation of chitin particles is mainly catalyzed by chitinase (Strobel et al., 2013).

Over the past 150 years, several zoos have tried to maintain pangolins. However, because of inadequate diets, these animals have not been successfully maintained for long periods by most zoos (Hua *et al.*, 2015). Some formulas for diets fed to *M. javanica* in captivity use a paste mixture of several kinds of food, such as hard-boiled chicken eggs, multivitamin liquid, horse meat, water, mealworms, insectivore pellets, salmon oil, and powdered termite mound (Vijayan, Yeong & Ling, 2009). Digestive disorders often appear in pangolins fed with artificial food, and the faeces of the animals become fluid. Several researchers have suggested that a certain proportion of chitin might be the key to artificial diets for pangolins (Ya-yong *et al.*, 1999; Yang *et al.*, 2007); however, an understanding of the molecular genetics that might provide a theoretical basis for raising *M. javanica* in captivity is lacking.

Genetic studies of endangered species have become increasingly widespread in the last two years (Choo *et al.*, 2016; Mwale *et al.*, 2017; Yusoff *et al.*, 2016; Zhihai *et al.*, 2016). In particular, the genomes and transcriptomes of *M. javanica* have been reported (Choo *et al.*, 2016; Yusoff *et al.*, 2016). Currently, increasing amounts of molecular information about *M. javanica* have become available as a result of the emergence of high-throughput next-generation sequencing (NGS) technologies. The complete genome sequencing of *M. javanica* and the transcriptome sequencing of eight organs, including the heart, liver, spleen, lung, kidney, thymus, cerebellum, and cerebrum, have progressively revealed unknown aspects of pangolin biology (Yusoff *et al.*, 2016). High-quality transcriptomes have been used for analyses of the functional and phylogenetic aspects of immunity biology (Yusoff *et al.*, 2016), but genetic research regarding myrmecophagy is still lacking. Here, two questions are raised. Are molecular pathways involved in the evolution of this dietary adaptation for myrmecophagy, and how do they affect the appearance of this feature? In terms of dietary adaptation, we wanted to investigate whether specific genes exist for digestive function. We selected the liver, small intestine, and salivary glands for transcriptome sequencing to analyze the genetic selection of potential candidate genes involved in myrmecophagy. Salivary glands are important amylase secretory organs although liver secretions also contain amylase and lipase, and the small intestine is involved in the absorption of nutrients. We aimed to analyze the unique feeding behavior of pangolins in the hope that our results may provide a new approach to the protection of this species.

## MATERIALS AND METHODS

### Ethics statement

All animal procedures were approved by the ethics committee for animal experiments at the Guangdong Institute of Applied Biological Resources (reference number: G2ABR20170523), and followed basic principles.

### Biological sample

Briefly, an adult female *M. javanica* was provided by the Dongguan Institute of Qingfengyuan Animal Medicine (Dongguan, Guangdong, China). The specimen was a wild individual that died in the process of rescue; the animal was 4.08 kg in weight before its natural death. The animal was dissected immediately after its natural death. During the

dissection, ants were found in its stomach. The salivary glands, liver, and small intestine were collected as soon as possible, frozen in liquid nitrogen, and stored at  $-80^{\circ}\text{C}$  until RNA extraction.

### **RNA isolation, cDNA library construction and Illumina sequencing**

Total RNA was extracted from the three tissues using RNAiso reagent (Takara, Otsu, Japan) and was treated with DNase I (Takara). RNA purity was checked using a NanoPhotometer spectrophotometer (IMPLEN, CA, USA). RNA concentration was measured using the Qubit RNA Assay Kit and a Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA), and RNA integrity was assessed using the RNA Nano 6000 Assay Kit and the Agilent Bioanalyzer 2100 system (Agilent Technologies, Santa Clara, CA, USA). RNA was frozen at  $-80^{\circ}\text{C}$  until cDNA library construction.

RNA samples were then mixed with fragmentation buffer (Ambion, Foster City, CA, USA) and fragmented into short fragments; the average insert size was 200 bp. mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. DNA contaminants were further removed through DNase enzyme digestion followed by rRNA removal. cDNA synthesis was then followed by PCR amplification to generate a complete cDNA library, which was sent for sequencing in a flowcell on the Illumina HiSeq<sup>TM</sup> 2000 platform using the TruSeq PE Cluster Kit V3-cBot-HS (Illumina PE-401-3001) and TruSeq SBS Kit-HS v3 200 cycles (Illumina FC-401-3001).

### **Data assembly and annotation**

Four groups of sequencing data were used for assembly and annotation and the data were annotated using a published paper housed at the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) accession number [SRR2561213](https://www.ncbi.nlm.nih.gov/sra/SRR2561213) (Yusoff *et al.*, 2016). The primary sequencing data were cleaned by (a) removing reads with adaptors; (b) removing reads where the proportion of unknown bases (N) exceeded 5% ( $-n = 0.05$ ); and (c) removing low-quality reads (reads in which the proportion of bases with quality  $<10$  exceeded 20%, i.e.,  $-l = 10$  and  $-q > 0.2$ ). Clean reads were aligned to reference sequences using a spliced read mapper for RNA-Seq-*TopHat2* (<http://ccb.jhu.edu/software/tophat/index.shtml>), which is based on Bowtie2 (<http://bowtie.cbcb.umd.edu/>). The parameter values of RNA-Seq-*TopHat2* were as follows;  $-\text{read-mismatches}$ : 2,  $-\text{read-edit-dist}$ : 2,  $-\text{max-intron-length}$ : 5000000,  $-\text{library-type}$ : fr-unstranded,  $-\text{GTF}$ : genome.gtf,  $-\text{mate-inner-dist}$ : 40,  $-\text{solexa1.3-quals}$ , and other parameters were set to the default values. There were three steps: (a) align the reads to the reference transcriptomes; (b) align the reads to the reference genome without the alignment described in (a); and (c) align the reads segmentally to the reference genome without the alignment described in (b). Following this, the alignment data were used to calculate the distribution and coverage of the reads on the reference genes.

To offer complementary insights, all transcripts ( $>200$  bp) were annotated on the basis of basic local alignment search tool (BLASTX) results with  $e$ -values of  $1e^{-5}$  against a total of seven databases, including the non-redundant protein database (NR, <ftp://ftp.ncbi.nih.gov/blast/db/>); a manually annotated, non-redundant protein

sequence database (Swiss-Prot, <http://www.uniprot.org/>); the Translated EMBL Nucleotide Sequence Data Library (TrEMBL, <http://www.bioinfo.pte.hu/more/TrEMBL.htm>), which complements the Swiss-Prot protein knowledgebase; the Kyoto Encyclopedia of Genes and Genomes (KEGG, <http://www.genome.jp/kegg/>), which is used to understand the high-level functions and utilities of biological system (such as the cell, the organism and the ecosystem) based on molecular-level information, especially for large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies; Gene Ontology (GO, <http://www.geneontology.org/>), which described biological processes, molecular functions, and cellular components; and Clusters of Orthologous Groups of proteins (KOG/COG: COG, <http://www.ncbi.nlm.nih.gov/COG/>, KOG, <http://www.ncbi.nlm.nih.gov/KOG/>). The best-aligned results were used to determine the sequence direction of the transcripts. If the results from the different databases were in conflict, the sequence direction of the transcript was determined using the prioritization order NR, Swiss-Prot, KEGG and COG.

### Analysis of differentially expressed genes

Gene expression levels were estimated based on the fragments per kilo base (kb) of transcript per million fragments mapped. The formula is as follows: FPKM = cDNA Fragments/Mapped Fragments (Millions)/Transcript Length (kb). Prior to differential gene expression analysis, for each sequenced library, the read counts were adjusted using the edgeR program package and a scaling normalized factor (Robinson, McCarthy & Smyth, 2010). The differential expression analysis of two samples was performed using the DESeq R package (Wang et al., 2010). *P*-value was adjusted using *q* value, and *q* value < 0.005 &  $|\log_2(\text{fold change})| \geq 1$  was set as the threshold for significantly differential expression (Storey & Tibshirani, 2003).

### Correlation between any two pangolin tissue transcriptomes

To examine the similarity between the different organ transcriptomes, the expression levels of the transcripts (FPKM) in the transcriptomes of each tissue were manipulated using the tool “RSEM-calculate-expression” in the RSEM pipeline (<http://deweylab.biostat.wisc.edu/rsem/README.html>), which performs accurate transcript quantification from RNA-Seq data with or without a reference genome. The reads for each tissue were mapped to the transcripts (Li & Dewey, 2011). Gene expression values, expressed as  $\log_{10}(\text{FPKM} + 1)$  for the transcriptomic data from each tissue, were plotted against one another to produce scatter plots.  $R^2$  values were then calculated from the scatter plots to assess the correlation between any two *M. javanica* transcriptomes.

## RESULTS

### Illumina sequencing and assembly

To obtain a comprehensive and representative transcriptome of *M. javanica*, 97,353,658 high-quality clean reads (for a total length of 2,920,609,740 bp) were generated from the three tissues after the removal of the adaptor sequences. All high-quality sequencing reads from *M. javanica* are available on the NCBI Gene Expression Omnibus (GEO)

**Table 1** Sequencing statistics for the transcriptomes obtained from the salivary glands, liver, and small intestine of adult female Malayan pangolin (*Manis javanica*).

Organ	Raw pairs (bp)	Clean pairs (bp)	Clean bases (bp)	GC content	% ≥ Q30%
Small intestine	20,296,915	15,117,014	4,535,104,200	53.71%	95.30%
Liver	33,288,751	22,073,184	6,621,955,200	53.15%	94.67%
Salivary glands	20,297,074	14,857,037	4,457,111,100	51.33%	95.58%
Referred liver	–	45,306,423	9,061,284,600	50.87%	90.35%

**Table 2** Summary of the sequencing data aligned to the Malayan pangolin (*Manis javanica*) whole genome reference sequence.

Organ	Total reads	Mapped reads	Unique mapped reads	Multiple map reads
Small intestine	30,234,028	22,277,733 (73.68%)	21,843,354 (72.25%)	434,379 (1.44%)
Liver	44,146,368	31,965,251 (72.41%)	31,694,559 (71.79%)	270,692 (0.61%)
Salivary glands	29,714,074	20,673,075 (69.57%)	19,911,660 (67.01%)	761,415 (2.56%)
Reference liver	90,612,846	80,913,231 (89.30%)	80,064,920 (88.36%)	848,311 (0.94%)

database under accession numbers [GSM2667949](#), [GSM2667950](#) and [GSM2667951](#). The average proportion of high-quality clean reads was 95.58% (Table 1). Clean reads were assembled into long assembled sequences (contigs) using *TopHat2*. The alignment efficiency between the sample and reference genome ranged from 69.57–89.30% (Table 2), while the alignment efficiency between the sample and the exons of the reference genome ranged from 76.11–85.16% (Fig. S1).

### Functional annotation

From the *M. javanica* transcriptome, 22,538 transcripts (93.05%) were annotated on the basis of the COG, GO, KEGG, KOG, Swiss-Prot, TrEMBL, and NR databases using BLAST. A total of 6,228 transcripts were annotated against the COG database, followed by 13,977, 14,115, 16,648, 17,135, and 20,964 transcripts annotated on the basis of the KEGG, KOG, GO, Swiss-Prot, and TrEMBL databases, respectively (Table S1). As expected, the majority of the 22,473 transcripts matched the NR databases ( $e$ -value  $< 10^{-5}$ ) (Fig. S2 and File S1). The *M. javanica* transcripts were annotated on the basis of the top BLASTX hits in the species distribution statistics. The top five organisms were *Ceratotherium simum* (2,382 transcripts, 10.60%), *Equus caballus* (1,402, 6.24%), *Canis lupus* (1,349, 6.01%), *Mustela putorius* (1,044, 4.65%) and *Odobenus rosmarus* (1,022, 4.55%) (Fig. S3).

### COG and KOG analysis

In the COG database, the largest category of *M. javanica* annotated transcripts was general function prediction only (R) (2,438 transcripts, 27.76%), followed by replication, recombination, and repair (L) (848, 9.65%); transcription (K) (845, 9.62%); signal transduction mechanisms (T) (790, 8.99%); and post-translational modification, protein

turnover, and chaperones (O) (484, 5.51%) (Fig. S4A). In the KOG database, the largest category of *M. javanica* annotated transcripts was general function prediction only (R) (2,910 transcripts, 18.3%), followed by signal transduction mechanisms (T) (2,744, 17.26%); post-translational modification, protein turnover, and chaperones (O) (1,259, 7.92%); function unknown (S) (1,182, 7.43%); and transcription (K) (1070, 6.73%) (Fig. S4B).

In both the COG and KOG analyses, several transcripts were involved in the transport and metabolism of the three major nutrients: 136 transcripts were related to carbohydrate transport and metabolism, 122 were related to lipid transport and metabolism, and 124 were related to amino acid transport and metabolism (File S2).

### Gene Ontology (GO)

Annotation of the *M. javanica* transcripts with the GO database classified 16,649 transcripts into 61 small classes in three ontologies: biological processes, molecular functions, and cellular components. A total of 44.36% of the transcripts were assigned to biological processes, 16.26% to molecular functions, and 39.38% to cellular components.

In the biological process ontology, the most highly represented terms were cellular processes (10,494, 63.03%), single-organism processes (9,554, 57.38%), and biological regulation (7,986, 47.97%). The fourth most represented term was metabolic process (7,173, 43.08%), which was followed by response to stimulus (4,915, 29.52%), multicellular organismal process (3,644, 21.89%), signaling (3,161, 18.99%), localization (3,001, 18.03%), developmental process (2,890, 17.36%), and cellular component organization or biogenesis (2,667, 16.02%). The terms associated with biological regulation and metabolic process might be indicative of *M. javanica* transcriptome involvement in various digestive activities.

For molecular functions, the sequences were mainly assigned to binding (9,371, 56.29%) and catalytic activity (5,347, 32.12%) followed by molecular transducer activity (1,750, 10.51%), receptor activity (1,681, 10.1%) and transporter activity (1,026, 6.16%), which might be involved in food digestion and absorption.

As anticipated, cell part (11,445, 68.74%) and cell (11,412, 68.54%) were the predominant terms assigned to the pangolin transcriptome in cellular components followed by organelle (8,014, 48.14%), membrane (5,980, 35.92%), membrane part (4,441, 26.67%), organelle part (4,079, 24.5%), macromolecular complex (3,413, 20.5%) and extracellular region (1,044, 6.27%) (Fig. S5, File S3). Overall, these results indicate that a broad range of biological activities were related to the expressed pangolin transcriptome, representing a pooled collection of the three digestive tissues sequenced.

### KEGG pathway analysis

To identify the pathways in which the *M. javanica* transcripts were involved, the transcripts were mapped on the basis of KEGG pathways. A total of 13,977 (57.71%) *M. javanica* transcripts were associated with 290 unique KEGG pathways, with a total of 15 unique KEGG pathways representing cellular processes, followed by 22, 27, 66, 68, and 89 representing genetic information passing, environmental information processing, organismal systems, human diseases, and metabolism, respectively (File S4).

The most-represented pathways in the *M. javanica* transcripts included olfactory transduction (969 transcripts) and pathways in cancer (444 transcripts), followed by the PI3K-Akt signaling pathway (391 transcripts), the MAPK signaling pathway (300 transcripts), and neuroactive ligand–receptor interaction (292 transcripts) (Fig. S6). The sense of smell is closely related to the biological activity of instinctive behaviors such as feeding, and the olfactory pathway plays a key role in the specific recognition of smells, thus leading the animal to different foods. A total of 969 genes were associated with olfactory transduction in the *M. javanica* transcripts, 942 transcripts of which were annotated as various kinds of olfactory receptors. These findings may explain the keen sense of smell in *M. javanica*.

### Metabolic pathway analysis

A total of 1,814 transcripts were associated with 89 unique KEGG metabolic pathways. Most transcripts were involved in lipid metabolism (431 transcripts), carbohydrate metabolism (365 transcripts), amino acid metabolism (321 transcripts), glycan biosynthesis and metabolism (271 transcripts), nucleotide metabolism (233 transcripts), and the metabolism of cofactors and vitamins (225 transcripts). Other transcripts were associated with global and overview maps (199 transcripts), energy metabolism (167 transcripts), the metabolism of other amino acids (119 transcripts), and xenobiotic biodegradation and metabolism (116 transcripts); a small number of the transcripts were associated with terpenoid and polyketide metabolism (27 transcripts) and secondary metabolite biosynthesis (13 transcripts) (Fig. 1A).

### Carbohydrate metabolism

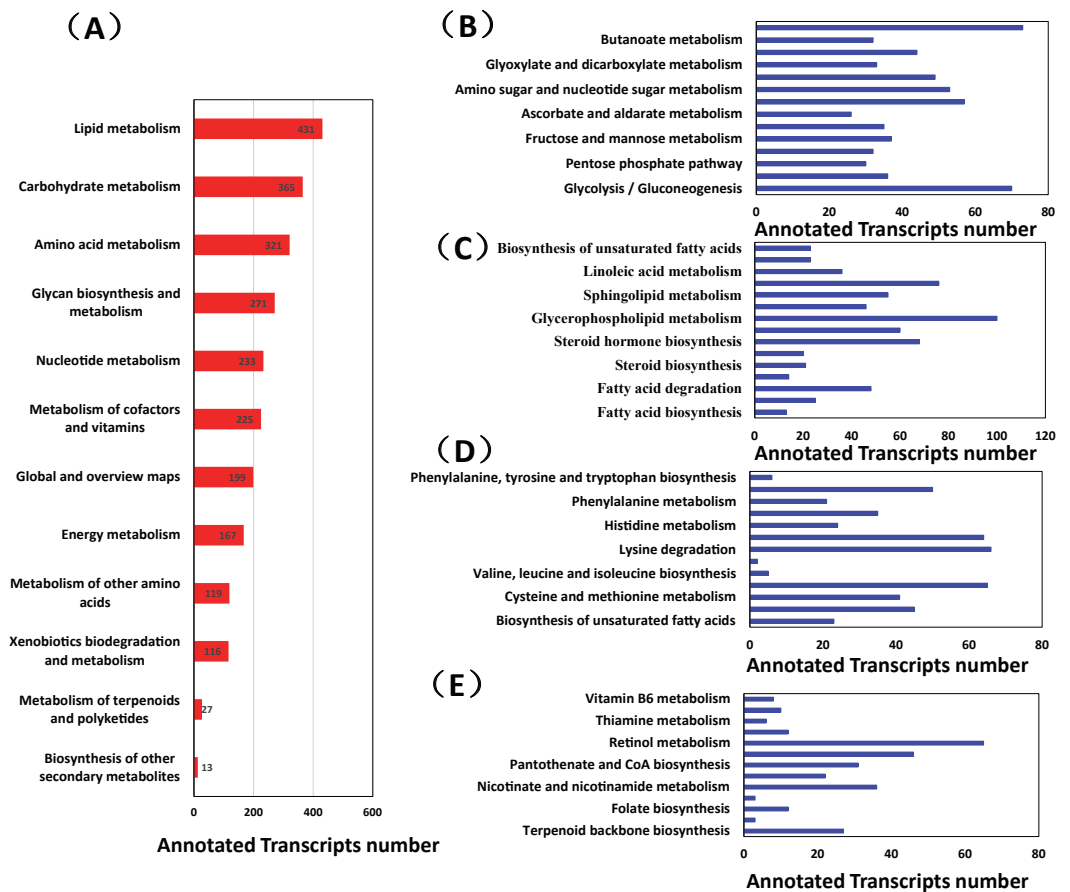
Inositol phosphate metabolism (73 transcripts), glycolysis/gluconeogenesis (70 transcripts), starch and sucrose metabolism (57 transcripts), and amino sugar and nucleotide sugar metabolism (53 transcripts) were at the top of the carbohydrate metabolic lists, whereas ascorbate and aldarate metabolism were at the bottom.

The chitin-degrading enzyme acidic mammalian chitinase (CHIA), which is involved in the degradation of the chitin in the insect cuticle and the peritrophic membrane of the dietary ant, was found in the amino sugar and nucleotide sugar metabolism pathway (KEGG: 00520), thus suggesting that this pathway may be directly involved in ant digestion by *M. javanica* (Fig. 1B).

### Lipid metabolism

Glycerophospholipid metabolism (100 transcripts), arachidonic acid metabolism (76 transcripts), steroid hormone biosynthesis (68 transcripts), and sphingolipid metabolism (55 transcripts) were at the top of the lipid metabolism list; in contrast, fatty acid biosynthesis (55 transcripts) was at the bottom. We identified transcripts from several pathways in unsaturated fatty acid metabolism, including arachidonic acid metabolism (76 transcripts), linoleic acid metabolism (36 transcripts), alpha-linolenic acid metabolism pathways (23 transcripts), and the biosynthesis of unsaturated fatty acids (23 transcripts) (Fig. 1C).





**Figure 1** The metabolic pathway analysis of transcripts from *Manis javanica*. (A) Metabolism, (B) Carbohydrate metabolism, (C) Lipid metabolism, (D) Amino acid metabolism, (E) Cofactors and vitamin metabolism and terpenoid and polyketide metabolism. The x-axis shows the numbers of annotated transcripts in one class, and the y-axis shows the KEGG function classes.

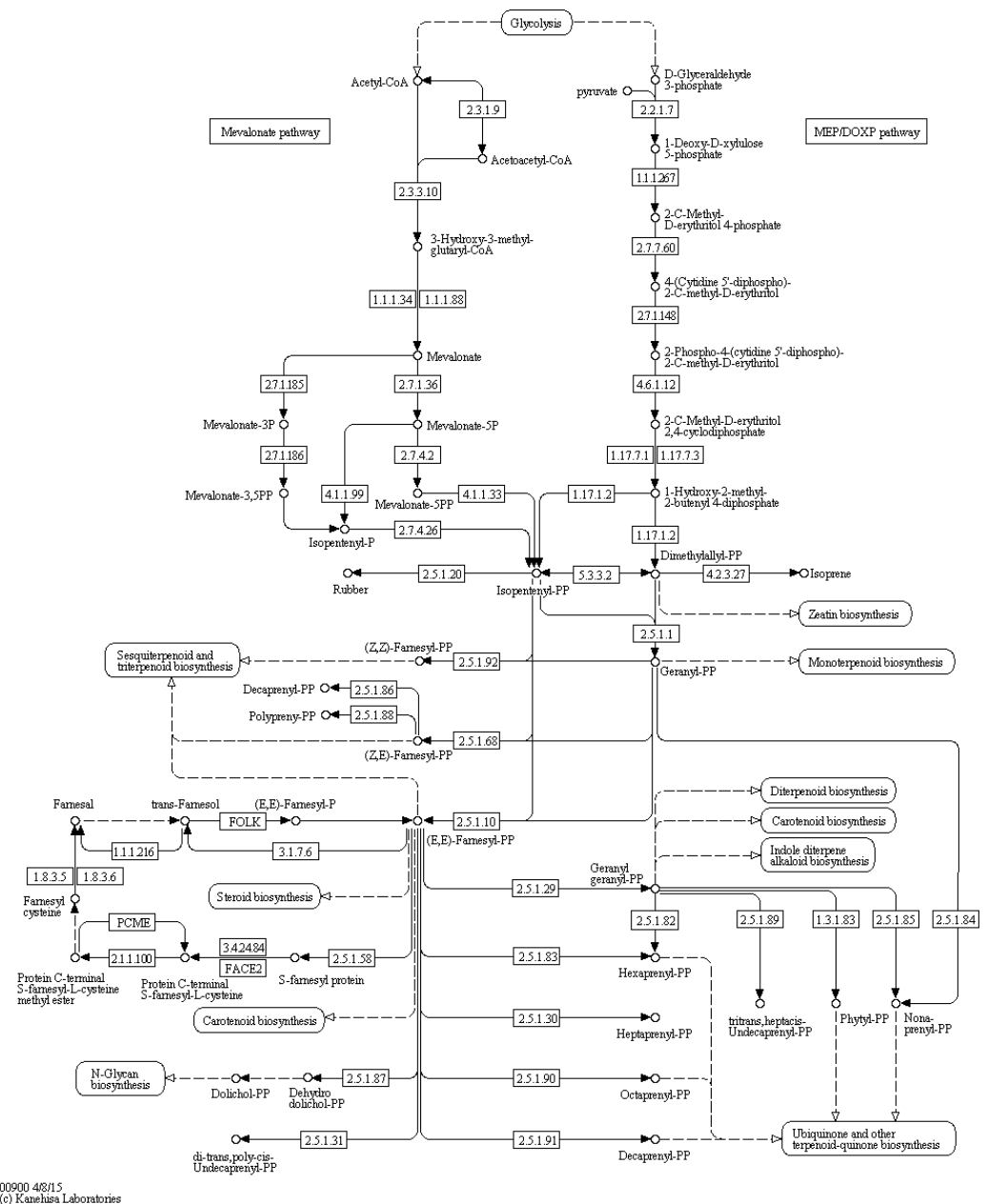
Full-size [DOI: 10.7717/peerj.4140/fig-1](https://doi.org/10.7717/peerj.4140/fig-1)

## Amino acid metabolism

Lysine degradation (66 transcripts), valine, leucine and isoleucine degradation (65 transcripts), arginine and proline metabolism (64 transcripts), and tryptophan metabolism (35 transcripts) were at the top of the amino acid metabolic lists. The biosynthesis pathways of some amino acids, such as phenylalanine, tyrosine, and tryptophan (6 transcripts); valine, leucine and isoleucine (5 transcripts); and lysine (2 transcripts) were at the bottom (Fig. 1D). None of the transcripts were found to be involved in arginine biosynthesis.

## Cofactors and vitamin metabolism and terpenoid and polyketide metabolism

Retinol metabolism (65 transcripts), porphyrin and chlorophyll metabolism (46 transcripts), nicotinate and nicotinamide metabolism (36 transcripts), and pantothenate and CoA biosynthesis (31 transcripts) were at the top of the cofactors and vitamin



**Figure 2** Terpenoid backbone biosynthesis (KEGG map 00900).

Full-size [DOI: 10.7717/peerj.4140/fig-2](https://doi.org/10.7717/peerj.4140/fig-2)

metabolism lists. There was only one list relating to terpenoids and polyketides; terpenoid backbone biosynthesis (27 transcripts) (Figs. 1E and 2).

### Annotation of the new transcripts

On the basis of the genome sequences of *M. javanica*, Cufflinks software was used to join the mapped reads, to compare them with the annotated information for the original genome, and to search for the gapped sequences, which were not annotated. A total of

**Table 3** Summary of genes related to the diet of the Malayan pangolin (*Manis javanica*).

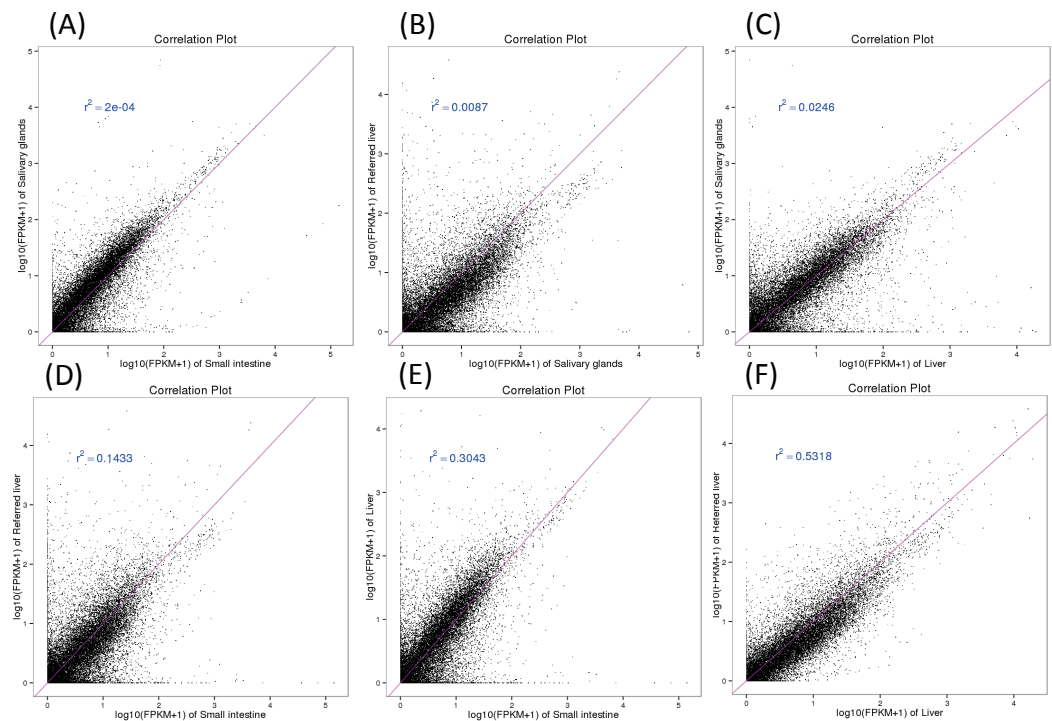
Type	Gene name
Opsin	GRK1, OPN1SW, OPN1LW, OPN4, PDE6D, PDE6G, PDE6H, RHO
Taste	TAS1R2, TAS1R3, TAS2R1, TAS2R4, TAS2R7, TAS2R10, TAS2R30, TAS2R38, TAS2R40
Olfactory	CNGA2, DTMT, OLF1, OLF2, OLF3, OLF4, OR1A1, OR1D2, OR1E1, OR1E2, OR1E5, OR1G1, OR3A1, OR3A2, OR3A3
Carbohydrases	AGL, AMY2, CHIA, CHI3L1, CHID1, GAA, GANAB, GANC, GBA3, GLB1, GLB1L, PRKCSH, SI
Lipases	ABHD6, ABHD12, CEL, CLPS, DDHD1, GPLD1, Group XV phospholipase A2, LIPA, LIPC, LIPE, LIPE, LIPH, LMF1, LMF2, LPL, LYPLAL1, NAPEPLD, PLA1A, PLA2G1B, PLA2G2A, PLA2G3, PLA2G4A, PLA2R1, PLB1, PLBD1, PLBD2, PLD3, PNLIP, PNLIPRP1, PNLIPRP2, PNPLA2, PNPLA8
Protease	Anionic trypsin, ANPEP, Cationic trypsin, CELA1, Chymotrypsin A chain C, CTB1, CTB2, CTB3, CTB4, CTB5, CTB6, CTB7, CTB8, CTB9, CTB10, CTB11, CTB12, CTB13, CTB14, CTB15, CTB16, CTB17, CTB18, CTB19, CTB20, CTB21, CTB22, CTB23, CTB24, CTB25, CTB26, CTB27, CTB28, CTB29, CTB30, CTB31, CTB32, CTB33, CTB34, CTB35, CTB36, CTB37, CTB38, CTB39, CTB40, CTB41, CTB42, CTB43, CTB44, CTB45, CTB46, CTB47, CTB48, CTB49, CTB50, CTB51, CTB52, CTB53, CTB54, CTB55, CTB56, CTB57, CTB58, CTB59, CTB60, CTB61, CTB62, CTB63, CTB64, CTB65, CTB66, CTB67, CTB68, CTB69, CTB70, CTB71, CTB72, CTB73, CTB74, CTB75, CTB76, CTB77, CTB78, CTB79, CTB80, CTB81, CTB82, CTB83, CTB84, CTB85, CTB86, CTB87, CTB88, CTB89, CTB90, CTB91, CTB92, CTB93, CTB94, CTB95, CTB96, CTB97, CTB98, CTB99, CTB100, CTB101, CTB102, CTB103, CTB104, CTB105, CTB106, CTB107, CTB108, CTB109, CTB110, CTB111, CTB112, CTB113, CTB114, CTB115, CTB116, CTB117, CTB118, CTB119, CTB120, CTB121, CTB122, CTB123, CTB124, CTB125, CTB126, CTB127, CTB128, CTB129, CTB130, CTB131, CTB132, CTB133, CTB134, CTB135, CTB136, CTB137, CTB138, CTB139, CTB140, CTB141, CTB142, CTB143, CTB144, CTB145, CTB146, CTB147, CTB148, CTB149, CTB150, CTB151, CTB152, CTB153, CTB154, CTB155, CTB156, CTB157, CTB158, CTB159, CTB160, CTB161, CTB162, CTB163, CTB164, CTB165, CTB166, CTB167, CTB168, CTB169, CTB170, CTB171, CTB172, CTB173, CTB174, CTB175, CTB176, CTB177, CTB178, CTB179, CTB180, CTB181, CTB182, CTB183, CTB184, CTB185, CTB186, CTB187, CTB188, CTB189, CTB190, CTB191, CTB192, CTB193, CTB194, CTB195, CTB196, CTB197, CTB198, CTB199, CTB200, CTB201, CTB202, CTB203, CTB204, CTB205, CTB206, CTB207, CTB208, CTB209, CTB210, CTB211, CTB212, CTB213, CTB214, CTB215, CTB216, CTB217, CTB218, CTB219, CTB220, CTB221, CTB222, CTB223, CTB224, CTB225, CTB226, CTB227, CTB228, CTB229, CTB230, CTB231, CTB232, CTB233, CTB234, CTB235, CTB236, CTB237, CTB238, CTB239, CTB240, CTB241, CTB242, CTB243, CTB244, CTB245, CTB246, CTB247, CTB248, CTB249, CTB250, CTB251, CTB252, CTB253, CTB254, CTB255, CTB256, CTB257, CTB258, CTB259, CTB260, CTB261, CTB262, CTB263, CTB264, CTB265, CTB266, CTB267, CTB268, CTB269, CTB270, CTB271, CTB272, CTB273, CTB274, CTB275, CTB276, CTB277, CTB278, CTB279, CTB280, CTB281, CTB282, CTB283, CTB284, CTB285, CTB286, CTB287, CTB288, CTB289, CTB290, CTB291, CTB292, CTB293, CTB294, CTB295, CTB296, CTB297, CTB298, CTB299, CTB300, CTB301, CTB302, CTB303, CTB304, CTB305, CTB306, CTB307, CTB308, CTB309, CTB310, CTB311, CTB312, CTB313, CTB314, CTB315, CTB316, CTB317, CTB318, CTB319, CTB320, CTB321, CTB322, CTB323, CTB324, CTB325, CTB326, CTB327, CTB328, CTB329, CTB330, CTB331, CTB332, CTB333, CTB334, CTB335, CTB336, CTB337, CTB338, CTB339, CTB340, CTB341, CTB342, CTB343, CTB344, CTB345, CTB346, CTB347, CTB348, CTB349, CTB350, CTB351, CTB352, CTB353, CTB354, CTB355, CTB356, CTB357, CTB358, CTB359, CTB360, CTB361, CTB362, CTB363, CTB364, CTB365, CTB366, CTB367, CTB368, CTB369, CTB370, CTB371, CTB372, CTB373, CTB374, CTB375, CTB376, CTB377, CTB378, CTB379, CTB380, CTB381, CTB382, CTB383, CTB384, CTB385, CTB386, CTB387, CTB388, CTB389, CTB390, CTB391, CTB392, CTB393, CTB394, CTB395, CTB396, CTB397, CTB398, CTB399, CTB400, CTB401, CTB402, CTB403, CTB404, CTB405, CTB406, CTB407, CTB408, CTB409, CTB410, CTB411, CTB412, CTB413, CTB414, CTB415, CTB416, CTB417, CTB418, CTB419, CTB420, CTB421, CTB422, CTB423, CTB424, CTB425, CTB426, CTB427, CTB428, CTB429, CTB430, CTB431, CTB432, CTB433, CTB434, CTB435, CTB436, CTB437, CTB438, CTB439, CTB440, CTB441, CTB442, CTB443, CTB444, CTB445, CTB446, CTB447, CTB448, CTB449, CTB450, CTB451, CTB452, CTB453, CTB454, CTB455, CTB456, CTB457, CTB458, CTB459, CTB460, CTB461, CTB462, CTB463, CTB464, CTB465, CTB466, CTB467, CTB468, CTB469, CTB470, CTB471, CTB472, CTB473, CTB474, CTB475, CTB476, CTB477, CTB478, CTB479, CTB480, CTB481, CTB482, CTB483, CTB484, CTB485, CTB486, CTB487, CTB488, CTB489, CTB490, CTB491, CTB492, CTB493, CTB494, CTB495, CTB496, CTB497, CTB498, CTB499, CTB500, CTB501, CTB502, CTB503, CTB504, CTB505, CTB506, CTB507, CTB508, CTB509, CTB510, CTB511, CTB512, CTB513, CTB514, CTB515, CTB516, CTB517, CTB518, CTB519, CTB520, CTB521, CTB522, CTB523, CTB524, CTB525, CTB526, CTB527, CTB528, CTB529, CTB530, CTB531, CTB532, CTB533, CTB534, CTB535, CTB536, CTB537, CTB538, CTB539, CTB540, CTB541, CTB542, CTB543, CTB544, CTB545, CTB546, CTB547, CTB548, CTB549, CTB550, CTB551, CTB552, CTB553, CTB554, CTB555, CTB556, CTB557, CTB558, CTB559, CTB560, CTB561, CTB562, CTB563, CTB564, CTB565, CTB566, CTB567, CTB568, CTB569, CTB570, CTB571, CTB572, CTB573, CTB574, CTB575, CTB576, CTB577, CTB578, CTB579, CTB580, CTB581, CTB582, CTB583, CTB584, CTB585, CTB586, CTB587, CTB588, CTB589, CTB590, CTB591, CTB592, CTB593, CTB594, CTB595, CTB596, CTB597, CTB598, CTB599, CTB600, CTB601, CTB602, CTB603, CTB604, CTB605, CTB606, CTB607, CTB608, CTB609, CTB610, CTB611, CTB612, CTB613, CTB614, CTB615, CTB616, CTB617, CTB618, CTB619, CTB620, CTB621, CTB622, CTB623, CTB624, CTB625, CTB626, CTB627, CTB628, CTB629, CTB630, CTB631, CTB632, CTB633, CTB634, CTB635, CTB636, CTB637, CTB638, CTB639, CTB640, CTB641, CTB642, CTB643, CTB644, CTB645, CTB646, CTB647, CTB648, CTB649, CTB650, CTB651, CTB652, CTB653, CTB654, CTB655, CTB656, CTB657, CTB658, CTB659, CTB660, CTB661, CTB662, CTB663, CTB664, CTB665, CTB666, CTB667, CTB668, CTB669, CTB670, CTB671, CTB672, CTB673, CTB674, CTB675, CTB676, CTB677, CTB678, CTB679, CTB680, CTB681, CTB682, CTB683, CTB684, CTB685, CTB686, CTB687, CTB688, CTB689, CTB690, CTB691, CTB692, CTB693, CTB694, CTB695, CTB696, CTB697, CTB698, CTB699, CTB700, CTB701, CTB702, CTB703, CTB704, CTB705, CTB706, CTB707, CTB708, CTB709, CTB710, CTB711, CTB712, CTB713, CTB714, CTB715, CTB716, CTB717, CTB718, CTB719, CTB720, CTB721, CTB722, CTB723, CTB724, CTB725, CTB726, CTB727, CTB728, CTB729, CTB730, CTB731, CTB732, CTB733, CTB734, CTB735, CTB736, CTB737, CTB738, CTB739, CTB740, CTB741, CTB742, CTB743, CTB744, CTB745, CTB746, CTB747, CTB748, CTB749, CTB750, CTB751, CTB752, CTB753, CTB754, CTB755, CTB756, CTB757, CTB758, CTB759, CTB760, CTB761, CTB762, CTB763, CTB764, CTB765, CTB766, CTB767, CTB768, CTB769, CTB770, CTB771, CTB772, CTB773, CTB774, CTB775, CTB776, CTB777, CTB778, CTB779, CTB780, CTB781, CTB782, CTB783, CTB784, CTB785, CTB786, CTB787, CTB788, CTB789, CTB790, CTB791, CTB792, CTB793, CTB794, CTB795, CTB796, CTB797, CTB798, CTB799, CTB800, CTB801, CTB802, CTB803, CTB804, CTB805, CTB806, CTB807, CTB808, CTB809, CTB810, CTB811, CTB812, CTB813, CTB814, CTB815, CTB816, CTB817, CTB818, CTB819, CTB820, CTB821, CTB822, CTB823, CTB824, CTB825, CTB826, CTB827, CTB828, CTB829, CTB830, CTB831, CTB832, CTB833, CTB834, CTB835, CTB836, CTB837, CTB838, CTB839, CTB840, CTB841, CTB842, CTB843, CTB844, CTB845, CTB846, CTB847, CTB848, CTB849, CTB850, CTB851, CTB852, CTB853, CTB854, CTB855, CTB856, CTB857, CTB858, CTB859, CTB860, CTB861, CTB862, CTB863, CTB864, CTB865, CTB866, CTB867, CTB868, CTB869, CTB870, CTB871, CTB872, CTB873, CTB874, CTB875, CTB876, CTB877, CTB878, CTB879, CTB880, CTB881, CTB882, CTB883, CTB884, CTB885, CTB886, CTB887, CTB888, CTB889, CTB890, CTB891, CTB892, CTB893, CTB894, CTB895, CTB896, CTB897, CTB898, CTB899, CTB900, CTB901, CTB902, CTB903, CTB904, CTB905, CTB906, CTB907, CTB908, CTB909, CTB910, CTB911, CTB912, CTB913, CTB914, CTB915, CTB916, CTB917, CTB918, CTB919, CTB920, CTB921, CTB922, CTB923, CTB924, CTB925, CTB926, CTB927, CTB928, CTB929, CTB930, CTB931, CTB932, CTB933, CTB934, CTB935, CTB936, CTB937, CTB938, CTB939, CTB940, CTB941, CTB942, CTB943, CTB944, CTB945, CTB946, CTB947, CTB948, CTB949, CTB950, CTB951, CTB952, CTB953, CTB954, CTB955, CTB956, CTB957, CTB958, CTB959, CTB960, CTB961, CTB962, CTB963, CTB964, CTB965, CTB966, CTB967, CTB968, CTB969, CTB970, CTB971, CTB972, CTB973, CTB974, CTB975, CTB976, CTB977, CTB978, CTB979, CTB980, CTB981, CTB982, CTB983, CTB984, CTB985, CTB986, CTB987, CTB988, CTB989, CTB990, CTB991, CTB992, CTB993, CTB994, CTB995, CTB996, CTB997
Transporters	SLC1A1, SLC1A3, SLC1A6, SLC1A4, SLC1A5, SLC7A8, SLC43A2, SLC6A15, SLC6A17, SLC6A19, SLC38A1, SLC38A2, SLC38A4, SLC38A5, SLC38A7, SLC38A10, SLC38A11, SLC7A2, SLC7A14, SLC7A11, SLC25A29, SLC2A1, SLC2A2, SLC2A3, SLC2A4, SLC2A5, SLC2A8, SLC2A9, SLC2A12, SLC35A4, SLC35A5, SLC50A1, SLC35A3, SLC35B4, SLC35D2, CLCN3, CLCN5, CLCN7, MFSD5, MAGT1, MGMT1, MRS2, NIPA2, NIPAL1, Sodium-independent sulfate anion transporter, SLC4A4, SLC20A1, SLC20A2, SLCO1C1, SLCO3A1, SLCO4C1, LMBRD1, SLC5A6, SLC19A3, SLC25A32, SLC52A2, SLC52A3, SLC5A1, SLC5A4, SLC5A10, SLC5A2, SLC28A1, SLC5A3, APOA1, APOA2, APOB, Apolipoprotein A-IV, APOC2, APOC3, APOC4, APOD, APOE, APOM, APOO, SLC6A2, SLC6A3, SLC6A4, SLC6A8, SLC6A9, SLC6A12, SLC6A13, SLC10A2, SLC5A12, SLC16A1, SLC16A9, SLC16A13, SLC17A6, SLC17A7, SLC26A2, SLC29A3, SLC44A2, SLC44A3, SLC44A4, SLC44A5, SLC45A2, SLC46A2

3,373 new transcripts were discovered (File S5), of which 1,459 transcripts were annotated on the basis of the COG, GO, KEGG, KOG, Swiss-Prot, TrEMBL and NR databases using BLAST (Table S2 and File S6).

In the Gene Ontology analysis, 75 new transcripts were involved in 22 metabolic categories. Some of the new genes were involved in inositol metabolism (GO: 0006020; e.g., *Manis javanica\_newGene\_958*), and some were involved in the linoleic acid metabolism (GO: 0043651; e.g., *Manis javanica\_newGene\_12722*). For KEGG, 114 new genes were related to metabolic function, including lipid metabolism (41 transcripts), carbohydrate metabolism (24 transcripts), cofactor and vitamin metabolism (22 transcripts), and amino acid metabolism (20 transcripts) (Fig. S7).

### Gene expression repertoire

Distributions of potential transcripts related to feeding among the three tissue libraries are shown in Table 3, Table S3 and File S7, including the 997 transcripts related to sensory perception. Of these transcripts, 972 were related to olfaction, and 11 and 14 transcripts were related to vision and taste, respectively. A total of 133 transcripts were related to digestive enzyme gene families; 70 of these transcripts were related to lipid degradation,



**Figure 3** Pairwise correlation between organs (A–F).

Full-size  DOI: [10.7717/peerj.4140/fig-3](https://doi.org/10.7717/peerj.4140/fig-3)

and 39 and 20 were related to the degradation of proteins and carbohydrates, respectively. These genes were considered to be involved in the profile of food choice, digestion and absorption, which might serve as a molecular mechanism in myrmecophagy. Among these transcripts, acidic mammalian chitinase (*CHIA*), chitinase-3-like protein 1 (*CHI3L1*), and chitinase domain-containing protein 1 (*CHID1*) were related to chitin degradation. A total of 199 transcripts were related to molecular transporters, including sugar transporters, amino acid transporters, apolipoprotein transporters, cationic/anion transporters, vitamin transporters, cotransporters and others; among these, the UDP-N-acetylglucosamine transporter (*SLC35A3*, *SLC35B4*, and *SLC35D2*) was related to the decomposer absorption of the chitin unit N-acetylglucosamine.

### Pairwise comparisons of different transcriptomic profiles

To examine the similarity among organ transcriptomes, we performed statistical correlation analysis for each pair of organs using  $\log_{10}(\text{FPKM} + 1)$  transformation to normalize the plots (Fig. 3). Our data showed that two liver transcriptome expression profiles were the most similar ( $R^2 = 0.53$ ), followed by those of the liver and small intestine ( $R^2 = 0.30$ ). The salivary glands and small intestine had the least similar transcriptomic profiles ( $R^2 = 2e^{-0.4}$ ). The low correlation between the liver transcriptomes found in our study and that of *Yusoff et al. (2016)* may reflect the varying complexity between the same organs in different individuals, possibly because one of the two specimens was pregnant. The

differences between each pair of compared organs reflect the different digestive functions of the three organs.

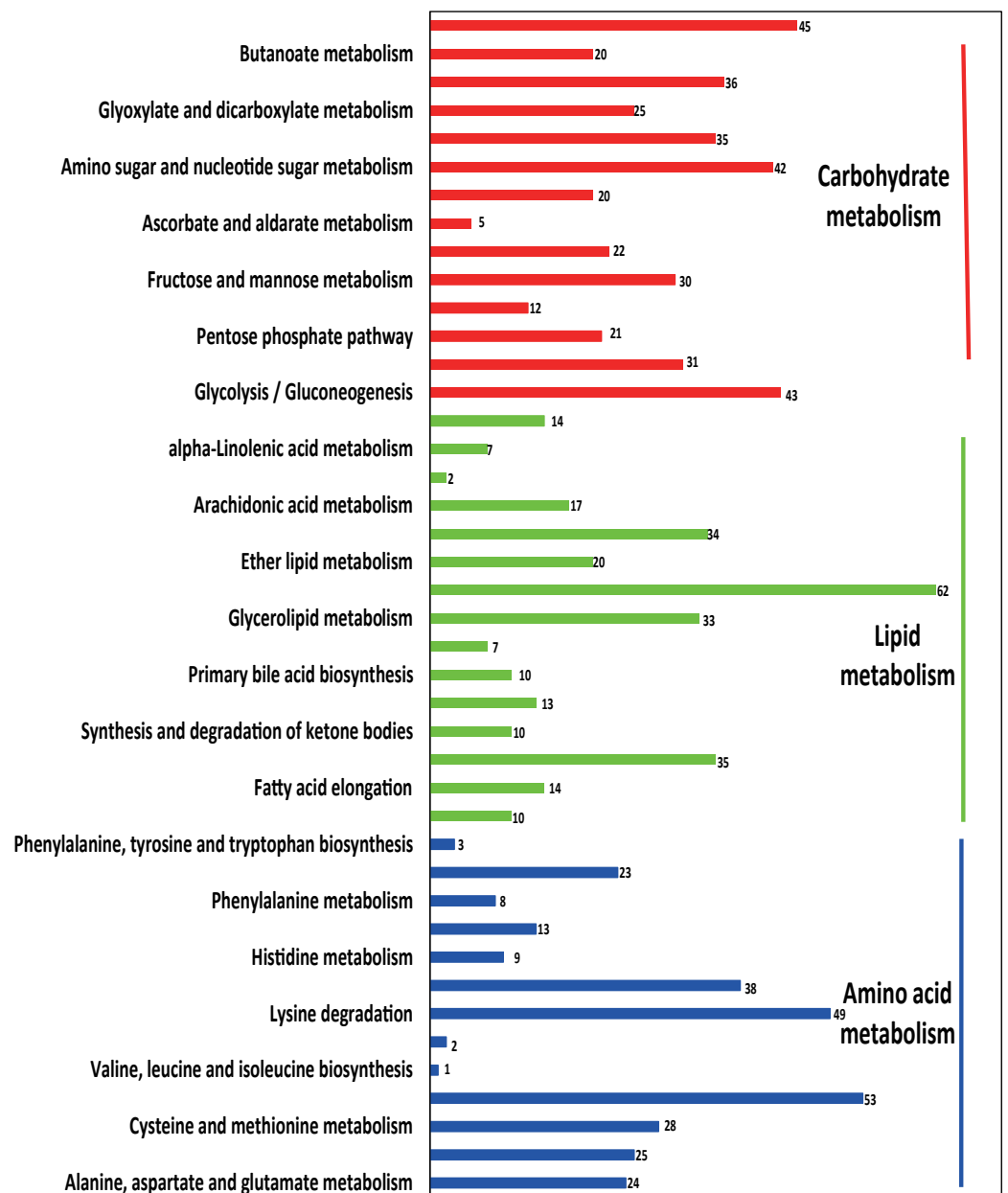
A total of 11,055 transcripts were expressed (FPKM >1.0) in all three tissues; of these, 3,947 were annotated in KEGG. Several transcripts were enriched in lipid metabolism (214 transcripts), carbohydrate metabolism (240 transcripts), and amino acid metabolism (188 transcripts) (Fig. 4). Other transcripts near the top of the metabolic lists included glycerophospholipid metabolism (62 transcripts); valine, leucine and isoleucine degradation (53 transcripts); lysine degradation (49 transcripts); and inositol phosphate metabolism (45 transcripts). The biosynthesis pathways of valine, leucine, and isoleucine (one transcript) and lysine (two transcripts) were less commonly represented.

Differentially expressed genes were identified among systems in the metabolic pathways of sugars, lipids, and amino acids. A total of 382 transcripts were differentially expressed between the small intestine and liver; this value compares with 258 transcripts that were differentially expressed between the salivary glands and small intestine. Twenty-three and 27 genes differed between the small intestine and liver for starch and sucrose metabolism and arginine and proline metabolism, respectively, and 31 genes for steroid hormone biosynthesis differed between the liver and the referred liver from (Yusoff *et al.*, 2016) (Fig. S8).

Several transcripts were specifically expressed in a single sample, including 21 transcripts in the small intestine that were involved in 19 pathways, five transcripts in the liver that were involved in 11 pathways, 36 transcripts in the referred liver that were involved in 27 pathways, and six transcripts in the salivary glands that were involved in 10 pathways (Table S4). The highest number of specific transcripts was eight, and these transcripts were involved in the arachidonic acid metabolism in the liver reported in the previous study (Yusoff *et al.*, 2016); six transcripts were involved in glycerophospholipid metabolism, in either lipid metabolism or arachidonic acid metabolism of the small intestine (Fig. 5).

## DISCUSSION AND CONCLUSION

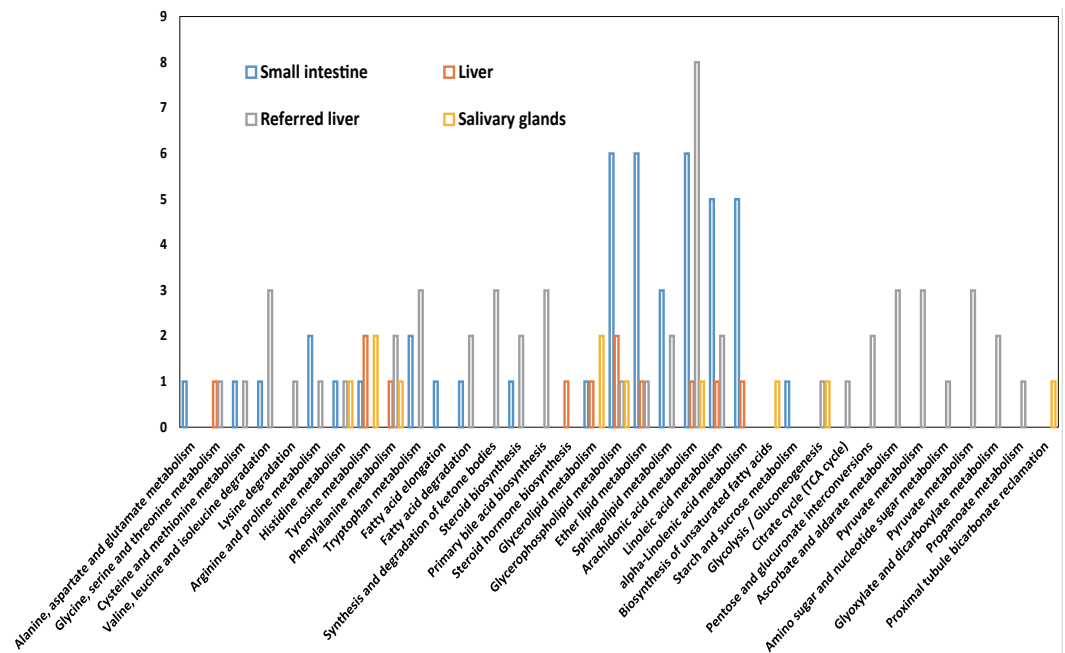
Based on these results, a total of 3,373 new genes were discovered in the transcriptomic datasets, of which 1,459 were annotated, and 75 new genes were involved in metabolism. The new genes provide new information for studying the myrmecophagous mechanisms of pangolins. A large number of genes were expressed in all three tissues, and several specific genes in the three systems played different roles in the metabolism of sugars, lipids, and amino acids. The digestive functions of the small intestine and salivary glands were similar in comparison with the differences between the small intestine and liver. The functions of the livers from the two different individuals differed in lipid metabolism pathways, suggesting that the ratio of lipid in the feed might be changed appropriately during pregnancy. Pangolin scales are considered an effective Chinese medicine and are thought to promote blood circulation, accelerate milk secretion, detumescence and apocenos (Hua *et al.*, 2015). The scales are helpful for the treatment of such human diseases. Interestingly, our results indicated that several KEGG pathways were related to human diseases, possibly indicating the mechanisms by which pangolin scales act on human diseases. However, further data are needed to elucidate the underlying mechanisms.



**Figure 4** The transcripts related to metabolic pathways expressed in all three tissues. The  $x$ -axis shows the number of transcripts with the KEGG function class, which was shown above the column, and the  $y$ -axis shows the KEGG function classes.

Full-size DOI: 10.7717/peerj.4140/fig-4

Ants are a protein-rich food (Tomotake, Katagiri & Yamato, 2010). They contain more than 50% crude protein, according to a nutritional value evaluation, and contain more than 20 amino acids; various microelements; special chemicals, such as formic acid and herbaceous acetaldehyde, which are triterpenoid compounds, and several vitamins (Pattarayingsakul et al., 2017). Many pangolin genes are likely to be involved



**Figure 5** The numbers of transcripts related to metabolic pathways that are specifically expressed in the salivary glands, liver, and small intestine of adult female Malayan pangolin (*Manis javanica*). The x-axis shows the KEGG function classes, and the y-axis shows the number of transcripts with the corresponding KEGG function class.

Full-size DOI: 10.7717/peerj.4140/fig-5

in the digestion of these materials because 27 of the transcripts related to the terpenoid backbone biosynthesis, and this might be one of the biological basis for their adaptations to myrmecophagy. However, no genes were found to participate in the arginine synthesis pathway according to the KEGG analysis, and only two transcripts were involved in lysine synthesis. We did find a few transcripts encoding proteins that are necessary for the biosynthesis of amino acids that are essential in humans, such as lysine, valine, leucine, isoleucine, phenylalanine, and tryptophan (Galili, Amir & Fernie, 2016; Zhenyukh et al., 2017). These results suggest that *M. javanica* might have at least a limited ability to synthesize those amino acids and that arginine might be the first limiting amino acid. Therefore, it might not be necessary to add the amino acids that pangolins have the potential to synthesis to manufactured feed.. However, the regulation of the relevant transcripts is not fully characterized and the ability of pangolins to flourish without those amino acids is unclear. This issue needs further research, which might be very useful for theoretical and practical purposes.

In our *M. javanica* transcriptomic analysis, the same enriched GO terms were found to be enriched in biological processes in response to a stimulus as those found by Yusoff et al. (2016), indicating that the *M. javanica* transcripts may be actively involved in stimulus or stress responses. However, several GO terms involving biological processes were found to be enriched in our study, including metabolic and developmental processes that might be involved in food digestion and absorption. Regarding molecular functions,

receptor activity was similarly well enriched. In the previous study, the other main GO term found was cytoskeletal protein binding, and this likely required the evolution of a sophisticated musculoskeletal system or the formation of pangolin scales, which are made of keratin (Yusoff *et al.*, 2016). In our study, some enriched GO terms were also related to metabolism, including catalytic activity, molecular transducer activity and transporter activity. The observed differences might be related to the different tissues used to obtain the transcriptome data.

Chitin, one of the main components of the epidermis of ants and termites, comprises of N-acetyl-D-glucosamines (GluNAc) connected by a  $\beta$ -1, 4 glycosidic bond. Chitin can be digested only with chitinase and acidic mammalian chitinase (*AMCase*); *AMCase* is widely found in the digestive organs of animals (Eurich *et al.*, 2009; Krykbaev *et al.*, 2010; Strobel *et al.*, 2013). The origin of the digestive processes might be closely related to the activity of *AMCase*, which determines the beginning of chitin. The transporter genes of UDP-N-acetyl glucosamine (*SLC35A3*, *SLC35B4*, and *SLC35D2*) might be directly related to the absorption of carbohydrate units during chitin metabolism (Gerardy-Schahn, Oelmann & Bakker, 2001). Therefore, chitin could be added to the formulated diets to aid in the digestion and absorption of nutrients. This suggestion is consistent with the referenced formulas mentioned in Vijayan, Yeong & Ling (2009).

Molecular adaptations consistent with the diet have also been shown in other studies. The tiger genome is particularly enriched in olfactory receptor activity, the G-protein-coupled receptor signaling pathway, signal transducer activity, amino-acid transport and protein metabolism. These markers of amino-acid metabolism have been associated with an obligatory carnivorous diet (Cho *et al.*, 2013). We also found highly enriched transcripts for olfactory receptor activity in *M. javanica*. In red panda, regarding gene ontology and KEGG enriched analyses, significant terms and pathways were involved in limb development and nutrient utilization, including appendage and limb development, cilium assembly, protein digestion and absorption, and retinol metabolism; these pathways might be related to their bamboo diet (Hu *et al.*, 2017). In the panda, loss-of-function is observed for *TAS1R1*, and this might prevent the panda from expressing a functional umami taste receptor, thus partly explaining why the panda diet is primarily herbivorous despite its taxonomic classification as a carnivore (Li *et al.*, 2010). The polar bear is adapted to cope with a diet rich in fatty acids. Gene Ontology analysis for putative genes under positive selection in the polar bear lineage has shown that genes associated with adipose tissue development are enriched, reflecting the crucial role that lipids play in the ecology and life history of polar bears (Liu *et al.*, 2014). However, transcripts involved in lipid metabolism were enriched in *M. javanica*, a finding that is generally consistent with the Gene Ontology analysis of the polar bear genome, while the markers of amino-acid metabolism are apparently associated with the diet in pandas and tigers.

Malayan pangolins are endangered mammals for which captive breeding provides an opportunity to study the molecular mechanisms of myrmecophagy. Our *M. javanica* transcriptomic datasets of the three representative tissues present the first attempt at uncovering the mysteries of the genetic mechanisms of digestion and reproduction in this rare and unique mammal. Here, the transcriptome sequencing of digestive organs



was performed to observe the metabolic pathways and functional genes related to myrmecophagy in an attempt to understand the molecular mechanisms involved. The results may provide an important theoretical basis for the successful captive breeding of this species. The transcriptomic data for the three organs were obtained with a high degree of confidence, and the transcripts were well annotated, thus providing a genomic and molecular basis for future study of this lesser-known mammalian species. However, the experimental design was constrained by the availability of samples, thus, further studies with more biological replicates are warranted. Other major organs involved in digestion (such as tongue, stomach, and pancreas) should be examined in further study, and these pangolin-specific genes may also be required to understand the unique traits or adaptations of pangolins (as compared to other mammals). Together, these results might shed light on the special diet of *M. javanica*.

## ACKNOWLEDGEMENTS

We acknowledge BMKCloud (Beijing, China) for the data analysis. We acknowledge the Dongguan Institute of Qingfengyuan Animal Medicinal (Dongguan, Guangdong, China) for their selfless offering of three female pangolin samples.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

This project was supported by the Planning Funds of Scientific and Technological of Guangdong Province (2016B070701016 and 2017A050503005) and the Training Fund of Guangdong Institute of Applied Biological Resources for Postdoctoral Researchers (NO.GIABR-pyjj201601), the Funds for Environment Construction and Capacity Building of GDAS' Research Platform (2016GDASPT-0107), GDAS Special Project of Science and Technology Development (2017GDASCX-0107), and the Natural Science Foundation of Guangdong Province, China (2017A030310425). There was no additional external funding received for this study. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Grant Disclosures

The following grant information was disclosed by the authors:

Planning Funds of Scientific and Technological of Guangdong Province: 2016B070701016, 2017A050503005.

Training Fund of Guangdong Institute of Applied Biological Resources for Postdoctoral Researchers: NO.GIABR-pyjj201601.

Funds for Environment Construction and Capacity Building of GDAS' Research Platform: 2016GDASPT-0107.

GDAS Special Project of Science and Technology Development: 2017GDASCX-0107.

Natural Science Foundation of Guangdong Province, China: 2017A030310425.

## Competing Interests

The authors declare there are no competing interests.

## Author Contributions

- Jing-E Ma conceived and designed the experiments, performed the experiments, wrote the paper, prepared figures and/or tables.
- Lin-Miao Li and Guan-Yu Li performed the experiments.
- Hai-Ying Jiang and Juan Li analyzed the data.
- Xiu-Juan Zhang contributed reagents/materials/analysis tools.
- Li-Hong Yuan contributed reagents/materials/analysis tools, reviewed drafts of the paper.
- Jun Wu reviewed drafts of the paper.
- Jin-Ping Chen conceived and designed the experiments, wrote the paper, reviewed drafts of the paper.

## Animal Ethics

The following information was supplied relating to ethical approvals (i.e., approving body and any reference numbers):

Guangdong Institute of Applied Biological Resources provided full approval for this study (reference number: G2ABR20170523).

## Data Availability

The following information was supplied regarding data availability:

The data are available on the NCBI Gene Expression Omnibus (GEO) database with the accession: [GSM2667949](#), [GSM2667950](#) and [GSM2667951](#).

## Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.4140#supplemental-information>.

## REFERENCES

- Ashokkumar M, Valsarajan D, Suresh MA, Kaimal AR, Chandy G. 2017. Stomach contents of the Indian Pangolin *Manis crassicaudata* (Mammalia: Pholidota: Manidae) in tropical forests of southern India. *Journal of Threatened Taxa* **9**(5):10246–10248.
- Challender D, Nguyen Van T, Shepherd C, Krishnasamy K, Wang A, Lee B, Panjang E, Fletcher L, Heng S, Seah Han Ming J, Olsson A, Nguyen The Truong A, Nguyen Van QC. 2014. *Manis javanica*. The IUCN Red List of Threatened Species 2014. Version 2017-1. Available at <http://www.iucnredlist.org/details/12763/0> (accessed on 15 August 2017).
- Cho YS, Hu L, Hou H, Lee H, Xu J, Kwon S, Oh S, Kim HM, Jho S, Kim S, Shin YA, Kim BC, Kim H, Kim CU, Luo SJ, Johnson WE, Koepfli KP, Schmidt-Kuntzel A, Turner JA, Marker L, Harper C, Miller SM, Jacobs W, Bertola LD, Kim TH, Lee S, Zhou Q, Jung HJ, Xu X, Gadhvi P, Xu P, Xiong Y, Luo Y, Pan S, Gou C, Chu

- X, Zhang J, Liu S, He J, Chen Y, Yang L, Yang Y, He J, Liu S, Wang J, Kim CH, Kwak H, Kim JS, Hwang S, Ko J, Kim CB, Kim S, Bayarlkhagva D, Paek WK, Kim SJ, O'Brien S, Wang J, Bhak J. 2013. The tiger genome and comparative analysis with lion and snow leopard genomes. *Nature Communications* 4:Article 2433 DOI 10.1038/ncomms3433.
- Choo SW, Rayko M, Tan TK, Hari R, Komissarov A, Wee WY, Yurchenko AA, Kliver S, Tamazian G, Antunes A, Wilson RK, Warren WC, Koepfli KP, Minx P, Krasheninikova K, Kotze A, Dalton DL, Vermaak E, Paterson IC, Dobrynin P, Sitam FT, Rovie-Ryan JJ, Johnson WE, Yusoff AM, Luo SJ, Karuppannan KV, Fang G, Zheng D, Gerstein MB, Lipovich L, O'Brien SJ, Wong GJ. 2016. Pangolin genomes and the evolution of mammalian scales and immunity. *Genome Research* 26:1312–1322 DOI 10.1101/gr.203521.115.
- Du Toit Z, Grobler JP, Kotzé A, Jansen R, Brettschneider H, Dalton DL. 2014. The complete mitochondrial genome of Temminck's ground pangolin (*Smutsia temminckii*; Smuts, 1832) and phylogenetic position of the Pholidota (Weber, 1904). *Gene* 551(1):49–54 DOI 10.1016/j.gene.2014.08.040.
- Eurich K, Segawa M, Toei-Shimizu S, Mizoguchi E. 2009. Potential role of chitinase 3-like-1 in inflammation-associated carcinogenic changes of epithelial cells. *World Journal of Gastroenterology* 15:5249–5259.
- Galili G, Amir R, Fernie AR. 2016. The regulation of essential amino acid synthesis and accumulation in plants. *Annual Review of Plant Biology* 67:153–178 DOI 10.1146/annurev-arplant-043015-112213.
- Gaudin TJ, Emry RJ, Wible JR. 2009. The phylogeny of living and extinct pangolins (Mammalia, Pholidota) and associated taxa: a morphology based analysis. *Journal of Mammalian Evolution* 16:235–305 DOI 10.1007/s10914-009-9119-9.
- Gerardy-Schahn R, Oelmann S, Bakker H. 2001. Nucleotide sugar transporters: biological and functional aspects. *Biochimie* 83:775–782 DOI 10.1016/S0300-9084(01)01322-0.
- Hu Y, Wu Q, Ma S, Ma T, Shan L, Wang X, Nie Y, Ning Z, Yan L, Xiu Y, Wei F. 2017. Comparative genomics reveals convergent evolution between the bamboo-eating giant and red pandas. *Proceedings of the National Academy of Sciences of the United States of America* 114:1081–1086 DOI 10.1073/pnas.1613870114.
- Hua L, Gong S, Wang F, Li W, Ge Y, Li X, Hou F. 2015. Captive breeding of pangolins: current status, problems and future prospects. *ZooKeys* 507:99–114 DOI 10.3897/zookeys.507.6970.
- Kawashima S, Ikehata H, Tada C, Ogino T, Kakizaki H, Ikeda M, Fukushima H, Matsumiya M. 2016. Stomach chitinase from Japanese Sardine *Sardinops melanostictus*: purification, characterization, and molecular cloning of chitinase isozymes with a long linker. *Marine Drugs* 14:Article 22 DOI 10.3390/md14010022.
- Krykbaev R, Fitz LJ, Reddy PS, Winkler A, Xuan D, Yang XK, Fleming M, Wolf SF. 2010. Evolutionary and biochemical differences between human and monkey acidic mammalian chitinases. *Gene* 452:63–71 DOI 10.1016/j.gene.2009.12.005.

- Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323 DOI 10.1186/1471-2105-12-323.
- Li R, Fan W, Tian G, Zhu H, He L, Cai J, Huang Q, Cai Q, Li B, Bai Y, Zhang Z, Zhang Y, Wang W, Li J, Wei F, Li H, Jian M, Li J, Zhang Z, Nielsen R, Li D, Gu W, Yang Z, Xuan Z, Ryder OA, Leung FC, Zhou Y, Cao J, Sun X, Fu Y, Fang X, Guo X, Wang B, Hou R, Shen F, Mu B, Ni P, Lin R, Qian W, Wang G, Yu C, Nie W, Wang J, Wu Z, Liang H, Min J, Wu Q, Cheng S, Ruan J, Wang M, Shi Z, Wen M, Liu B, Ren X, Zheng H, Dong D, Cook K, Shan G, Zhang H, Kosiol C, Xie X, Lu Z, Zheng H, Li Y, Steiner CC, Lam TT, Lin S, Zhang Q, Li G, Tian J, Gong T, Liu H, Zhang D, Fang L, Ye C, Zhang J, Hu W, Xu A, Ren Y, Zhang G, Bruford MW, Li Q, Ma L, Guo Y, An N, Hu Y, Zheng Y, Shi Y, Li Z, Liu Q, Chen Y, Zhao J, Qu N, Zhao S, Tian F, Wang X, Wang H, Xu L, Liu X, Vinar T, Wang Y, Lam TW, Yiu SM, Liu S, Zhang H, Li D, Huang Y, Wang X, Yang G, Jiang Z, Wang J, Qin N, Li L, Li J, Bolund L, Kristiansen K, Wong GK, Olson M, Zhang X, Li S, Yang H, Wang J, Wang J. 2010. The sequence and de novo assembly of the giant panda genome. *Nature* 463:311–317 DOI 10.1038/nature08696.
- Liu S, Lorenzen ED, Fumagalli M, Li B, Harris K, Xiong Z, Zhou L, Korneliussen TS, Somel M, Babbitt C, Wray G, Li J, He W, Wang Z, Fu W, Xiang X, Morgan CC, Doherty A, O’Connell MJ, McInerney JO, Born EW, Dalen L, Dietz R, Orlando L, Sonne C, Zhang G, Nielsen R, Willerslev E, Wang J. 2014. Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. *Cell* 157:785–794 DOI 10.1016/j.cell.2014.03.054.
- Mahmood T, Jabeen K, Hussain I, Kayani AR. 2013. Plant species association, burrow characteristics and the diet of the Indian Pangolin, *Manis crassicaudata*, in the Potohar Plateau, Pakistan. *Pakistan Journal of Zoology* 45:1533–1539.
- Meyer W, Liumsiricharoen M, Suprasert A, Fleischer LG, Hewicker-Trautwein M. 2013. Immunohistochemical demonstration of keratins in the epidermal layers of the Malayan pangolin (*Manis javanica*), with remarks on the evolution of the integumental scale armour. *European Journal of Histochemistry* 57:172–177 DOI 10.4081/ejh.2013.e27.
- Mwale M, Dalton DL, Jansen R, De Bruyn M, Pietersen D, Mokgokong PS, Kotze A. 2017. Forensic application of DNA barcoding for identification of illegally traded African pangolin scales. *Genome* 60:272–284 DOI 10.1139/gen-2016-0144.
- Pattarayingsakul W, Nilavongse A, Reamtong O, Chittavanich P, Mungsantisuk I, Mathong Y, Prasitwuttisak W, Panbangred W. 2017. Angiotensin-converting enzyme inhibitory and antioxidant peptides from digestion of larvae and pupae of Asian weaver ant, *Oecophylla smaragdina*, Fabricius. *Journal of the Science of Food and Agriculture* 97:3133–3140 DOI 10.1002/jsfa.8155.
- Pietersen DW, Symes CT, Woodborne S, McKechnie AE, Jansen R. 2016. Diet and prey selectivity of the specialist myrmecophage, Temminck’s ground pangolin. *Journal of Zoology* 298:198–208 DOI 10.1111/jzo.12302.

- Richer R, Coulson IAN, Heath M. 1997.** Foraging behaviour and ecology of the Cape pangolin (*Manis temminckii*) in north-western Zimbabwe. *African Journal of Ecology* 35(4):361–369 DOI 10.1111/j.1365-2028.1997.101-89101.x.
- Robinson MD, McCarthy DJ, Smyth GK. 2010.** edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140 DOI 10.1093/bioinformatics/btp616.
- Storey JD, Tibshirani R. 2003.** Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America* 100:9440–9445 DOI 10.1073/pnas.1530509100.
- Strobel S, Roswag A, Becker NI, Trenczek TE, Encarnacao JA. 2013.** Insectivorous bats digest chitin in the stomach using acidic mammalian chitinase. *PLOS ONE* 8:e72770 DOI 10.1371/journal.pone.0072770.
- Swart JM, Richardson PRK, Ferguson JWH. 1999.** Ecological factors affecting the feeding behaviour of pangolins (*Manis temminckii*). *Journal of Zoology* 247:281–292 DOI 10.1111/j.1469-7998.1999.tb00992.x.
- Tomotake H, Katagiri M, Yamato M. 2010.** Silkworm pupae (*Bombyx mori*) are new sources of high quality protein and lipid. *Journal of Nutritional Science and Vitaminology* 56:446–448.
- Trageser S, Ghose A, Faisal M, Mro P, Mro P, Rahman SC. 2017.** Pangolin distribution and conservation status in Bangladesh. *PLOS ONE* 12:e0175450 DOI 10.1371/journal.pone.0175450.
- Vijayan M, Yeong C, Ling D. 2009.** Captive management of Malayan pangolins *Manis javanica* in the Night Safari. In: *Workshop on trade and conservation of pangolins native to south and southeast Asia*, 119.
- Wang L, Feng Z, Wang X, Wang X, Zhang X. 2010.** DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* 26:136–138 DOI 10.1093/bioinformatics/btp612.
- Ya-yong K, Hong C, Shi-bao W, Qian L, Gan-xin F. 1999.** A study on Chinese pangolin's main food nutrition. *Zoological Research* 20:394–395.
- Yang CW, Chen S, Chang CY, Lin MF, Block E, Lorentsen R, Chin JS, Dierenfeld ES. 2007.** History and dietary husbandry of pangolins in captivity. *Zoo Biology* 26:223–230 DOI 10.1002/zoo.20134.
- Yusoff A, Tan TK, Hari R, Koepfli KP, Wee WY, Antunes A, Sitam FT, Rovie-Ryan JJ, Karuppannan KV, Wong GJ, Lipovich L, Warren WC, O'Brien SJ, Choo SW. 2016.** De novo sequencing, assembly and analysis of eight different transcriptomes from the Malayan pangolin. *Scientific Reports* 6:28199 DOI 10.1038/srep28199.
- Zhenyukh O, Civantos E, Ruiz-Ortega M, Sanchez MS, Vazquez C, Peiro C, Egidio J, Mas S. 2017.** High concentration of branched-chain amino acids promotes oxidative stress, inflammation and migration of human peripheral blood mononuclear cells via mTORC1 activation. *Free Radical Biology and Medicine* 104:165–177 DOI 10.1016/j.freeradbiomed.2017.01.009.
- Zhihai H, Jiang X, Shuiming X, Baosheng L, Yuan G, Chaochao Z, Xiaohui Q, Wen X, Shilin C. 2016.** Comparative optical genome analysis of two pangolin species: *Manis pentadactyla* and *Manis javanica*. *Gigascience* 5:1–5 DOI 10.1093/gigascience/giw001.