Transcriptomic analysis identifies genes and pathways related to myrmecophagy in the Malayan pangolin

Jing E Ma [1], **Lin Miao Li** [1], **Hai Ying Jiang** [1], **Xiu Juan Zhang** [1], **Juan Eyre** Corresp., [1], **Guan Yu Li** [1], **Li Hong Yuan** [1], **Jun Wu** [2], **Jin Ping Chen** Corresp. [3]

[1] Guangdong Key Laboratory of Animal Conservation and Resource Utilization, Guangdong Institute of Applied Biological Resources, Guangzhou, Guangdong, China
[2] Wildlife disease surveillance and molecular ecology research Center, Nanjing Institute of Environmental Sciences under Ministry of Environmental Protection, Guangzhou, Guangdong, China
[3] xingangxi, Guangdong Institute of Applied Biological Resources, Guangzhou, Guangdong, China

Corresponding Authors: Juan Eyre, Jin Ping Chen
Email address: lijuan315@mails.ucas.ac.cn, chenjp@gdei.gd.cn

The Malayan pangolin, an unusual mammal that is a scale-covered, toothless specialist myrmecophage, is maintained primarily through captive breeding in China. Maintaining this species in captivity is a significant challenge partly because little known of the molecular mechanisms on its digestive system up till now. Here, the first large-scale sequencing analysis of transcriptomes from salivary glands, liver and small intestine with Malayan pangolin genome was performed, compared with published data sets involving liver transcriptome profiles from a pregnant Malayan pangolin. A total of 24,452 transcripts were obtained, among which 22,538 transcripts 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. Several of the annotated transcripts were involved in digestive functions: 997 transcripts were related to sensory perception, 129 transcripts belonged to digestive enzyme gene families, and 199 transcripts were related to molecular transporters. One transcript of acidic mammalian chitinase was found in the annotated data, and these might be closely related to the unique digestive function of pangolins. These pathways and transcripts are involved in specialization processes related to myrmecophagy and carbohydrate, protein and lipid digestive pathways, thus probably reflecting adaptations to myrmecophage. Our study has firstly revealed the molecular mechanism of myrmecophagy in Malayan pangolin that may play a role in the protection of the pangolin.

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

Jing-e Ma[1,2,3], Lin-miao Li [1,2,3], Hai-ying Jiang [1,2,3], Xiu-juan Zhang [1,2,3], Juan Li [1,2,3], Guan-yu Li[1,2,3], Li-hong Yuan [1,2,3], Jun Wu[4], Jin-ping Chen [1,2,3,*]

1Guangdong Key Laboratory of Animal Conservation and Resource Utilization,

2 Guangdong Public Laboratory of Wild Animal Conservation and Utilization,

3 Guangdong Institute of Applied Biological Resources, Guangzhou, China,

4 Nanjing Institute of Environmental Sciences under Ministry of Environmental Protection, Nanjing 210042, China.

*Corresponding author

E-mail: chenjp@giabr.gd.cn

Abstract

The Malayan pangolin (*Manis javanica*), an unusual mammal that is a scale-covered, toothless specialist myrmecophage, is maintained primarily through captive breeding in China. Maintaining this species in captivity is a significant challenge partly because little known of the molecular mechanisms on its digestive system up till now. Here, the first large-scale sequencing analysis of transcriptomes from salivary glands, liver and small intestine with *M. javanica* genome was performed, compared with published data sets involving liver transcriptome profiles from a pregnant *M. javanica*. A total of 24,452 transcripts were obtained, among which 22,538 transcripts 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. Several of the annotated transcripts were involved in digestive functions: 997 transcripts were related to sensory perception, 129 transcripts belonged to digestive enzyme gene families, and 199 transcripts were related to molecular transporters. One transcript of acidic mammalian chitinase was found in the annotated data, and these might be closely related to the unique digestive function of pangolins. These pathways and transcripts are involved in specialization processes related to myrmecophagy (insectivory) and carbohydrate, protein and lipid digestive pathways, thus probably reflecting adaptations to myrmecophage. Our study has firstly revealed the molecular mechanism of myrmecophagy in *M. javanica* that may play a role in the protection of the pangolin.

Keywords: pangolin, conservation, digestion, myrmecophagy.

Introduction

Pangolins, also known as scaly anteaters, are eutherians and unique placental mammals. Eight pangolin species are recognized: four from Asia, *M. javanica*, *M. pentadactyla*, *M. crassicaudata,* and *M. culionensis,* and four from Africa, *M. tricuspis*, *M. tetradactyla*, *M. gigantea,* and *M. temminckii* (Choo et al. 2016). *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. 2015; Yang et al. 2007). Pangolins have been reported to consume four kinds of ants (*Anoplolepis steingroeveri*, *Camponotus fulvopilosus,* and two *Crematogaster spp.*) and one kind of termite species (*Trinervitermes trinervoides*) (Pietersen et al. 2015). They also have a well-developed muscular system for fossorial or arboreal behavior and a remarkable olfactory system. As predators of ants and termites, pangolins have a specialized diet and perform an important ecological role in regulating insect populations. Individual adult pangolins have been estimated to consume more than 70 million insects annually and have a significant effect on the control of forest termites (Hua et al. 2015; Pietersen et al. 2015; 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 a delicacy, and their scales which are used in traditional medicine (Trageser et al. 2017).

Manis javanica is classified as critically endangered by The International Union for Conservation of Nature (IUCN) Red List of Threatened Species, (2017; Prop), which has been classified in an appendix I species by the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES I). The number of *M. javanica* in the wild is dramatically declining for several reasons. A major threat is the rapid loss and deterioration of their natural habitat, owing to deforestation activities; illegal hunting, which might reflect the pangolin's economic value; and human agricultural expansion. 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. Pangolins have adapted to a highly specialized diet of ants and termites, thus making it difficult to replace their natural food completely with 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 composed of N-acetyl-beta-glucosaminidase with a beta-1, 4 glycosidic bond, which is one of substances contained in insect exoskeletons and the peritrophic membranes of ants. Internal degradation of chitin particles is mainly performed 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 egg (hard boiled), multivitamin liquid, horse meat, water, mealworms, insectivore pellets, salmon oil, and powdered termite mound (Vijayan et al. 2009).

Digestive disorders often appear in pangolins fed with artificial food, and the feces 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), but an understanding of the molecular genetics is lacking, which might provide a theoretical basis for raising *M. javanica* in captivity.

Genetic studies of endangered species have become increasingly widespread in the recent two years (Choo et al. 2016; Mohamed Yusoff et al. 2016; Mwale et al. 2017; Zhihai et al. 2016). Especially, the genomes and transcriptomes of *M. javanica* have been sequentially reported (Choo et al. 2016; Mohamed Yusoff et al. 2016) . Nowadays, more and more molecular information about *M. javanica* can be available for the high-throughput next-generation sequencing (NGS) technologies popping up like mushrooms. The complete genome sequencing of *M. javanica* and transcriptome sequencing of eight organs, including the heart, liver, spleen, lung, kidney, thymus, cerebellum, and cerebrum, have progressively revealed unknown aspects of pangolin biology. The high-quality transcriptomes have been used for analyses of the functional and phylogenetic aspects of immunity biology (Mohamed Yusoff et al. 2016), but genetic research regarding myrmecophagy is still lacking. This observation led us to consider the molecular specificity of pathways in myrmecophagy. What specific molecular pathways are involved in the evolution of this dietary adaption, and how do they affect the appearance of this feature? In terms of dietary adaptation, do specific genes exist for digestive function? Here, we selected liver, small intestine, and salivary glands for transcriptome sequencing and analysis of the genetic selection or potential candidate genes involved in myrmecophagy, which play an important role in the digestive system,

salivary glands is one of the important amylase secretory organs, there are a lot of amylase and lipase contained in the secretions of liver, and the small intestine is involved in the absorption of the nutriment, to analyze the unique feeding behavior of pangolins and to provide a new approach to the protection of the pangolin.

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, one female adult *M. javanica* sample were provided by the Dongguan Institute of Qingfengyuan Animal Medicine (Dongguan, Guangdong, China). The specimen was dissected immediately after their natural death. The salivary glands, liver, and small intestine were collected as soon as possible, frozen in liquid nitrogen, and stored at -80°C until RNA extraction.

RNA isolation, cDNA library construction and Illumina sequencing

Total RNA of three tissues was extracted from tissues with RNAiso reagent (Takara, Otsu, Japan) and was treated with DNase I (Takara). RNA purity was checked using the NanoPhotometer spectrophotometer (IMPLEN, CA, USA). RNA concentration was measured using Qubit RNA Assay Kit in Qubit 2.0 Flurometer (Life Technologies, CA, USA). RNA integrity was assessed using the RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 system (Agilent

Technologies, CA, USA). RNA was frozen at -80°C until cDNA library construction.

RNA samples from three organs of sample 1 were sent for preprocessing, and 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. Then cDNA synthesis was performed, and was followed by PCR amplification to generate a complete cDNA library, which was sent for sequencing using the Illumina HiSeq™ 2000 platform.

Data assembly and annotation

Four groups of sequencing data were used for assembling and annotating, and one group was selected from the sequenced liver from a published paper with National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) accession No.SRR2561213 (Mohamed Yusoff et al. 2016). Primary sequencing data (raw reads) from the Illumina HiSeq ™ 2000 were subjected to quality control (QC) through in-house perl scripts to determine whether a resequencing step was needed. Followed as: a) remove reads with adaptors, b) remove reads in which unknown bases (N) are more than 5% (Parameters:-n 0.05), c) remove low quality reads (we define the low quality read as the percentage of base which quality is lesser than 10 is greater than 20% in a read) (Parameters:-l 10 -q 0.2). Clean reads were aligned to reference sequences with a spliced read mapper for RNA-Seq-*TopHat2* (http://ccb.jhu.edu/software/tophat/index.shtml), the reference sequences were the assembled whole-genome sequences of *Malayan pangolin* which had been deposited at GenBank under the accessions JSZB00000000.1. Then, the alignment data were used to calculate the distribution and

coverage of the reads on the reference genes.

Next, all the transcripts (>200 bp) were annotated on the basis of basic local alignment search tool (BLASTX) results with e-values of $1e^{-5}$ against 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 Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.genome.jp/kegg/), Cluster of Orthologous Groups (COG, http://www.ncbi.nlm.nih.gov/COG/), Gene Ontology (GO, http://www.geneontology.org/), the database of Clusters of Protein homology (KOG, http://www.ncbi.nlm.nih.gov/KOG/) and the Translated EMBL Nucleotide Sequence Data Library (TrEMBL, http://www.bioinfo.pte.hu/more/TrEMBL.htm) database. Furthermore, gene expression analysis was performed, which included the gene expression level and differential gene expression between every two groups of data.

Correlation between any two pangolin tissue transcriptomes

To examine the close relatedness of *M. javanica* organ transcriptomes, the expression levels of the transcripts (FPKM) in the transcriptomes of each tissue were manipulated. By using the tool "RSEM-calculate-expression" in the RSEM pipeline (http://deweylab.biostat.wisc.edu/RSEM), which is an accurate transcript quantification software from RNA-Seq data with or without a reference genome. The reads of each tissue were mapped to the transcripts (Li & Dewey 2011). Gene expression values, expressed as log 10 (FPKM+1) for the transcriptomic data from each tissue were plotted against one another to produce scatter plots. R2 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) database with the accession: GSM2667949, GSM2667950 and GSM2667951. The average proportion of high-quality clean reads was 95.58% (Table 1). The clean reads were assembled into long assembled sequences (contigs) with *TopHat2*. The alignment efficiency between the sample and reference genome ranged from 69.57% to 89.30% (Table 2). The ratio of the transcripts ranged from 76.11% to 85.16%, as compared with the exons (S1 Figure).

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 (S1 Table). As expected, the majority of the 22,473 transcripts matched the NR databases (e-value < 10^{-5}) (S2 Figure and S1 File). 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,

192 6.24%), *Canis lupus* (1,349, 6.01%), *Mustela putorius* (1,044, 4.65%) and *Odobenus rosmarus*
193 (1,022, 4.55%) (S3 Figure).

194 COG and KOG analysis

195 In the COG database, the largest category of *M. javanica* annotated transcripts was general
196 function prediction only (R) (2,438 transcripts, 27.76%), which was followed by replication,
197 recombination, and repair (L) (848, 9.65%); transcription (K) (845, 9.62%); signal transduction
198 mechanisms (T) (790, 8.99%); and post-translational modification, protein turnover, and
199 chaperones (O) (484, 5.51%) (S4A Figure). In the KOG database, the largest category of *M.*
200 *javanica* annotated transcripts was general function prediction only (R) (2,910 transcripts, 18.3%),
201 followed by signal transduction mechanisms (T) (2,744, 17.26%); post-translational modification,
202 protein turnover, and chaperones (O) (1,259, 7.92%); function unknown (S) (1,182, 7.43%); and
203 transcription (K) (1070, 6.73%) (S4B Figure).

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

208 Gene Ontology (GO)

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

213 In the biological process ontology, the most highly represented terms were cellular processes
214 (10,494, 63.03%), single-organism processes (9,554, 57.38%), and biological regulation (7,986,
215 47.97%). The fourth top represented term was metabolic process (7,173, 43.08%), which was
216 followed by response to stimulus (4,915, 29.52%), multicellular organismal process (3,644,
217 21.89%), signaling (3,161, 18.99%), localization (3,001, 18.03%), developmental process (2,890,
218 17.36%), and cellular component organization or biogenesis (2,667, 16.02%). The terms
219 associated with biological regulation and metabolic process might be indicative of the involvement
220 of the *M. javanica* transcriptome in various digestive activities.

221 For molecular functions, the sequences were mainly assigned to binding (9,371, 56.29%) and
222 catalytic activity (5347, 32.12%). These were followed by molecular transducer activity (1,750,
223 10.51%), receptor activity (1,681, 10.1%) and transporter activity (1,026, 6.16%), which might be
224 involved in food digestion and absorption.

225 As anticipated, cell part (11,445, 68.74%) and cell (11,412, 68.54%) were the predominant
226 terms assigned to the pangolin transcriptome in cellular components and were followed by
227 organelle (8,014, 48.14%), membrane (5,980, 35.92%), membrane part (4,441, 26.67%), organelle
228 part (4,079, 24.5%), macromolecular complex (3,413, 20.5%) and extracellular region (1,044,
229 6.27%) (S5 Figure and S3 File). Overall, these results indicate the broad range of biological
230 activities related to the expressed pangolin transcriptome, representing a pooled collection of the
231 three digestive tissues sequenced.

232 KEGG pathway analysis

233 To identify the pathways in which the transcripts of *M. javanica* 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 representing cellular processes, followed by 22, 27, 66, 68, and 89 unique KEGG pathways representing genetic information passing, environmental information processing, organismal systems, human diseases, and metabolism, respectively (S4 File).

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), MAPK signaling pathway (300 transcripts), and neuroactive ligand-receptor interaction (292 transcripts) (S6 Figure). The sense of smell is closely related to the biological activity of instinctive behavior such as feeding. The olfactory pathway plays a key role in the specific recognition ability of smell, thus leading the animal to different foods. A total of 969 genes were associated with olfactory transduction in *M. javanica* transcripts, 942 transcripts of which were annotated as various kinds of olfactory receptors. These finding might lead to a keen sense of smell in *M. javanica*.

Metabolic pathway analysis

A total of 1,814 transcripts were associated with 89 unique metabolic pathways of KEGG. Most transcripts were enriched 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 metabolism of cofactors and vitamins (225 transcripts). Other transcripts were associated with global and overview maps (199 transcripts), energy metabolism (167 transcripts), metabolism of other amino acids (119

255 transcripts), and xenobiotics biodegradation and metabolism (116 transcripts), and the smallest
256 number of the transcripts were associated with terpenoid and polyketide metabolism (27
257 transcripts) and secondary metabolite biosynthesis (13 transcripts) (Fig 1A).

258 Carbohydrate metabolism

259 Inositol phosphate metabolism (73 transcripts), glycolysis/gluconeogenesis (70 transcripts),
260 starch and sucrose metabolism (57 transcripts), and amino sugar and nucleotide sugar metabolism
261 (53 transcripts) were at the top of the carbohydrate metabolic lists, whereas ascorbate and aldarate
262 metabolism were at the bottom.

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

267 Lipid metabolism

268 Glycerophospholipid metabolism (100 transcripts), arachidonic acid metabolism (76
269 transcripts), steroid hormone biosynthesis (68 transcripts), and sphingolipid metabolism (55
270 transcripts) were at the top of the lipid metabolic list; in contrast, fatty acid biosynthesis (55
271 transcripts) was at the bottom. We identified transcripts from several pathways in unsaturated fatty
272 acid metabolism, including arachidonic acid metabolism (76 transcripts), linoleic acid metabolism
273 (36 transcripts), alpha-linolenic acid metabolism pathways (23 transcripts), and biosynthesis of
274 unsaturated fatty acids (23 transcripts) (Fig 1C).

275 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 biosynthesis (6 transcripts); valine, leucine and isoleucine biosynthesis (5 transcripts); and lysine (2 transcripts) were at the bottom (Fig 1D). None of the transcripts were found to be involved in arginine biosynthesis.

Metabolism of cofactors and vitamins, and terpenoids and polyketides

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 metabolism of cofactors and vitamins lists. There was only one list in the terpenoids and polyketides, which was terpenoid backbone biosynthesis (27 transcripts) (Fig 1E), as shown in Fig 2.

Annotation of the new transcripts

On the basis of the genome sequences of *M. javanica*, Cufflinks software was used for joining the mapped reads together, comparing them with the annotated information for the original genome, and searching for the gapped sequences, which were not annotated. A total of 3,373 new transcripts were discovered in the results (S5 File), from which 1,459 transcripts were annotated on the basis of the COG, GO, KEGG, KOG, Swiss-Prot, TrEMBL and NR databases by using BLAST (S2 Table and S6 File).

In Gene Ontology analysis, 75 new transcripts were involved in 22 metabolic categories. Some of the new genes were involved in the inositol metabolic process (GO: 0006020), for

297 example, Manis_javanica_newGene_958, and some were involved in the linoleic acid metabolic
298 process (GO: 0043651), for example, Manis_javanica_newGene_12722. For KEGG, 114 new
299 genes were related to metabolic function, including lipid metabolism (41 transcripts), carbohydrate
300 metabolism (24 transcripts), cofactor and vitamin metabolism (22 transcripts), and amino acid
301 metabolism (20 transcripts) (S7 Figure).

302 Gene expression repertoire

303 Distributions of potential transcripts related to feeding among the three tissue libraries are
304 shown in Table 3, S3 Table and S7 File, including the 997 transcripts related to sensory perception.
305 972 of these transcripts were related to olfaction, and 11 and 14 transcripts were related to vision
306 and taste, respectively. A total of 133 transcripts were related to digestive enzyme gene families,
307 and 70 of these transcripts were related to lipid degradation, and 39 and 20 transcripts were related
308 to the degradation of proteins and carbohydrates, respectively. These genes were considered to be
309 involved in the profile of food choice, digestion and absorption, which might serve as a molecular
310 mechanism in myrmecophagy. Among these transcripts, acidic mammalian chitinase (*CHIA*),
311 chitinase-3-like protein 1 (*CHI3L1*), and chitinase domain-containing protein 1 (*CHID1*) were
312 related to chitin degradation. A total of 199 transcripts were related to molecular transporters,
313 including sugar transporters, aa transporters, apolipoprotein transporters, cationic/anion
314 transporters, vitamin transporters, cotransporters and others, among which, the UDP-N-
315 acetylglucosamine transporter (*SLC35A3*, *SLC35B4*, and *SLC35D2*) was related to the decomposer
316 absorption of the chitin unit which was 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 by using log10 (FPKM+1) to normalize the plots (Fig 3). Our data showed that the two liver transcriptomes had the most similar expression profiles (coefficient of determination, R^2=0.53), followed by the liver and gut (R^2=0.30). The salivary glands and gut had the least similar transcriptomic profiles with $R^2=2e^{-0.4}$. The results reflected the varying complexity between the same organs in different individuals, possibly because one of the two specimens was pregnant. The differences between every two organs compared reflected the different digestive functions of the three organs.

A total of 11,055 transcripts were expressed (FPKM>1.0) in all three tissues, of which 3,947 transcripts 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 biosynthesis (1 transcripts) and lysine (2 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, compared with 258 transcripts differentially expressed between the salivary glands and small intestine. The numbers of the different genes were 23 and 27 in the starch

and sucrose metabolism and arginine and proline metabolism between the small intestine and liver, respectively, whereas 31 genes were in the steroid hormone biosynthesis between the liver and the referred liver (S8 Figure).

Several transcripts were specifically expressed in the 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 (S4 Table). The highest number of specific transcripts was eight, and these transcripts were involved in the arachidonic acid metabolism of the transcriptional results of the liver from the previous study; six transcripts were involved in glycerophospholipid metabolism, in either lipid metabolism or arachidonic acid metabolism of the small intestine (Fig 5).

Discussion

Malaysian pangolins are unique endangered mammals. Captive breeding of this species provides an opportunity to study the molecular mechanisms of myrmecophagy. Here, transcriptome sequencing of digestive organs was performed to observe the metabolic pathways and functional genes related to the myrmecophagy in an attempt to understand the molecular mechanisms involved in myrmecophagy. The results may provide an important theoretical basis for the successful captive breeding of the species. The transcriptomic data of the three organs showed 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 unique mammalian species.

359 Functional annotation of the *M. javanica* transcripts revealed the involvement of molecular
360 mechanisms in various essential KEGG pathways, such as the olfactory transduction, amino sugar
361 and nucleotide sugar metabolism, lipid metabolism and terpenoid and polyketide metabolism
362 pathways, which may support the myrmecophagous feeding habits of this mammalian species,
363 including diet selection, digestion and absorption.

364 Ants are a high-lipid, high-protein food (Tomotake et al. 2010). They contain more than 50%
365 crude protein, as determined according to a nutritional value evaluation, and contain more than 20
366 kinds of amino acids; varied microelements; special chemicals, such as formic acid and
367 herbaceous acetaldehyde, which are all triterpenoid compounds; and multivitamins
368 (Pattarayingsakul et al. 2017). Many pangolin genes are likely to be involved in the digestion of
369 these materials, because there were 27 transcripts related to the biosynthesis of terpenoid
370 backbone, which might be one of biological basis for their adaptions to ant-eating habits found in
371 our transcriptomic results. However, no genes were found to participate in the pathway of
372 arginine synthesis, according to the KEGG analysis, and only two transcripts were involved in the
373 synthesis of lysine. Few transcripts were involved in the synthesis pathways of the essential amino
374 acids in humans, such as, lysine, valine, leucine, isoleucine, phenylalanine, and tryptophan (Galili
375 et al. 2016; Zhenyukh et al. 2017). These results suggested that *M. javanica* may synthesize
376 arginine de novo. Therefore, both essential amino acids and arginine must be added to
377 manufactured feed.

378 Chitin, one of the main components of the epidermis of ants and termites, is made of N-acetyl-
379 D-glucosamines (GluNAc) connected by a β-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; Pietersen et al. 2015; Strobel et al. 2013). The origination of the digestive processes might be closely related to the activity of AMCase, which determines the start of the chitin decomposition. The transporter genes of UDP-N-acetyl glucosamine (*SLC35A3*, *SLC35B4*, and *SLC35D2*) might be directly related to the absorption of the carbohydrate units during the process of chitin digestion. Therefore, chitin should be added to the formulated diets to aid in digestion and absorption of nutrients. This suggestion is consistent with the referenced formulas mentioned in Vijayan *et al* (Vijayan et al. 2009).

A total of 3,373 new genes were discovered in the transcriptomic datasets, of which 1459 were annotated, and 75 new genes were involved in metabolism. The new genes provided new information for studying the myrmecophagous mechanisms of pangolins. A large number of genes were expressed in all three tissues, whereas several specific genes in the three systems played different roles in the metabolism of sugars, lipids, and amino acids, and the digestive functions of the small intestine and salivary gland were relatively similar, in contrast to the differences between the small intestine and liver. The functions of the livers from the two different individuals were relatively different in the pathways of lipid metabolism, thus suggesting that the ratio of lipid in the feed should be changed appropriately during pregnancy.

Overall, the smell of the formulated diet may be important for *M. javanica* with a strong sense of smell. Diets with high-fat and high-protein content are conducive to pangolin management. In addition, lysine, arginine and chitin could be added to the formulas to aid in the digestion and absorption of the nutrients. Some indications suggested that changes in the fat content might be

appropriate during pregnancy.

In conclusion, *M. javanica* transcriptomic datasets of the three representative tissues would provide the first-hand knowledge for uncovering the mysteries of the genetic mechanisms of digestion and reproduction in this rare and unique mammal. For further, more and more organizations involved in digestion, such as tongue, stomach, and pancreas, would be collected for studying myrmecophagy. All these results would be gathered to unveil the mysteries of the special diet of *M. javanica*.

Acknowledgments

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.

Funding Sources: This project was supported by the Planning Funds of Scientific and Technological of Guangdong Province (2016B070701016) 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 (xm-2015081), GDAS Special Project of Science and Technology Development (2017GDASCX-0107).

References

2017. The IUCN Red List of Threatened Species. Version 2017-1. www.iucnredlist.org.

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 K-P, Minx P, Krasheninnikova K, Kotze A, Dalton DL, Vermaak E, Paterson IC, Dobrynin P, Sitam FT, Rovie-Ryan JJ, Johnson WE, Yusoff AM, Luo S-J, Karuppannan KV, Fang G, Zheng D, Gerstein MB, Lipovich L, O'Brien SJ, and Wong GJ. 2016. Pangolin genomes and the evolution of mammalian scales and immunity. *Genome Research* 26:1312-1322.

Eurich K, Segawa M, Toei-Shimizu S, and Mizoguchi E. 2009. Potential role of chitinase 3-like-1 in inflammation-associated carcinogenic changes of epithelial cells. *World J Gastroenterol* 15:5249-5259.

Galili G, Amir R, and Fernie AR. 2016. The Regulation of Essential Amino Acid Synthesis and Accumulation in Plants. *Annual Review of Plant Biology* 67:153-178.

Hua L, Gong S, Wang F, Li W, Ge Y, Li X, and Hou F. 2015. Captive breeding of pangolins: current status, problems and future prospects. *ZooKeys* 507:99-114.

Krykbaev R, Fitz LJ, Reddy PS, Winkler A, Xuan D, Yang X, Fleming M, and Wolf SF. 2010. Evolutionary and biochemical differences between human and monkey acidic mammalian chitinases. *Gene* 452:63-71.

Li B, and Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323.

Meyer W, Liumsiricharoen M, Suprasert A, Fleischer LG, and 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. *Eur J Histochem* 57:e27.

Mohamed Yusoff A, Tan TK, Hari R, Koepfli K-P, Wee WY, Antunes A, Sitam FT, Rovie-Ryan JJ, Karuppannan KV, Wong GJ, Lipovich L, Warren WC, O Brien SJ, and Choo SW. 2016. De novo sequencing, assembly and analysis of eight different transcriptomes from the Malayan pangolin. *Scientific Reports* 6:28199.

Mwale M, Dalton DL, Jansen R, De Bruyn M, Pietersen D, Mokgokong PS, and Kotze A. 2017. Forensic application of DNA barcoding for identification of illegally traded African pangolin scales. *Genome* 60:272-284.

Pattarayingsakul W, Nilavongse A, Reamtong O, Chittavanich P, Mungsantisuk I, Mathong Y, Prasitwuttisak W, and 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.

Pietersen DW, Symes CT, Woodborne S, McKechnie AE, and Jansen R. 2015. Diet and prey selectivity of the specialist myrmecophage, Temminck's ground pangolin. *Journal of Zoology*.

Prop CC. 13. 2000. Transfer of Manis crassicaudata, M. pentadactyla, M. javanica from Appendix II to Appendix I. India, Nepal, Sri Lanka and the United States of America.

Strobel S, Roswag A, Becker NI, Trenczek TE, and Encarna O JA. 2013. Insectivorous Bats Digest Chitin in the Stomach Using Acidic Mammalian Chitinase. *PLoS ONE* 8:e72770.

Tomotake H, Katagiri M, and Yamato M. 2010. Silkworm pupae (Bombyx mori) are new sources of high quality protein and lipid. *J Nutr Sci Vitaminol (Tokyo)* 56:446-448.

Trageser SJ, Ghose A, Faisal M, Mro P, Mro P, and Rahman SC. 2017. Pangolin distribution and conservation status in Bangladesh. *PLoS ONE* 12:e0175450.

Vijayan M, Yeong C, and Ling D. 2009. Captive management of Malayan pangolins Manis javanica in the Night Safari. WORKSHOP ON TRADE AND CONSERVATION OF PANGOLINS NATIVE TO SOUTH AND SOUTHEAST ASIA. p 119.

Ya-yong K, Hong C, Shi-bao W, Qian L, and 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, and Dierenfeld ES. 2007. History and dietary husbandry of pangolins in captivity. *Zoo Biol* 26:223-230.

Zhenyukh O, Civantos E, Ruiz-Ortega M, Sánchez MS, Vázquez C, Peiró C, Egido J, and 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.

Zhihai H, Jiang X, Shuiming X, Baosheng L, Yuan G, Chaochao Z, Xiaohui Q, Wen X, and Shilin C. 2016. Comparative optical genome analysis of two pangolin species: Manis pentadactyla and Manis javanica. *Gigascience* 5:1-5.

Figure 1

The metabolic pathway analysis of transcripts from *M. javanica*.

The x-axis shows the numbers of annotated transcripts in one class, and the y-axis shows the KEGG function classes.

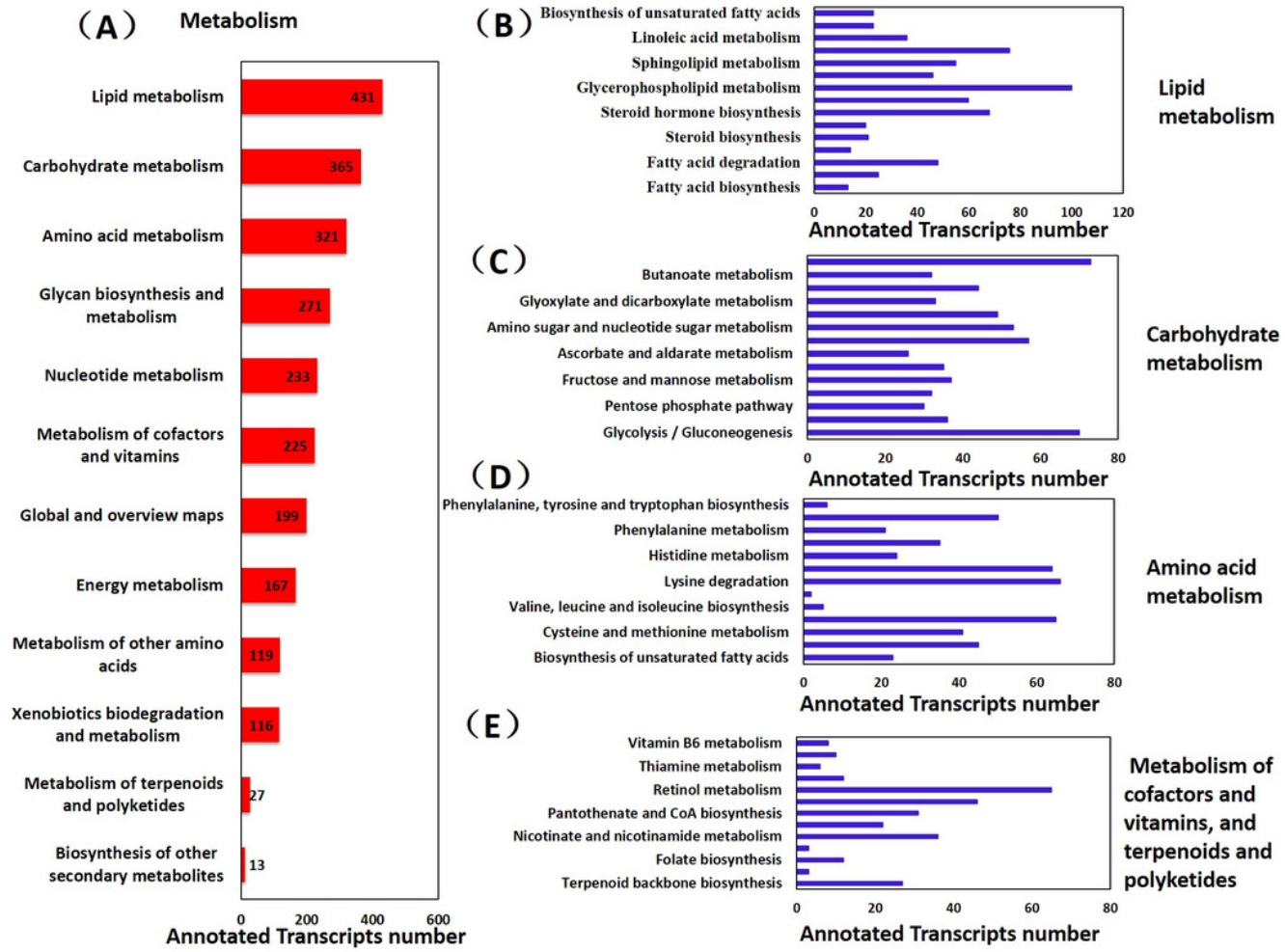

Figure 2

Terpenoid backbone biosynthesis (KEGG map 00900).

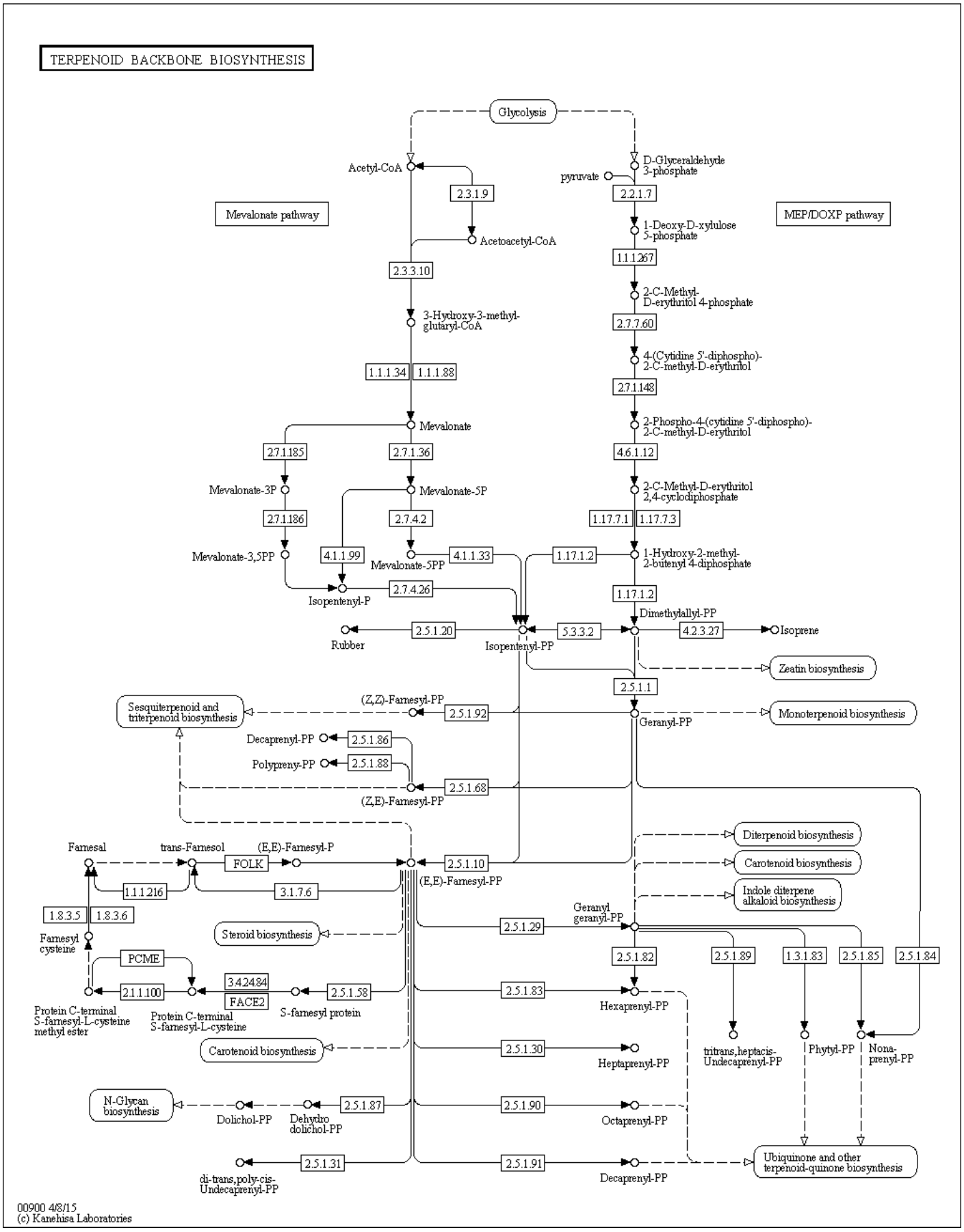
TERPENOID BACKBONE BIOSYNTHESIS
Glycolysis
Acetyl-CoA
pyruvate
D-Glyceraldehyde 3-phosphate
2.3.1.9
2.2.1.7
Mevalonate pathway
MEP/DOXP pathway
Acetoacetyl-CoA
1-Deoxy-D-xylulose 5-phosphate
2.3.3.10
1.1.1.267
2-C-Methyl-D-erythritol 4-phosphate
3-Hydroxy-3-methyl-glutaryl-CoA
2.7.7.60
4-(Cytidine 5'-diphospho)-2-C-methyl-D-erythritol
1.1.1.34
1.1.1.88
2.7.1.148
Mevalonate
2-Phospho-4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol
2.7.1.185
2.7.1.36
4.6.1.12
Mevalonate-3P
Mevalonate-5P
2-C-Methyl-D-erythritol 2,4-cyclodiphosphate
2.7.1.186
2.7.4.2
1.17.7.1
1.17.7.3
Mevalonate-3,5PP
4.1.1.99
Mevalonate-5PP
4.1.1.33
1.17.1.2
1-Hydroxy-2-methyl-2-butenyl 4-diphosphate
2.7.4.26
Isopentenyl-P
1.17.1.2
Dimethylallyl-PP
Rubber
2.5.1.20
Isopentenyl-PP
5.3.3.2
4.2.3.27
Isoprene
Zeatin biosynthesis
2.5.1.1
Sesquiterpenoid and triterpenoid biosynthesis
(Z,Z)-Farnesyl-PP
2.5.1.92
Geranyl-PP
Monoterpenoid biosynthesis
Decaprenyl-PP
2.5.1.86
Polyprenyl-PP
2.5.1.88
2.5.1.68
(Z,E)-Farnesyl-PP
Diterpenoid biosynthesis
Farnesal
trans-Farnesol
(E,E)-Farnesyl-P
FOLK
2.5.1.10
Carotenoid biosynthesis
(E,E)-Farnesyl-PP
1.1.1.216
3.1.7.6
Indole diterpene alkaloid biosynthesis
1.8.3.5
1.8.3.6
Geranyl geranyl-PP
Steroid biosynthesis
2.5.1.29
Farnesyl cysteine
2.5.1.82
2.5.1.89
1.3.1.83
2.5.1.85
2.5.1.84
PCME
3.4.24.84
2.1.1.100
FACE2
2.5.1.58
2.5.1.83
Hexaprenyl-PP
Protein C-terminal S-farnesyl-L-cysteine methyl ester
Protein C-terminal S-farnesyl-L-cysteine
S-farnesyl protein
tritrans,heptacis-Undecaprenyl-PP
Phytyl-PP
Nona-prenyl-PP
Carotenoid biosynthesis
2.5.1.30
Heptaprenyl-PP
N-Glycan biosynthesis
2.5.1.87
2.5.1.90
Dolichol-PP
Dehydro dolichol-PP
Octaprenyl-PP
Ubiquinone and other terpenoid-quinone biosynthesis
2.5.1.31
2.5.1.91
di-trans,poly-cis-Undecaprenyl-PP
Decaprenyl-PP
00900 4/8/15
(c) Kanehisa Laboratories

Figure 3

Correlation between any two organs.

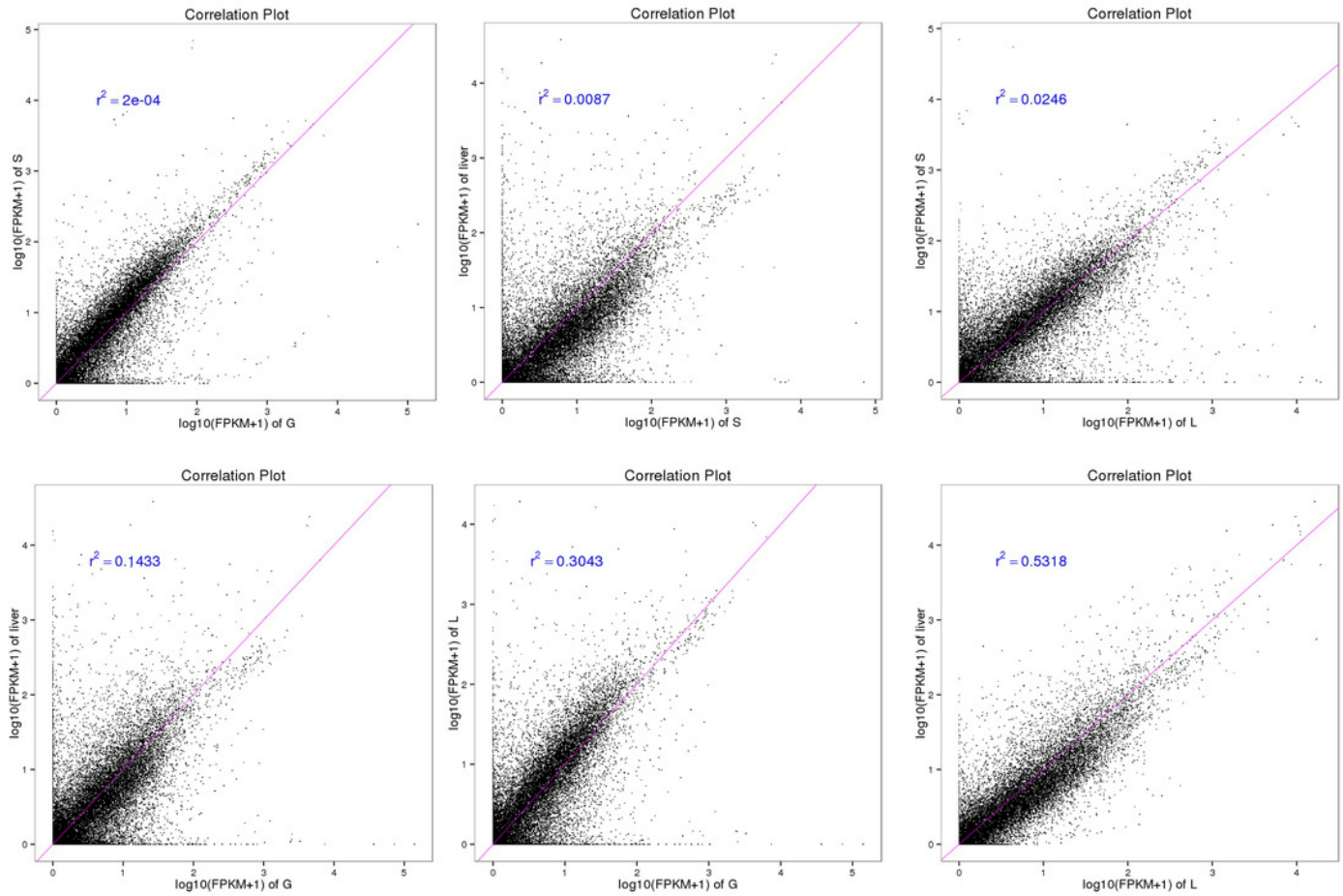

Figure 4

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.

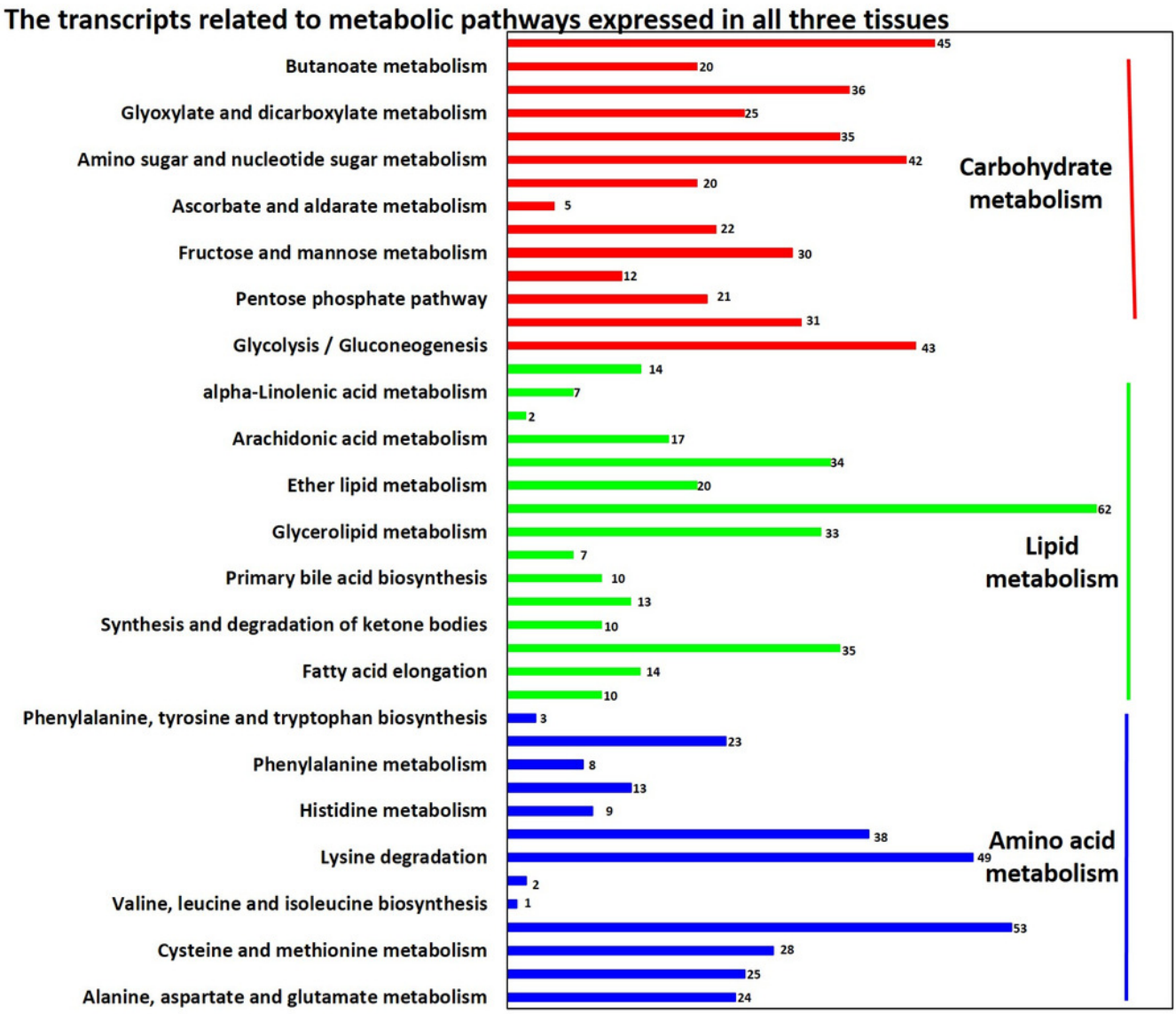

Figure 5

The number of transcripts related to metabolic pathways specifically expressed in the three tissues.

The x-axis shows the KEGG function classes, and the y-axis shows the number of transcripts with the corresponding KEGG function class.

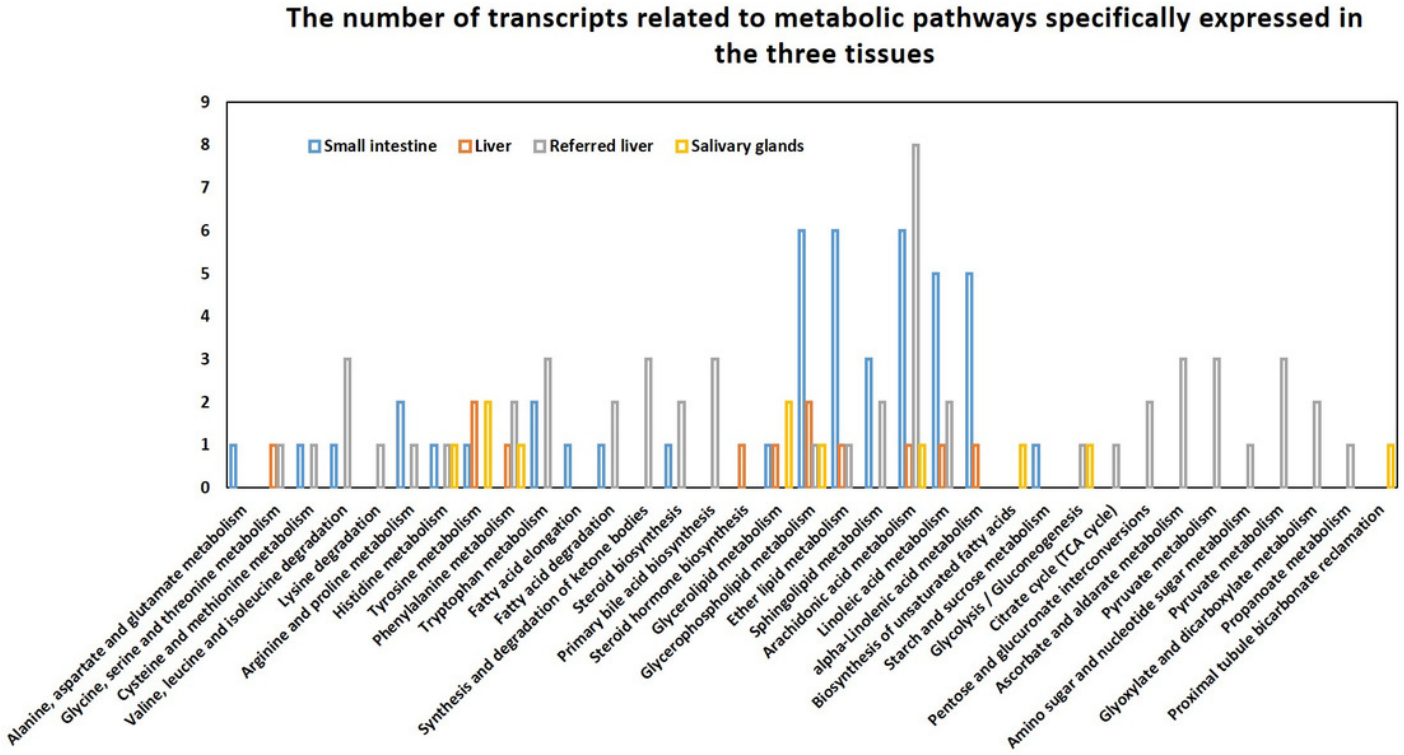

Table 1(on next page)

utput statistics of sequencing.

1
2

ID	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(on next page)

The results of sequencing data aligned to the *Manis.javanica* genome.

ID	Total Reads	Mapped Reads	Uniq 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%)
Referred liver	90,612,846	80,913,231 (89.30%)	80,064,920 (88.36%)	848,311 (0.94%)

Table 3(on next page)

Genes related to the diet of *M.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, LIPF, 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, CTRB1, CTRC, DPP6, DNPEP, ENPEP, ERAP2, LAP3, METAP1, METAP2, NPEPL1, PGC, PRSS12, Trypsin, XPNPEP1, XPNPEP2, XPNPEP3
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, MMGT1, 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