GigaScience

Transcriptome Analysis of the Response of Burmese Python to Digestion
--Manuscript Draft--

Manuscript Number:	GIGA-D-16-00152R1
Full Title:	Transcriptome Analysis of the Response of Burmese Python to Digestion
Article Type:	Research
Abstract:	Background: The exceptional and extreme feeding behaviour makes the Burmese python a unique and interesting model to study physiological remodelling and metabolic adaptation in response to feeding after prolonged starvation. With outset in specific hypotheses based on in vivo physiological responses, we use transcriptome sequencing of five visceral organs and three digestive stages to unravel the patterns of changes in the gene expression of Burmese python upon ingestion of a large meal. We first used the combined data to perform a de novo assembly of the transcriptome. We supplemented with a proteomic survey of enzymes in the gastric juice, stomach secretome and plasma during digestion assisted by our transcriptome sequence database. Results: We constructed a high-quality transcriptome with 34,423 transcripts of which 19,713 (57%) were annotated. Among highly expressed genes (FPKM>100 in one tissue) we found differential expression for 43 genes in heart, 206 genes in liver, 114 genes in stomach, 89 genes in pancreas and 158 genes in intestine. We interrogated the function of these genes in order to test previous hypothesis on the response to feeding. We also used the transcriptome to identify 314 secreted proteins in the python gastric juice. Conclusions: We provide comprehensive transcriptome data of multiple organs and various digestive time of Burmese python and address specific hypothesis on certain pathways known to related digestion process. We also identify, for the first time, stomach-related proteins from a digesting individual and thereby demonstrate that the sensitivity of modern LC-MS/MS equipment allows the identification of gastric juice proteins that are present during digestion thereby providing novel insight into the digestion mechanism.
Response to Reviewers:	Dear Editor, Thank you for returning the constructive and useful comments from the two reviewers that kindly evaluated our manuscript entitled "Transcriptome Analysis of the Response of Burmese Python to Digestion" that we submitted for publication in GigaScience. Both reviewers provided positive overall assessments, but also raised a number of specific queries to be addressed in the revision. We are pleased to return a revised manuscript where we have followed all the advice given by the two reviewers. The responses to each query is listed in a separate PDF file entitled "Answer sheet to reviewers comments" attached as personal cover where you can see our responses describing the changes we have made to the manuscript. We greatly appreciate these comments and feel the manuscript has been improved in this review-process. We hope you will find the revised manuscript acceptable for publication in GigaScience and we are looking forward to hearing from you in due course. Please do not hesitate to contact me in case you need additional information. Sincerely yours Jinjie Duan (on behalf of all the authors)
Additional Information:	
Question	**Response**
Are you submitting this manuscript to a special series or article collection?	No
Experimental design and statistics Full details of the experimental design and statistical methods used should be given	Yes

in the Methods section, as detailed in our Minimum Standards Reporting Checklist. Information essential to interpreting the data presented should be made available in the figure legends. Have you included all the information requested in your manuscript?	
Resources A description of all resources used, including antibodies, cell lines, animals and software tools, with enough information to allow them to be uniquely identified, should be included in the Methods section. Authors are strongly encouraged to cite Research Resource Identifiers (RRIDs) for antibodies, model organisms and tools, where possible. Have you included the information requested as detailed in our Minimum Standards Reporting Checklist?	Yes
Availability of data and materials All datasets and code on which the conclusions of the paper rely must be either included in your submission or deposited in publicly available repositories (where available and ethically appropriate), referencing such data using a unique identifier in the references and in the "Availability of Data and Materials" section of your manuscript. Have you have met the above requirement as detailed in our Minimum Standards Reporting Checklist?	Yes

Transcriptome Analysis of the Response of

Burmese Python to Digestion

Jinjie Duan[1], Kristian Wejse Sanggaard[2,3], Leif Schauser[5], Sanne Enok Lauridsen[4], Jan J. Enghild[2,3], Mikkel Heide Schierup[1,4] and Tobias Wang[4]

1. Bioinformatics Research Center, Aarhus University, Denmark

2. Department of Molecular Biology and Genetics, Aarhus University, Denmark

3. Interdisciplinary Nanoscience Center (iNANO), Aarhus University, Denmark

4. Department of Bioscience, Aarhus University, Denmark

5. QIAGEN Aarhus, Silkeborgvej 2, 8000 Aarhus C, Denmark

Corresponding authors: Jinjie Duan, Mikkel Heide Schierup & Tobias Wang

Email address:

Jinjie Duan: jjduan@birc.au.dk

Kristian Wejse Sanggaard: kristian@wejsesanggaard.dk

Leif Schauser: Leif.Schauser@qiagen.com

Sanne Enok Lauridsen: sanne.lauridsen@inano.au.dk

Jan J. Enghild: jje@mbg.au.dk

Mikkel Heide Schierup: mheide@birc.au.dk

Tobias Wang: tobias.wang@bios.au.dk

Abstract

23

24 Background:

25 ~~The~~ exceptional and extreme feeding behaviour makes the Burmese python a unique and

26 interesting model to study physiological remodelling and metabolic adaptation in response to

27 feeding after prolonged starvation. With outset in specific hypotheses based on in vivo

28 physiological responses, we use transcriptome sequencing of five visceral organs and three

29 digestive stages to unravel the patterns of changes in the gene expression of Burmese python

30 upon ingestion of a large meal. We first used the combined data to perform a *de novo*

31 assembly of the transcriptome. We supplemented with a proteomic survey of enzymes in the

32 gastric juice, stomach secretome and plasma during digestion assisted by our transcriptome

33 sequence database.

34 Results:

35 We constructed a high-quality transcriptome with 34,423 transcripts of which 19,713 (57%)

36 were annotated. Among highly expressed genes (FPKM>100 in one tissue) we found

37 differential expression for 43 genes in heart, 206 genes in liver, 114 genes in stomach, 89

38 genes in pancreas and 158 genes in intestine. We interrogated the function of these genes in

39 order to test previous hypothesis on the response to feeding. We also used the transcriptome

40 to identify 314 secreted proteins in the python gastric juice.

41 Conclusions:

42 We provide comprehensive transcriptome data of multiple organs and various digestive time

43 of Burmese python and address specific hypothesis on certain pathways known to related

44 digestion process. We also identify, for the first time, stomach-related proteins from a

2

45 digesting individual and thereby demonstrate that the sensitivity of modern LC-MS/MS

46 equipment allows the identification of gastric juice proteins that are present during digestion

47 thereby providing novel insight into the digestion mechanism.

48 Keywords:

49 Burmese Python, transcriptome, tissue expression, digestion, pathway, proteome

Background

All animals exhibit dynamic changes in the size and functional capacities of bodily organs and tissues to match energetic maintenance costs to the prevailing physiological demands [1]. This phenotypic flexibility is particularly pronounced for the digestive organs in animals that naturally experience prolonged periods of fasting, but capable of ingesting large prey items at irregular intervals. The Burmese python is an iconic example of this extreme phenotype [1]. Many species of pythons easily endure months of fasting, but remain capable of subduing and ingesting very large meals. In Burmese pythons, digestion is attended by a large and rapid rise in mass and/or functional capacities of the intestine, stomach, liver, heart and kidneys [2-4] in combination with a stimulation of secretory processes and an activation of enzymes and transporter proteins. These physiological responses are associated with a many-fold rise in aerobic metabolism. Hence, the Burmese python is an excellent model to study the mechanisms underlying extreme metabolic transitions and physiological remodelling in response to altered demand [1, 3, 5-10].

The postprandial changes in the morphology and physiology of the intestine, heart and other organs have been described in some detail in pythons [1, 5, 8, 9, 11], but only a few studies [12-14] have addressed the underlying transcriptional changes of this interesting biological response. Transcriptome sequencing technology now allows comprehensive surveys [15, 16], and we therefore decided to use transcriptome sequencing of heart, liver, stomach, pancreas and intestine in snakes that had fasted for one month and at 24 and 48h into the postprandial period. These organs were chosen because a number of earlier studies reveal profound phenotypic changes during the postprandial period [1-4, 17], and are therefore likely to exhibit large changes in gene expression. Differential gene expression in some of these organs have previously been reported [12-14], but we provide new data on 48h into the

4

74 digestive period and the first descriptions of gene expression in the stomach and the pancreas.

75 As the Burmese python reference genome assembly [12] ~~currently~~ is relatively fragmented

76 (contig size N50 ~10kb), we found it impractical to use re-sequencing approaches and opted

77 instead to use our high coverage data to build a *de novo* transcriptome assembly to identify

78 differentially expressed genes (DEGs). To identify the enzymes involved in the digestion

79 process, we initiated digestion, then isolated the digestive fluid and characterized the protein

80 composition using a proteomics-based approached. This also allowed us to identify the major

81 hydrolytic enzymes used to digest the large and un-masticated meals.

82 **Analyses**

83 Data summary

84 277,485,924 raw paired reads (2*101 bp, insert size 180 bp) were obtained from Illumina Hi-

85 Seq 2000 sequencing of 15 non-normalized cDNA libraries derived from five tissues (heart,

86 liver, stomach, pancreas and intestine) at three time points (fasted for one month, 24h and 48h

87 post-feeding) and 10 DSN-normalized cDNA libraries (see methods) (Supplementary Table

88 S1). After removal of low-quality reads (See methods), 213,806,111 (77%), high-quality

89 paired reads were retained. These reads contained a total 43,146,073,200 bp nucleotides with

90 a mean Phred quality higher than 37 (Q37). To develop a comprehensive transcriptomics

91 resource for the Burmese python (Fig. 1), we pooled these high-quality reads from 25

92 libraries for subsequent *de novo* assembly.

93 *de novo* transcriptome assembly and evaluation

94 As short k-mers have a higher propensity to generate misassembled transcripts when using a

95 de Bruijn graph-based *de novo* assembler, such as Velvet [18], we conservatively chose an

96 assembly generated using long k-mers for subsequent analysis, at the cost of some sensitivity

97 regarding assembled isoforms. Thus, balancing key metrics (Supplementary Table S2), we

98 used an assembly based on the longest k-mer = 95 (Table 1), as it had the fewest

99 scaffolds/transcripts (34,423), but represented a very large proportion (74%) of all reads. The

100 scaffold N50 of this assembly was 1,673 bp.

101 To evaluate the accuracy of the transcriptome assembly, we compared it with the

102 Burmese python reference genome (GenBank assembly accession: GCA_000186305.2) and

103 corresponding gene set in NCBI database using rnaQUAST v1.4.0 [19]. The transcriptome

104 assembly had 34,423 transcripts in total. 34,040 (98%) of these transcripts had at least one

6

105 significant alignment to the reference genome, and 31,102 (91%) out of 34,040 were uniquely

106 aligned (Supplementary table S3). Average aligned fraction (i.e. total number of aligned

107 bases in the transcript divided by the total transcript length) was 0.975 (Supplementary table

108 S3). The high concordance between the *de novo* transcript assembly and genome reference

109 strengthened our confidence in using *de novo* assembly as our reference, and shows that the

110 individual fragments were accurate although the reference genome assembly is fragmented.

111 By aligning assembled sequences back to reference genome, we checked the chimeric

112 assembled sequences which have discordant best-scored alignment (partial alignments that

113 are either mapped to different strands/different chromosomes/in reverse order/too far away)

114 and found 1,974 (5.7%) misassembled (chimeric) transcripts (Supplementary table S3) which

115 sequences were stored in a supplementary FASTA file. The comparison of assembled

116 sequences and reference gene sequences (Supplementary table S3) showed that 26,320

117 (77.3%) assembled transcripts cover at least one isoform from the reference gene set and the

118 mean fraction of transcript matched is 67.8%, suggesting there is a good concordance but also

119 some differences which can be due to errors in either the reference genome

120 assembly/annotation or our assembly. In addition, we assessed the completeness of our

121 transcriptome assembly with the Benchmarking Universal Single-Copy Orthologs (BUSCO)

122 strategy. Results showed 55.2% (1,428 out of 2,586) complete BUSCOs, 19.8% (512)

123 fragmented BUSCOs and 25% (646) missing BUSCOs. These results are consistent with the

124 survey [20] of assessment completeness of 28 transcriptomes from 18 vertebrates. In this

125 survey, most of transcriptomes from species with close phylogenetic relationship to snake

126 contain less than 50% complete BUSCOs and more than 40% missing BUSCOs. Therefore,

127 we conclude the quality of our transcriptome assembly was ~~well~~ acceptable.

Transcriptome annotation

129 19,713 transcripts (57% of 34,423) were annotated using transfer of blastx hit annotation

7

130 against the non-redundant (nr) NCBI peptide database [21]. To assign proper annotation for

131 each transcript, we chose the first best hit that was not represented in uninformative

132 descriptions (Supplementary Table S4). The most closely related species with an annotated

133 genome, *Anolis carolinensis* was able to annotate 10,704 transcripts (54% of all annotated

134 transcripts). Burmese Python and *Anolis carolinensis* both belong to the reptilian Squamata

135 order, and are separated by approximately 120 million years of evolution [22].

136   Blast2GO [23] then annotated these 19,713 transcripts, and 16,992 of them could be

137 assigned by one or more GO terms and putative functional roles were described. The

138 distributions of the most frequently identified GO terms categories for biological process

139 (BP), molecular function (MF) and cellular component (CC) are shown in Fig. S1. Moreover,

140 we used the functionality of InterPro [24] annotations in Blast2GO to retrieve domain/motif

141 information for our transcripts, and 21,023 transcripts were annotated by the InterPro

142 database.

143 Gene expression analysis and principal component analysis

144 For comparisons between genes, expression profiles were obtained by mapping high quality

145 reads to the reference transcriptome and the expression level was given by fragments per kilo

146 base per million sequenced reads (FPKM) [25]. For the study of expression profiles, we

147 chose to investigate 1862 highly expressed genes (FPKM $>= 100$ in at least one tissue of 15),

148 as it is known that for highly expressed genes, the biological variation among biological

149 replicates in the same tissue at the same stage is lower than for genes showing low expression

150 levels [26]. The majority (~64%) of these 1862 genes were expressed in all tissues, and only

151 ~18% were expressed solely in one tissue (Supplementary Fig. S2). The liver had the highest

152 number of uniquely expressed genes, which may reflect its particular role in metabolism and

153 excretion of waste products.

154   We used principal component analysis (PCA) to reveal overall differences in gene

155 expression patterns among tissues and time points within the digestive period. The first three

156 principal components (PCs) accounted for ~58% of the variation (Supplementary Fig. S3).

157 Despite the large overlap in expressed genes (Supplementary Fig. S2), the different tissues

158 exhibited distinct transcriptional signatures shown by the PCA in Figure 2, showing a

159 tendency for 24h to represent an intermediate position between fasting and 48h. Liver,

160 intestine and stomach displayed greater shifts in the PCA plots compared to heart and

161 pancreas, and the largest changes occurred between fasting and 24h in the stomach and

162 intestine. This fits well with the expectation that the stomach and intestine respond early in

163 digestion [3]. The dramatic changes in gene expression in the liver are also consistent with

164 previous observations on pythons [12].

165 Pattern of transcriptional responses to feeding

166 The postprandial response involves thousands of genes and large changes in gene expression.

167 To restrict the analysis of these many genes, we used a conservative approach where we

168 selected genes that are both highly and differentially expressed with two strict thresholds (see

169 methods). Application of these two thresholds yielded 43 genes for heart, 206 genes for liver,

170 114 genes for stomach, 89 genes for pancreas and 158 genes for intestine, respectively, that

171 were differentially expressed in response to digestion (Fig. 3). To illustrate in greater detail,

172 we enlarged the five sub-clusters with the most prominent increase in expression. These sub-

173 clusters, labelled a - e in Figure 3, are shown with full annotation in Figures 4-8. To unravel

174 the functional implications of these responses, we searched for genes encoding for proteins

175 involved in processes of tissue re-organization, cellular metabolism and digestion within

176 these sub-clusters for each organ.

177 GO enrichment analysis and colored KEGG pathway maps

9

178   To get a broader biological insight, compared to the strict threshold set used in the above

179   clustering analysis, we applied a looser threshold set (Table 2) of defining DEG and highly

180   expressed genes for functional annotation analysis. The summary of number of DEGs during

181   digestion in each tissue is illustrated in Table 3. In each organ, most of genes ($> 76\%$) have

182   low expression (max FPKM < 10). Around 1% of the genes are highly expressed (max

183   FPKM $>= 200$). The number of upregulated genes is approximately 3% in each organ, except

184   for the heart where only 0.57% of the genes were upregulated in response to feeding. This

185   suggests that during digestion, the digestive organs, like liver, stomach, intestine and

186   pancreas show more pronounced post feeding response than the heart. To dissect the

187   functions of DEGs, we performed GO enrichment analysis with upregulated genes and highly

188   expressed genes respectively for each organ (Supplementary Figs. S4-S8). As an example,

189   the most significantly associated GO term to upregulated genes in stomach was

190   "mitochondrial respiratory chain complex 1", "endoplasmic reticulum membrane" and

191   "cytosol" (Supplementary Fig. S4A).

192   To specifically identify the pathways associated to DEGs and highly expressed genes,

193   we mapped genes to KEGG [27, 28] human pathway maps and colored the mapped entries

194   with trend of gene expression during digestion (Table 2). We identified upregulated genes

195   and highly expressed genes, respectively, involved in three selected pathways

196   (glycolysis/gluconeogenesis, citrate cycle (TCA cycle), and oxidative phosphorylation) for

197   each tissue (Supplementary table S5), and we performed the same identification for two main

198   pathway categories in the KEGG pathway database (1.3 lipid metabolism and 1.5 amino acid

199   metabolism; Supplementary table S6). The glycolysis/gluconeogenesis pathway,

200   glyceraldehyde-3 phosphate dehydrogenase showed high expression in all organs.

201   Identification of the python gastric juice proteome

10

202 We identified the secretome of the python stomach during digestion (Fig. 9). The resulting

203 mass spectrometry data (containing 122538 MS/MS spectra) was used to interrogate our

204 python transcriptome database, which includes transcriptome from stimulated stomach tissue.

205 In total, 549 python proteins were identified using this approach. Afterwards, all

206 identifications based on a single tryptic peptide were removed reducing the number of

207 identified python proteins to 314 (Supplementary Table S7).

208 Five classical types of pepsinogens exist, namely pepsinogen A, B, and F,

209 progastricsin (or pepsinogen C), and prochymosin [29]. Of these, our analyses

210 (Supplementary table S8 and S9) show that pythons primarily rely on progastricsin for

211 proteolytic digestion, as the five most abundant proteases identified in the gastric juice are

212 annotated as progastricsin-like. Alignment of the sequences of the various transcripts for

213 gastricsin-like proteins shows considerable differences in sequence, which indicate the

214 presence of numerous different proteins with similar functions. This annotation is based on

215 accession XP_003220378.1 and XP_003220378.1 from *Anolis carolinensis.* Alignment of the

216 python sequences with the two anole sequences, as well as with the well-characterized human

217 gastricsin variant, shows that both the active site residues, as well as cysteine bridges, are

218 conserved. It demonstrates the similarity between these enzymes and suggests that the

219 identified python sequences indeed represent catalytically active proteolytic enzymes

220 (Supplementary Fig. S9). The last identified pepsinogen-like python sequence

221 (m.31615_Py95) was annotated based on the predicted embryonic pepsinogen-like sequence

222 (XP_003220239.1), also from *Anolis carolinensis.* Here, the annotation originates from an

223 embryonic pepsinogen identified in chicken [30]. This protease was identified in the python's

224 gastric juice with a lower emPAI value than the gastricsin sequences indicating a lower

225 concentration of this enzyme (Supplementary table S8), although the transcript displays the

226 highest concentration of the analysed pepsinogens in the post-prandial period (Supplementary

11

227 table S9). As the name indicate it is exclusively expressed during the embryonic period [30,

228 31], and phylogenic analysis of the sequence suggest that its closest homolog, among the

229 classical pepsinogens, is prochymosin [30]. Also, prochymosin displays a temporal

230 expression pattern and is, in mammals, mainly expressed in new-born species. However, the

231 identified python snake embryonic chicken pepsinogen homolog does not display a similar

232 development-related temporal expression pattern and is, as shown, used among adult species

233 for digestion. However, it does not exclude that the protease is expressed during the python's

234 embryonic phase.

235 Identification of prey proteins and the python plasma proteome

236 Many of the obtained MS/MS spectra were expected to correspond to abundant mice

237 proteins, such as collagen. To facilitate the downstream analyses of the python proteins, we

238 produced a list of background proteins related to the prey. Hence, interrogation of the mass

239 spectrometry data against the 16693 mouse protein sequences in the Swiss-Prot database

240 resulted in the identification of 212 mouse proteins, after removing hits based on single

241 peptides (Supplementary table S10). To produce a list of identified python proteins, most

242 likely present in the digestive fluid samples due to blood contaminations during collection,

243 we characterized the python plasma proteome. The most abundant plasma proteins are

244 produced by the liver. Consequently, our python transcriptome sequence database, which

245 encompasses liver transcriptomes, is expected to contain the protein sequences of the python

246 plasma proteins. Thus, our python plasma LC-MS/MS data was used to interrogate our

247 python sequence database. It provided an overview of the most abundant python plasma

248 proteins (Supplementary table S11). In total, 64 plasma proteins were identified with

249 minimum two tryptic peptides. The result supports the liver transcriptome data, since the

250 abundant (based on emPAI) plasma proteins correlate with the transcripts that are detected at

251 high concentration in the liver tissue. The overall protein composition is similar to the

252 composition in humans with albumin, fibrinogen, alpha-2-macroglobulin, immunoglobulins,

253 complement factors and apolipoproteins being the dominating proteins. One protein that

254 stands out is the anti-haemorrhagic factor cHLP-B (m.27_Py95), which appeared in high

255 concentration in the plasma of these snakes. This is a protease inhibitor of the haemorrhagic-

256 causing metalloproteinases present in snake venom and these inhibitors have previously been

257 purified from serum of venomous snakes and thoroughly characterized [32, 33]. The role of

258 such a protease inhibitor in non-venomous pythons is not obvious, but it has been proposed

259 that they inhibit the deleterious action of venom enzymes in non-venomous snakes [34].

260 Identification of the python stomach secretome

261 To identify the python stomach secretome, the list of python proteins, identified in the

262 digestive fluid (Supplementary table S7) was analysed further. We assumed no overlap

263 between abundant plasma proteins and proteins secreted by the stomach. Thus, plasma

264 proteins, identified in the gastric juice, were assumed to be contaminations from blood and

265 therefore the 64 identified plasma proteins were, when present, removed from the list.

266 Subsequently, python proteins that most likely were identified based on prey proteins

267 homology (*e.g.* python collagens and keratins, as well as conserved intracellular household

268 proteins) were removed. These two steps reduced the list of proteins identified in the stomach

269 samples from 314 to 114 proteins (Supplementary table S12). It cannot be excluded that a

270 few proteins belonging to the python stomach secretome also were removed.

271 To identify the secretome, the 114 identified proteins were manually analysed as

272 described in the method section (Supplementary table S12). In addition to household proteins,

273 the identified intracellular proteins also included intracellular stomach-specific proteins (*e.g.*

274 the stomach specific calpain 9 cysteine protease [35]), underlining the specificity of the

13

275 proteomics analysis. In total, 37 proteins constituted the putative python stomach secretome

276 (Supplementary table S8). These could be divided into 18 gastric mucosal-related proteins

277 (e.g. mucin homologous and gastrokine), seven proteolytic enzymes (mainly pepsin

278 homologous), four other hydrolytic enzymes (e.g. phospholipases), and eight other proteins

279 (e.g. gastric intrinsic factor) (Supplementary table S8). Previous gastric juice proteomics

280 analyses were performed on samples obtained from fasting humans, most likely to avoid the

281 complex prey-protein background. In our study, we identify, for the first time, stomach-

282 related proteins from a digesting individual and thereby demonstrate that the sensitivity of

283 modern LC-MS/MS equipment allows the identification of gastric juice proteins that are

284 present during digestion.

14

Discussion

285

286 A primary motivation for our description of the temporal changes in gene expression profiles

287 as the visceral organs of Burmese pythons made the transition from fasting to digestion was

288 to identify key regulatory genes and pathways responsible for the pronounced tissue

289 restructuring and the increased functional capacity during the postprandial period. An equally

290 important motivation was to address specific hypothesis on the upregulation of certain

291 pathways known to be involved in the secretion of digestive juices and enzymes as well as

292 the absorption of the nutrients as digestion proceed. We achieved these goals by identifying

293 the biochemical and physiological roles of the highly expressed genes with increased

294 expression during digestion and by using KEGG analysis of specific pathways underlying

295 physiological responses known to be stimulated by digestion. We also present GO

296 enrichment analyses of both up-regulated genes and highly expressed genes in all organs

297 (Supplementary Figs. S4-S8), showing that "biological process" is the most common

298 enriched category.

299 The influence of digestion on gene expression profiles in heart, liver, kidney and small

300 intestine has been studied previously in pythons [12-14]. These earlier studies reported

301 thousands of genes being either up- or downregulated within the first day of digestion [12-

302 14], and we confirm these substantial changes in gene expression at 24h and 48h. However,

303 we merely identified hundreds of genes, probably because we selected a more stringent

304 threshold for calling the differential expression. Given the differences in the selection of

305 thresholds and analysis strategy for differential expression and differences in times of

306 sampling, it is difficult to make a direct comparison between our study and that of Castoe et

307 al (2013). Nevertheless, for heart, liver and small intestine, both studies have determined a

308 number of upregulated genes at 24h where we identified 15, 93 and 61 upregulated genes,

15

309 respectively. Comparing upregulated genes between two studies (see supplementary material

310 for detailed method and result), we found there was ==good overlap== in identifying upregulated

311 genes in the liver where more than half of the 93 genes identified in our study were identified

312 as upregulated genes by Castoe et al (2013). However, there was less overlap for heart and

313 the small intestine. These differences may be due to the use of different quantification

314 methods for gene expression in the various studies, but may also be a result of the limited

315 biological replicates in our study. ==Nevertheless, genes identified as being upregulated in both==

316 ==studies, are probably of high confidence.==

317 Physiological interpretation of the upregulated genes in the stomach

318 The considerable changes in gene expression in the stomach were reflected in a pronounced

319 rise in expression of ribosomal 40S and 60S proteins (Fig. 4) that is likely to have attended a

320 rise in protein synthesis required for the marked transition from a quiescent fasting state to

321 the activated digestive state. This is also supported by the presence of ribosomal functions in

322 the enriched GO analysis of the stomach of the highly-expressed genes (Supplementary Fig.

323 S4B). During fasting, gastric acid secretion and presumably also the secretion of digestive

324 enzymes and lysozymes, is halted, such that the gastric juice has a neutral pH, whilst

325 ingestion of prey is followed by an immediate activation of gastric acid secretion [36, 37].

326 The stimulation of the secretory actions of the stomach is attended by an increased mass of

327 the stomach, where particularly the mucosa expands already within the first 24h [38].

328 The KEGG analysis, however, shows that the genes encoding for the gastric H,K

329 ATPase, the active and ATP consuming ion-transporter responsible for gastric acid secretion,

330 are highly expressed in fasting animals, and not additionally elevated in the postprandial

331 period (Fig. 10). This strongly indicates that the enzymatic machinery for gastric acid

332 secretion is maintained during fasting, a trait that may enable fast activation of acid secretion,

333 at modest energetic expenditure, to kill bacteria and match gastric pH to the optimum value

334 for pepsin. This interpretation is consistent with a number of recent studies indicating a rather

335 modest contribution of gastric acid secretion to the specific dynamic action (SDA) response

336 in pythons [39, 40], but we also did observe a high prevalence of ATP synthase subunits (Fig.

337 4) amongst the highly upregulated genes, which does indicate a rise in aerobic metabolism

338 (see also supplementary Fig. S4). Furthermore, the upregulation of the gene encoding for

339 creatine kinase (Fig. 4) indicate increased capacity for aerobic respiration required costs of

340 acid secretion and the stimulation of the accompanying gastric functions. It has been

341 proposed that gastric processes account for more than half of the rise in total metabolism

342 during digestion [36], and aerobic metabolism of isolated gastric strips *in vitro* increased

343 during digestion [41]. However, while metabolism of the stomach certainly must increase

344 during the postprandial period, more recent studies indicate a considerably smaller

345 contribution of gastric acid secretion to the total SDA response is considerable lower than

346 50% [39, 40, 42].

347 Our KEGG analysis also showed a large rise in expression of the gene encoding for

348 carbonic anhydrase (Fig. 10), the enzyme that hydrates CO_2 and provide protons for gastric

349 acid secretion. Gastric acid secretion, therefore, does not appear to under transcriptional

350 regulation, but is likely to involve translocation of existing H,K ATPases in vesicles from

351 intracellular vacuoles to the apical membrane of the oxyntopeptic cells that are responsible

352 for both gastric acid secreting as well as the release of pepsinogen in reptiles [43]. An

353 activation of the processes involved in vesicle transport is further supported by increased

354 transcription of the gene encoding for CD63 (Fig. 4), which belongs to the tetraspanin family

355 and mediate signal transduction events.

356 In contrast to acid secretion, expression of several genes encoding for digestive

17

357 enzymes (embryonic pepsinogen-like, gastricsin precursor and gastricsin-like) (Fig. 4) were

358 upregulated, which is consistent with *de novo* synthesis of the enzymes responsible gastric

359 protein degradation. Also, there was good overlap between the upregulation of the relevant

360 genes encoding for the proteins identified in the stomach secretome, such as gastrokines,

361 pepsin homologous, phospholipases and gastric intrinsic factor (Supplementary table S8). In

362 this context, it is also interesting that mucin 6 (Fig. 4), the gene coding for the large

363 glycoprotein (gastric mucin) that protects the gastric mucosa from the acidic and

364 proteolytically active chyme in the stomach lumen was upregulated. Thus, as gastric acid

365 secretion is activated, probably in response to increased levels of the gastrin as well as

366 luminal factors, there is an accompanying activation of the protective mucus layer that

367 prevents auto-digestion of the gastric mucosa. It is also noteworthy that the genes for both

368 gastrokine 1 and 2 were upregulated during digestion (Fig. 4). Gastrokines are constitutively

369 produced proteins in the gastric mucosa in mammals and chickens, and while the

370 physiological function remains somewhat elusive, they appear to upregulated during mucosal

371 remodelling in response to inflammation (*e.g.* in connection with ulcers) and often

372 downregulated in cancers. Thus, it is likely that the gastrokines are involved in regulating the

373 restructuring of the mucosa during digestion in pythons.

374 In addition to analysing the gene expression profiles of the stomach, we also used a

375 proteomics approach, assisted by our python transcriptome sequence database, to identify the

376 hydrolytic enzymes in the gastric juice secreted during digestion. We identified python

377 proteins on a complex background of highly abundant mice proteins. Python's digested food

378 is, when it enters the duodenum, overall similar to digested food in e.g. humans. Thus, the

379 digestive enzymes secreted by the pancreas are probably functional similar to known

380 hydrolytic enzymes from other species. Consequently, the enzymes that facilitate the extreme

381 digestion process and allow for have to be present in the stomach's digestive fluid.

18

382   We hypothesized that relative aggressive proteolytic digestive enzymes in the gastric

383   juice facilitate digestion of large and un-masticated whole prey items [8]. In our analysis, six

384   out of the seven identified proteolytic enzymes were pepsinogens homologous (Peptidase

385   subfamily A1A), and these were also the most abundant hydrolytic enzymes in the gastric

386   juice according to the emPAI values (Supplementary table S8). Most likely other pepsinogen

387   isoforms exist in the gastric juice, as our approach predominantly target the most abundant

388   proteolytic enzymes. The importance of the proteomics-identified pepsinogens was also

389   substantiated by the transcriptomics data (Supplementary table S9). Here, we found that the

390   six different pepsinogens were upregulated between 2.2 and 22.2 fold from the fasting

391   animals to 48 hours after ingestion of mice. In average the pepsinogen transcripts were

392   upregulated 10.7 fold. It supports that these proteases play a substantial role in the aggressive

393   digestion process performed by the python.

394   Our proteomic analysis also suggested the identification of the pepsinogens as the

395   major digestive proteolytic enzymes is similar to all other vertebrate species. Thus, our

396   results indicate that it is not unique (with respect to protease class) and hitherto

397   uncharacterized proteases that facilitate the aggressive digestion process. Instead, pepsins,

398   homologous to pepsins among other species, digest the intact swallowed prey. The general

399   condition in the stomach during digestion (e.g. pH) is also similar to other species. Thus, it is

400   likely that these pepsins variants are among the most effective and aggressive pepsins

401   identified so far and the provided sequence information facilitate future cloning, expression,

402   and characterization of these potential industrial relevant enzymes.

403   Physiological interpretation of the upregulated genes in the intestine

404   The small intestine of pythons undergoes a remarkable and fast expansion during digestion

405   where both wet and dry mass more than doubles within the first 24 hours. The expansion

406   stems primarily from increased mucosal mass, achieved by swelling of the individual

407   enterocytes [44], while the smooth muscle in the gut wall is much less responsive [45].

408   Earlier studies on gene expression profiles during digestion in the python intestine revealed

409   massive upregulation of more than one thousand genes, commencing within the first six

410   hours after ingestion [12, 13]. Importantly, this previous study [13] identified a number of

411   genes that are likely to be involved in the restructuring of the microvilli, cell division and

412   apoptosis, as well as brush-border transporter proteins. In line with these earlier findings, our

413   GO enrichment analysis also highlights functions pertaining to mitotic cell division, which

414   supports a contribution to growth by hyperplasia faster cell turnover (Supplementary Fig. S5).

415   The expansion of the individual enterocytes is accompanied by pronounced elongation of the

416   microvilli [46] and the resulting rise in surface area of the intestinal lining is accompanied by

417   an ten-fold increase in intestinal transport capacity for amino acids and other nutrients [1, 4,

418   47].

419   Earlier studies provided strong evidence for an upregulation of genes coding for nutrient

420   transporter proteins, such as D-glucose, L-proline and L-leucine [13]. In this context, it is

421   noteworthy that there were no nutrient transporters amongst the highly expressed and

422   upregulated genes in the intestine (Fig. 5), but our KEGG analysis nevertheless showed

423   increased expression of the serosal L-type amino acid transporter. Clearly, it would be

424   worthwhile to quantitatively analyse the extent to which *de novo* synthesis of the various

425   nutrient transporters, particularly those for amino acids, is increased during digestion and

426   how much such synthesis contribute to absorptive capacity. It would seem adaptive if many

427   of the transporters merely have to be activated, either by insertion within the luminal

428   membrane or exposed as the enterocytes expand, to allow for an energetically cheap manner

429   of matching intestinal performance to the sudden appearance of nutrients in the intestine after

430   a meal. The GO enrichment analysis also pointed to an enrichment of various metabolic

20

431 processes during digestion, particularly for the upregulated genes (Supplementary Fig. S5). It

432 is noteworthy that the expression of genes for glutathione S-transferase, peroxiredoxin and

433 selenoprotein increased during digestion (Fig. 5). These three proteins are involved in cellular

434 defence, particularly as antioxidants as a likely protection of reactive oxygen species

435 resulting from increased aerobic metabolism.

436 There is consensus that the anatomical and structural responses underlying this

437 phenotypic flexibility of intestinal function occur at modest energetic expenditure [17, 36,

438 48], but our expression profile does show increased expression of the gene coding for

439 Cytochrome P450 pointing to increased aerobic and mitochondrial metabolism. An increased

440 expression of genes involved in oxidative phosphorylation was also reported in earlier studies

441 on pythons [12, 13]. This rise in metabolism may be driven primarily by the massive rise in

442 secondary active transport to absorb the amino acids and smaller peptides rather than the

443 structural changes [48]. Nevertheless, the structural changes may be reflected in increased

444 expression of galectin 1 (Fig. 5), which mediate numerous function including cell–cell

445 interactions, cell–matrix adhesion and transmembrane signalling.

446 Fig. 5 reveals the importance of lipid absorption and the subsequent transport by the

447 cardiovascular and lymph systems, and it is also possible that several of the expressed

448 proteins play a role in the incorporation of lipid droplets within the enterocytes. Thus, the

449 presence of numerous apolipoproteins, and their precursor apoe protein, amongst the list of

450 highly expressed and highly expressed genes (Fig. 5) are probably needed to transport the

451 absorbed lipids in plasma and lymph, but the apolipoproteins could also act enzyme

452 cofactors, receptor ligands, and lipid transfer carriers in the regulation of lipoprotein

453 metabolism and cellular uptake. Diazepam-binding inhibitor (Fig. 5), a protein involved in

454 lipid metabolism and under hormonal regulation mostly within nervous tissue, is also likely

455 to reflect the increased lipid absorption and metabolism in the postprandial period, and there

456 was also a rise phospholipases (Fig. 5) that are likely to be involved in lipid degradation.

457 Also, the capacity for protein metabolism clearly increased in the intestine during digestion

458 (meprin A and endopeptidase that cleaves peptides, as well as 4-aminobutyrate

459 aminotransferase, 4-trimethylaminobutyraldehyde dehydrogenase and diamine

460 acetyltransferase) and there was a rise in the ammonium transporter protein Rh (Fig. 5).

461 Finally, a number of proteins involved in calcium uptake and metabolism, such as calbindin

462 and calmodulin (Fig. 5), could be important to handle the break-down of the bone in a normal

463 rodent, and it was recently shown the enterocytes of pythons contain small particles of bone

464 already 24 hours after ingestion [46].

465 Physiological interpretation of the upregulated genes in the heart

466 The large metabolic response to digestion is tailored by a doubling of heart rate and stroke of

467 the heart such that cardiac output remains elevated for many days during digestion [49, 50].

468 This cardiovascular response plays a pivotal role in securing adequate oxygen delivery to the

469 various organs and serves to ensure an appropriate convective transport of the nutrients taken

470 up by the intestine. The tachycardia is mediated by a release of vagal tone and the presence of

471 a non-adrenergic-non-cholinergic stimulation of the heart, which has been speculated to be

472 released from the gastrointestinal organs during digestion [51, 52]. The increased heart rate,

473 and the rise in the amount of blood pumped with each beat, must be supported by increased

474 metabolism of the myocardium and we observed an upregulation of malate dehydrogenase,

475 cytochromes and ATPase linked enzymes (Fig. 6) that are likely to be related to an increased

476 oxidative phosphorylation within the individual myocytes (see also the prevalence of

477 enriched GO terms associated with aerobic metabolism in Supplementary Fig. S8). Previous

478 gene expression studies on the python heart also yielded evidence for increased oxidative

22

479 capacity in postprandial period [53] and cytochrome oxidase activity is almost doubled

480 during digestion [54], and we confirm that transcription for heat shock proteins may be

481 increased [53], possibly to protect against oxidative damage as result of the increased

482 metabolism. As in earlier studies [53], our observation of increased ATP synthase

483 lipid−binding protein and fatty acid binding protein 3 (Fig. 6) provide evidence for increased

484 fatty acid metabolism, which may reflect the substantial rise in circulating fatty acids in the

485 plasma.

486   It was originally suggested that the postprandial rise in stroke volume could be

487 ascribed to an impressive and swift growth of the heart [10], possibly triggered lipid-

488 signalling [53]. However, a number of recent studies, primarily from our laboratory, have

489 shown that increased cardiac mass is not an obligatory postprandial response amongst

490 pythons [54-56], and that stroke volume may be increased in response to increased venous

491 return rather than cardiac hypertrophy [54]. It is nevertheless, noteworthy that our and the

492 previous studies show a clear increase in the expression of contractile proteins (e.g. myosin

493 and actin) as well as tubulin (Fig. 6), which may reflect increased protein-turnover in

494 response to increased myocardial workload rather than cell proliferation or hypertrophy. The

495 enriched GO analyses also point to major changes in the extracellular space as well as both

496 elastin and collagen, which may indicate some level of cardiac reorganization at the cellular

497 or subcellular level that may alter compliance of the myocardial wall and influence cardiac

498 filling (Supplementary Fig. S8). It is noteworthy that the increased expression of BNP may

499 serve a signalling function as described in response to the cardiac hypertrophy that attends

500 hypertension.

501 Physiological interpretation of the genes in the liver

502 The liver exhibited a diverse expression profile in response to digestion that is likely to

23

503 reflect its many metabolic functions in connection with metabolism, synthesis and

504 detoxification during the postprandial period. This pattern is also evident from the many

505 metabolic functions identified in the enriched GO analysis (Supplementary Fig. S7). There

506 were marked upregulations of the P450 system (Fig. 7), which stems well with a rise in

507 synthesis and breakdown of hormones and signalling molecules, cholesterol synthesis in

508 response to lipid absorption and possibly also an increased metabolism of potentially toxic

509 compounds in the prey. A rise in cholesterol metabolism was supported by increased

510 expression apolipoproteins (Fig. 7). The hepatic involvement in lipid metabolism was also

511 supported by the increased expression of genes for Alpha-2-macroglobulin and serum

512 albumin (Fig. 7). The increased expression of albumin obviously also fits nicely with the

513 proteomic analysis of plasma proteins and it is likely that the postprandial rise in plasma

514 albumin serves a functional role in the lipid transport between the intestine and the liver as

515 well as other metabolically active organs

516   It is also noteworthy that a number of genes associated with the protection of

517 oxidative stress, such as catalase, heat shock protein and glutathionine transferase were

518 markedly upregulated (Fig. 7). It was recently argued that snakes digesting large meals

519 experience oxidative damage due to reactive oxygen metabolites requiring increased

520 antioxidant responses to protect cellular functions [57].

521 Physiological interpretation of the genes in the pancreas

522 We sampled the entire pancreas for our analysis of gene expression and our data therefore

523 reflect both endocrine and exocrine pancreatic functions. The vast majority of the upregulated

524 genes concerned the exocrine pancreas, and we found ample evidence for upregulated

525 expression of genes associated with the digestive functions, such as lipases, trypsin,

526 chymotrypsin and elastase and other enzymes for digestion of protein and lipid (Fig. 8). This

24

527 general upregulation of secretory processes is likely to explain the prevalence of processes

528 associated with protein synthesis in the enriched GO analysis (Supplementary Fig. S6). There

529 was even an increased expression of amylase (Fig. 8) that breaks down polysaccharides. In

530 connection with this latter function, the increased expression of insulin (Fig. 8) from the

531 endocrine pancreas is likely to reflect increased cellular signalling for postprandial uptake of

532 both glucose and amino acids. As in the other organs, we found increased expression of

533 cytochrome oxidase (Fig. 8) indicative of increased metabolism during digestion, and the rise

534 in heat shock protein expression may reflect a response to formation of reactive oxygen-

535 species as metabolism is stimulated by increased secretion of the pancreas.

Conclusions

537 Our study confirms that the extensive physiological and anatomical reorganization of the

538 visceral organs of pythons during the postprandial period is driven by differential expression

539 of hundreds or even thousands of genes. Many of the upregulated functions pertain to energy

540 production to support the rise in aerobic metabolism associated with digestion and absorption

541 of the large meals. In terms of the gastrointestinal organs, the gene expression profiles also

542 support the view that many of the digestive functions, such as gastric acid secretion and

543 nutrient absorption, can be stimulated with little gene expression indicating that the proteins

544 involved in these processes are merely need to be activated during the postprandial period,

545 and thus avoiding the energy and time-consuming processes associated with *de novo*

546 synthesis. This digestive strategy may, at least in part, explain how intermittent feeders, such

547 as snakes, retain the capacity for fast and reliable upregulation of the digestive processes

548 immediately after ingestion.

26

Methods

Stimulation of the postprandial response, collection of tissue biopsies and purification of

RNA for mRNA-seq analyses

Six *Python molurus* (Tiger Python/Burmese Python) with a body mass ranging from 180 to

700 g (average 373 g) were obtained from a commercial supplier and housed in vivaria with a

heating system providing temperatures of 25-32 °C. The animals were fed rodents once a

week and fresh water was always available. The animals appeared healthy and all

experiments were performed according to Danish Federal Regulations. All six individuals

were fasted for one month and divided in three groups. Four animals were fed a rodent meal

of 25 % of body weight and euthanized with an intra-peritoneal injection of pentobarbital (50

mg kg^{-1}; Mebumal) at 24h (N = 2) or 48h after feeding (N = 2). The remaining two snakes

served as fasted controls. During deep anaesthesia, two biopsies were obtained from each

snake from each of the following tissues: The heart (ventricles), liver, stomach, intestine, and

pancreas. In regard to the stomach tissue samples, one sample was obtained from the

proximal part of the stomach and one sample was obtained from the distal part. In total, 60

biopsies were collected. The samples were taken from the same part of the different tissues in

all individuals. After sampling, the biopsies were weighted and immediately snap frozen in

liquid nitrogen; stomach and intestinal tissues were rinsed in sterile saline solution before

weighting to avoid contamination with rodent tissue from the ingested meal. Subsequently,

all 60 biopsies were homogenized in liquid nitrogen and the four biological replicates (two

biopsies from each individual) were pooled in a 1:1 manner based on mass. It resulted in 15

samples (five tissues X three time points). From these samples, total RNA was purified using

the Nucleospin RNA II kit (Machery-Nagel GmbH & Co.), as recommended by the

manufacturer. The RNA concentration and quality were assessed by Nanodrop ND 1000

27

573 Spectrophotometer (Thermo Scientific) analyses, agarose gel-electrophoreses, and Agilent

574 BioAnalyzer (Agilent) analyses.

575 Library production and sequencing

576 Poly-A transcripts were enriched and the transcripts broken in the presence of Zn^{2+}.

577 Subsequently, double-stranded cDNA was synthesized using random primers and RNase H.

578 After end repair and purification, the fragments were ligated with bar-coded paired-end

579 adapters, and fragments with insert sizes of approximately 150-250 bp were isolated from an

580 agarose gel. Each of the 15 samples derived from five tissues (heart, liver, stomach, pancreas

581 and intestine) at the three time points (fasted for one month, 24h and 48h post-feeding) were

582 amplified by PCR to generate DNA colonies template libraries and the libraries were then

583 purified. In addition, to sample as broadly from transcriptome as possible, we also produced

584 normalized libraries for each tissue in order to capture the reads from lowly expressed, tissue-

585 specific genes. Here, a part of the samples, which originating from the same tissue, were

586 pooled before the PCR analyses, i.e. in total five pooled samples were generated. These five

587 samples were split in two and after PCR amplification and library purification they were

588 normalized using two different normalization protocols, i.e. in total 10 normalized libraries

589 were prepared. Library quality of all 25 samples was then assessed by a titration-run (1 x 50

590 bp) on an Illumina HiSeq 2000 instrument. Finally, the sequencing was performed on the

591 same instrument using paired-reads (2×101 bp). One channel was used for the 15 non-

592 normalized libraries and one channel was used for the 10 normalized libraries.

593 Data pre-processing and *de novo* transcriptome assembly

594 To reduce the amount of erroneous data, the raw paired reads were processed by i) removing

595 reads that contained the sequencing adaptor, ii) removing reads that contained ambiguous

596 characters (Ns), and iii) trimming bases that had ~~the~~ low average quality (Q<20) within a

597 sliding window of length 10.

598　　To develop a comprehensive transcriptomics resource for the Burmese python, all

599 high-quality reads from 25 libraries were pooled together for *de novo* assembly. To determine

600 the optimal assembly, *de novo* assembly was performed using Velvet（version 1.2.03) [18]

601 and Oases (version 0.2.06) [58] with different k-mer parameters. The performance of these

602 assemblies was assessed according to number of transcripts, total length of transcripts, N50

603 length, mean length, proportion of mapped reads and number of transcripts which length is

604 larger than N50 (Supplementary Table S2).

605 Assessment of the transcriptome assembly

606 The transcriptome assembly was evaluated by rnaQUAST 1.4.0 with default parameters

607 supplying reference genome sequences and genome annotation of Burmese python (GenBank

608 assembly accession: GCA_000186305.2).

609　　BUSCO_v2 [20] was used to test the completeness of transcriptome assembly with

610 dependencies NCBI BLAST+ 2.4.0 [59] and HMMER 3.1b2 [60]. The vertebrata lineage set

611 was used and accessed on 28 Nov 2016.

612 Transcriptome annotation

613 To assess the identity of the most closely related gene in other organisms, the assembled

614 transcripts were compared with the sequences in the National Center for Biotechnology

615 Information (NCBI) non-redundant protein (nr) database using blastx [61] with an e-value

616 cut-off of 0.01. The nr annotation term of each transcript was assigned with the first best hit,

617 which was not represented in uninformative description (e.g., 'hypothetical protein', 'novel

618 protein', 'unnamed protein product', 'predicted protein' or 'Uncharacterized protein')

29

619 (Supplementary Table S4). To assign functional annotations of transcripts, Blast2GO was

620 used (e-value threshold = 0.01) to return GO annotation, Enzyme code annotation with

621 KEGG maps and InterPro annotation.

622 Estimation of gene expression values

623 For each 15 non-normalized libraries, the paired-end reads were firstly mapped back to

624 assembled transcriptome using Bowtie2 [62]with default parameters, the raw counts then

625 were calculated based on the alignment results using RSEM (version 1.1.20) [63] for each

626 transcript. To quantify the gene expression level, for genes with alternative splicing

627 transcripts, the longest transcript was selected to represent the gene, and a gene's abundance

628 estimate was the sum of its transcripts' abundance estimates. Finally, the raw expression

629 counts were normalized into FPKM with custom Perl scripts.

630 PCA

631 To facilitate graphical interpretation of tissue relatedness, R function prcomp was used to

632 perform PCA with genes which the maximum FPKM of 15 samples was greater than 100.

633 Identification of DEGs and clustering analysis

634 For each tissue, DEGs were selected with two thresholds, 1) FPKM is greater than or equal to

635 400 in at least one time point and 2) fold change (FC) is greater than or equal to two in at

636 least one pairwise comparison among three time points. FPKM values of DEGs were log2-

637 transformed and median-centered, then hierarchical clustering was performed using R

638 command hclust with method = 'average' and distance = 'Spearman correlation' and results

639 were displayed using R command heatmap.2.

640

30

641   Colored KEGG Pathway and GO enrichment analysis

642   For each tissue, all assembled genes were mapped to KEGG human pathway maps using

643   KOBAS 2.0 [64] with e-value 1e-50. Then genes were colored by representing FPKM value

644   and trend of differential expression value (Table 2).

645   Blast2GO was used to implement GO enrichment analysis (Fisher's exact test) with

646   threshold of FDR 0.001. The reference set is the whole transcripts with GO slim annotation.

647   For each organ, the selected test set is either upregulated or highly expressed genes defined in

648   Table 2. Finally, we performed Blast2GO to reduce to most specific GO terms.

649   Isolation of samples for proteomics analyses

650   Two Burmese pythons (weighing 400 and 800 g, respectively) were fed a rodent meal

651   corresponding to approximately 25% of their body mass. Approximately 24 h into the

652   postprandial period the animals were euthanized with an overdose of pentobarbital (100 mg

653   kg^{-1}, i.m.). Immediately afterwards, an incision was made to expose the stomach, which was

654   then ligated at the lower oesophagus and the pylorus, before the intact stomach was excised

655   by a cleavage just below the two sutures resulting in the stomach being released from the rest

656   of the animal. All undigested mouse remains were manually removed by forceps and 25

657   ml/kg tris-buffered saline (TBS) was injected into the stomach. The stomach was then ligated

658   at the opened end, rinsed by gently shaking the tissue, and finally the digestive fluid-

659   containing solution was collected and stored on ice. To ensure collection of all gastric fluid,

660   the stomach was rinsed additional two-three times with 12 ml/kg TBS. Subsequently, the

661   samples were filtered and centrifuged, and the supernatant stored at -80 °C. We also obtained

662   two samples of gastric juice from a third individual (200 g) that had been fed 4 g peptone

663   (Sigma Aldrich), suspended in water. Peptone is a mixture of small peptides and amino acids

664 and the solution was injected directly into the stomach and after three hours the snake was

665 euthanized by an overdose of pentobarbital. The stomach was removed, rinsed with TBS, and

666 a single sample collected and stored, as described above. We analysed two samples from each

667 of the three individuals, resulting in a total of six digestive fluid samples being analysed by

668 MS/MS. In addition, we obtained a single plasma sample from each snake by direct cardiac

669 puncture followed by centrifugation and storage for later analysis.

670 Sample preparation for mass spectrometry analyses

671 The proteins in the six obtained python digestive fluid samples were recovered by

672 trichloroacetic acid precipitation. The resulting pellets were resuspended in 8 M Urea, 5 mM

673 DTT, 0.1 M ammonium bicarbonate pH 8.0 and incubated for 30 minutes at room

674 temperature in order to denature and reduce the proteins. Subsequently, the proteins were

675 alkylated by the addition of iodoacetamide to a final concentration of 25 mM. The samples

676 were incubated for additional 20 minutes at room temperature and then diluted five times

677 with a 50 mM ammonium bicarbonate, pH 8.0 buffer before the addition of approximately 2

678 μg sequencing grade modified trypsin (Promega) per 50 μg protein in the sample.

679 Subsequently, the samples were incubated at 37 °C for approximately 16 h. The proteins in

680 the plasma sample were denatured, reduced, alkylated, and digested with trypsin, as described

681 for the digestive fluid samples. Finally, the resulting peptides in all samples were

682 micropurified and stored at -20 C until the LC-MS/MS analyses.

683 Liquid chromatography-tandem mass spectrometry analyses

684 Nano-liquid chromatography-tandem mass spectrometry (LC–MS/MS) analyses were

685 performed on a nanoflow HPLC system (Thermo Scientific, EASY-nLC II) connected to a

686 mass spectrometer (TripleTOF 5600, AB Sciex) equipped with an electrospray ionization

687 source (NanoSpray III, AB Sciex) and operated under Analyst TF 1.6 control. The samples

688 were dissolved in 0.1% formic acid, injected, trapped and desalted isocratically on a

689 precolumn whereupon the peptides were eluted and separated on an analytical column (16 cm

690 × 75 μm i.d.) packed in-house with ReproSil-Pur C18-AQ 3 μm resin (Dr. Marisch GmbH).

691 The peptides were eluted at a flow rate of 250 nL/min using a 50 min gradient from 5 % to 35

692 % phase B (0.1 % formic acid and 90 % acetonitrile). An information dependent acquisition

693 method was employed allowing up to 25 MS/MS spectra per cycle of 2.8 s.

694 Protein identification and filtering of data

695 The six collected MS files, related to digested fluid, were converted to Mascot generic format

696 (MGF) using the AB SCIEX MS Data Converter beta 1.3 (AB SCIEX) and the "proteinpilot

697 MGF" parameters. Subsequently, the files were merged to a single MGF-file using Mascot

698 daemon. The resulting file (encompassing 122538 MS/MS queries) was used to interrogate

699 the 16693 *Mus musculus* sequences in the Swiss-Prot database (version 2014_10) and the

700 generated python database encompassing 21131 protein sequences using Mascot 2.5.0

701 (Matrix Science)[65]. Trypsin, with up to one missed cleavage allowed, was selected as

702 enzyme; carbamidomethyl was employed as fixed modification, and oxidation of methionine

703 and proline was selected as variable modifications. The instrument setting was specified as

704 ESI-QUAD-TOF, the mass accuracy of the precursor and product ions was 15 ppm and 0.2

705 da respectively, and the significance threshold (p) was set to 0.01 and an expect cut-off at

706 0.005. The data obtained by the LC-MS/MS-analysis of the python plasma proteome was

707 analysed as described for the digestive fluid samples, except that the *Mus musculus* sequences

708 were not interrogated. This dataset contains 9224 MS/MS queries. All obtained results were

709 subsequently parsed using MS Data Miner v. 1.3.0 [66], and protein hits were only accepted

710 if they were identified based on two unique peptides. Semi-quantitative proteomics data was

711 obtained using the emPAI-values given by the Mascot 2.5.0 software after analysis of the

712 MS/MS data [67].

713   To identify the proteins secreted into the python stomach, identified python plasma

714 proteins, as well as the mouse protein homologs were removed from the list of identified

715 python digestive fluid proteins. With regard to the removal of prey protein homologs, the

716 overall mouse protein names were used to search the list of python proteins (e.g. "collagen"

717 was used as search term, not "collagen alpha-1(I) chain") and to identify python proteins that

718 were identified based on homology with mouse. These proteins were removed from the list of

719 stomach-secreted python proteins. For each identified protein remaining on the list, we

720 reassessed the annotation of the python sequence, i.e. sequence comparisons were performed

721 using blastp version 2.2.30, and in addition, UniProt and NCBI protein databases, as well as

722 PubMed and SignalP 4.1, were interrogated to identify functional properties and cellular

723 location of the identified proteins. Plasma proteins, remaining collagen homologous,

724 intracellular proteins, and membrane proteins were discarded from the list of identified

725 python stomach secretome proteins.

726 **List of abbreviations**

727 DEG   differentially expressed genes

728 FC   fold change

729 FPKM fragments per kilo base per million sequenced reads

730 PCA   principal component analysis

731 **Declarations**

732 Ethics approval and consent to participate

733 Not applicable

734

735 Consent for publication

736 Not applicable

737

738 Availability of data and material

742

743 Competing interests

744 The authors declare that they have no competing interests.

745

746 Funding

749

Authors' contributions

JD, KWS, TW and MHS designed the study. JD performed the transcriptome data analysis with input from LS and was a major contributor in writing the manuscript. SEL performed RNA-Seq lab experiment. KWS and JE performed the proteomics experiment and data analysis. WT interpreted the transcriptome data regarding digestion. All authors read and approved the final manuscript.

756

Acknowledgements

Figure and table legends

761 **Fig. 1. The workflow of Python RNA-Seq data analysis**. The diagram shows the main

762 steps and bioinformatics tools used in the study.

763 **Fig. 2. PCA plots of FPKM of 1862 genes.** PC, principal component. PC1 represents 25%,

764 PC2 represents 18% and PC3 represents 16% of total variation in the data. The name of the

765 label consists two part, one capital letter plus one number. Letter H, S, I, L, P represent

766 heart, stomach, intestine, liver and pancreas respectively. Number 0, 1, 2 represent fasting

767 for one month, 24h/1d after feeding and 48h/2d after feeding respectively.

768 **Fig. 3. Heat maps from hierarchical clustering of DEGs in each tissue.** Heat maps

769 showing the hierarchically clustered Spearman correlation matrix resulting from comparing

770 the normalized FPKM value for each pair of genes. Heat map columns represent samples

771 and rows correspond to genes. Expression values (FPKM) are $\log_2$-transformed and then

772 median-centered by gene. Relative levels of gene expression are represented by colors.

773 Pale colour is low expression and darker blue is high expression. Five sub-clusters labelled

774 a to e are shown with full annotation in Fig. 4-8.

775 **Fig. 4. The cluster of upregulated genes with NCBI nr annotation in stomach**. It shows

776 the cluster e in Fig. 3. Heat map columns represent samples and rows correspond to genes.

777 Expression values (FPKM) are $\log_2$-transformed and then median-centered by gene.

778 Relative levels of gene expression are represented by colors. Pale colour is low expression

779 and darker blue is high expression.

780 **Fig. 5. The cluster of upregulated genes with NCBI nr annotation in intestine**. It shows

781 the cluster b in Fig. 3. Heat map columns represent samples and rows correspond to genes.

782 Expression values (FPKM) are $\log_2$-transformed and then median-centered by gene.

783 Relative levels of gene expression are represented by colors. Pale colour is low expression

784 and darker blue is high expression.

785 **Fig. 6. The cluster of upregulated genes with NCBI nr annotation in heart**. It shows the

786 cluster a in Fig. 3. Heat map columns represent samples and rows correspond to genes.

787 Expression values (FPKM) are $\log_2$-transformed and then median-centered by gene.

788 Relative levels of gene expression are represented by colors. Pale colour is low expression

789 and darker blue is high expression.

790 **Fig. 7. The cluster of upregulated genes with NCBI nr annotation in liver.** It shows the

791 cluster c in Fig. 3. Heat map columns represent samples and rows correspond to genes.

792 Expression values (FPKM) are $\log_2$-transformed and then median-centered by gene.

793 Relative levels of gene expression are represented by colors. Pale colour is low expression

794 and darker blue is high expression.

795 **Fig. 8. The cluster of upregulated genes with NCBI nr annotation in pancreas.** It shows

796 the cluster d in Fig. 3. Heat map columns represent samples and rows correspond to genes.

797 Expression values (FPKM) are $\log_2$-transformed and then median-centered by gene.

798 Relative levels of gene expression are represented by colors. Pale colour is low expression

799 and darker blue is high expression.

800 **Fig. 9. The workflow used to identify the python's stomach secretome during**

801 **digestion. 1)** Initially pythons were feed with mice, or a peptide mixture, and later the gastric

802 juice samples were obtained and mice debris were removed. **2)** The proteins were

803 precipitated, denatured and digested with trypsin. **3)** The resulting tryptic peptides were

804 analysed by LC-MS/MS analyses and the data merged into a single file. **4)** The file was used

805 to interrogate the in-house generated python protein sequence database (based on the

806 transcriptomic data) and python proteins were identified. **5)** The data was filtered to remove

807 mice proteins and plasma proteins. Subsequently, the annotation of the remaining proteins

808 was reassessed and the secretome identified.

809 **Fig. 10. Cartoon depiction of colored KEGG pathway of gastric acid secretion in**

810 **stomach**. Entry in red represents upregulated during digestion; Entry in purple for highly

811 expressed. H/K is H+/K+-exchanging ATPase alpha polypeptide. CA is carbonic anhydrase.

812 AE is solute carrier family 26 (anion exchange transporter).

813 **Table 1. Summary of transcriptome assembly of Burmese Python.**

Table 2. Colour coding of genes in KEGG pathway maps. Three criteria are used to classify and colour genes. First, i) whether the maximum FPKM of the gene among fasting, 24h and 48h is over 10, then ii) whether the gene is differential expressed in at least one of the pairwise comparison among fasting, 24h and 48h with FC over 4. Finally, iii) for those genes expressed, but not differential expressed, whether it is highly expressed with maximum FPKM among three time points over 200. The term expression trend indicates the trend of gene expression across fasting, 24h and 48h. e.g. The trend up means the gene is upregulated from either fasting to 24h, fasting to 48h or 24h to 48h. The trend up-then-down means the gene is firstly upregulated from fasting to 24h, then downregulated from 24h to 48h.

Table 3. The number of DEGs across fasting, 24h and 48h in each tissue. The expression trend is consistent with definition in Table 2.

826 **Reference**

827　1.　Secor SM, Diamond J: **A vertebrate model of extreme physiological**
828　　　**regulation.** *Nature* 1998, **395:**659-662.

829　2.　Cox CL, Secor SM: **Matched regulation of gastrointestinal performance in the**
830　　　**Burmese python, Python molurus.** *J Exp Biol* 2008, **211:**1131-1140.

831　3.　Secor SM, Diamond J: **Adaptive responses to feeding in Burmese pythons:**
832　　　**pay before pumping.** *J Exp Biol* 1995, **198:**1313-1325.

833　4.　Secor SM, Diamond JM: **Evolution of regulatory responses to feeding in**
834　　　**snakes.** *Physiological and Biochemical Zoology* 2000, **73:**123-141.

835　5.　Aubret F, Shine R, Bonnet X: **Evolutionary biology: adaptive developmental**
836　　　**plasticity in snakes.** *Nature* 2004, **431:**261-262.

837　6.　Cohn MJ, Tickle C: **Developmental basis of limblessness and axial patterning**
838　　　**in snakes.** *Nature* 1999, **399:**474-479.

839　7.　Ott BD, Secor SM: **Adaptive regulation of digestive performance in the genus**
840　　　**Python.** *J Exp Biol* 2007, **210:**340-356.

841　8.　Secor SM: **Digestive physiology of the Burmese python: broad regulation of**
842　　　**integrated performance.** *J Exp Biol* 2008, **211:**3767-3774.

843　9.　Zaar M, Overgaard J, Gesser H, Wang T: **Contractile properties of the**
844　　　**functionally divided python heart: two sides of the same matter.** *Comp*
845　　　*Biochem Physiol A Mol Integr Physiol* 2007, **146:**163-173.

846　10.　Andersen JB, Rourke BC, Caiozzo VJ, Bennett AF, Hicks JW: **Physiology:**
847　　　**postprandial cardiac hypertrophy in pythons.** *Nature* 2005, **434:**37-38.

848　11.　Vidal N, Hedges SB: **Molecular evidence for a terrestrial origin of snakes.**
849　　　*Proc Biol Sci* 2004, **271 Suppl 4:**S226-229.

850　12.　Castoe TA, de Koning AP, Hall KT, Card DC, Schield DR, Fujita MK, Ruggiero RP,
851　　　Degner JF, Daza JM, Gu W, et al: **The Burmese python genome reveals the**
852　　　**molecular basis for extreme adaptation in snakes.** *Proc Natl Acad Sci U S A*
853　　　2013, **110:**20645-20650.

854　13.　Andrew AL, Card DC, Ruggiero RP, Schield DR, Adams RH, Pollock DD, Secor SM,
855　　　Castoe TA: **Rapid changes in gene expression direct rapid shifts in intestinal**
856　　　**form and function in the Burmese python after feeding.** *Physiol Genomics*
857　　　2015, **47:**147-157.

858　14.　Castoe TA, Fox SE, Jason de Koning A, Poole AW, Daza JM, Smith EN, Mockler TC,
859　　　Secor SM, Pollock DD: **A multi-organ transcriptome resource for the**
860　　　**Burmese Python (Python molurus bivittatus).** *BMC Res Notes* 2011, **4:**310.

861　15.　Wang Z, Gerstein M, Snyder M: **RNA-Seq: a revolutionary tool for**
862　　　**transcriptomics.** *Nat Rev Genet* 2009, **10:**57-63.

863　16.　Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B: **Mapping and**
864　　　**quantifying mammalian transcriptomes by RNA-Seq.** *Nat Methods* 2008,
865　　　**5:**621-628.

866　17.　Wang T, Hung CC, Randall DJ: **The comparative physiology of food**
867　　　**deprivation: from feast to famine.** *Annu Rev Physiol* 2006, **68:**223-251.

868　18.　Zerbino DR, Birney E: **Velvet: algorithms for de novo short read assembly**
869　　　**using de Bruijn graphs.** *Genome Res* 2008, **18:**821-829.

870　19.　Bushmanova E, Antipov D, Lapidus A, Suvorov V, Prjibelski AD: **rnaQUAST: a**
871　　　**quality assessment tool for de novo transcriptome assemblies.**
872　　　*Bioinformatics* 2016, **32:**2210-2212.

873   20.   Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM: **BUSCO:**
874   **assessing genome assembly and annotation completeness with single-copy**
875   **orthologs.** *Bioinformatics* 2015:btv351.
876   21.   Pruitt KD, Tatusova T, Maglott DR: **NCBI reference sequences (RefSeq): a**
877   **curated non-redundant sequence database of genomes, transcripts and**
878   **proteins.** *Nucleic Acids Res* 2007, **35:**D61-65.
879   22.   Benton MJ: **Phylogeny of the major tetrapod groups: morphological data and**
880   **divergence dates.** *J Mol Evol* 1990, **30:**409-424.
881   23.   Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M: **Blast2GO: a**
882   **universal tool for annotation, visualization and analysis in functional**
883   **genomics research.** *Bioinformatics* 2005, **21:**3674-3676.
884   24.   Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U,
885   Daugherty L, Duquenne L, et al: **InterPro: the integrative protein signature**
886   **database.** *Nucleic Acids Res* 2009, **37:**D211-215.
887   25.   Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg
888   SL, Wold BJ, Pachter L: **Transcript assembly and quantification by RNA-Seq**
889   **reveals unannotated transcripts and isoform switching during cell**
890   **differentiation.** *Nat Biotechnol* 2010, **28:**511-515.
891   26.   Fan HP, Di Liao C, Fu BY, Lam LC, Tang NL: **Interindividual and interethnic**
892   **variation in genomewide gene expression: insights into the biological**
893   **variation of gene expression and clinical implications.** *Clin Chem* 2009,
894   **55:**774-785.
895   27.   Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M: **KEGG for integration and**
896   **interpretation of large-scale molecular data sets.** *Nucleic Acids Res* 2012,
897   **40:**D109-114.
898   28.   Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes.** *Nucleic*
899   *Acids Res* 2000, **28:**27-30.
900   29.   Kageyama T: **Pepsinogens, progastricsins, and prochymosins: structure,**
901   **function, evolution, and development.** *Cell Mol Life Sci* 2002, **59:**288-306.
902   30.   Hayashi K, Agata K, Mochii M, Yasugi S, Eguchi G, Mizuno T: **Molecular cloning**
903   **and the nucleotide sequence of cDNA for embryonic chicken pepsinogen:**
904   **phylogenetic relationship with prochymosin.** *J Biochem* 1988, **103:**290-296.
905   31.   Watanuki K, Yasugi S: **Analysis of transcription regulatory regions of**
906   **embryonic chicken pepsinogen (ECPg) gene.** *Dev Dyn* 2003, **228:**51-58.
907   32.   Aoki N, Deshimaru M, Terada S: **Active fragments of the antihemorrhagic**
908   **protein HSF from serum of habu (Trimeresurus flavoviridis).** *Toxicon* 2007,
909   **49:**653-662.
910   33.   Perales J, Neves-Ferreira AG, Valente RH, Domont GB: **Natural inhibitors of**
911   **snake venom hemorrhagic metalloproteinases.** *Toxicon* 2005, **45:**1013-1020.
912   34.   Tomihara Y, Yonaha K, Nozaki M, Yamakawa M, Kawamura Y, Kamura T, Toyama
913   S: **Purification of an antihemorrhagic factor from the serum of the non-**
914   **venomous snake Dinodon semicarinatus.** *Toxicon* 1988, **26:**420-423.
915   35.   Lee HJ, Sorimachi H, Jeong SY, Ishiura S, Suzuki K: **Molecular cloning and**
916   **characterization of a novel tissue-specific calpain predominantly**
917   **expressed in the digestive tract.** *Biol Chem* 1998, **379:**175-183.
918   36.   Secor SM: **Gastric function and its contribution to the postprandial**
919   **metabolic response of the Burmese python Python molurus.** *Journal of*
920   *Experimental Biology* 2003, **206:**1621-1630.

921 37. Bessler SM, Secor SM: **Effects of feeding on luminal pH and morphology of**
922 **the gastroesophageal junction of snakes.** *Zoology (Jena)* 2012, **115:**319-329.
923 38. Helmstetter C, Pope RK, T'Flachebba M, Secor SM, Lignot J-H: **The effects of**
924 **feeding on cell morphology and proliferation of the gastrointestinal tract of**
925 **juvenile Burmese pythons (Python molurus).** *Canadian Journal of Zoology*
926 2009, **87:**1255-1267.
927 39. Enok S, Simonsen LS, Wang T: **The contribution of gastric digestion and**
928 **ingestion of amino acids on the postprandial rise in oxygen consumption,**
929 **heart rate and growth of visceral organs in pythons.** *Comp Biochem Physiol A*
930 *Mol Integr Physiol* 2013, **165:**46-53.
931 40. Nørgaard S, Andreassen K, Malte CL, Enok S, Wang T: **Low cost of gastric acid**
932 **secretion during digestion in ball pythons.** *Comparative Biochemistry and*
933 *Physiology Part A: Molecular & Integrative Physiology* 2016, **194:**62-66.
934 41. Secor SM, Taylor JR, Grosell M: **Selected regulation of gastrointestinal acid-**
935 **base secretion and tissue metabolism for the diamondback water snake**
936 **and Burmese python.** *J Exp Biol* 2012, **215:**185-196.
937 42. Andrade DV, De Toledo LF, Abe AS, Wang T: **Ventilatory compensation of the**
938 **alkaline tide during digestion in the snake Boa constrictor.** *J Exp Biol* 2004,
939 **207:**1379-1385.
940 43. Koelz HR: **Gastric acid in vertebrates.** *Scand J Gastroenterol Suppl* 1992, **193:**2-
941 6.
942 44. Starck JM, Beese K: **Structural flexibility of the intestine of Burmese python**
943 **in response to feeding.** *Journal of Experimental Biology* 2001, **204:**325-335.
944 45. Holmberg A, Kaim J, Persson A, Jensen J, Wang T, Holmgren S: **Effects of**
945 **digestive status on the reptilian gut.** *Comparative Biochemistry and Physiology*
946 *a-Molecular and Integrative Physiology* 2002, **133:**499-518.
947 46. Lignot JH, Helmstetter C, Secor SM: **Postprandial morphological response of**
948 **the intestinal epithelium of the Burmese python (Python molurus).** *Comp*
949 *Biochem Physiol A Mol Integr Physiol* 2005, **141:**280-291.
950 47. Secor SM, Whang EE, Lane JS, Ashley SW, Diamond J: **Luminal and systemic**
951 **signals trigger intestinal adaptation in the juvenile python.** *American Journal*
952 *of Physiology-Gastrointestinal and Liver Physiology* 2000, **279:**G1177-G1187.
953 48. Overgaard J, Andersen JB, Wang T: **The effects of fasting duration on the**
954 **metabolic response to feeding in Python molurus: an evaluation of the**
955 **energetic costs associated with gastrointestinal growth and upregulation.**
956 *Physiol Biochem Zool* 2002, **75:**360-368.
957 49. Secor SM, White SE: **Prioritizing blood flow: cardiovascular performance in**
958 **response to the competing demands of locomotion and digestion for the**
959 **Burmese python, Python molurus.** *J Exp Biol* 2010, **213:**78-88.
960 50. Secor SM, Hicks JW, Bennett AF: **Ventilatory and cardiovascular responses of**
961 **a python (Python molurus) to exercise and digestion.** *J Exp Biol* 2000,
962 **203:**2447-2454.
963 51. Skovgaard N, Møller K, Gesser H, Wang T: **Histamine induces postprandial**
964 **tachycardia through a direct effect on cardiac H2-receptors in pythons.**
965 *American Journal of Physiology - Regulatory, Integrative and Comparative*
966 *Physiology* 2009, **296:**R774-R785.
967 52. Enok S, Simonsen LS, Pedersen SV, Wang T, Skovgaard N: **Humoral regulation**
968 **of heart rate during digestion in pythons (Python molurus and Python**
969 **regius).** *Am J Physiol Regul Integr Comp Physiol* 2012, **302:**R1176-1183.

43

53. Riquelme CA, Magida JA, Harrison BC, Wall CE, Marr TG, Secor SM, Leinwand LA: **Fatty acids identified in the Burmese python promote beneficial cardiac growth.** *Science* 2011, **334:**528-531.

54. Enok S, Leite G, Leite C, Gesser H, Hedrick MS, Wang T: **Improved cardiac filling facilitates the postprandial elevation of stroke volume in Python regius.** *J Exp Biol* 2016.

55. Slay CE, Enok S, Hicks JW, Wang T: **Reduction of blood oxygen levels enhances postprandial cardiac hypertrophy in Burmese python (Python bivittatus).** *J Exp Biol* 2014, **217:**1784-1789.

56. Jensen B, Larsen CK, Nielsen JM, Simonsen LS, Wang T: **Change of cardiac function, but not form, in postprandial pythons.** *Comp Biochem Physiol A Mol Integr Physiol* 2011, **160:**35-42.

57. Butler MW, Lutz TJ, Fokidis HB, Stahlschmidt ZR: **Eating increases oxidative damage in a reptile.** *J Exp Biol* 2016, **219:**1969-1973.

58. Schulz MH, Zerbino DR, Vingron M, Birney E: **Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels.** *Bioinformatics* 2012, **28:**1086-1092.

59. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL: **BLAST+: architecture and applications.** *BMC Bioinformatics* 2009, **10:**421.

60. Eddy SR: **A probabilistic model of local sequence alignment that simplifies statistical significance estimation.** *PLoS Comput Biol* 2008, **4:**e1000069.

61. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215:**403-410.

62. Langmead B, Salzberg SL: **Fast gapped-read alignment with Bowtie 2.** *Nat Methods* 2012, **9:**357-359.

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

64. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, Kong L, Gao G, Li CY, Wei L: **KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases.** *Nucleic Acids Res* 2011, **39:**W316-322.

65. Perkins DN, Pappin DJC, Creasy DM, Cottrell JS: **Probability-based protein identification by searching sequence databases using mass spectrometry data.** *Electrophoresis* 1999, **20:**3551-3567.

66. Dyrlund TF, Poulsen ET, Scavenius C, Sanggaard KW, Enghild JJ: **MS Data Miner: A web-based software tool to analyze, compare, and share mass spectrometry protein identifications.** *Proteomics* 2012, **12:**2792-2796.

67. Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, Mann M: **Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein.** *Mol Cell Proteomics* 2005, **4:**1265-1272.

44

Table 1

Parameter	*De novo* assembly
Total transcripts	34,423
Annotated transcripts with nr NCBI	19,713
Annotated transcripts with GO term	16,992
Minimum transcript size (nt)	100
Medium transcrpt size (nt)	605
Mean transcript size (nt)	1,034
Largest transcript (nt)	26,010
N50	6,240
N50 size (nt)	1,673
Total assembled bases (Mb)	35.6

Table 2

Expression level	Fold change level	Expression trend (fasting -> 24h -> 48h)	Color code
max FPKM over 10	FC over 4	Up-regulated	Red
		Down-regulated	Blue
		Up-then-down regulated	Yellow
		Down-then-up regulated	Brown
	FC below 4	Highly expressed (max FPKM over 200)	Purple
		Moderately expressed (max FPKM below 200)	Pink
max FPKM below 10	-	Lowly expressed	Darkgrey

Table 3

Expression trend (fasting -> 24h -> 48h)	Stomach	Intestine	Pancreas	Liver	Heart
Up-regulated	932 (2.9%)	1,131 (3.5%)	859 (2.6%)	1,047 (3.2%)	184 (0.6%)
Up-then-down regulated	28 (0.1%)	31 (0.1%)	150 (0.5%)	61 (0.2%)	6 (0.0%)
Down-regulated	869 (2.7%)	625 (1.9%)	567 (1.7%)	618 (1.9%)	168 (0.5%)
Down-then-up regulated	36 (0.1%)	45 (0.1%)	127 (0.4%)	90 (0.3%)	16 (0.1%)
Highly expressed	199 (0.6%)	211 (0.7%)	225 (0.7%)	354 (1.1%)	232 (0.7%)
Moderately expressed	5,541 (17.0%)	5,582 (17.2%)	4,933 (15.2%)	5,385 (16.5%)	6,044 (18.6%)
Lowly expressed	24,926 (76.6%)	24,906 (76.5%)	25,670 (78.9%)	24,976 (76.8%)	25,881 (79.5%)
Total	32,531 (100%)	32,531 (100%)	32,531 (100%)	32,531 (100%)	32,531 (100%)

Figure 1

Raw data
277.5 M paired reads
56 Gb in total

Filtering and trimming
213.8 M paired reads
43 Gb in total

***de novo* assembly**
Velvet/Oases

Annotation
Blastx, Blast2GO, KOBAS

Expression abundance
RSEM

Principal component analysis

Identification of differential expressed genes

Enriched pathway analysis
KOBAS
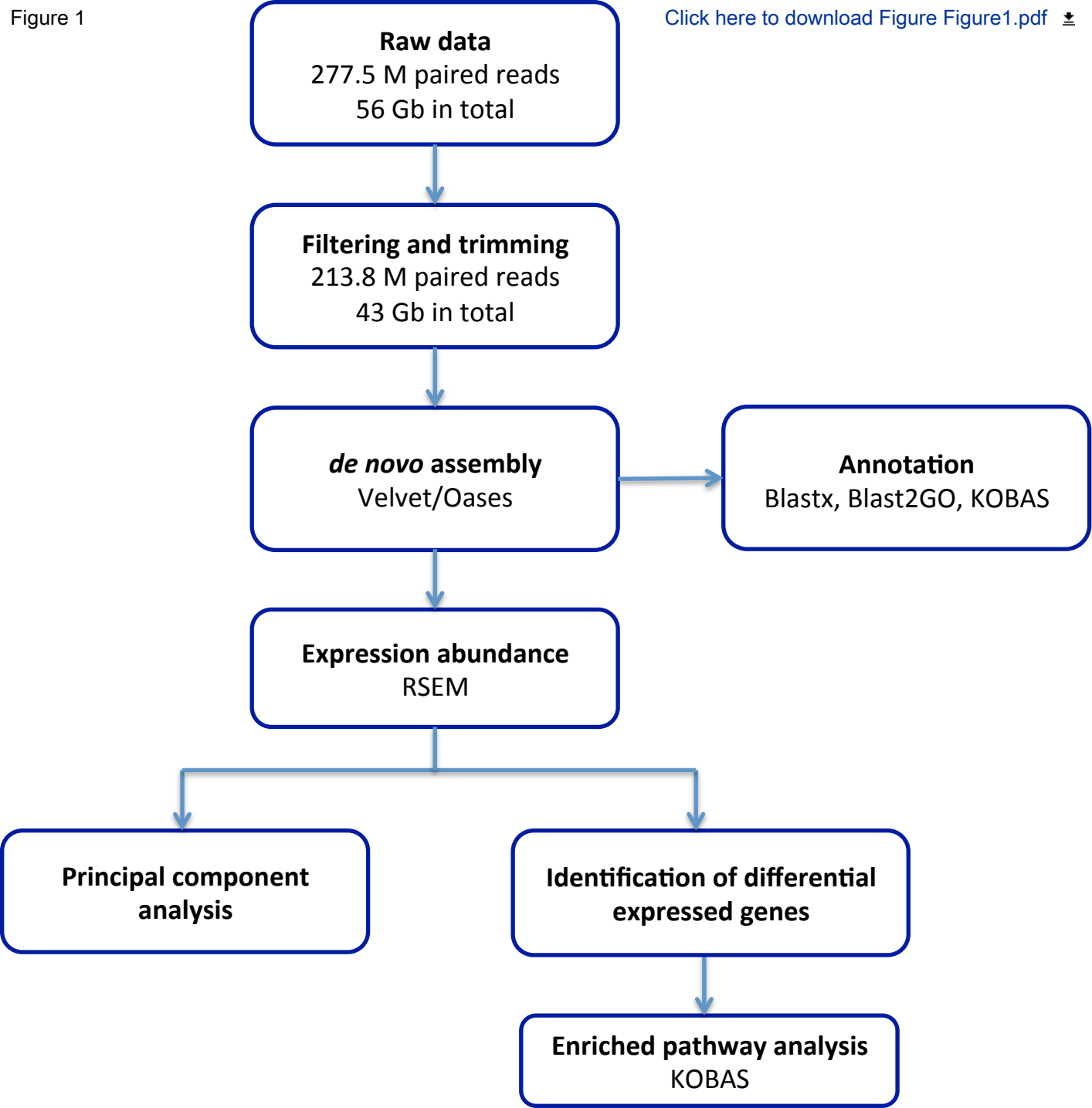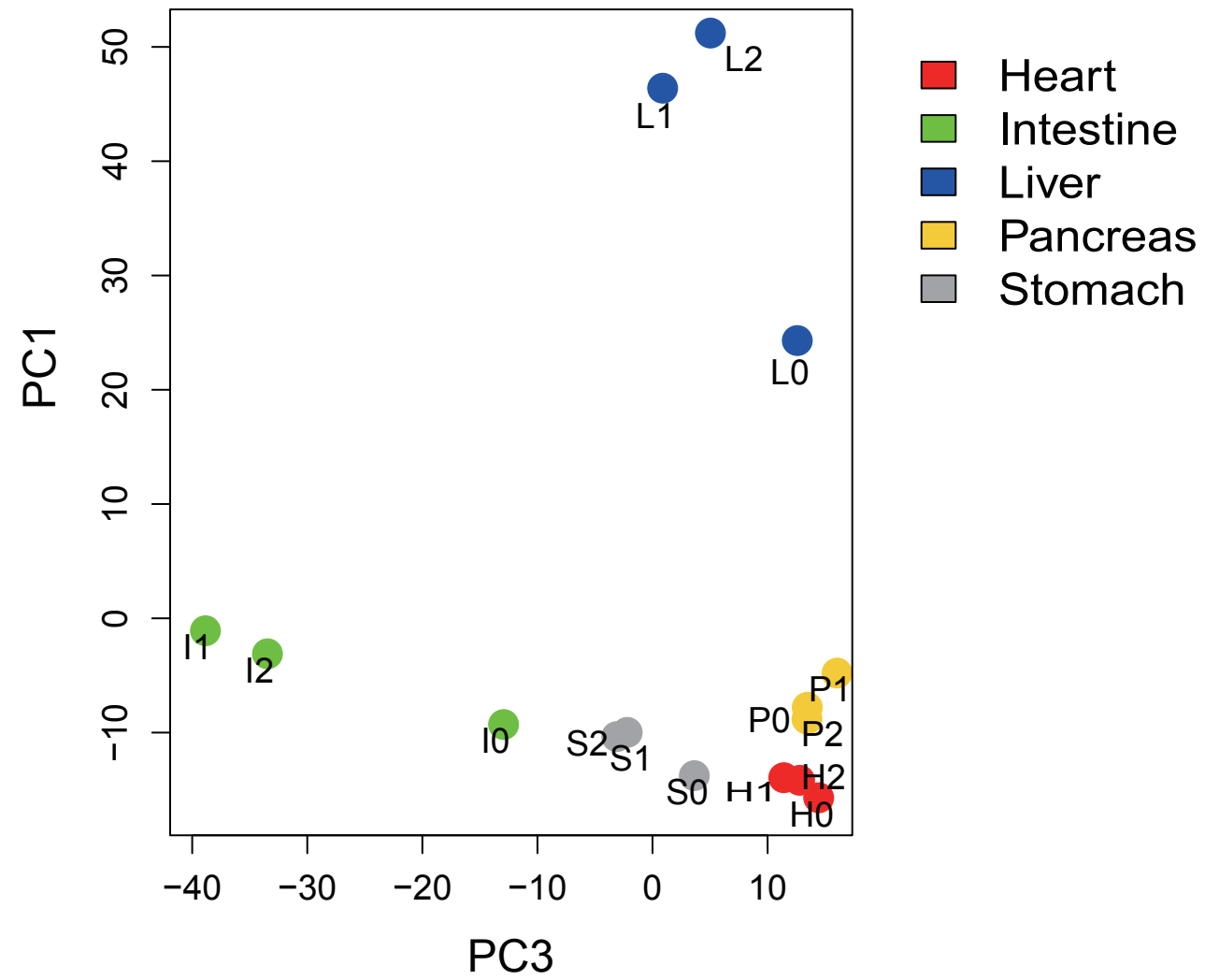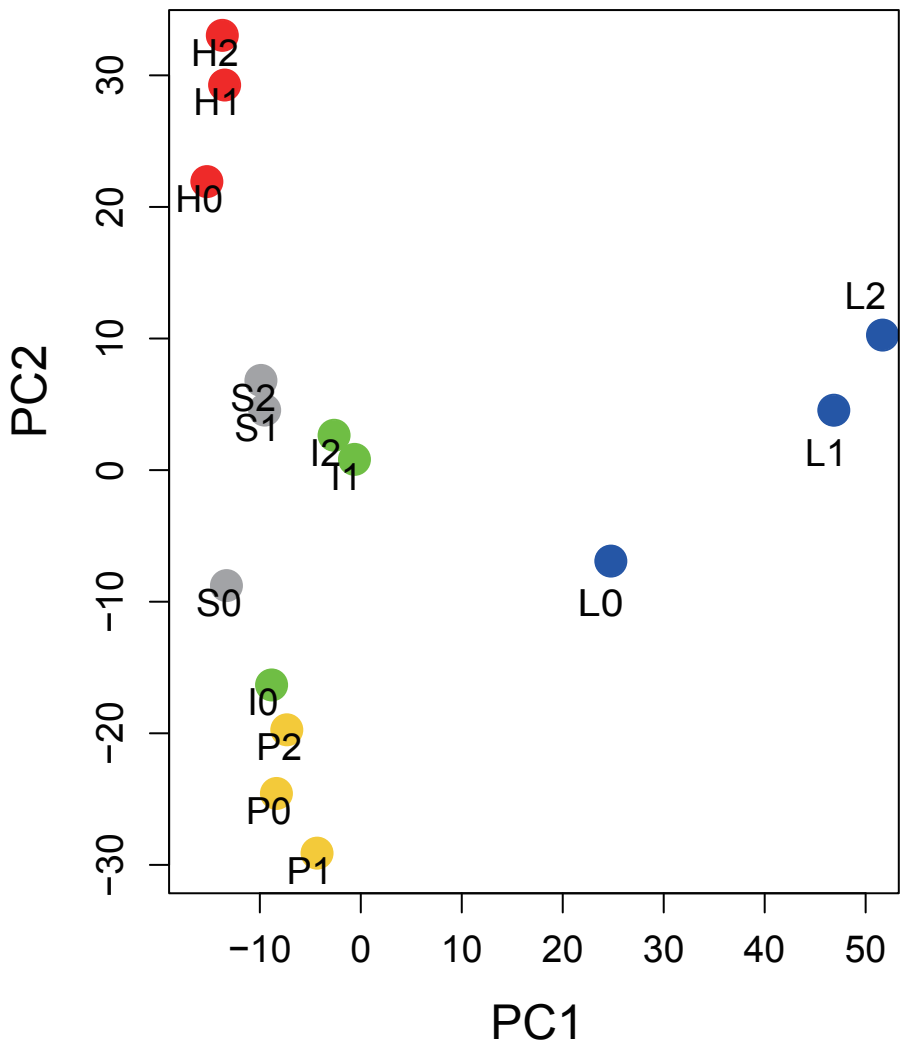
Figure 2

Figure 3

Heart · Intestine · Liver · Pancreas · Stomach

days after feeding (0, 1, 2)

Color Key (−1, 0, 1)

Figure 4

Stomach

−1   0   1

protease, serine, 3 isoform 3
None
peroxiredoxin–6–like
calponin–1–like
ribosomal protein S14–2
creatine kinase B–type
nucleoside diphosphate kinase
None
None
anterior gradient protein 2 homolog
cystatin precursor
UBIQP_XENLA (Polyubiquitin)
gastricsin–like
gastrokine–2–like
CD63 antigen–like
embryonic pepsinogen–like
ATP synthase subunit beta, mitochondrial–like
gastricsin–like
ATP synthase lipid–binding protein, mitochondrial–like
None
gastricsin–like
LOW QUALITY PROTEIN: carbonic anhydrase 2–like
cytochrome c oxidase subunit 7C, mitochondrial–like
hypothetical protein LOC100619418
integral membrane transporter protein
gastricsin precursor
ATP synthase subunit g, mitochondrial–like
gastricsin–like
gastricsin precursor
ATP synthase subunit alpha, mitochondrial–like
pepsin A–like isoform 2
ATP synthase lipid–binding protein, mitochondrial–like
gastrokine–1–like, partial
gastricsin–like
gastricsin–like
actin, gamma–enteric smooth muscle isoform 1 precursor
protein S100–A6
mucin 6, oligomeric mucus/gel–forming
Senescence–associated protein
Aa1–330
CDH1–D
60S ribosomal protein L9–like
ubiquitin–40S ribosomal protein S27a–like
40S ribosomal protein S23
60S ribosomal protein L37a
polyubiquitin
60S ribosomal protein L31 isoform 3
40S ribosomal protein S5–like
60S ribosomal protein L38

0   1   2

days after feeding

Figure 5

Intestine

(left column, top to bottom)
- alpha−crystallin B chain−like
- actin, beta, partial
- fatty acid−binding protein, intestinal−like
- calreticulin−like
- None
- retinol−binding protein 2
- myosin regulatory light polypeptide 9−like
- ferritin heavy chain A−like
- deleted in malignant brain tumors 1 protein
- fatty acid−binding protein, liver−like
- deleted in malignant brain tumors 1 protein
- calbindin D28K
- apoe protein
- natterin−4−like
- selenoprotein P precursor
- cytochrome P450 2G1−like
- placenta−specific gene 8 protein−like
- glutamine synthetase−like, partial
- cytosol aminopeptidase
- sulfotransferase family cytosolic 1B member 1−like
- microsomal glutathione S−transferase 1−like isoform 1
- A Chain A, Refined Solution Structure Of Calmodulin C
- cytosolic non−specific dipeptidase−like
- ferritin heavy chain−like
- peroxiredoxin−1−like
- 3'-phosphoadenosine 5'-phosphosulfate synthase 2
- keratin, type I cytoskeletal 18−like
- creatine kinase B−type
- spermine oxidase−like
- retinal dehydrogenase 1−like isoform 2
- sulfotransferase 6B1−like
- large neutral amino acids transporter small subunit 2
- None
- keratin, type II cytoskeletal 8−like
- sulfotransferase 1C1−like
- putative ornithine decarboxylase antizyme variant 1

(right column, top to bottom)
- sulfotransferase 6B1−like isoform 1
- diamine acetyltransferase 2
- 4−trimethylaminobutyraldehyde dehydrogenase−like
- calmodulin−like
- keratin, type II cytoskeletal cochleal−like isoform 1
- uncharacterized protein ENSP00000244321 homolog
- diazepam binding inhibitor
- 4−aminobutyrate aminotransferase
- galectin−1−like
- similar to reverse transcriptase−like protein
- sulfotransferase 1C1−like
- glutathione S−transferase A1−like isoform 1
- acyl−CoA−binding protein−like
- phospholipase A2 inhibitor subunit gamma B−like
- ammonium transporter Rh type B−like
- phospholipase A2 inhibitor 31 kDa subunit−like
- None
- amiloride binding protein 1−like
- ectonucleotide pyrophosphatase/phosphodiesterase
- arylsulfatase A
- meprin A subunit alpha−like
- deleted in malignant brain tumors 1
- apolipoprotein A−I preproprotein
- apolipoprotein A−IV−like
- galectin−4−like
- SON protein
- UBIQP_XENLA (Polyubiquitin)
- apolipoprotein B−100−like
- prostasin−like, partial
- peroxiredoxin−6−like
- zonadhesin variant 6
- CDH1−D
- Aa1−330
- integral membrane transporter protein
- Senescence−associated protein
- polyubiquitin
- cysteine−rich protein 1−like
- selenium binding protein 1
- zonadhesin precursor

0 1 2
days after feeding

0 1 2
days after feeding

Figure 6

Heart

None
heat shock protein 90a
adipocyte fatty acid–binding protein
B–type natriuretic peptide precursor
myosin regulatory light polypeptide 9–like
ATP synthase lipid–binding protein
protease, serine, 3 isoform 3
heat shock protein beta–7–like
ATP synthase subunit alpha
four and a half LIM domains protein 2–like
actin, alpha cardiac muscle 1 proprotein
CYC_CROAD (Cytochrome C)
actin, alpha skeletal muscle B–like
tubulin, alpha, ubiquitous (predicted)
actin, alpha skeletal muscle
desmin–like
myosin regulatory light polypeptide 9–like
SJCHGC06004 protein
coil–helix domain–containing protein 10
fatty acid binding protein 3
ATP synthase lipid–binding protein
malate dehydrogenase

−1   0   1

0 1 2
days after feeding

Figure 7

Liver

Left column (top to bottom):
- dehydrogenase/reductase (SDR family) member 7
- None
- calreticulin–like
- amine sulfotransferase–like
- None
- amine oxidase [flavin–containing] B–like
- amine sulfotransferase–like
- hypothetical protein LOC100567022
- sulfotransferase 1A1–like
- isocitrate dehydrogenase [NADP] cytoplasmic
- sulfotransferase 1C1
- ammonium transporter Rh type B–like
- cytochrome P450 2G1–like
- coil–helix domain–containing protein 10
- protein disulfide–isomerase A6–like
- peroxiredoxin–1–like
- angiotensinogen–like
- alcohol dehydrogenase 1–like
- cytochrome P450 2G1–like
- heat shock cognate 71 kDa protein–like
- translocating chain–associated membrane protein 1
- sulfotransferase 6B1–like
- cytochrome b–c1 complex subunit 7–like
- protein disulfide–isomerase A4–like
- 10 kDa heat shock protein, mitochondrial–like
- stomatin
- cytoplasmic aconitate hydratase
- ADH1_NAJNA (Alcohol dehydrogenase 1)
- sulfotransferase family cytosolic 1B member 1
- endoplasmin–like
- transmembrane protein 14A–like
- FABPL_ANOPU (Fatty acid–binding protein)
- ADP–sugar pyrophosphatase–like
- hydroxymethylglutaryl–CoA synthase
- phospholipase A2 inhibitor beta
- AF232771_1 antitoxic factor PLI
- pterin–4–alpha–carbinolamine dehydratase 2
- LOC496671 protein, partial
- apolipoprotein A–II–like
- non–specific lipid–transfer protein–like
- Eukaryotic translation initiation factor 5A–1
- ATP synthase subunit alpha, mitochondrial
- apolipoprotein M–like
- alpha–2–macroglobulin–like
- methyltransferase–like protein 7A–like
- PLILP_ELAQU (PLIalpha–like protein)
- apolipoprotein A–IV–like
- small serum protein–5
- ENOA_PYTRG (Phosphopyruvate hydratase)
- phosphoglycerate mutase 1–like
- 4–trimethylaminobutyraldehyde dehydrogenase
- alcohol dehydrogenase 1B–like isoform 1
- aldehyde dehydrogenase, mitochondrial–like
- None
- putative lymphocyte G0/G1 switch protein 2–like
- glutamate dehydrogenase 1, mitochondrial–like
- protein disulfide–isomerase A3 precursor
- glutathione S–transferase Mu 1–like
- vitamin K epoxide reductase complex subunit 1
- argininosuccinate lyase
- microsomal glutathione S–transferase 1
- glutathione S–transferase 2–like
- fructose–bisphosphate aldolase B–like
- betaine––homocysteine S–methyltransferase 1
- phospholipid hydroperoxide glutathione peroxidase
- acyl–CoA–binding protein–like

Right column (top to bottom):
- arylacetamide deacetylase–like
- LOW QUALITY PROTEIN: dihydropyrimidinase
- hydroxysteroid 11–beta–dehydrogenase 1
- actin, beta, partial
- transmembrane 7 superfamily member 4
- retinal dehydrogenase 1–like isoform 2
- superoxide dismutase [Cu–Zn]–like
- cytochrome P450 2G1–like
- glutathione S–transferase A1–like isoform 1
- mannose–binding protein C–like
- cytochrome P450 2A13–like
- cytochrome P450 2G1–like
- cytochrome P450 2G1–like
- transferrin precursor
- cholinesterase–like
- 3–ketoacyl–CoA thiolase, mitochondrial
- glyoxylate reductase/hydroxypyruvate reductase
- peptidyl–prolyl cis–trans isomerase B–like
- None
- cytochrome P450 2F3–like
- putative glutathione S–transferase
- cytochrome P450 3A9–like
- apovitellenin–1–like
- CDH1–D
- phosphoserine aminotransferase–like
- 17–beta–hydroxysteroid dehydrogenase type 6
- desmocollin 2–like
- dimethylaniline monooxygenase
- glycerol–3–phosphate dehydrogenase [NAD+]
- Senescence–associated protein
- integral membrane transporter protein
- Aa1–330
- uncharacterized oxidoreductase C663.09c–like
- hypothetical protein LOC100562305
- multifunctional protein ADE2–like
- N–acylsphingosine amidohydrolase
- uncharacterized protein ENSP00000244321
- fibrinogen alpha chain–like
- complement factor D
- cyclic AMP–dependent transcription factor ATF
- alpha–2–macroglobulin–like
- three finger toxin–like precursor
- fibrinogen gamma chain–like
- small serum protein–2
- complement C4–B–like, partial
- catalase–like
- None
- hemopexin–like
- Selenoprotein P, plasma, 1b
- coagulation factor IX–like, partial
- 2,4–dienoyl–CoA reductase
- alpha–2–macroglobulin–like
- pterin–4–alpha–carbinolamine dehydratase
- None
- retinol dehydrogenase 7–like
- alpha–2–macroglobulin–like
- alpha–2–macroglobulin–like, partial
- sphingosine–1–phosphate lyase 1–like
- alpha–2–antiplasmin–like
- protein G7c
- ceruloplasmin–like
- synaptotagmin–2–like
- phosphatidylethanolamine–binding protein 1
- prostatic acid phosphatase
- serum albumin
- serum albumin

0 1 2
days after feeding

0 1 2
days after feeding

Figure 8

Pancreas

Figure 8. Pancreas. Heatmap with legend scale −1, 0, 1.

Row labels (top to bottom):
- phosphatidylethanolamine−binding protein 1−like
- hCG1647491−like
- zymogen granule membrane protein 16−like isoform 1
- trypsin−1−like isoform 2
- alpha−amylase 1 isoform 3
- putative transposase
- colipase−like
- cationic trypsin−3−like
- chymotrypsin−like protease CTRL−1−like
- pancreatic alpha−amylase−like
- endonuclease domain−containing 1 protein−like
- UPF0762 protein C6orf58 homolog
- chymotrypsin−like elastase family member 1−like
- natterin−4−like
- trypsin inhibitor CITI−1
- UHRF1−binding protein 1−like
- natterin−4−like
- None
- insulin−like
- cytochrome c oxidase subunit 6B1−like
- LOC100170417 protein
- alpha−crystallin B chain−like
- trypsin I−P1 precursor
- None
- calreticulin−like
- bile salt−activated lipase−like
- peptidyl−prolyl cis−trans isomerase B−like
- cysteine−rich with EGF−like domain
- hypothetical protein LOC100619418
- acyl−CoA−binding protein−like
- probable proline dehydrogenase 2−like
- cytochrome c oxidase subunit 7C, mitochondrial−like
- translocon−associated protein subunit gamma−like
- None
- chymotrypsin B, partial
- heat shock 70kDa protein 5
- pancreatic lipase−related protein 1−like
- VSP_PHIOL (Venom serine protease)

0 1 2

days after feeding

1. Sampling of digestive fluid

2. Recovery, denaturation, and trypsin treatment of digestive fluid proteins

3. LC-MS/MS analyses

4. Protein identification

5. Identification of the python stomach secretome

Figure 10
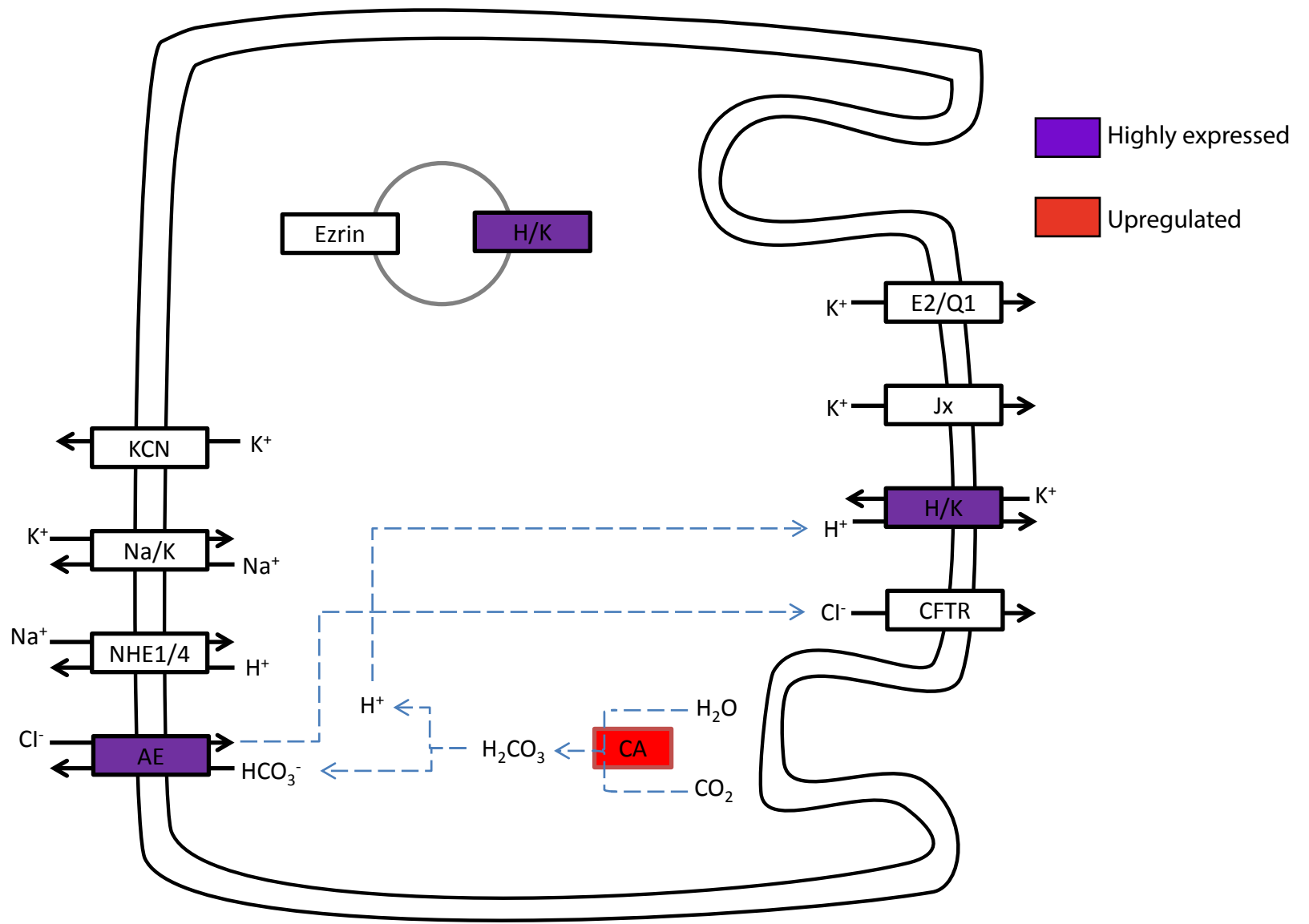
Figure 10

Figure S4

Click here to access/download
Supplementary Material
FigureS4.pdf
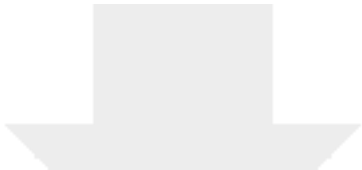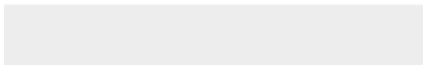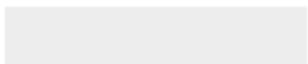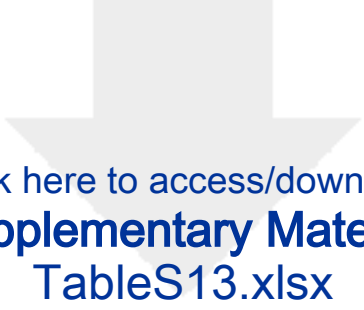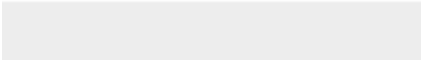
Figure S5

Click here to access/download
Supplementary Material
FigureS5.pdf

Figure S6

Click here to access/download
Supplementary Material
FigureS6.pdf
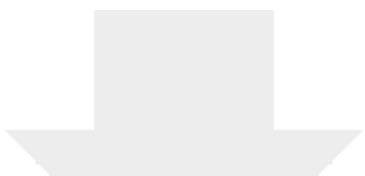
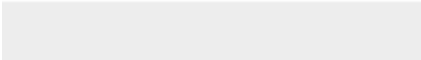Figure S7

Figure S8

Click here to access/download
Supplementary Material
FigureS8.pdf

Figure S9

Click here to access/download
Supplementary Material
FigureS9.pdf

Table S3

Click here to access/download
Supplementary Material
TableS3.xlsx
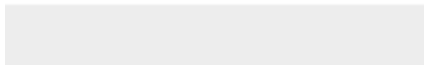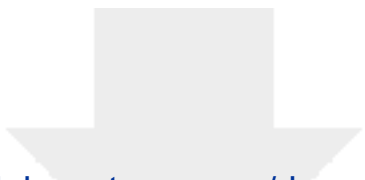
Table S5

Click here to access/download
Supplementary Material
TableS6.xlsx

Table S7

Click here to access/download
Supplementary Material
TableS7.xlsx
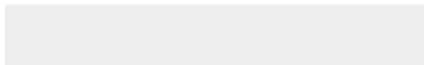
Table S8

Click here to access/download

Supplementary Material

TableS8.xlsx
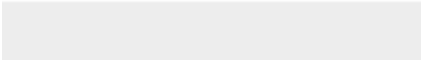
Table S9

Click here to access/download
Supplementary Material
TableS9.xlsx
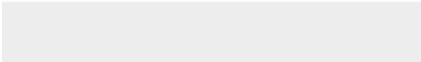
Table S10

Click here to access/download
Supplementary Material
TableS10.xlsx

Click here to access/download
Supplementary Material
TableS11.xlsx

Table S12

Click here to access/download
Supplementary Material
TableS12.xlsx

Click here to access/download
Supplementary Material
TableS13.xlsx

Table S14

Click here to access/download
Supplementary Material
supplementary material.docx

FASTA sequences of misassembled transcripts

Click here to access/download
Supplementary Material
AU_assembly.misassembled.fasta
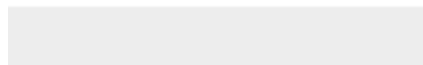
Dear Editor,

Thank you for returning the constructive and useful comments from the two reviewers that kindly evaluated our manuscript entitled "Transcriptome Analysis of the Response of Burmese Python to Digestion" that we submitted for publication in GigaScience. Both reviewers provided positive overall assessments, but also raised a number of specific queries to be addressed in the revision. We are pleased to return a revised manuscript where we have followed all the advice given by the two reviewers. The responses to each query is listed in a separate PDF file where you can see our responses describing the changes we have made to the manuscript. We greatly appreciate these comments and feel the manuscript has been improved in this review-process. We hope you will find the revised manuscript acceptable for publication in GigaScience and we are looking forward to hearing from you in due course. Please do not hesitate to contact me in case you need additional information.

Sincerely yours

Jinjie Duan (on behalf of all the authors)

Reviewer #1:

The authors used a transcriptomic time-series of five different organs and complementary proteomic surveys to characterize changes in expression following a feeding event in the Burmese Python.

I recommend the acceptance of this manuscript pending revisions. The only major criticism I have for the current manuscript is the lack of comparison to previous studies on this system and believe a section should be added to the manuscript explicitly comparing the time-series transcriptome data of this paper with others (see below). Comments below.

Introduction

After reading the introduction, it was unclear what knowledge this paper would add considering time-series transcriptome sequencing of particular organs (e.g., heart, liver) has already been performed (e.g., Castoe et al., 2013 PNAS). There is novel work being done in the form of additional tissues and, particularly, the proteomics, and I think the authors should state this explicitly to make it clear to the reader what is novel.

This is a good point and we now emphasize that we have repeated measurements in some tissues, but bring new data regarding the pancreas and the stomach as well as a new time point (48h) into the digestive period. These changes have been made in the second paragraph of the section entitled "background".

Methods

Page 26 lines 526-527: Why were the biopsies pooled? This reduced the n from two to one.

At the time (2011) it was for technical and economic reasons, two samples were pooled to average out some of the variability among biological replicates. Yet, we regret not having more biological replicates since this restricts our analysis to highly expressed genes. Nevertheless, by virtue of our three time points (fasting, 24h and 48h, respectively), we do present some biological replication.

Page 27 lines 541-542: Why were samples pooled by tissue? What was the justification?

We included some normalized libraries to sample as broadly from the transcriptome as possible. This also included sampling over the different tissues, such that we would obtain some reads also from lowly expressed, tissue-specific genes. We have added an explanation in line 583 "In addition, to sample as broadly from transcriptome as possible, we also produced normalized libraries for each tissue in order to capture the reads from lowly expressed, tissue-specific genes."

Page 28 lines 554-560: Although the authors used relatively long k-mers for assembly, they should still perform a specific check for mis-assembled chimeric sequences, especially considering the reads from all libraries were pooled to assemble a reference (see Yang and Smith 2013 BMC Genomics), and this reference was the basis for all subsequent transcriptomic analyses.

To check the mis-assembled chimeric sequences, we have now compared our assembly with the reference genome (Castoe et al) and gene sequences using rnaQUAST, and amended the result (lines 101-115) and method (lines 606-608) sections, correspondingly. The result shows our transcriptome assembly had 34,423 transcripts in total. 34,040 (98%) transcripts had at least 1 significant alignment to the reference genome and 31,102 out of 34,040 were uniquely aligned. Average aligned fraction (i.e. total number of aligned bases in the transcript divided by the total transcript length) was 0.975. The total number of misassembled (chimeric) transcripts, which have discordant best-scored alignment (partial alignments that are either mapped to different strands/different chromosomes/in reverse order/too far away) was 1,974 (5.7%). The FASTA sequences of these misassembled transcripts are attached in Supplementary material.

Page 29 lines 578-579: Why was T-coffee used specifically for albumin-like genes? Justification should be added to the section.

We used T-coffee because it is well recommended for better accuracy of multiple sequences alignment (Thompson et al, 2011, Plos One; Pais et al,2014, Algorithms Mol Biol). Due to the improved analysis of the albumin-like sequences, we decided to move the "albumin-story" to the supplementary material (A more detailed explanation

is presented below under the responses to comments raised to our "analyses"). The justification of using T-coffee was added in the supplementary material in line 10: "We did multiple sequences alignment of these paralogues genes together with predicted ORF sequences of our five sequences using T-coffee (version 11.00) [2] with default parameters which is well recommended for better accuracy of multiple sequence alignment [3, 4]".

Page 31 line 623: I failed to follow the text and come up with 6 samples. Were two snakes at 400 g and two snakes at 800 g fed a rodent? If so, these samples, along with the peptone control, equals five. If only one snake at 400 g and one snake at 800 g were fed a rodent, these samples, again along with the control, equal three. I do not see how the authors collected six samples. Was there also an n of two for each of these groups? Additionally, how was plasma collected?

We apologize for the confusing manner in which we originally described these procedures. We have altered to the text in line 650 and 661 to clarify that samples were obtained in duplicate from three individuals snakes (two upon digestion of a rodent meal, as well as one snake that had been fed peptone). We have also added a description in line 668 to explain that the blood (plasma) samples were obtained by cardiac puncture.

Analyses

Page 8 lines 124-129: How dissimilar were the transcripts (i.e.,sequence divergence)? Were the six albumin-like proteins identified in the MS analysis the six most highly-expressed albumin-like transcripts? In other words, was there a detection bias in your MS analyses against low-abundance transcripts? I have seen this in my work (e.g., Rokyta et al., 2015 G3). Also, can the authors be sure that these are different copies and not alternatively spliced transcripts?

We appreciate this concern. After performing additional analyses (phylogenetic analysis and alignment against reference genome), we conclude these albumin-like sequences are most likely alternatively spliced transcripts, rather than paralogues. Therefore, we feel that this part of our results do not longer present sufficient new advance to be discussed in the main text. We accordingly decided to move the albumin results to supplementary material in lines 2-26. In addition, we did observe a

Line 197: "the five most abundant proteases identified in the gastric juice": How was protein quantitation performed? The methods do not mention protein quantitation. Are these simply based on spectral counts? If the authors are attempting to quantify the proteome, a more complete transcriptome-proteome comparison is warranted.

Lines 201, 208: carolinensis should not be capitalized

Perhaps the largest gap in the current study was the lack of a comparison to previous, extremely similar work on this system (e.g., Castoe et al., 2013 PNAS and Andrew et al.,2015 Physiol. Genomics). How do the authors' results compare to those of previous studies? Were they largely congruent? A section explicitly comparing the current study to previously published works should be added.

Figures and Tables

Information in table 2 should be provided with the KEGG pathway figures.

How were the sub-clusters in Figure 3 chosen? Do these represent all of the DEGs for that tissue?

The heat maps in Figure 3 show, for each tissue, all the genes that are both highly and differentially expressed with strict thresholds (defined in section "identification of DEGs and clustering analysis in Method section). We chose those sub-clusters because they represent a cluster of all upregulated genes, which are expected to be involved many functional changes during digestion.

Other

Small grammatical errors throughout, particularly in the discussion.

We have edited the manuscript carefully and hope we have corrected all grammatical mistakes.

Reviewer #2:

Duan et al. conduct a broad study using transcriptomic and proteomic methods to understand the molecular underpinnings of extreme physiological responses to feeding in Burmese pythons. Overall, the data collected are extensive and reasonably analyzed, and the manuscript is well written. The lack of replication and thorough analyses substantially limit the conclusions and novelty of the study, although generally I do believe that the manuscript is reasonable and valid in its current form. As such, given the aims of the journal, I do believe this manuscript does fit within its scope, as a sound descriptive study associated with a large amount of data that benefits from having these data directly linked to the paper. Below I note a handful of concerns and suggestions that would improve the ms.

I found it interesting that the authors chose to use de novo transcript assemblies rather than the annotated gene set available for the Burmese python genome. The authors make the case that the genome is somewhat fragmentary, which is true, and that this justified the use of a de novo assembly. While I don't completely agree, I do believe that their use of the de novo transcript assembly for mapping RNAseq data is reasonable, and what they find seems quite sensible. I am surprised, however, that they did not compare their annotations in any way to the annotated gene set on NCBI.

We appreciate these good comments. We have compared our assembly with annotated gene set in NCBI using rnaQUAST, and have updated the corresponding result and method section. It now reads in lines 115-120 "The comparison of assembled sequences and reference gene sequences (Supplementary table S3) showed that 26,320 (77.3%) assembled transcripts cover at least one isoform from the reference gene set and the mean fraction of transcript matched is 67.8%, suggesting there is a good concordance but also some differences which can be due to errors in either the reference genome assembly/annotation or our assembly".

Unfortunately, the authors did not have any replication in their RNAseq or proteomic data, and therefore any meaningful statistical comparisons are made difficult - for example, it is difficult to get decent estimates of how many genes are statistically differentially expressed across time points for organ-specific time course analyses. I

assume this is why the authors instead use arbitrary cutoffs: "1) FPKM is greater than or equal to 400 in at least one time point and 2) fold change is greater than or equal to 2 in at least one pairwise comparison among three time points." Without replication, I suppose the authors are somewhat limited in what they can do, and I do accept what they did as reasonable. However, they should avoid any instances of using the word "significant" throughout the text, which they use several times (e.g., LINE: 279: genes with significantly increased expression during digestion"). Honestly, they don't really have the power to detect significance with these data.

We agree that we should avoid the word significant when discussing the results since it is likely read as meaning statistically significant which we cannot know. Consequently, we have moved all "significant" throughout the manuscript.

I am concerned about what might be an over-interpretation of the findings from serum proteomics studies. The authors claim to have found a peptide that they identify in the serum as the protease inhibitor "anti-haemorrhagic factor cHLP-B (m.27_Py95)", and go on to conclude that "Our data supports older studies that identify these inhibitors of the deleterious action of venom enzymes in non-venomous snakes [32]."… My sense is that they should tone down their conclusion because 1) the python isn't venomous (and thus has no need for such proteins), and 2) the inference is simply based on blast homology with what is likely available online (venomous snake blood peptides). I think the finding is interesting and notable, but their inference of the function of this peptide being directly linked to resistance to venom is quite far fetched - more likely it may be indicative of a class of plasma peptides that could have been recruited in venomous snakes for self-defense against self-envenomation.

We agree with the reviewer's point and have changed the words accordingly in lines 255-259 such that it now reads "This is a protease inhibitor of the haemorrhagic-causing metalloproteinases present in snake venom and these inhibitors have previously been purified from serum of venomous snakes and thoroughly characterized [32, 33]. The role of such a protease inhibitor in non-venomous pythons is not obvious, but it has been proposed that they inhibit the deleterious action of venom enzymes in non-venomous snakes [34]."

Discussion section "Physiological interpretation of the upregulated genes in the intestine" - this section is noticeably lacking any citations or linking of results to a previously published in-depth transcriptional study of the python intestine (citation #13). There are also a number of incorrect claims made here (e.g., LINE 379: "It remains, however, unknown to what extent the increased capacity for nutrient uptake is also driven by increased synthesis of nutrient transporters".) that in fact have been clearly demonstrated in citation 13 - these links and statements made in this section need to be carefully re-written to more meaningfully incorporate this previous work.

We agree and appreciate this criticism. In the revised manuscript, we now give more credit to the previous studies in postprandial gene expression and we point more specifically to where there are differences between their findings and those reported by us. We hope you find the revised manuscript to be better balanced.

The figures should be improved for reading as a printed article. For example, there are multiple heat maps that are enormous, and are not printable in any reasonable way that would allow the labels to be read (e.g., Fig. 7). Simply spanning these over multiple columns would at least help with this. Also, while I realize that Gigascience is an online journal, the use of 15 in-text figures seems to be counter-productive for having there be clear points conveyed by the MS, and make the manuscript appear more like a massive data dump rather than a paper.

We have splitted the long heatmaps (Figures 5 and 7) into two columns to enable the reading of the labels.

We have moved the original Figures 9-13,15 and Tables 4-5 to supplementary material.

Copy Edits:

I suggest searching throughout the manuscript and writing out any numbers less than 10. For example: writing out four rather than 4.

Thanks for the comment. We have corrected them throughout the text.

Line 201 (and elsewhere) - change to: Anolis carolinensis (here and throughout the MS so that specific name is lower case)

Thanks for the comment. We have corrected them throughout the text.