

1 Transcriptome Analysis of the Response of

2 Burmese Python to Digestion

3 Jinjie Duan¹, Kristian Wejse Sanggaard^{2,3}, Leif Schauser⁵, Sanne Enok Lauridsen⁴, Jan J.

4 Enghild^{2,3}, Mikkel Heide Schierup^{1,4} and Tobias Wang⁴

5
6 1. Bioinformatics Research Center, Aarhus University, Denmark

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

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

9 4. Department of Bioscience, Aarhus University, Denmark

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

11
12 Corresponding authors: Jinjie Duan, Mikkel Heide Schierup & Tobias Wang

13
14
15 Email address:

16 Jinjie Duan: jjduan@birc.au.dk

17 Kristian Wejse Sanggaard: kristian@wejsesanggaard.dk

18 Leif Schauser: Leif.Schauser@qiagen.com

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

20 Jan J. Enghild: jje@mbg.au.dk, ORCID: 0000-0001-9292-9172

21 Mikkel Heide Schierup: mheide@birc.au.dk, ORCID: 0000-0002-5028-1790

22 Tobias Wang: tobias.wang@bios.au.dk, ORCID: 0000-0002-4350-3682

1 **23 Abstract**

2
3
4
5 **24 Background:**

6
7
8 **25** Exceptional and extreme feeding behaviour makes the Burmese python (*Python bivittatus*) an
9
10 **26** interesting model to study physiological remodelling and metabolic adaptation in response to
11
12 **27** refeeding after prolonged starvation. In this study, we used transcriptome sequencing of five
13
14 **28** visceral organs during fasting as well as 24h and 48h after ingestion of a large meal to
15
16 **29** unravel the postprandial changes in Burmese pythons. We first used the pooled data to
17
18 **30** perform a *de novo* assembly of the transcriptome and supplemented this with a proteomic
19
20 **31** survey of enzymes in the plasma and gastric fluid.
21
22

23
24
25
26 **32 Results:**

27
28
29 **33** We constructed a high-quality transcriptome with 34,423 transcripts of which 19,713 (57%)
30
31 **34** were annotated. Among highly expressed genes (FPKM>100 in one tissue) we found the
32
33 **35** transition from fasting to digestion was associated with differential expression of 43 genes in
34
35 **36** the heart, 206 genes in the liver, 114 genes in the stomach, 89 genes in the pancreas and 158
36
37 **37** genes in the intestine. We interrogated the function of these genes to test previous hypotheses
38
39 **38** on the response to feeding. We also used the transcriptome to identify 314 secreted proteins
40
41 **39** in the gastric fluid of the python.
42
43

44
45
46
47
48 **40 Conclusions:**

49
50
51 **41** Digestion was associated with an upregulation of genes related to metabolic processes, and
52
53 **42** translational changes therefore appears to support the postprandial rise in metabolism. We
54
55 **43** identify stomach-related proteins from a digesting individual and demonstrate that the
56
57 **44** sensitivity of modern LC-MS/MS equipment allows the identification of gastric juice proteins
58
59
60
61
62
63
64
65

45 that are present during digestion.

46 Keywords:

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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 **48 Background**

2
3
4
5 49 All animals exhibit dynamic changes in the size and functional capacities of bodily organs
6
7 50 and tissues to match energetic maintenance costs to prevailing physiological demands [1].

8
9 51 This phenotypic flexibility is particularly pronounced in the digestive organs in animals that
10
11 52 naturally experience prolonged periods of fasting, but are capable of ingesting large prey
12
13 53 items at irregular intervals. The Burmese python is an iconic example of this extreme
14
15 54 phenotype [1]. Many species of pythons easily endure months of fasting, while remaining
16
17 55 capable of subduing and ingesting very large meals. In Burmese pythons, digestion is
18
19 56 attended by a large and rapid rise in mass and/or functional capacity of the intestine, stomach,
20
21 57 liver, heart and kidneys [2-4] in combination with a stimulation of secretory processes and an
22
23 58 activation of enzymes and transporter proteins. These physiological responses are associated
24
25 59 with a many-fold rise in aerobic metabolism. Hence, the Burmese python is an excellent
26
27 60 model to study the mechanisms underlying extreme metabolic transitions and physiological
28
29 61 remodelling in response to altered demand [1, 3, 5-10].
30
31
32
33
34
35
36

37 62 The postprandial changes in the morphology and physiology of the intestine, heart and other
38
39 63 organs have been described in some detail in pythons [1, 5, 8, 9, 11], but only a few studies
40
41 64 [12-14] have addressed the underlying transcriptional changes of this interesting biological
42
43 65 response. Transcriptome sequencing technology now allows comprehensive surveys [15, 16],
44
45 66 prompting our use of transcriptome sequencing of heart, liver, stomach, pancreas and
46
47 67 intestine in snakes that had fasted for one month and those at 24 and 48h into the postprandial
48
49 68 period. These organs were chosen because a number of earlier studies have revealed their
50
51 69 profound phenotypic changes during the postprandial period [1-4, 17], and they are therefore
52
53 70 likely to exhibit large changes in gene expression. Differential gene expression in some of
54
55 71 these organs has previously been reported [12-14], but we provide new data on 48h into the
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

72 digestive period and the first descriptions of gene expression in the stomach and the pancreas.
73 As the Burmese python reference genome assembly [12] is relatively fragmented (contig size
74 N50 ~10kb), we found it impractical to use re-sequencing approaches and opted instead to
75 use our high coverage data to build a *de novo* transcriptome assembly to identify
76 differentially expressed genes (DEGs). To identify the enzymes involved in the digestion
77 process, we isolated the digestive fluid and characterized the protein composition using a
78 proteomics-based approach. This also allowed us to identify the major hydrolytic enzymes
79 used to digest the large and un-masticated meals.

80 **Analyses**

81 **Data summary**

82 277,485,924 raw paired reads (2*101 bp, insert size 180 bp) were obtained from Illumina Hi-
83 Seq 2000 sequencing of 15 non-normalized cDNA libraries derived from five tissues (heart,
84 liver, stomach, pancreas and intestine) at three time points (fasted for one month, 24h and 48h
85 postprandial) and 10 DSN-normalized cDNA libraries (see methods and Supplementary
86 Table S1). After removal of low-quality reads (See methods), 213,806,111 (77%), high-
87 quality paired reads were retained. These reads contained a total of 43,146,073,200 bp
88 nucleotides with a mean Phred quality higher than 37 (Q37). To develop a comprehensive
89 transcriptomics resource for the Burmese python (Fig. 1), we pooled these high-quality reads
90 from 25 libraries for subsequent *de novo* assembly.

91 *De novo* transcriptome assembly and evaluation

92 As short k-mers have a higher propensity to generate misassembled transcripts when using a
93 de Bruijn graph-based *de novo* assembler, such as Velvet [18], we conservatively chose an
94 assembly generated using long k-mers for subsequent analysis, at the cost of some sensitivity
95 regarding assembled isoforms. Thus, balancing key metrics (Supplementary Table S2), we
96 used an assembly based on the longest k-mer = 95 (Table 1), as it had the fewest
97 scaffolds/transcripts (34,423), but represented a very large proportion (74%) of all reads. The
98 scaffold N50 of this assembly was 1,673 bp.

99 To evaluate the accuracy of the transcriptome assembly, we compared it with the
100 Burmese python reference genome (GenBank assembly accession: GCA_000186305.2) and
101 corresponding gene set in NCBI database using rnaQUAST v1.4.0 [19]. The transcriptome
102 assembly had 34,423 transcripts in total. 34,040 (98%) of these transcripts had at least one

103 significant alignment to the reference genome, and 31,102 (91%) out of 34,040 were uniquely
104 aligned (Supplementary table S3). Average aligned fraction (i.e. total number of aligned
105 bases in the transcript divided by the total transcript length) was 0.975 (Supplementary table
106 S3). The high concordance between the de novo transcript assembly and the genome
107 reference strengthened our confidence in using the de novo assembly as our reference, and
108 showed that the individual fragments were accurate although the reference genome assembly
109 is fragmented. By aligning assembled sequences back to the reference genome, we reviewed
110 the chimeric assembled sequences which had discordant best-scored alignment (partial
111 alignments that are either mapped to different strands/different chromosomes/in reverse
112 order/too distant) and found 1,974 (5.7%) misassembled (chimeric) transcripts
113 (Supplementary table S3). Considered that some of these sequences could be potentially
114 correct, so we included all sequences in our subsequent analysis, but also provided chimeric
115 sequences in a supplementary FASTA file. The comparison of assembled sequences and
116 reference gene sequences (Supplementary table S3) showed that 26,320 (77.3%) assembled
117 transcripts cover at least one isoform from the reference gene set and the mean fraction of
118 transcript matched is 67.8%, suggesting there is a good concordance but also some
119 differences which can be due to errors in either the reference genome assembly/annotation or
120 our assembly. In addition, we assessed the completeness of our transcriptome assembly with
121 the Benchmarking Universal Single-Copy Orthologs (BUSCO) strategy. Results comprised
122 55.2% (1,428 out of 2,586) complete BUSCOs, 19.8% (512) fragmented BUSCOs and 25%
123 (646) missing BUSCOs. These results are consistent with a survey [20] of assessment
124 completeness of 28 transcriptomes from 18 vertebrates. In this survey, most of transcriptomes
125 from species with close phylogenetic relationship to snakes contain less than 50% complete
126 BUSCOs and more than 40% missing BUSCOs. Therefore, we conclude the quality of our
127 transcriptome assembly was acceptable.

128 Transcriptome annotation

1
2
3 129 19,713 transcripts (57% of 34,423) were annotated using transfer of blastx hit annotation
4
5 130 against the non-redundant (nr) NCBI peptide database [21]. To assign proper annotation for
6
7
8 131 each transcript, we chose the first best hit that was not represented in uninformative
9
10 132 descriptions (Supplementary Table S4). The most closely related species with an annotated
11
12
13 133 genome, *Anolis carolinensis* was able to annotate 10,704 transcripts (54% of all annotated
14
15 134 transcripts). Burmese Python and *Anolis carolinensis* both belong to the reptilian order
16
17 135 Squamata, and diverged from each other approximately 160 million years ago [22].
18
19
20

21 136 Blast2GO was used to [23] annotate these 19,713 transcripts, of which 16,992 could
22
23
24 137 be assigned to one or more GO terms and their putative functional roles described. The
25
26 138 distributions of the most frequently identified GO term categories for biological processes
27
28 139 (BP), molecular function (MF) and cellular component (CC) are shown in Fig. S1. Moreover,
29
30
31 140 we used the functionality of InterPro [24] annotations in Blast2GO to retrieve domain/motif
32
33 141 information for our transcripts, and 21,023 transcripts were annotated by the InterPro
34
35
36 142 database.
37
38

39 143 Gene expression analysis and principal component analysis

40
41
42
43 144 For comparisons between genes, expression profiles were obtained by mapping high quality
44
45 145 reads to the reference transcriptome and the expression level was given by fragments per kilo
46
47
48 146 base per million sequenced reads (FPKM) [25]. For the study of expression profiles, we
49
50
51 147 chose to investigate 1862 highly expressed genes (FPKM \geq 100 in at least one of 15 tissues),
52
53 148 as it is known that for highly expressed genes, the biological variation among biological
54
55
56 149 replicates in the same tissue at the same stage is lower than for genes showing low expression
57
58 150 levels [26]. The majority (~64%) of these 1862 genes were expressed in all tissues, and only
59
60
61
62
63
64
65

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 the 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 in 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 numerous 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 this in greater
172 detail, we enlarged the five sub-clusters with the most prominent increase in expression.
173 These sub-clusters, labelled a - e in Figure 3, are shown with full annotation in Figures 4-8.
174 To unravel the functional implications of these responses, we searched for genes encoding for

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

176 within these sub-clusters for each organ.

177 GO enrichment analysis and coloured KEGG pathway maps

178 To gain broader biological insight, in functional annotation analysis we applied a looser

179 threshold set (Table 2) to define DEGs as both maximum FPKM (of three time points) over

180 10 and fold change (FC) over 4 (along with digestion) and highly expressed genes as both

181 maximum FPKM over 200 and FC below 4. The summary of number of genes differentially

182 expressed during digestion in each tissue is illustrated in Table 3. In each organ, most genes

183 (> 76%) have low expression (max FPKM < 10). Around 1% of the genes are highly

184 expressed (max FPKM \geq 200). The number of upregulated genes is approximately 3% in

185 each organ, except for the heart where only 0.57% of the genes were upregulated in response

186 to feeding. This suggests that during digestion, the digestive organs, like liver, stomach,

187 intestine and pancreas show more pronounced post feeding response than the heart. To

188 dissect the functions of DEGs, we performed GO enrichment analysis with upregulated genes

189 and highly expressed genes respectively for each organ (Supplementary Figs. S4-S8). As an

190 example, the GO terms most significantly associated with upregulated genes in the stomach

191 were “mitochondrial respiratory chain complex 1”, “endoplasmic reticulum membrane” and

192 “cytosol” (Supplementary Fig. S4A).

193 To specifically identify the pathways associated with DEGs and highly expressed

194 genes, we mapped genes to KEGG [27, 28] human pathway maps and coloured the mapped

195 entries with trends of gene expression during digestion (Table 2). We identified upregulated

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

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

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

199 pathway categories in the KEGG pathway database (1.3 lipid metabolism and 1.5 amino acid
1
2 200 metabolism; Supplementary table S6). The glycolysis/gluconeogenesis pathway,
3
4
5 201 glyceraldehyde-3 phosphate dehydrogenase, showed high expression in all organs.
6
7

8 202 Identification of the python gastric juice proteome 9

10
11
12 203 We identified the secretome of the python stomach during digestion (Fig. 9). The resulting
13
14 204 mass spectrometry data (containing 122538 MS/MS spectra) was used to interrogate our
15
16
17 205 python transcriptome database, which included transcriptome from stimulated stomach tissue.
18
19 206 In total, 549 python proteins were identified using this approach. Afterwards, all
20
21
22 207 identifications based on a single tryptic peptide were removed, reducing the number of
23
24 208 identified python proteins to 314 (Supplementary Table S7).
25
26
27

28 209 Five classical types of pepsinogens exist, namely pepsinogen A, B, and F,
29
30 210 progastricsin (or pepsinogen C), and prochymosin [29]. Of these, our analyses
31
32
33 211 (Supplementary table S8) show that pythons primarily rely on progastricsin for proteolytic
34
35 212 digestion, as the five most abundant proteases identified in the gastric juice are annotated as
36
37
38 213 progastricsin-like. We aligned the six gastricsin-like transcript sequences using webPRANK
39
40 214 [30] on amino acid level and calculate the pairwise distance between sequences using
41
42
43 215 Tajima-Nei model (Supplementary table S9) in MEGA7 [31]. The mean pairwise distance
44
45 216 1.16 suggests considerable differences in their sequences, which indicate the presence of
46
47
48 217 numerous different proteins with similar functions. This annotation is based on accession
49
50 218 XP_003220378.1 and XP_003220378.1 from *Anolis carolinensis*. Alignment of the python
51
52 219 sequences with the two anole sequences, as well as with the well-characterized human
53
54
55 220 gastricsin variant, shows that both the active site residues, as well as cysteine bridges, are
56
57 221 conserved. It demonstrates the similarity between these enzymes and suggests that the
58
59
60 222 identified python sequences indeed represent catalytically active proteolytic enzymes
61
62
63
64
65

223 (Supplementary Fig. S9). The last identified pepsinogen-like python sequence
1
2 224 (m.31615_Py95) was annotated based on the predicted embryonic pepsinogen-like sequence
3
4 225 (XP_003220239.1), also from *Anolis carolinensis*. Here, the annotation originates from an
5
6
7 226 embryonic pepsinogen identified in chicken [32]. This protease was identified in the python's
8
9
10 227 gastric juice with a lower emPAI value than the gastricsin sequences indicating a lower
11
12 228 concentration of this enzyme (Supplementary table S8), although the transcript displays the
13
14 229 highest concentration of the analysed pepsinogens in the post-prandial period (Supplementary
15
16
17 230 table S9). As the name indicates, it is exclusively expressed during the embryonic period in
18
19 231 chickens [32, 33], and phylogenetic analysis of the sequence suggests that its closest
20
21
22 232 homolog, among the classical pepsinogens, is prochymosin [32]. Prochymosin also displays a
23
24 233 temporal expression pattern and is, in mammals, mainly expressed in new-born species.
25
26
27 234 However, the identified python embryonic-chicken-pepsinogen homolog does not display a
28
29 235 similar development-related temporal expression pattern and is, as shown, produced among
30
31
32 236 adult specimens during digestion. However, this does not exclude the protease also being
33
34 237 expressed during the python's embryonic phase.

37 238 Identification of prey proteins and the python plasma proteome

39
40
41 239 Many of the obtained MS/MS spectra were expected to correspond to abundant mice
42
43
44 240 proteins, such as collagen. To facilitate the downstream analyses of python proteins, we
45
46 241 produced a list of background proteins related to the prey. Hence, cross examination of the
47
48
49 242 mass spectrometry data with the 16693 mouse protein sequences in the Swiss-Prot database
50
51 243 was performed, resulting in the identification of 212 mouse proteins, after removing hits
52
53
54 244 based on single peptides (Supplementary table S10). To produce a list of identified python
55
56 245 proteins, most likely present in the digestive fluid samples due to blood contamination during
57
58
59 246 collection, we characterized the python plasma proteome. The most abundant plasma proteins

247 are produced by the liver. Consequently, our python transcriptome sequence database, which
1
2 248 encompasses liver transcriptomes, is expected to contain the protein sequences of the python
3
4 249 plasma proteins. Thus, our python plasma LC-MS/MS data was used to interrogate our
5
6
7 250 python sequence database. It provided an overview of the most abundant python plasma
8
9
10 251 proteins in Supplementary table S11. In total, 64 plasma proteins were identified with
11
12 252 minimum two tryptic peptides. We observed a limited correlation $R^2=0.13$ fitted with a linear
13
14 253 model (Supplementary table S11) between these abundant (based on emPAI) plasma protein
15
16
17 254 expression and corresponding mRNA expression levels (based on FPKM value at 1 day post-
18
19 255 feeding in liver). One protein that stands out is the anti-haemorrhagic factor cHLP-B
20
21
22 256 (m.27_Py95), which appeared in high concentrations in the plasma of these snakes. This is a
23
24 257 protease inhibitor of the haemorrhagic-causing metalloproteinases present in snake venom
25
26
27 258 and these inhibitors have previously been purified from serum of venomous snakes [34, 35]
28
29 259 and have been proposed to inhibit deleterious actions of venom enzymes in non-venomous
30
31
32 260 snakes [36]. It is, however, also possible that it is an ancestral gene with a function not related
33
34 261 to venom production.

35 36 37 262 Identification of the python stomach secretome

38
39
40
41 263 To identify the python stomach secretome, the list of python proteins, identified in the
42
43 264 digestive fluid (Supplementary table S7) was analysed further. We assumed no overlap
44
45
46 265 between abundant plasma proteins and proteins secreted by the stomach. Thus, plasma
47
48
49 266 proteins, identified in the gastric juice, were assumed to be contaminations from blood and
50
51 267 therefore the 64 identified plasma proteins were, when present, removed from the list.
52
53
54 268 Subsequently, python proteins that most likely were identified based on prey proteins
55
56 269 homology (*e.g.*, mouse collagens and keratins, as well as conserved intracellular household
57
58 270 proteins) were removed. These two steps reduced the list of proteins identified in the stomach
59
60
61
62
63
64
65

271 samples from 314 to 114 proteins (Supplementary table S12). It cannot be excluded that a
1
2 272 few proteins belonging to the python stomach secretome also were removed.
3
4

5
6 273 To identify the secretome, the 114 identified proteins were manually analysed as
7
8 274 described in the method section (Supplementary table S12). In addition to household proteins,
9
10 275 the identified intracellular proteins also included intracellular stomach-specific proteins (*e.g.*
11
12 276 the stomach specific calpain 9 cysteine protease [37]), underlining the specificity of the
13
14 277 proteomics analysis. In total, 37 proteins constituted the putative python stomach secretome
15
16 278 (Supplementary table S8). These could be divided into 18 gastric mucosal-related proteins
17
18 279 (*e.g.* mucin homologous and gastrokine), seven proteolytic enzymes (mainly pepsin
19
20 280 homologous), four other hydrolytic enzymes (*e.g.* phospholipases), and eight other proteins
21
22 281 (*e.g.* gastric intrinsic factor) (Supplementary table S8). Here, we identify stomach-related
23
24 282 proteins from a digesting individual and thereby demonstrate that the sensitivity of modern
25
26 283 LC-MS/MS equipment allows the identification of gastric juice proteins that are present
27
28 284 during digestion.
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

285 **Discussion**

1
2
3
4 286 As a primary motivation, we wished to describe the temporal changes in gene expression in
5
6 287 the visceral organs of Burmese pythons during the transition from fasting to digestion, and
7
8 288 identify key regulatory genes and pathways responsible for the pronounced tissue
9
10 289 restructuring, increased metabolism and the increased functional capacity during the
11
12
13 290 postprandial period. We achieved these goals by identifying the biochemical and
14
15
16 291 physiological roles of highly expressed genes with increased expression during digestion and
17
18 292 by using KEGG analysis of the specific pathways underlying physiological responses known
19
20
21 293 to be stimulated by digestion. We also present GO enrichment analyses of both up-regulated
22
23 294 genes and highly expressed genes in all organs (Supplementary Figs. S4-S8), showing that
24
25 295 “biological process” is the most common enriched category.
26
27
28

29 296 The influence of digestion on gene expression profiles in heart, liver, kidney and small
30
31 297 intestine has been studied previously in pythons [12-14]. These earlier studies reported
32
33
34 298 thousands of genes being either up- or downregulated within the first day of digestion [12-
35
36 299 14], and we confirm these substantial changes in gene expression at 24h and 48h. However,
37
38
39 300 we merely identified hundreds of genes, probably because we selected a more stringent
40
41 301 threshold for defining differential expression. Given the differences in the selection
42
43
44 302 thresholds and analysis strategies for differential expression and differences in the times of
45
46 303 sampling, it is difficult to make a direct comparison between our study and that of Castoe et
47
48
49 304 al (2013). Nevertheless, for heart, liver and small intestine, both studies have determined a
50
51 305 number of upregulated genes at 24h, where we identified 15, 93 and 61 upregulated genes,
52
53 306 respectively. Comparing upregulated genes between two studies (see supplementary material
54
55
56 307 for detailed method and results), we found in liver more than half of the 93 upregulated genes
57
58 308 identified in our study were also identified as upregulated genes by Castoe et al (2013).
59
60
61
62
63
64
65

309 However, there was less overlap for the heart and small intestine. These differences may be
1
2 310 due to the use of different quantification methods for gene expression in the two studies, but
3
4
5 311 may also be a result of the limited biological replicates in our study. Nevertheless, genes
6
7 312 identified as being upregulated in both studies, can be referred to with high confidence.

11 313 Physiological interpretation of the upregulated genes in the stomach

12
13
14 314 The considerable changes in gene expression in the stomach were reflected in a pronounced
15
16
17 315 rise in expression of ribosomal 40S and 60S proteins (Fig. 4) which is likely to have attended
18
19 316 the rise in protein synthesis required for the marked transition from a quiescent fasting state
20
21
22 317 to the activated digestive state. This is also supported by the presence of ribosomal functions
23
24 318 in the enriched GO analysis of the highly-expressed genes in the stomach (Supplementary
25
26
27 319 Fig. S4B). During fasting, gastric acid secretion and presumably also the secretion of
28
29 320 digestive enzymes and lysozymes, is halted, such that the gastric fluid has a neutral pH,
30
31
32 321 whilst ingestion of prey is followed by an immediate activation of gastric acid secretion [38,
33
34 322 39]. The stimulation of the secretory actions of the stomach is attended by an increased mass
35
36 323 of the stomach, where particularly the mucosa expands within the first 24h [40].

37
38
39
40 324 The KEGG analysis, however, shows that the genes encoding for the gastric H,K
41
42 325 ATPase, the active and ATP consuming ion-transporter responsible for gastric acid secretion,
43
44
45 326 are highly expressed in fasting animals, and not additionally elevated in the postprandial
46
47 327 period (Fig. 10). This strongly indicates that the enzymatic machinery for gastric acid
48
49
50 328 secretion is maintained during fasting, a trait that may enable fast activation of acid secretion,
51
52 329 at modest energetic expenditure, to kill bacteria and match gastric pH to the optimum value
53
54
55 330 for pepsin. This interpretation is consistent with a number of recent studies indicating a rather
56
57 331 modest contribution of gastric acid secretion to the specific dynamic action (SDA) response
58
59
60 332 in pythons [41, 42], but we also did observe a high prevalence of ATP synthase subunits (Fig.

333 4) amongst the highly upregulated genes, which does indicate a rise in aerobic metabolism
1
2 334 (see also supplementary Fig. S4). Furthermore, the upregulation of the gene encoding for
3
4 335 creatine kinase (Fig. 4) indicate increased capacity for aerobic respiration required costs of
5
6
7 336 acid secretion and the stimulation of the accompanying gastric functions. It has been
8
9
10 337 proposed that gastric processes account for more than half of the rise in total metabolism
11
12 338 during digestion [38], and aerobic metabolism of isolated gastric strips *in vitro* increased
13
14 339 during digestion [43]. However, while metabolism of the stomach certainly must increase
15
16
17 340 during the postprandial period, more recent studies indicate a considerably smaller
18
19 341 contribution of gastric acid secretion to the total SDA response, meaning that gastric acid
20
21
22 342 comprises considerably lower than 50% of the SDA [41, 42, 44].
23
24

25 343 Our KEGG analysis also showed a large rise in expression of the gene encoding for
26
27 344 carbonic anhydrase (Fig. 10), the enzyme that hydrates CO₂ and provide protons for gastric
28
29
30 345 acid secretion. Gastric acid secretion, therefore, does not appear to undergo transcriptional
31
32
33 346 regulation, but is likely to involve translocation of existing H,K ATPases in vesicles from
34
35 347 intracellular vacuoles to the apical membrane of the oxyntopeptic cells that are responsible
36
37
38 348 for both gastric acid secretion as well as the release of pepsinogen in reptiles [45]. An
39
40 349 activation of the processes involved in vesicle transport is further supported by increased
41
42
43 350 transcription of the gene encoding for CD63 (Fig. 4), which belongs to the tetraspanin family
44
45 351 and mediates signal transduction events.
46
47

48
49 352 In contrast to acid secretion, expression of several genes encoding digestive enzymes
50
51 353 (embryonic pepsinogen-like, gastricsin precursor and gastricsin-like peptides) (Fig. 4) were
52
53 354 upregulated, which is consistent with *de novo* synthesis of the enzymes responsible for gastric
54
55
56 355 protein degradation. Also, there was good overlap between the upregulation of the relevant
57
58
59 356 genes encoding the proteins identified in the stomach secretome, such as gastrokines, pepsin
60
61
62
63
64
65

357 homologs, phospholipases and gastric intrinsic factor (Supplementary table S8). In this
1
2 358 context, it is also interesting that mucin 6 (Fig. 4), the gene coding for the large glycoprotein
3
4
5 359 (gastric mucin) that protects the gastric mucosa from the acidic and proteolytically active
6
7 360 chyme in the stomach lumen was upregulated. Thus, as gastric acid secretion is activated,
8
9
10 361 probably in response to increased levels of the gastrin as well as luminal factors, there is an
11
12 362 accompanying activation of the protective mucus layer that prevents auto-digestion of the
13
14
15 363 gastric mucosa. It is also noteworthy that the genes for both gastrokine 1 and 2 were
16
17 364 upregulated during digestion (Fig. 4). Gastrokines are constitutively produced proteins in the
18
19 365 gastric mucosa in mammals and chickens, and while their physiological function remains
20
21
22 366 somewhat elusive, they appear to be upregulated during mucosal remodelling in response to
23
24 367 inflammation (*e.g.*, in connection with ulcers) and often downregulated in cancers. Thus, it is
25
26
27 368 likely that the gastrokines are involved in regulating the restructuring of the mucosa during
28
29 369 digestion in pythons.

33 370 In addition to analysing the gene expression profiles of the stomach, we also used a
34
35 371 proteomics approach, assisted by our python transcriptome sequence database, to identify the
36
37
38 372 hydrolytic enzymes in the gastric juice secreted during digestion. We identified python
39
40 373 proteins on a complex background of highly abundant mice proteins. Thus, the digestive
41
42
43 374 enzymes secreted by the pancreas are probably functionally similar to known hydrolytic
44
45 375 enzymes from other species.

48
49 376 We hypothesized that relatively aggressive proteolytic digestive enzymes in the
50
51 377 gastric juice facilitate digestion of large and un-masticated whole prey items [8]. In our
52
53
54 378 analysis, six out of the seven identified proteolytic enzymes were pepsinogen homologs
55
56 379 (Peptidase subfamily A1A), and these were also the most abundant hydrolytic enzymes in the
57
58
59 380 gastric juice according to the emPAI values (Supplementary table S8). It is likely that other

381 pepsinogen isoforms exist in the gastric juice, as our approach predominantly targeted the
1
2 382 most abundant proteolytic enzymes. The importance of the proteomics-identified pepsinogens
3
4
5 383 was also substantiated by the transcriptomics data (Supplementary table S9). Here, we found
6
7 384 that the six different pepsinogens were upregulated between 2.2 and 22.2 fold from the
8
9
10 385 fasting animals to 48 hours after ingestion of mice. On average the pepsinogen transcripts
11
12 386 were upregulated 10.7 fold. It supports that these proteases play a substantial role in the
13
14
15 387 aggressive digestion process performed by the python.

16
17
18 388 Our proteomic analysis also suggested the identification of the pepsinogens as the
19
20
21 389 major digestive proteolytic enzymes, similar to in all other vertebrate species. Thus, our
22
23 390 results indicate that the pepsinogen is not unique (with respect to protease class) and hitherto
24
25 391 uncharacterized proteases do not facilitate the aggressive digestion process. Instead, pepsins,
26
27
28 392 homologous to pepsins among other species, digest the intact swallowed prey. As in other
29
30 393 vertebrates, pythons have a low gastric pH during digestion [38, 42], and it is likely that these
31
32
33 394 pepsins variants are among the most effective and aggressive pepsins identified so far and our
34
35 395 sequence information facilitate future cloning, expression, and characterization of these
36
37
38 396 potentially industrial relevant enzymes.

41 397 Physiological interpretation of the upregulated genes in the intestine

42
43
44
45 398 The small intestine of pythons undergoes a remarkable and fast expansion during digestion
46
47 399 where both wet and dry mass more than double within the first 24 hours. The expansion
48
49
50 400 stems primarily from increased mucosal mass, achieved by swelling of the individual
51
52 401 enterocytes [46], while the smooth muscle in the gut wall is much less responsive [47].
53
54
55 402 Earlier studies on gene expression profiles during digestion in the python intestine revealed
56
57 403 massive upregulation of more than one thousand genes, commencing within the first six
58
59
60 404 hours after ingestion [12, 13]. Importantly, this previous study [13] identified a number of

405 genes that are likely to be involved in the restructuring of the microvilli, cell division and
1
2 406 apoptosis, as well as brush-border transporter proteins. In line with these earlier findings, our
3
4
5 407 GO enrichment analysis also highlights functions pertaining to mitotic cell division, which
6
7 408 supports a contribution to growth by hyperplasia faster cell turnover (Supplementary Fig. S5).
8
9
10 409 The expansion of the individual enterocytes is accompanied by pronounced elongation of the
11
12 410 microvilli [48] and the resulting rise in surface area of the intestinal lining is accompanied by
13
14
15 411 a ten-fold increase in intestinal transport capacity for amino acids and other nutrients [1, 4,
16
17 412 49].
18
19
20
21 413 Earlier studies provided strong evidence for an upregulation of genes coding for nutrient
22
23 414 transporter proteins, such as D-glucose, L-proline and L-leucine [13]. In this context, it is
24
25
26 415 noteworthy that there were no nutrient transporters amongst the highly expressed and
27
28 416 upregulated genes in the intestine (Fig. 5), but our KEGG analysis nevertheless showed
29
30
31 417 increased expression of the serosal L-type amino acid transporter. Clearly, it would be
32
33 418 worthwhile to quantitatively analyse the extent to which *de novo* synthesis of the various
34
35
36 419 nutrient transporters, particularly those for amino acids, is increased during digestion and
37
38 420 how much such synthesis contributes to absorptive capacity. It would seem adaptive if many
39
40
41 421 of the transporters merely have to be activated, either by insertion within the luminal
42
43 422 membrane or exposed as the enterocytes expand: an energetically cheap manner of matching
44
45 423 intestinal performance to the sudden appearance of nutrients in the intestine after a meal. The
46
47
48 424 GO enrichment analysis also pointed to an enrichment of various metabolic processes during
49
50
51 425 digestion, particularly for the upregulated genes (Supplementary Fig. S5). It is noteworthy
52
53 426 that the expression of genes coding for glutathione S-transferase, peroxiredoxin and
54
55 427 selenoprotein increased during digestion (Fig. 5). These three proteins are involved in cellular
56
57 428 defence, particularly as antioxidants as a likely protection from the reactive oxygen species
58
59
60 429 that result from increased aerobic metabolism.

430 There is consensus that the anatomical and structural responses underlying this
1
2 431 phenotypic flexibility of intestinal function occur at modest energetic expenditure [17, 38,
3
4 432 50], but our expression profile does show increased expression of the gene coding for
5
6
7 433 cytochrome P450, pointing to increased aerobic and mitochondrial metabolism. An increased
8
9
10 434 expression of the genes involved in oxidative phosphorylation was also reported in earlier
11
12 435 studies on pythons [12, 13]. This rise in metabolism may be driven primarily by the massive
13
14 436 rise in secondary active transport to absorb the amino acids and smaller peptides rather than
15
16
17 437 the structural changes [50]. Nevertheless, the structural changes may be reflected in increased
18
19 438 expression of galectin 1 (Fig. 5), which mediates numerous functions including cell–cell
20
21
22 439 interactions, cell–matrix adhesion and transmembrane signalling.
23
24

25 440 Fig. 5 reveals the importance of lipid absorption and its subsequent transport by the
26
27
28 441 cardiovascular and lymph systems, and it is also possible that several of the expressed
29
30 442 proteins play a role in the incorporation of lipid droplets within the enterocytes. Thus, the
31
32
33 443 presence of numerous apolipoproteins, and their precursor apoe protein, amongst the list of
34
35 444 highly expressed and highly expressed genes (Fig. 5) are probably required to transport the
36
37
38 445 absorbed lipids in plasma and lymph, but the apolipoproteins could also act as enzyme
39
40 446 cofactors, receptor ligands, and lipid transfer carriers in the regulation of lipoprotein
41
42
43 447 metabolism and cellular uptake. The presence of diazepam-binding inhibitor (Fig. 5), a
44
45 448 protein involved in lipid metabolism and under hormonal regulation mostly within nervous
46
47 449 tissue, is also likely to reflect the increased lipid absorption and metabolism in the
48
49
50 450 postprandial period, and there was also a rise in phospholipases (Fig. 5) that are likely to be
51
52 451 involved in lipid degradation. Also, the capacity for protein metabolism clearly increased in
53
54
55 452 the intestine during digestion (seen in *e.g.*, meprin A and endopeptidase that cleave peptides,
56
57 453 as well as 4-aminobutyrate aminotransferase, 4-trimethylaminobutyraldehyde dehydrogenase
58
59
60 454 and diamine acetyltransferase) and there was a rise in the ammonium transporter protein Rh
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

455 (Fig. 5). Finally, a number of proteins involved in calcium uptake and metabolism, such as
456 calbindin and calmodulin (Fig. 5), could be important to handle the break-down of the bone
457 in a normal rodent, and it was recently shown the enterocytes of pythons already contain
458 small particles of bone at 24 hours after ingestion [48].

459 Physiological interpretation of the upregulated genes in the heart

460 The large metabolic response to digestion is accompanied by a doubling of heart rate and
461 stroke volume of the heart such that cardiac output remains elevated for many days during
462 digestion [51, 52]. This cardiovascular response plays a pivotal role in securing adequate
463 oxygen delivery to the various organs and serves to ensure an appropriate convective
464 transport of the nutrients taken up by the intestine. The tachycardia is mediated by a release
465 of vagal tone and the presence of a non-adrenergic-non-cholinergic factor which stimulates
466 the heart, which has been speculated to be released from the gastrointestinal organs during
467 digestion [53, 54]. The increased heart rate, and the rise in the volume of blood pumped with
468 each beat, must be supported by increased myocardial metabolism and we observed an
469 upregulation of malate dehydrogenase, cytochromes and ATPase linked enzymes (Fig. 6) that
470 are likely to be related to an increased oxidative phosphorylation within the individual
471 myocytes (see also the prevalence of enriched GO terms associated with aerobic metabolism
472 in Supplementary Fig. S8). Previous gene expression studies on the python heart also yielded
473 evidence for its increased oxidative capacity in postprandial period [55] and cytochrome
474 oxidase activity is almost doubled during digestion [56]. We confirm that transcription for
475 heat shock proteins may be increased [55], possibly to protect against oxidative damage as
476 result of the increased metabolism. As in earlier studies [55], our observation of increased
477 ATP synthase lipid-binding protein and fatty acid binding protein 3 (Fig. 6) provide evidence
478 for increased fatty acid metabolism, which may reflect the substantial rise in circulating fatty

479 acids in the plasma.

1
2
3
4 480 It was originally suggested that the postprandial rise in stroke volume could be
5
6 481 ascribed to an impressive and swift growth of the heart [10], possibly triggered by lipid-
7
8 482 signalling [55]. However, a number of recent studies, primarily from our laboratory, have
9
10
11 483 shown that increased cardiac mass is not an obligatory postprandial response amongst
12
13 484 pythons [56-58], and that stroke volume may be increased in response to increased venous
14
15 485 return rather than cardiac hypertrophy [56]. It is nevertheless, noteworthy that our and the
16
17
18 486 previous studies show a clear increase in the expression of contractile proteins (e.g. myosin
19
20
21 487 and actin) as well as tubulin (Fig. 6), which may reflect increased protein-turnover in
22
23 488 response to increased myocardial workload rather than cell proliferation or hypertrophy. The
24
25 489 enriched GO analyses also point to major changes in the extracellular space as well as both
26
27
28 490 elastin and collagen, which may indicate some level of cardiac reorganization at the cellular
29
30
31 491 or subcellular level that may alter compliance of the myocardial wall and influence cardiac
32
33 492 filling (Supplementary Fig. S8). It is noteworthy that the increased expression of BNP may
34
35 493 serve a signalling function as described in response to the cardiac hypertrophy that attends
36
37
38 494 hypertension.

39
40
41 495 Physiological interpretation of the genes in the liver

42
43
44
45 496 The liver exhibited a diverse expression profile in response to digestion that is likely to
46
47 497 reflect its many metabolic functions in connection with metabolism, synthesis and
48
49
50 498 detoxification during the postprandial period. This pattern is also evident from the many
51
52 499 metabolic functions identified in the enriched GO analysis (Supplementary Fig. S7). There
53
54
55 500 were marked upregulations of the P450 system (Fig. 7), which fits well with a rise in
56
57 501 synthesis and breakdown of hormones and signalling molecules, cholesterol synthesis in
58
59
60 502 response to lipid absorption and possibly also an increased metabolism of potentially toxic

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

503 compounds in the prey. A rise in cholesterol metabolism was supported by increased
504 expression apolipoproteins (Fig. 7). The hepatic involvement in lipid metabolism was also
505 supported by the increased expression of genes for Alpha-2-macroglobulin and serum
506 albumin (Fig. 7). The increased expression of albumin obviously also corresponds nicely with
507 the proteomic analysis of plasma proteins and it is likely that the postprandial rise in plasma
508 albumin serves a functional role in the lipid transport between the intestine and the liver as
509 well as other metabolically active organs.

18 510 It is also noteworthy that a number of genes associated with the protection of
19
20
21 511 oxidative stress, such as catalase, heat shock protein and glutathione transferase were
22
23 512 markedly upregulated (Fig. 7). It was recently argued that snakes digesting large meals
24
25 513 experience oxidative damage due to reactive oxygen metabolites requiring increased
26
27
28 514 antioxidant responses to protect cellular functions [59].

31
32 515 Physiological interpretation of the genes in the pancreas

33
34
35 516 We sampled the entire pancreas for our analysis of gene expression and our data therefore
36
37
38 517 reflect both endocrine and exocrine pancreatic functions. We found ample evidence for
39
40 518 upregulated expression of genes associated with the digestive functions, such as lipases,
41
42
43 519 trypsin, chymotrypsin and elastase and other enzymes for digestion of protein and lipid (Fig.
44
45 520 8). This general upregulation of secretory processes is likely to explain the prevalence of
46
47
48 521 processes associated with protein synthesis in the enriched GO analysis (Supplementary Fig.
49
50 522 S6). There was even an increased expression of amylase (Fig. 8) which breaks down
51
52 523 polysaccharides. In connection with this latter function, the increased expression of insulin
53
54
55 524 (Fig. 8) from the endocrine pancreas is likely to reflect increased cellular signalling for
56
57 525 postprandial uptake of both glucose and amino acids. As in the other organs, we found
58
59
60 526 increased expression of cytochrome oxidase (Fig. 8) indicative of increased metabolism

1
2 527 during digestion, and the rise in heat shock protein expression may reflect a response to
3 528 formation of reactive oxygen-species as metabolism is stimulated by increased secretion of
4
5 529 the pancreas.
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

530 **Conclusions**

1
2
3
4 531 Our study confirms that the extensive physiological and anatomical reorganization of the
5
6 532 visceral organs of pythons during the postprandial period is driven by differential expression
7
8 533 of hundreds or even thousands of genes. Many of the upregulated functions pertain to energy
9
10 534 production to support the rise in aerobic metabolism associated with the digestion and
11
12 535 absorption of large meals. In terms of the gastrointestinal organs, the gene expression profiles
13
14 536 also support the view that many of the digestive functions, such as gastric acid secretion and
15
16 537 nutrient absorption, can be stimulated with little change to gene expression, indicating that
17
18 538 the proteins involved in these processes merely need to be activated during the postprandial
19
20 539 period, and thus avoiding the energy and time-consuming processes associated with *de novo*
21
22 540 synthesis. This digestive strategy may, at least in part, explain how intermittent feeders, such
23
24 541 as snakes, retain the capacity for rapid and reliable upregulation of the digestive processes
25
26 542 immediately after prey ingestion.
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

543 **Methods**

1
2
3
4 544 Stimulation of the postprandial response, collection of tissue biopsies and purification of
5
6 545 RNA for mRNA-seq analyses
7
8

9
10 546 Six *Python bivittatus* (Tiger Python/Burmese Python) with a body mass ranging from 180 to
11
12 547 700 g (average 373 g) were obtained from a commercial supplier and housed in vivaria with a
13
14 548 heating system providing temperatures of 25-32 °C. The animals were fed rodents once a
15
16
17 549 week and fresh water was always available. The animals appeared healthy and all
18
19 550 experiments were performed according to Danish Federal Regulations. All six individuals
20
21
22 551 were fasted for one month and divided in three groups. Four animals were fed a rodent meal
23
24 552 of 25 % of body weight and euthanized with an intra-peritoneal injection of pentobarbital (50
25
26 553 mg kg⁻¹; Mebumal) at 24h (N = 2) or 48h after feeding (N = 2). The remaining two snakes
27
28
29 554 served as fasted controls. During deep anaesthesia, two biopsies were obtained from each
30
31 555 snake from each of the following tissues: heart (ventricles), liver, stomach, intestine, and
32
33
34 556 pancreas. In regard to the stomach tissue samples, one sample was obtained from the
35
36 557 proximal part of the stomach and one sample was obtained from the distal part. In total, 60
37
38
39 558 biopsies were collected. The samples were taken from the same part of the different tissues in
40
41 559 all individuals. After sampling, the biopsies were weighed and immediately snap frozen in
42
43
44 560 liquid nitrogen; stomach and intestinal tissues were rinsed in sterile saline solution before
45
46 561 weighting to avoid contamination with rodent tissue from the ingested meal. Subsequently,
47
48
49 562 all 60 biopsies were homogenized in liquid nitrogen and the four biological replicates (two
50
51 563 biopsies from each individual) were pooled in a 1:1 manner based on mass. This resulted in
52
53
54 564 15 samples (five tissues X three time points). From these samples, total RNA was purified
55
56 565 using the Nucleospin RNA II kit (Machery-Nagel GmbH & Co.), as recommended by the
57
58 566 manufacturer. The RNA concentration and quality were assessed by Nanodrop ND 1000
59
60
61
62
63
64
65

1 567 Spectrophotometer (Thermo Scientific) analyses, agarose gel-electrophoreses, and Agilent
2 568 BioAnalyzer (Agilent) analyses.
3
4
5
6 569 Library production and sequencing
7
8
9
10 570 Poly-A transcripts were enriched and the transcripts broken in the presence of Zn²⁺.
11
12 571 Subsequently, double-stranded cDNA was synthesized using random primers and RNase H.
13
14 572 After end repair and purification, the fragments were ligated with bar-coded paired-end
15
16 573 adapters, and fragments with insert sizes of approximately 150-250 bp were isolated from an
17
18 574 agarose gel. Each of the 15 samples derived from five tissues (heart, liver, stomach, pancreas
19
20
21 575 and intestine) at the three time points (fasted for one month, 24h and 48h post-feeding) were
22
23
24 576 amplified by PCR to generate DNA colonies template libraries and the libraries were then
25
26 577 purified. In addition, to sample as broadly from each transcriptome as possible, we also
27
28
29 578 produced normalized libraries for each tissue in order to capture the reads from lowly
30
31 579 expressed, tissue-specific genes. Here, a part of the samples, which originating from the same
32
33 580 tissue, were pooled before the PCR analyses, i.e. in total five pooled samples were generated.
34
35
36 581 These five samples were split in two and after PCR amplification and library purification they
37
38
39 582 were normalized using two different normalization protocols, i.e. in total 10 normalized
40
41 583 libraries were prepared. Library quality of all 25 samples was then assessed by a titration-run
42
43 584 (1 x 50 bp) on an Illumina HiSeq 2000 instrument. Finally, the sequencing was performed on
44
45
46 585 the same instrument using paired-reads (2 × 101 bp). One channel was used for the 15 non-
47
48
49 586 normalized libraries and one channel was used for the 10 normalized libraries.
50
51
52 587 Data pre-processing and *de novo* transcriptome assembly
53
54
55
56 588 To reduce the amount of erroneous data, the raw paired reads were processed by i) removing
57
58 589 reads that contained the sequencing adaptor, ii) removing reads that contained ambiguous
59
60
61
62
63
64
65

1
2 590 characters (Ns), and iii) trimming bases that had the low average quality (Q<20) within a
3 sliding window of length 10.
4

5
6 592 To develop a comprehensive transcriptomics resource for the Burmese python, all
7 high-quality reads from 25 libraries were pooled together for *de novo* assembly. To determine
8 593 the optimal assembly, *de novo* assembly was performed using Velvet (version 1.2.03)(
9 Velvet, RRID:SCR_010755) [18] and Oases (Oases, RRID:SCR_011896) (version 0.2.06)
10 [60] with different k-mer parameters. The performance of these assemblies was assessed
11 594 according to number of transcripts, total length of transcripts, N50 length, mean length,
12 proportion of mapped reads and number of transcripts which length is larger than N50
13 595 (Supplementary Table S2).
14
15
16
17
18
19
20
21
22
23
24
25
26

27 600 Assessment of the transcriptome assembly

28
29
30
31 601 The transcriptome assembly was evaluated by rnaQUAST 1.4.0 with default parameters
32 supplying reference genome sequences and genome annotation of Burmese python (GenBank
33 602 assembly accession: GCA_000186305.2).
34
35
36
37
38

39 604 BUSCO_v2 (BUSCO , RRID:SCR_015008) [20] was used to test the completeness of
40 transcriptome assembly with dependencies NCBI BLAST+ 2.4.0 [61] and HMMER 3.1b2
41 605 (Hmmer, RRID:SCR_005305)[62]. The vertebrata lineage set was used and accessed on 28
42 606 Nov 2016.
43
44
45
46
47
48
49

50 608 Transcriptome annotation

51
52
53 609 To assess the identity of the most closely related gene in other organisms, the assembled
54 transcripts were compared with the sequences in the National Center for Biotechnology
55 610 Information (NCBI) non-redundant protein (nr) database using blastx (BLASTX ,
56 611 RRID:SCR_001653) [63] with an e-value cut-off of 0.01. The nr annotation term of each
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

613 transcript was assigned with the first best hit, which was not represented in uninformative
614 description (e.g., 'hypothetical protein', 'novel protein', 'unnamed protein product', 'predicted
615 protein' or 'Uncharacterized protein') (Supplementary Table S4). To assign functional
616 annotations of transcripts, Blast2GO was used (e-value threshold = 0.01) to return GO
617 annotation, Enzyme code annotation with KEGG maps and InterPro annotation.

618 Estimation of gene expression values

619 For each 15 non-normalized libraries, the paired-end reads were firstly mapped back to
620 assembled transcriptome using Bowtie2 (Bowtie , RRID:SCR_005476) [64]with default
621 parameters, the raw counts then were calculated based on the alignment results using RSEM
622 (version 1.1.20) [65] for each transcript. To quantify the gene expression level, for genes with
623 alternative splicing transcripts, the longest transcript was selected to represent the gene, and a
624 gene's abundance estimate was the sum of its transcripts' abundance estimates. Finally, the
625 raw expression counts were normalized into FPKM with custom Perl scripts.

626 PCA

627 To facilitate graphical interpretation of tissue relatedness, R function prcomp was used to
628 perform PCA with genes which the maximum FPKM of 15 samples was greater than 100.

629 Identification of DEGs and clustering analysis

630 For each tissue, DEGs were selected with two thresholds, 1) FPKM is greater than or equal to
631 400 in at least one time point and 2) FC is greater than or equal to two in at least one pairwise
632 comparison among three time points. FPKM values of DEGs were log₂-transformed and
633 median-centered, then hierarchical clustering was performed using R command hclust with
634 method = 'average' and distance = 'Spearman correlation' and results were displayed using R
635 command heatmap.2.

636

1
2 637 Coloured KEGG Pathway and GO enrichment analysis

3

4

5 638 For each tissue, all assembled genes were mapped to KEGG human pathway maps using

6

7 639 KOBAS 2.0 [66] with e-value $1e-50$. Then genes were coloured by representing FPKM value

8

9 640 and trend of differential expression value (Table 2).

10

11

12 641 Blast2GO was used to implement GO enrichment analysis (Fisher's exact test) with

13

14 642 threshold of FDR 0.001. The reference set is the whole transcripts with GO slim annotation.

15

16 643 For each organ, the selected test set is either upregulated or highly expressed genes defined in

17

18 644 Table 2. Finally, we performed Blast2GO to reduce to most specific GO terms.

19

20

21

22 645 Isolation of samples for proteomics analyses

23

24

25 646 Two Burmese pythons (weighing 400 and 800 g, respectively) were fed a rodent meal

26

27 647 corresponding to approximately 25% of their body mass. Approximately 24 h into the

28

29 648 postprandial period the animals were euthanized with an overdose of pentobarbital (100 mg

30

31 649 kg^{-1} , i.m.). Immediately afterwards, an incision was made to expose the stomach, which was

32

33 650 then ligated at the lower oesophagus and the pylorus, before the intact stomach was excised

34

35 651 by a cleavage just below the two sutures resulting in the stomach being released from the rest

36

37 652 of the animal. All undigested mouse remains were manually removed by forceps and 25

38

39 653 ml/kg tris-buffered saline (TBS) was injected into the stomach. The stomach was then ligated

40

41 654 at the opened end, rinsed by gently shaking the tissue, and finally the digestive fluid-

42

43 655 containing solution was collected and stored on ice. To ensure collection of all gastric fluid,

44

45 656 the stomach was rinsed additional two-three times with 12 ml/kg TBS. Subsequently, the

46

47 657 samples were filtered and centrifuged, and the supernatant stored at $-80\text{ }^{\circ}\text{C}$. We also obtained

48

49 658 two samples of gastric juice from a third individual (200 g) that had been fed 4 g peptone

50

51

52

53

54

55

56

57

58

659 (Sigma Aldrich), suspended in water. Peptone is a mixture of small peptides and amino acids
1
2 660 and the solution was injected directly into the stomach and after three hours the snake was
3
4
5 661 euthanized by an overdose of pentobarbital. The stomach was removed, rinsed with TBS, and
6
7 662 a single sample collected and stored, as described above. We analysed two samples from each
8
9
10 663 of the three individuals, resulting in a total of six digestive fluid samples being analysed by
11
12 664 MS/MS. In addition, we obtained a single plasma sample from each snake by direct cardiac
13
14
15 665 puncture followed by centrifugation and storage for later analysis.

16 17 18 666 Sample preparation for mass spectrometry analyses

19
20
21
22 667 The proteins in the six obtained python digestive fluid samples were recovered by
23
24 668 trichloroacetic acid precipitation. The resulting pellets were resuspended in 8 M Urea, 5 mM
25
26
27 669 DTT, 0.1 M ammonium bicarbonate pH 8.0 and incubated for 30 minutes at room
28
29 670 temperature in order to denature and reduce the proteins. Subsequently, the proteins were
30
31
32 671 alkylated by the addition of iodoacetamide to a final concentration of 25 mM. The samples
33
34 672 were incubated for additional 20 minutes at room temperature and then diluted five times
35
36 673 with a 50 mM ammonium bicarbonate, pH 8.0 buffer before the addition of approximately 2
37
38
39 674 μg sequencing grade modified trypsin (Promega) per 50 μg protein in the sample.
40
41 675 Subsequently, the samples were incubated at 37 °C for approximately 16 h. The proteins in
42
43
44 676 the plasma sample were denatured, reduced, alkylated, and digested with trypsin, as described
45
46 677 for the digestive fluid samples. Finally, the resulting peptides in all samples were
47
48
49 678 micropurified and stored at -20 C until the LC-MS/MS analyses.

50 51 52 679 Liquid chromatography-tandem mass spectrometry analyses

53
54
55
56 680 Nano-liquid chromatography-tandem mass spectrometry (LC-MS/MS) analyses were
57
58
59 681 performed on a nanoflow HPLC system (Thermo Scientific, EASY-nLC II) connected to a
60
61
62
63
64
65

682 mass spectrometer (TripleTOF 5600, AB Sciex) equipped with an electrospray ionization
683 source (NanoSpray III, AB Sciex) and operated under Analyst TF 1.6 control. The samples
684 were dissolved in 0.1% formic acid, injected, trapped and desalted isocratically on a
685 precolumn whereupon the peptides were eluted and separated on an analytical column (16 cm
686 × 75 µm i.d.) packed in-house with ReproSil-Pur C18-AQ 3 µm resin (Dr. Marisch GmbH).
687 The peptides were eluted at a flow rate of 250 nL/min using a 50 min gradient from 5 % to 35
688 % phase B (0.1 % formic acid and 90 % acetonitrile). An information dependent acquisition
689 method was employed allowing up to 25 MS/MS spectra per cycle of 2.8 s.

690 Protein identification and filtering of data

691 The six collected MS files, related to digested fluid, were converted to Mascot generic format
692 (MGF) using the AB SCIEX MS Data Converter beta 1.3 (AB SCIEX) and the “proteinpilot
693 MGF” parameters. Subsequently, the files were merged to a single MGF-file using Mascot
694 daemon. The resulting file (encompassing 122538 MS/MS queries) was used to interrogate
695 the 16693 *Mus musculus* sequences in the Swiss-Prot database (version 2014_10) and the
696 generated python database encompassing 21131 protein sequences using Mascot 2.5.0
697 (Matrix Science)[67]. Trypsin, with up to one missed cleavage allowed, was selected as
698 enzyme; carbamidomethyl was employed as fixed modification, and oxidation of methionine
699 and proline was selected as variable modifications. The instrument setting was specified as
700 ESI-QUAD-TOF, the mass accuracy of the precursor and product ions was 15 ppm and 0.2
701 da respectively, and the significance threshold (p) was set to 0.01 and an expect cut-off at
702 0.005. The data obtained by the LC-MS/MS-analysis of the python plasma proteome was
703 analysed as described for the digestive fluid samples, except that the *Mus musculus* sequences
704 were not interrogated. This dataset contains 9224 MS/MS queries. All obtained results were
705 subsequently parsed using MS Data Miner v. 1.3.0 [68], and protein hits were only accepted

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

706 if they were identified based on two unique peptides. Semi-quantitative proteomics data was
707 obtained using the emPAI-values given by the Mascot 2.5.0 software after analysis of the
708 MS/MS data [69].

8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

709 To identify the proteins secreted into the python stomach, identified python plasma
710 proteins, as well as the mouse protein homologs were removed from the list of identified
711 python digestive fluid proteins. With regard to the removal of prey protein homologs, the
712 overall mouse protein names were used to search the list of python proteins (e.g. “collagen”
713 was used as search term, not “collagen alpha-1(I) chain”) and to identify python proteins that
714 were identified based on homology with mouse. These proteins were removed from the list of
715 stomach-secreted python proteins. For each identified protein remaining on the list, we
716 reassessed the annotation of the python sequence, i.e. sequence comparisons were performed
717 using blastp version 2.2.30, and in addition, UniProt and NCBI protein databases, as well as
718 PubMed and SignalP 4.1, were interrogated to identify functional properties and cellular
719 location of the identified proteins. Plasma proteins, remaining collagen homologous,
720 intracellular proteins, and membrane proteins were discarded from the list of identified
721 python stomach secretome proteins.

722 The mass spectrometry proteomics data have been deposited to the ProteomeXchange
723 Consortium via the PRIDE [70] partner repository with the dataset identifier PXD006665.

724 **List of abbreviations**

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

725 DEG differentially expressed genes

726 FC fold change

727 FPKM fragments per kilo base per million sequenced reads

728 PCA principal component analysis

1 729 **Declarations**

2
3

4
5 730 Availability of data and materials

6
7

8 731 The raw RNA-Seq sequencing data that support the findings of this study have been

9
10

11 732 deposited in the NCBI BioProject database (accession no. PRJNA343735).

12
13

13 733 <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA343735>

14
15

16
17 734 The mass spectrometry proteomics data have been deposited to the ProteomeXchange

18
19

19 735 Consortium via the PRIDE [70] partner repository with the dataset identifier PXD006665.

20
21

22 736 Supporting data is also available from the *GigaScience* GigaDB repository[71].

23
24

25 737 Competing interests

26
27

28
29 738 The authors declare that they have no competing interests.

30
31

32
33 739

34
35

36 740 Funding

37
38

39
40 741 This work was supported by a grant (Novenia) from the Danish Research Council for

41
42

42 742 Strategic Research (grant identification number: 09-067076).

43
44

45
46 743

47
48

49
50 744 Authors' contributions

51
52

53 745 JD, KWS, TW and MHS designed the study. JD performed the transcriptome data analysis

54
55

56 746 with input from LS and was a major contributor in writing the manuscript. SEL performed

57
58

58 747 RNA-Seq lab experiment. KWS and JE performed the proteomics experiment and data

59
60

61
62

63
64
65

748 analysis. WT interpreted the transcriptome data regarding digestion. All authors read and

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

749 approved the final manuscript.

750

751 Acknowledgements

752 We thank Tania A. Nielsen (Aarhus, Denmark) for valuable assistance with RNA purification

753 and FASTERIS SA (Switzerland) for library preparation and Illumina sequencing.

754 **Figure and table legends**

1
2
3 755 **Fig. 1. The workflow of Python RNA-Seq data analysis.** The diagram shows the main
4
5
6 756 steps and bioinformatics tools used in the study.

7
8 757 **Fig. 2. PCA plots of FPKM of 1862 genes.** PC, principal component. PC1 represents 25%,
9
10 758 PC2 represents 18% and PC3 represents 16% of total variation in the data. The name of the
11
12 759 label consists of two parts: one capital letter plus one number. Letter H, S, I, L, P represent
13
14 760 heart, stomach, intestine, liver and pancreas respectively. Numbers 0, 1, 2 represent fasting
15
16
17 761 for one month, 24h/1d after feeding and 48h/2d after feeding respectively.

18
19 762 **Fig. 3. Heat maps from hierarchical clustering of DEGs in each tissue.** Heat maps
20
21 763 showing the hierarchically clustered Spearman correlation matrix resulting from comparing
22
23 764 the normalized FPKM value for each pair of genes. Heat map columns represent samples
24
25
26 765 and rows correspond to genes. Expression values (FPKM) are \log_2 -transformed and then
27
28 766 median-centered by gene. Relative levels of gene expression are represented by colours.
29
30 767 Pale colour is low expression and darker blue is high expression. Five sub-clusters labelled
31
32 768 a to e are shown with full annotation in Fig. 4-8.

33
34
35 769 **Fig. 4. The cluster of upregulated genes with NCBI nr annotation in stomach.** Figure
36
37 770 represents the cluster e in Fig. 3. Heat map columns represent samples and rows
38
39 771 correspond to genes. Expression values (FPKM) are \log_2 -transformed and then median-
40
41 772 centered by gene. Relative levels of gene expression are represented by colours. Pale
42
43 773 colour is low expression and darker blue is high expression.

44
45
46 774 **Fig. 5. The cluster of upregulated genes with NCBI nr annotation in intestine.** Figure
47
48 775 represents the cluster b in Fig. 3. Heat map columns represent samples and rows
49
50 776 correspond to genes. Expression values (FPKM) are \log_2 -transformed and then median-
51
52 777 centered by gene. Relative levels of gene expression are represented by colours. Pale
53
54 778 colour is low expression and darker blue is high expression.

55
56
57 779 **Fig. 6. The cluster of upregulated genes with NCBI nr annotation in heart.** Figure
58
59 780 represents the cluster a in Fig. 3. Heat map columns represent samples and rows
60
61
62
63
64
65

781 correspond to genes. Expression values (FPKM) are \log_2 -transformed and then median-
782 centered by gene. Relative levels of gene expression are represented by colours. Pale
783 colour is low expression and darker blue is high expression.

784 **Fig. 7. The cluster of upregulated genes with NCBI nr annotation in liver.** Figure
785 represents the cluster c in Fig. 3. Heat map columns represent samples and rows
786 correspond to genes. Expression values (FPKM) are \log_2 -transformed and then median-
787 centered by gene. Relative levels of gene expression are represented by colours. Pale
788 colour is low expression and darker blue is high expression.

789 **Fig. 8. The cluster of upregulated genes with NCBI nr annotation in pancreas.** Figure
790 represents the cluster d in Fig. 3. Heat map columns represent samples and rows
791 correspond to genes. Expression values (FPKM) are \log_2 -transformed and then median-
792 centered by gene. Relative levels of gene expression are represented by colours. Pale
793 colour is low expression and darker blue is high expression.

794 **Fig. 9. The workflow used to identify the python's stomach secretome during**
795 **digestion. 1)** Initially pythons were fed with mice, or a peptide mixture, and later the gastric
796 juice samples were obtained and mice debris was removed. **2)** The proteins were
797 precipitated, denatured and digested with trypsin. **3)** The resulting tryptic peptides were
798 analysed by LC-MS/MS analyses and the data merged into a single file. **4)** The file was used
799 to interrogate the in-house generated python protein sequence database (based on the
800 transcriptomic data) and python proteins were identified. **5)** The data was filtered to remove
801 mice proteins and plasma proteins. Subsequently, the annotation of the remaining proteins
802 was reassessed and the secretome identified.

803 **Fig. 10. Cartoon depiction of coloured KEGG pathway of gastric acid secretion in**
804 **stomach.** Entry in red represents upregulated during digestion; Entry in purple for highly
805 expressed. H/K is H⁺/K⁺-exchanging ATPase alpha polypeptide. CA is carbonic anhydrase.
806 AE is solute carrier family 26 (anion exchange transporter).

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

808 **Table 2. Colour coding of genes in KEGG pathway maps.** Three criteria are used to
1
2 809 classify and colour genes. First, i) whether the maximum FPKM of the gene among fasting,
3
4 810 24h and 48h is over 10, then ii) whether the gene is differential expressed in at least one of
5
6 811 the pairwise comparison among fasting, 24h and 48h with FC over 4. Finally, iii) for those
7
8 812 genes expressed, but not differential expressed, whether it is highly expressed with
9
10 813 maximum FPKM among three time points over 200. The term expression trend indicates the
11
12 814 trend of gene expression across fasting, 24h and 48h. e.g. The trend up means the gene is
13
14 815 upregulated from either fasting to 24h, fasting to 48h or 24h to 48h. The trend up-then-down
15
16 816 means the gene is firstly upregulated from fasting to 24h, then downregulated from 24h to
17
18 817 48h.
19
20
21

22 818 **Table 3. The number of DEGs across fasting, 24h and 48h in each tissue.** The
23
24 819 expression trend is consistent with definition in Table 2.
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 820 **Reference**

- 2
3
4
5 821 1. Secor SM, Diamond J: **A vertebrate model of extreme physiological**
6 822 **regulation.** *Nature* 1998, **395**:659-662.
7 823 2. Cox CL, Secor SM: **Matched regulation of gastrointestinal performance in the**
8 824 **Burmese python, *Python molurus*.** *J Exp Biol* 2008, **211**:1131-1140.
9 825 3. Secor SM, Diamond J: **Adaptive responses to feeding in Burmese pythons:**
10 826 **pay before pumping.** *J Exp Biol* 1995, **198**:1313-1325.
11 827 4. Secor SM, Diamond JM: **Evolution of regulatory responses to feeding in**
12 828 **snakes.** *Physiological and Biochemical Zoology* 2000, **73**:123-141.
13 829 5. Aubret F, Shine R, Bonnet X: **Evolutionary biology: adaptive developmental**
14 830 **plasticity in snakes.** *Nature* 2004, **431**:261-262.
15 831 6. Cohn MJ, Tickle C: **Developmental basis of limblessness and axial patterning**
16 832 **in snakes.** *Nature* 1999, **399**:474-479.
17 833 7. Ott BD, Secor SM: **Adaptive regulation of digestive performance in the genus**
18 834 ***Python*.** *J Exp Biol* 2007, **210**:340-356.
19 835 8. Secor SM: **Digestive physiology of the Burmese python: broad regulation of**
20 836 **integrated performance.** *J Exp Biol* 2008, **211**:3767-3774.
21 837 9. Zaar M, Overgaard J, Gesser H, Wang T: **Contractile properties of the**
22 838 **functionally divided python heart: two sides of the same matter.** *Comp*
23 839 *Biochem Physiol A Mol Integr Physiol* 2007, **146**:163-173.
24 840 10. Andersen JB, Rourke BC, Caiozzo VJ, Bennett AF, Hicks JW: **Physiology:**
25 841 **postprandial cardiac hypertrophy in pythons.** *Nature* 2005, **434**:37-38.
26 842 11. Vidal N, Hedges SB: **Molecular evidence for a terrestrial origin of snakes.**
27 843 *Proc Biol Sci* 2004, **271 Suppl 4**:S226-229.
28 844 12. Castoe TA, de Koning AP, Hall KT, Card DC, Schield DR, Fujita MK, Ruggiero RP,
29 845 Degner JF, Daza JM, Gu W, et al: **The Burmese python genome reveals the**
30 846 **molecular basis for extreme adaptation in snakes.** *Proc Natl Acad Sci U S A*
31 847 2013, **110**:20645-20650.
32 848 13. Andrew AL, Card DC, Ruggiero RP, Schield DR, Adams RH, Pollock DD, Secor SM,
33 849 Castoe TA: **Rapid changes in gene expression direct rapid shifts in intestinal**
34 850 **form and function in the Burmese python after feeding.** *Physiol Genomics*
35 851 2015, **47**:147-157.
36 852 14. Castoe TA, Fox SE, Jason de Koning A, Poole AW, Daza JM, Smith EN, Mockler TC,
37 853 Secor SM, Pollock DD: **A multi-organ transcriptome resource for the**
38 854 **Burmese Python (*Python molurus bivittatus*).** *BMC Res Notes* 2011, **4**:310.
39 855 15. Wang Z, Gerstein M, Snyder M: **RNA-Seq: a revolutionary tool for**
40 856 **transcriptomics.** *Nat Rev Genet* 2009, **10**:57-63.
41 857 16. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B: **Mapping and**
42 858 **quantifying mammalian transcriptomes by RNA-Seq.** *Nat Methods* 2008,
43 859 **5**:621-628.
44 860 17. Wang T, Hung CC, Randall DJ: **The comparative physiology of food**
45 861 **deprivation: from feast to famine.** *Annu Rev Physiol* 2006, **68**:223-251.
46 862 18. Zerbino DR, Birney E: **Velvet: algorithms for de novo short read assembly**
47 863 **using de Bruijn graphs.** *Genome Res* 2008, **18**:821-829.
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

- 864 19. Bushmanova E, Antipov D, Lapidus A, Suvorov V, Prjibelski AD: **rnaQUAST: a**
1 865 **quality assessment tool for de novo transcriptome assemblies.**
2 866 *Bioinformatics* 2016, **32**:2210-2212.
- 3 867 20. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM: **BUSCO:**
4 868 **assessing genome assembly and annotation completeness with single-copy**
5 869 **orthologs.** *Bioinformatics* 2015:btv351.
- 6 870 21. Pruitt KD, Tatusova T, Maglott DR: **NCBI reference sequences (RefSeq): a**
7 871 **curated non-redundant sequence database of genomes, transcripts and**
8 872 **proteins.** *Nucleic Acids Res* 2007, **35**:D61-65.
- 9 873 22. Pyron RA: **Novel Approaches for Phylogenetic Inference from Morphological**
10 874 **Data and Total-Evidence Dating in Squamate Reptiles (Lizards, Snakes, and**
11 875 **Amphisbaenians).** *Syst Biol* 2017, **66**:38-56.
- 12 876 23. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M: **Blast2GO: a**
13 877 **universal tool for annotation, visualization and analysis in functional**
14 878 **genomics research.** *Bioinformatics* 2005, **21**:3674-3676.
- 15 879 24. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U,
16 880 Daugherty L, Duquenne L, et al: **InterPro: the integrative protein signature**
17 881 **database.** *Nucleic Acids Res* 2009, **37**:D211-215.
- 18 882 25. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg
19 883 SL, Wold BJ, Pachter L: **Transcript assembly and quantification by RNA-Seq**
20 884 **reveals unannotated transcripts and isoform switching during cell**
21 885 **differentiation.** *Nat Biotechnol* 2010, **28**:511-515.
- 22 886 26. Fan HP, Di Liao C, Fu BY, Lam LC, Tang NL: **Interindividual and interethnic**
23 887 **variation in genomewide gene expression: insights into the biological**
24 888 **variation of gene expression and clinical implications.** *Clin Chem* 2009,
25 889 **55**:774-785.
- 26 890 27. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M: **KEGG for integration and**
27 891 **interpretation of large-scale molecular data sets.** *Nucleic Acids Res* 2012,
28 892 **40**:D109-114.
- 29 893 28. Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes.** *Nucleic*
30 894 *Acids Res* 2000, **28**:27-30.
- 31 895 29. Kageyama T: **Pepsinogens, progastricsins, and prochymosins: structure,**
32 896 **function, evolution, and development.** *Cell Mol Life Sci* 2002, **59**:288-306.
- 33 897 30. Loytynoja A, Goldman N: **webPRANK: a phylogeny-aware multiple sequence**
34 898 **aligner with interactive alignment browser.** *BMC Bioinformatics* 2010,
35 899 **11**:579.
- 36 900 31. Kumar S, Stecher G, Tamura K: **MEGA7: Molecular Evolutionary Genetics**
37 901 **Analysis Version 7.0 for Bigger Datasets.** *Mol Biol Evol* 2016, **33**:1870-1874.
- 38 902 32. Hayashi K, Agata K, Mochii M, Yasugi S, Eguchi G, Mizuno T: **Molecular cloning**
39 903 **and the nucleotide sequence of cDNA for embryonic chicken pepsinogen:**
40 904 **phylogenetic relationship with prochymosin.** *J Biochem* 1988, **103**:290-296.
- 41 905 33. Watanuki K, Yasugi S: **Analysis of transcription regulatory regions of**
42 906 **embryonic chicken pepsinogen (ECPg) gene.** *Dev Dyn* 2003, **228**:51-58.
- 43 907 34. Aoki N, Deshimaru M, Terada S: **Active fragments of the antihemorrhagic**
44 908 **protein HSF from serum of habu (Trimeresurus flavoviridis).** *Toxicon* 2007,
45 909 **49**:653-662.
- 46 910 35. Perales J, Neves-Ferreira AG, Valente RH, Domont GB: **Natural inhibitors of**
47 911 **snake venom hemorrhagic metalloproteinases.** *Toxicon* 2005, **45**:1013-1020.

- 912 36. Tomihara Y, Yonaha K, Nozaki M, Yamakawa M, Kawamura Y, Kamura T, Toyama
1 913 **S: Purification of an antihemorrhagic factor from the serum of the non-**
2 914 **venomous snake *Dinodon semicarinatus*. *Toxicon* 1988, 26:420-423.**
- 3 915 37. Lee HJ, Sorimachi H, Jeong SY, Ishiura S, Suzuki K: **Molecular cloning and**
4 916 **characterization of a novel tissue-specific calpain predominantly**
5 917 **expressed in the digestive tract. *Biol Chem* 1998, 379:175-183.**
- 6 918 38. Secor SM: **Gastric function and its contribution to the postprandial**
7 919 **metabolic response of the Burmese python *Python molurus*. *Journal of***
8 920 ***Experimental Biology* 2003, 206:1621-1630.**
- 9 921 39. Bessler SM, Secor SM: **Effects of feeding on luminal pH and morphology of**
10 922 **the gastroesophageal junction of snakes. *Zoology (Jena)* 2012, 115:319-329.**
- 11 923 40. Helmstetter C, Pope RK, T'Flachebba M, Secor SM, Lignot J-H: **The effects of**
12 924 **feeding on cell morphology and proliferation of the gastrointestinal tract of**
13 925 **juvenile Burmese pythons (*Python molurus*). *Canadian Journal of Zoology***
14 926 **2009, 87:1255-1267.**
- 15 927 41. Enok S, Simonsen LS, Wang T: **The contribution of gastric digestion and**
16 928 **ingestion of amino acids on the postprandial rise in oxygen consumption,**
17 929 **heart rate and growth of visceral organs in pythons. *Comp Biochem Physiol A***
18 930 ***Mol Integr Physiol* 2013, 165:46-53.**
- 19 931 42. Nørgaard S, Andreassen K, Malte CL, Enok S, Wang T: **Low cost of gastric acid**
20 932 **secretion during digestion in ball pythons. *Comparative Biochemistry and***
21 933 ***Physiology Part A: Molecular & Integrative Physiology* 2016, 194:62-66.**
- 22 934 43. Secor SM, Taylor JR, Grosell M: **Selected regulation of gastrointestinal acid-**
23 935 **base secretion and tissue metabolism for the diamondback water snake**
24 936 **and Burmese python. *J Exp Biol* 2012, 215:185-196.**
- 25 937 44. Andrade DV, De Toledo LF, Abe AS, Wang T: **Ventilatory compensation of the**
26 938 **alkaline tide during digestion in the snake *Boa constrictor*. *J Exp Biol* 2004,**
27 939 **207:1379-1385.**
- 28 940 45. Koelz HR: **Gastric acid in vertebrates. *Scand J Gastroenterol Suppl* 1992, 193:2-**
29 941 **6.**
- 30 942 46. Starck JM, Beese K: **Structural flexibility of the intestine of Burmese python**
31 943 **in response to feeding. *Journal of Experimental Biology* 2001, 204:325-335.**
- 32 944 47. Holmberg A, Kaim J, Persson A, Jensen J, Wang T, Holmgren S: **Effects of**
33 945 **digestive status on the reptilian gut. *Comparative Biochemistry and Physiology***
34 946 ***a-Molecular and Integrative Physiology* 2002, 133:499-518.**
- 35 947 48. Lignot JH, Helmstetter C, Secor SM: **Postprandial morphological response of**
36 948 **the intestinal epithelium of the Burmese python (*Python molurus*). *Comp***
37 949 ***Biochem Physiol A Mol Integr Physiol* 2005, 141:280-291.**
- 38 950 49. Secor SM, Whang EE, Lane JS, Ashley SW, Diamond J: **Luminal and systemic**
39 951 **signals trigger intestinal adaptation in the juvenile python. *American Journal***
40 952 ***of Physiology-Gastrointestinal and Liver Physiology* 2000, 279:G1177-G1187.**
- 41 953 50. Overgaard J, Andersen JB, Wang T: **The effects of fasting duration on the**
42 954 **metabolic response to feeding in *Python molurus*: an evaluation of the**
43 955 **energetic costs associated with gastrointestinal growth and upregulation.**
44 956 ***Physiol Biochem Zool* 2002, 75:360-368.**
- 45 957 51. Secor SM, White SE: **Prioritizing blood flow: cardiovascular performance in**
46 958 **response to the competing demands of locomotion and digestion for the**
47 959 **Burmese python, *Python molurus*. *J Exp Biol* 2010, 213:78-88.**

- 960 52. Secor SM, Hicks JW, Bennett AF: **Ventilatory and cardiovascular responses of**
1 961 **a python (*Python molurus*) to exercise and digestion.** *J Exp Biol* 2000,
2 962 **203:2447-2454.**
- 3 963 53. Skovgaard N, Møller K, Gesser H, Wang T: **Histamine induces postprandial**
4 964 **tachycardia through a direct effect on cardiac H2-receptors in pythons.**
5 965 *American Journal of Physiology - Regulatory, Integrative and Comparative*
6 966 *Physiology* 2009, **296:R774-R785.**
- 7 967 54. Enok S, Simonsen LS, Pedersen SV, Wang T, Skovgaard N: **Humoral regulation**
8 968 **of heart rate during digestion in pythons (*Python molurus* and *Python***
9 969 **regius). *Am J Physiol Regul Integr Comp Physiol* 2012, **302:R1176-1183.****
- 10 970 55. Riquelme CA, Magida JA, Harrison BC, Wall CE, Marr TG, Secor SM, Leinwand LA:
11 971 **Fatty acids identified in the Burmese python promote beneficial cardiac**
12 972 **growth.** *Science* 2011, **334:528-531.**
- 13 973 56. Enok S, Leite G, Leite C, Gesser H, Hedrick MS, Wang T: **Improved cardiac filling**
14 974 **facilitates the postprandial elevation of stroke volume in *Python regius*.** *J*
15 975 *Exp Biol* 2016.
- 16 976 57. Slay CE, Enok S, Hicks JW, Wang T: **Reduction of blood oxygen levels enhances**
17 977 **postprandial cardiac hypertrophy in Burmese python (*Python bivittatus*).** *J*
18 978 *Exp Biol* 2014, **217:1784-1789.**
- 19 979 58. Jensen B, Larsen CK, Nielsen JM, Simonsen LS, Wang T: **Change of cardiac**
20 980 **function, but not form, in postprandial pythons.** *Comp Biochem Physiol A Mol*
21 981 *Integr Physiol* 2011, **160:35-42.**
- 22 982 59. Butler MW, Lutz TJ, Fokidis HB, Stahlschmidt ZR: **Eating increases oxidative**
23 983 **damage in a reptile.** *J Exp Biol* 2016, **219:1969-1973.**
- 24 984 60. Schulz MH, Zerbino DR, Vingron M, Birney E: **Oases: robust de novo RNA-seq**
25 985 **assembly across the dynamic range of expression levels.** *Bioinformatics*
26 986 2012, **28:1086-1092.**
- 27 987 61. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL:
28 988 **BLAST+: architecture and applications.** *BMC Bioinformatics* 2009, **10:421.**
- 29 989 62. Eddy SR: **A probabilistic model of local sequence alignment that simplifies**
30 990 **statistical significance estimation.** *PLoS Comput Biol* 2008, **4:e1000069.**
- 31 991 63. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment**
32 992 **search tool.** *J Mol Biol* 1990, **215:403-410.**
- 33 993 64. Langmead B, Salzberg SL: **Fast gapped-read alignment with Bowtie 2.** *Nat*
34 994 *Methods* 2012, **9:357-359.**
- 35 995 65. Li B, Dewey CN: **RSEM: accurate transcript quantification from RNA-Seq data**
36 996 **with or without a reference genome.** *BMC Bioinformatics* 2011, **12:323.**
- 37 997 66. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, Kong L, Gao G, Li CY, Wei L: **KOBAS**
38 998 **2.0: a web server for annotation and identification of enriched pathways**
39 999 **and diseases.** *Nucleic Acids Res* 2011, **39:W316-322.**
- 40 1000 67. Perkins DN, Pappin DJC, Creasy DM, Cottrell JS: **Probability-based protein**
41 1001 **identification by searching sequence databases using mass spectrometry**
42 1002 **data.** *Electrophoresis* 1999, **20:3551-3567.**
- 43 1003 68. Dyrland TF, Poulsen ET, Scavenius C, Sanggaard KW, Enghild JJ: **MS Data Miner:**
44 1004 **A web-based software tool to analyze, compare, and share mass**
45 1005 **spectrometry protein identifications.** *Proteomics* 2012, **12:2792-2796.**
- 46 1006 69. Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, Mann M:
47 1007 **Exponentially modified protein abundance index (emPAI) for estimation of**
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1008
11009
21010
31011
41012
51013
61014
71015
81016
91017
101018
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics* 2005, **4**:1265-1272.

70. Vizcaino JA, Csordas A, del-Toro N, Dianes JA, Griss J, Lavidas I, Mayer G, Perez-Riverol Y, Reisinger F, Ternent T, et al: **2016 update of the PRIDE database and its related tools.** *Nucleic Acids Res* 2016, **44**:D447-456.

71. Duan, J; Sanggaard, K, W; Schausser, L; Lauridsen, S, E; Enghild, J, J; Schierup, M, H; Wang, T (2017): **Supporting data for "Transcriptome Analysis of the Response of Burmese Python to Digestion"** GigaScience Database. <http://dx.doi.org/10.5524/100320>

| Parameter | <i>De novo</i> 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 transcript 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 |

| 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 |

| 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%) |

Figure 1

[Click here to download Figure Figure1.pdf](#)

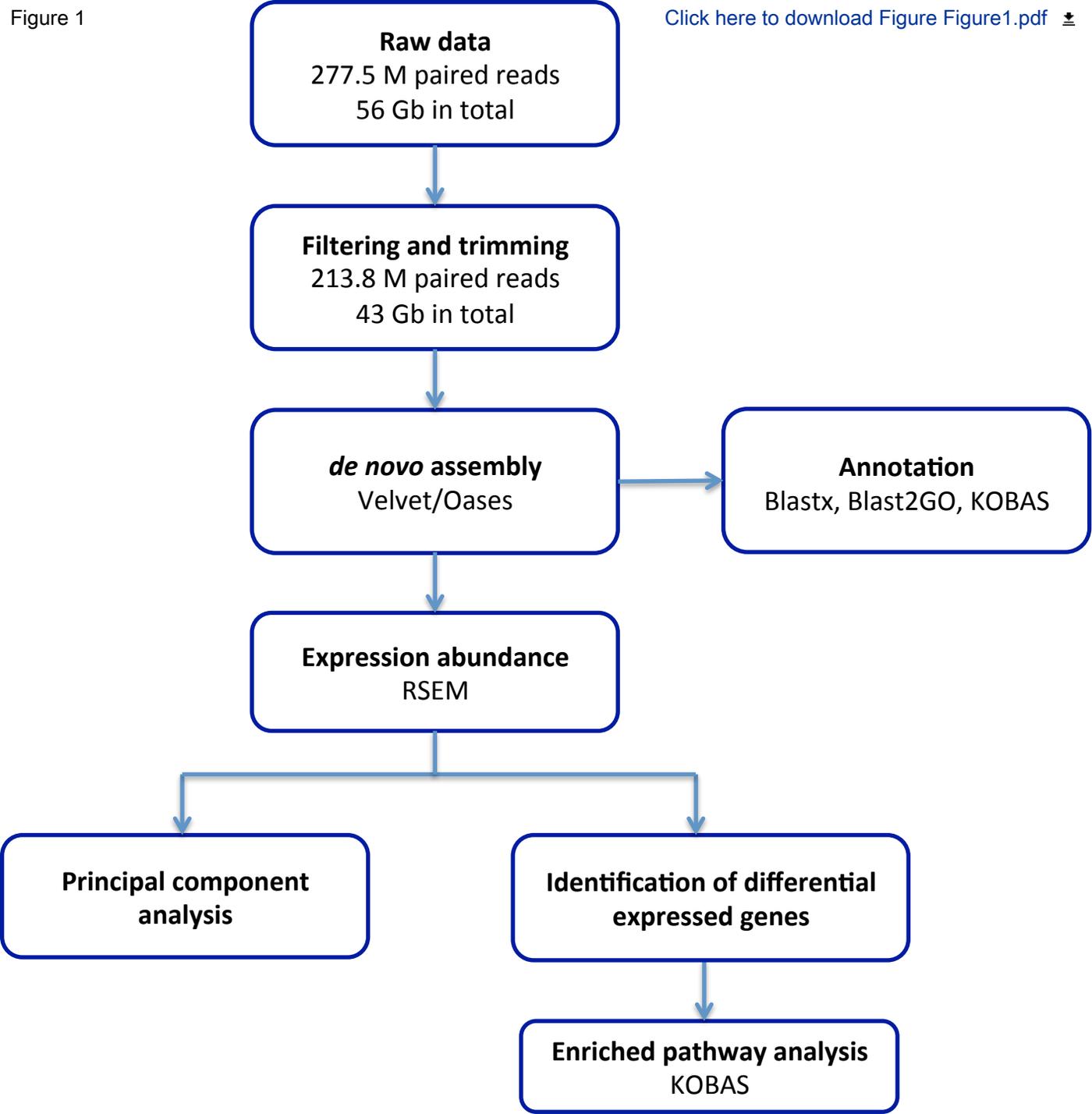


Figure 2

[Click here to download Figure2.pdf](#)

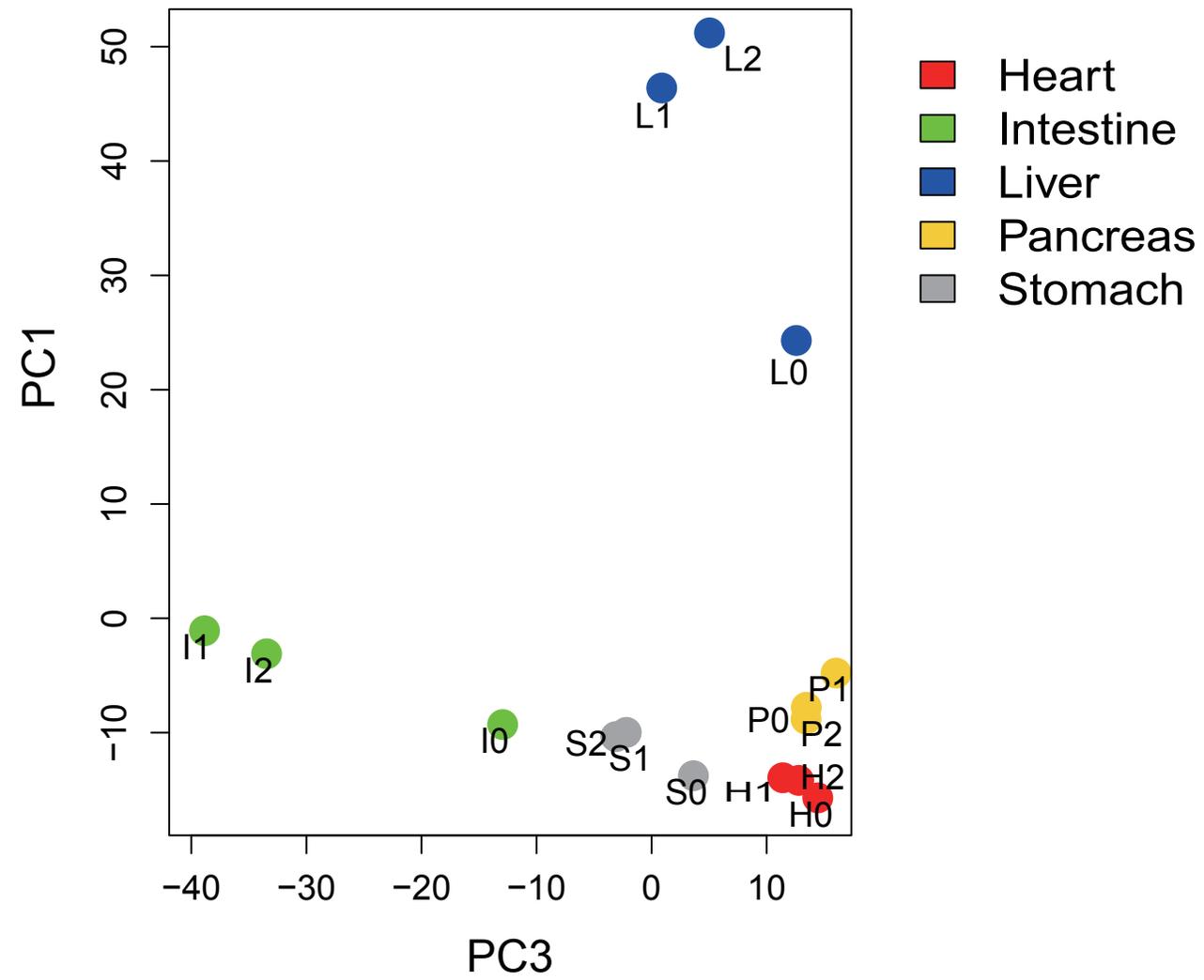
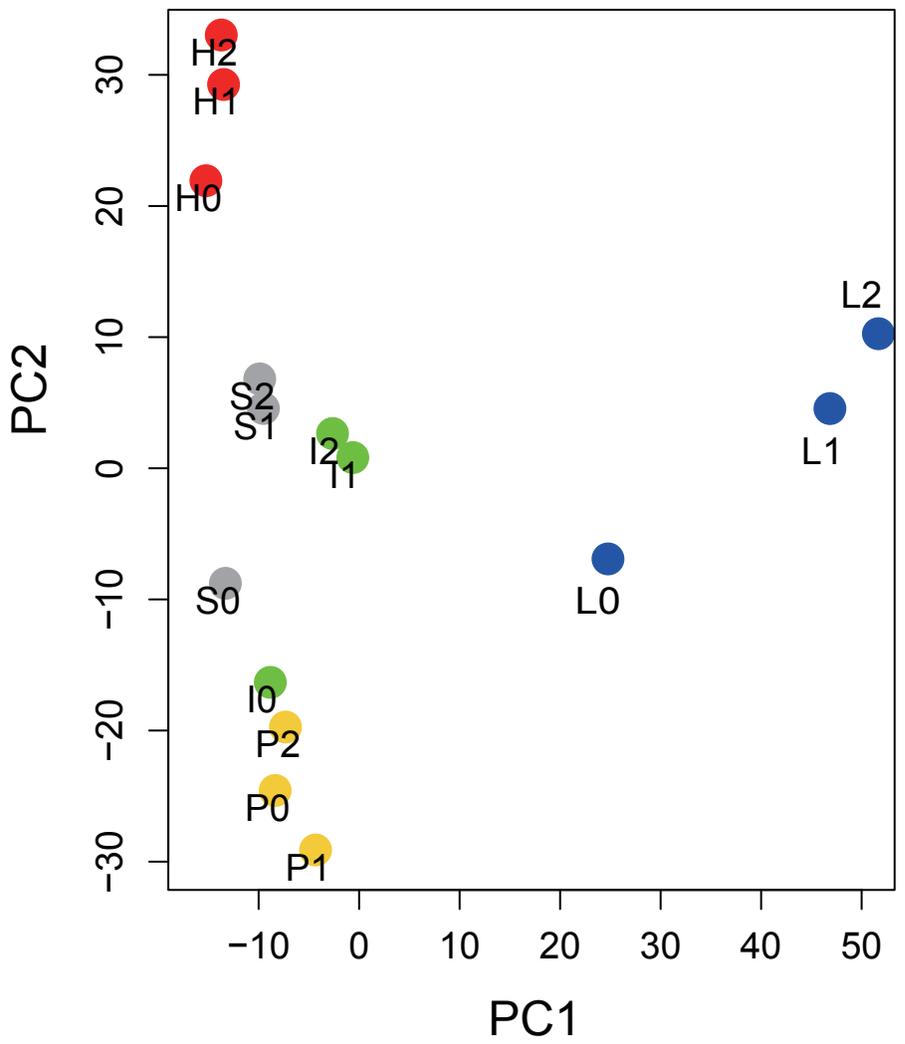


Figure 3

[Click here to download Figure Figure3.pdf](#)

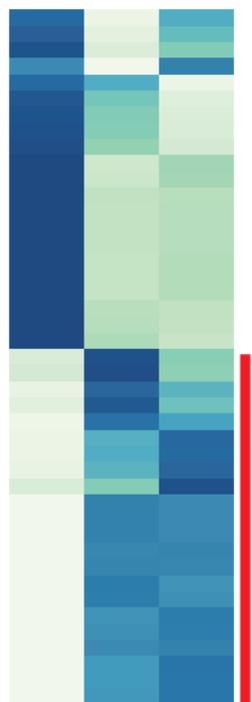
Heart

Intestine

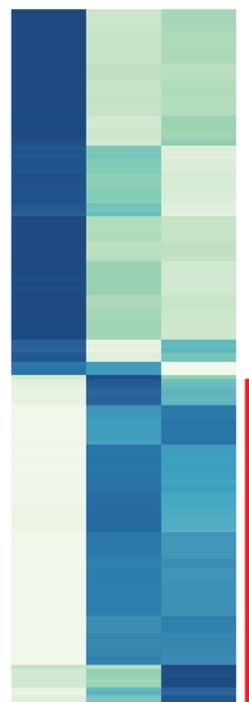
Liver

Pancreas

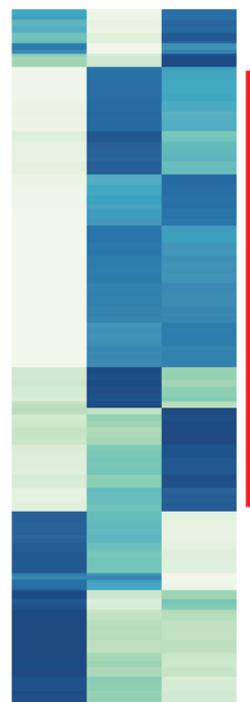
Stomach



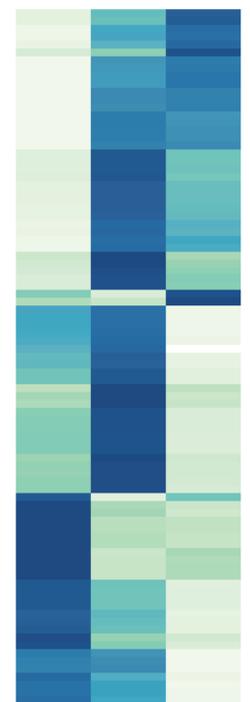
a



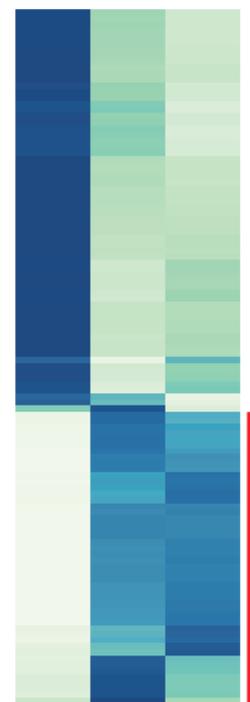
b



c



d



e

Color Key



-1 0 1

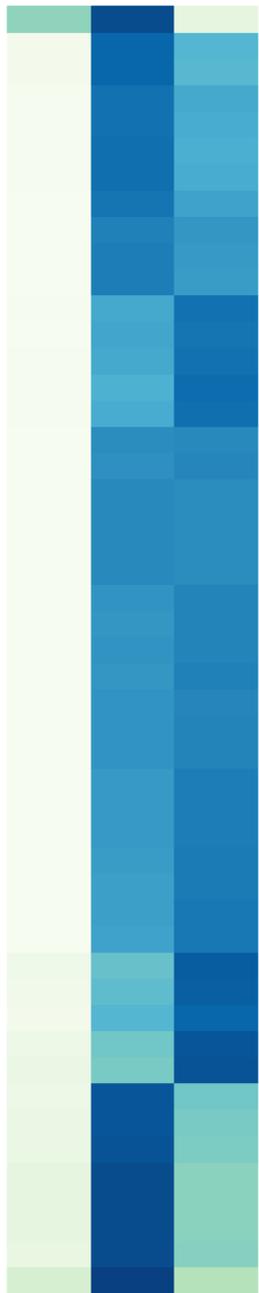
0 1 2

days after feeding

Figure 4

[Click here to download Figure4.pdf](#)

Stomach



- 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

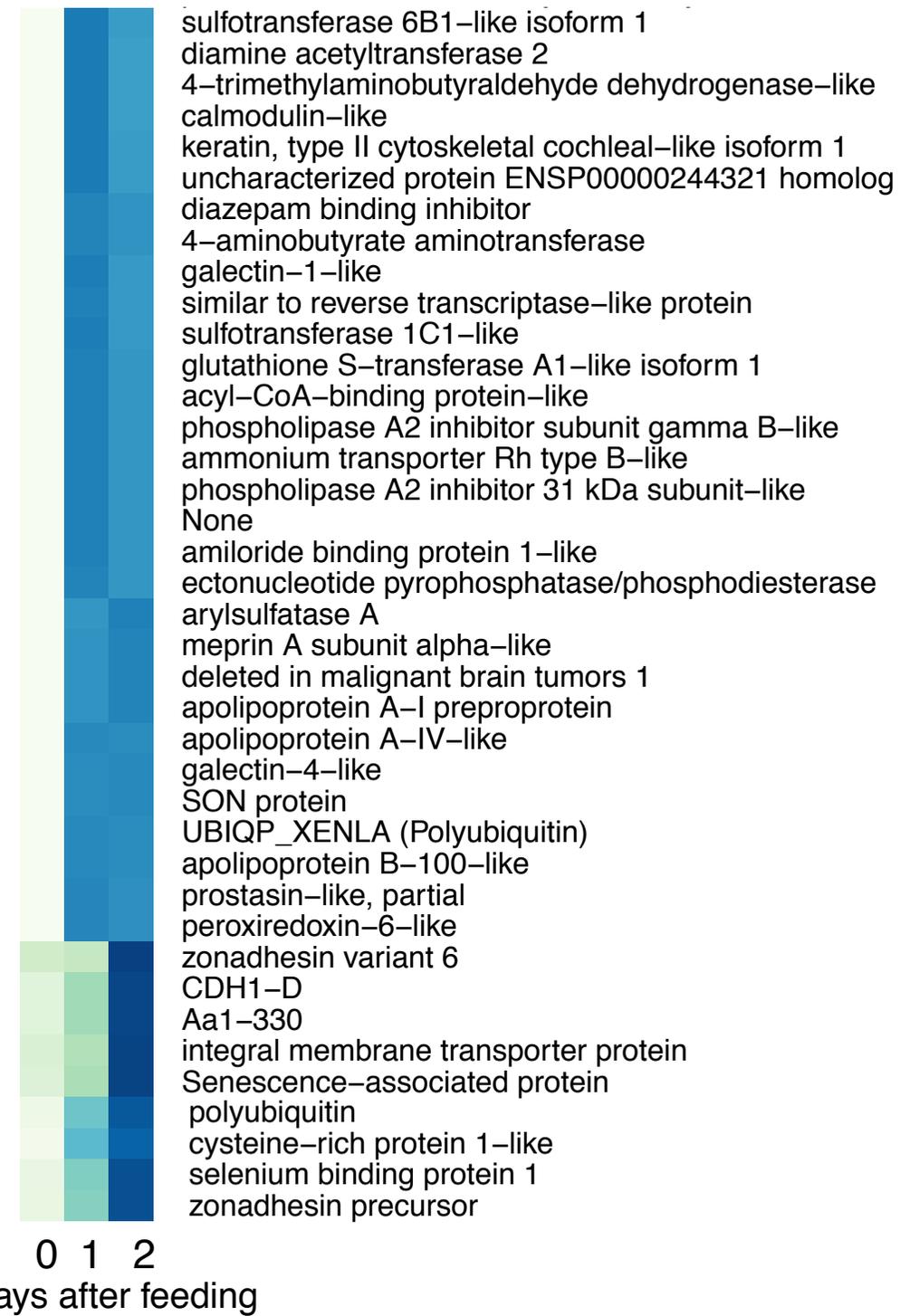
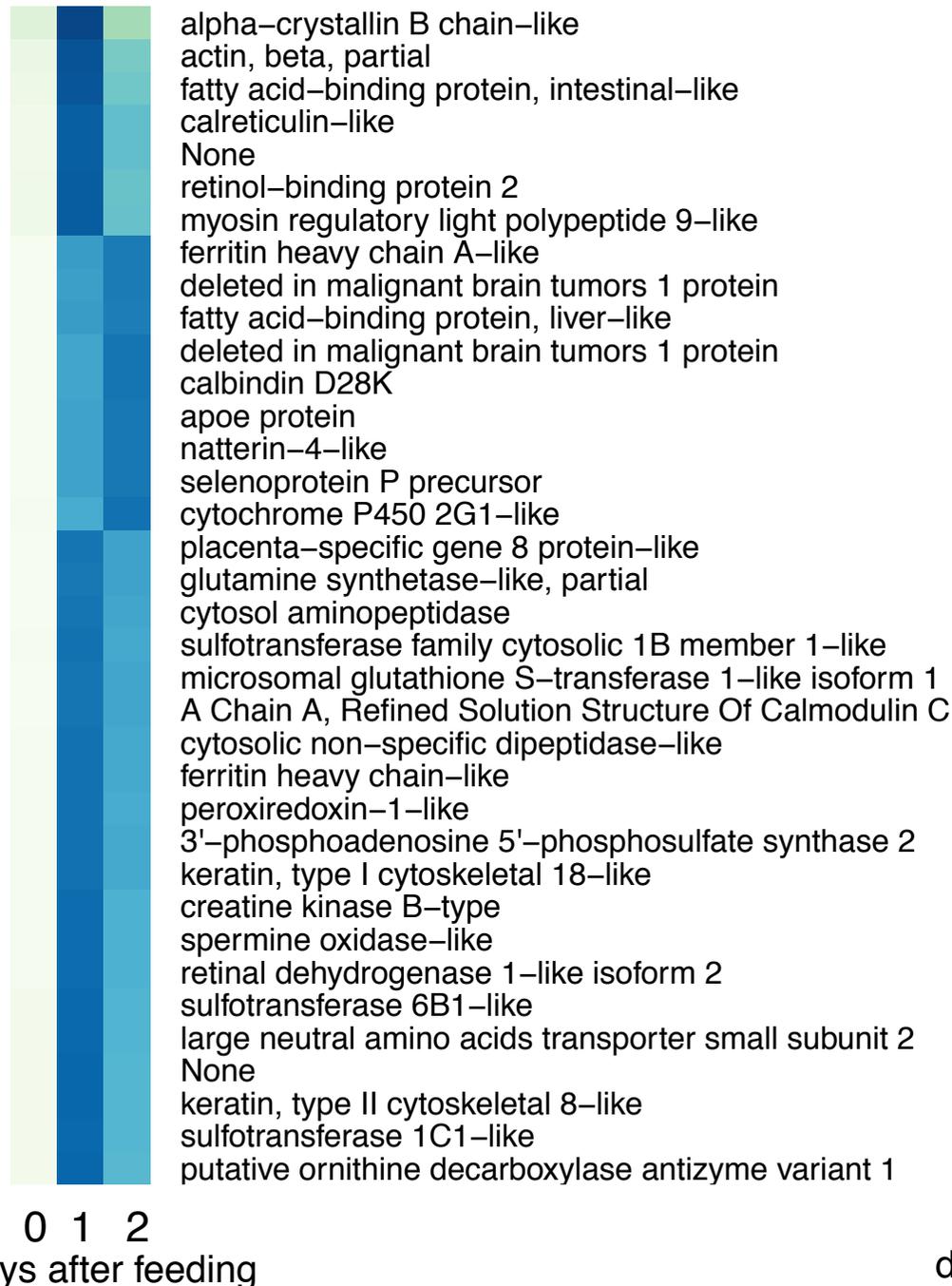
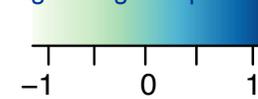
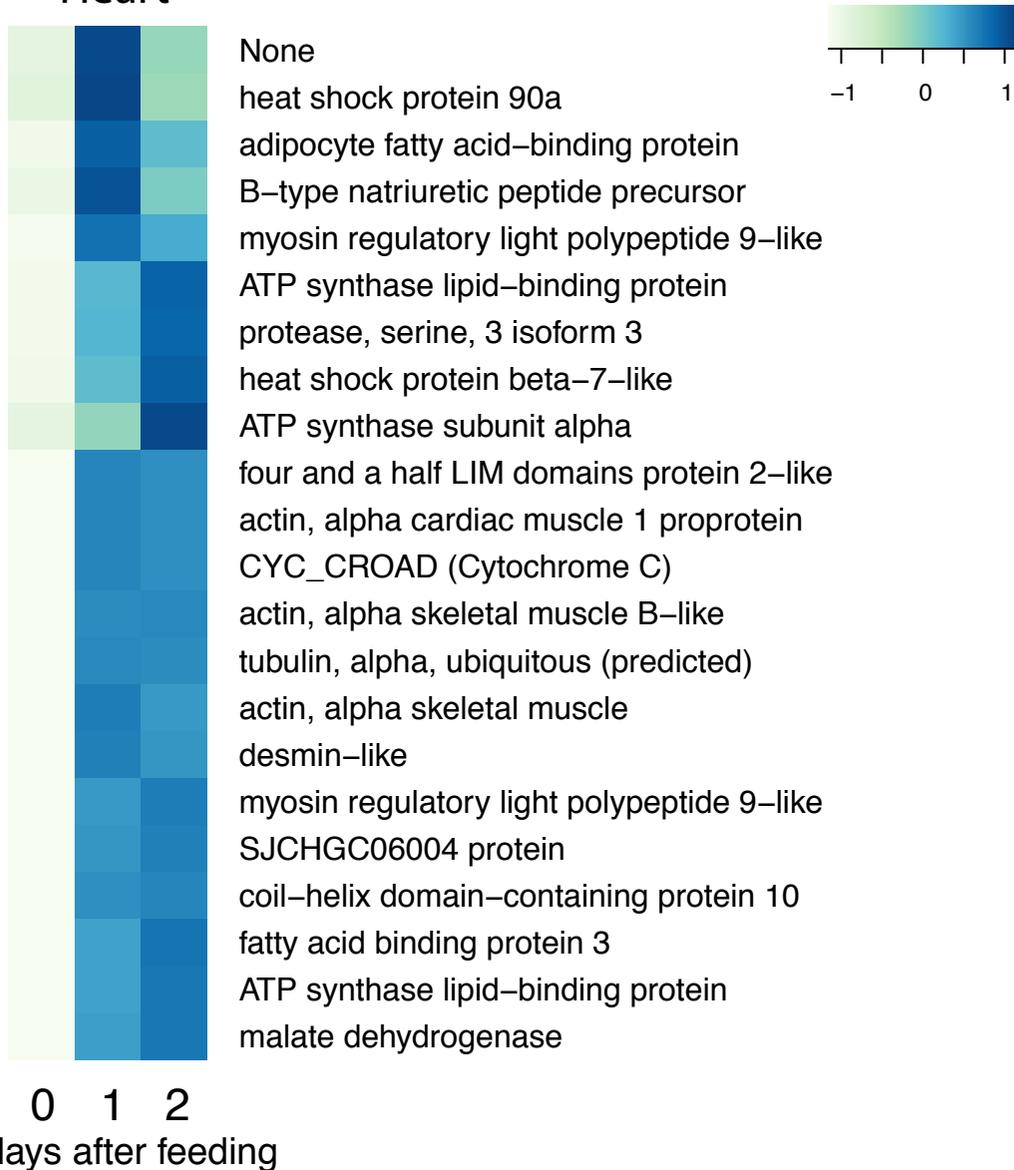
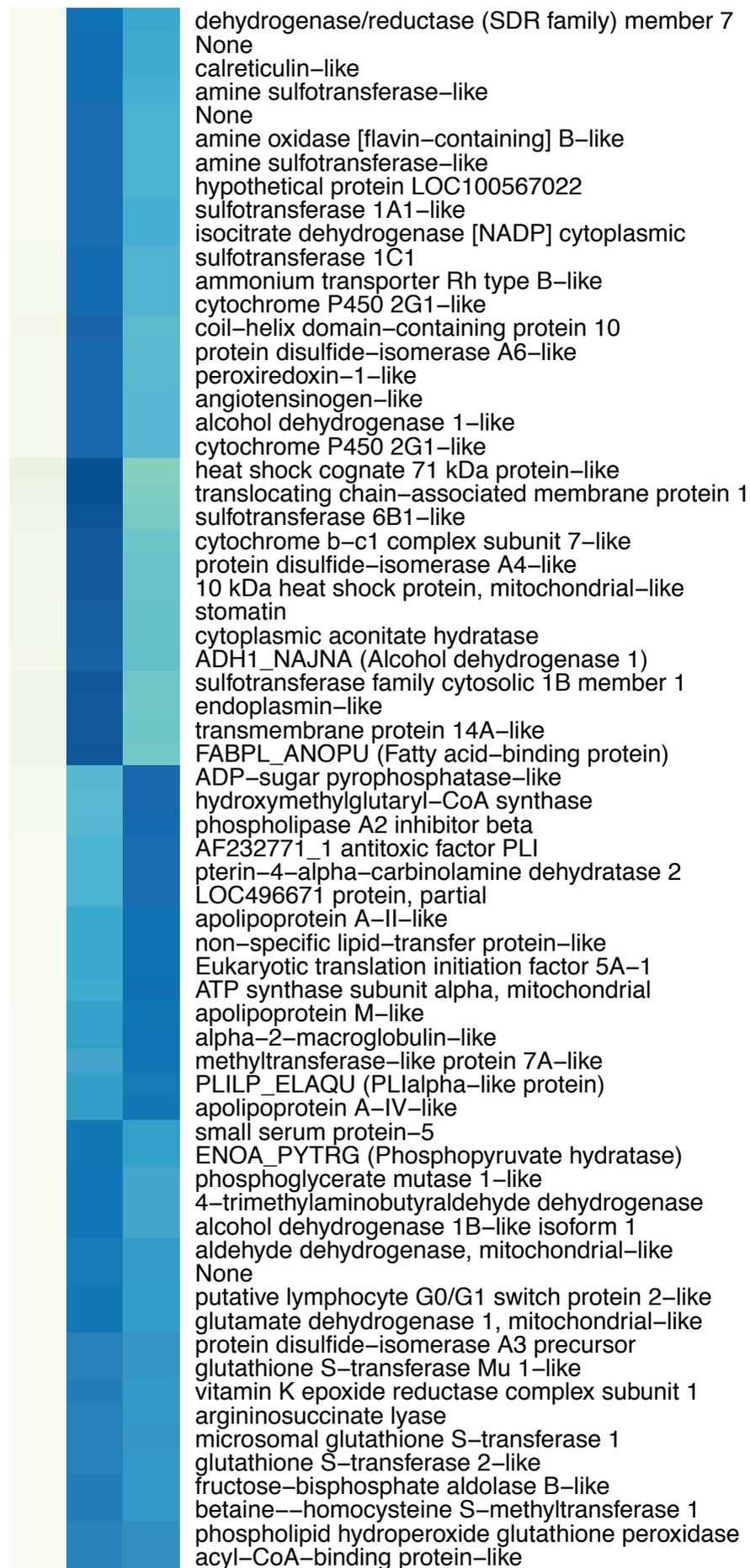
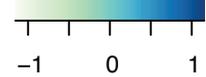
[Click here to download Figure5.pdf](#)

Figure 6

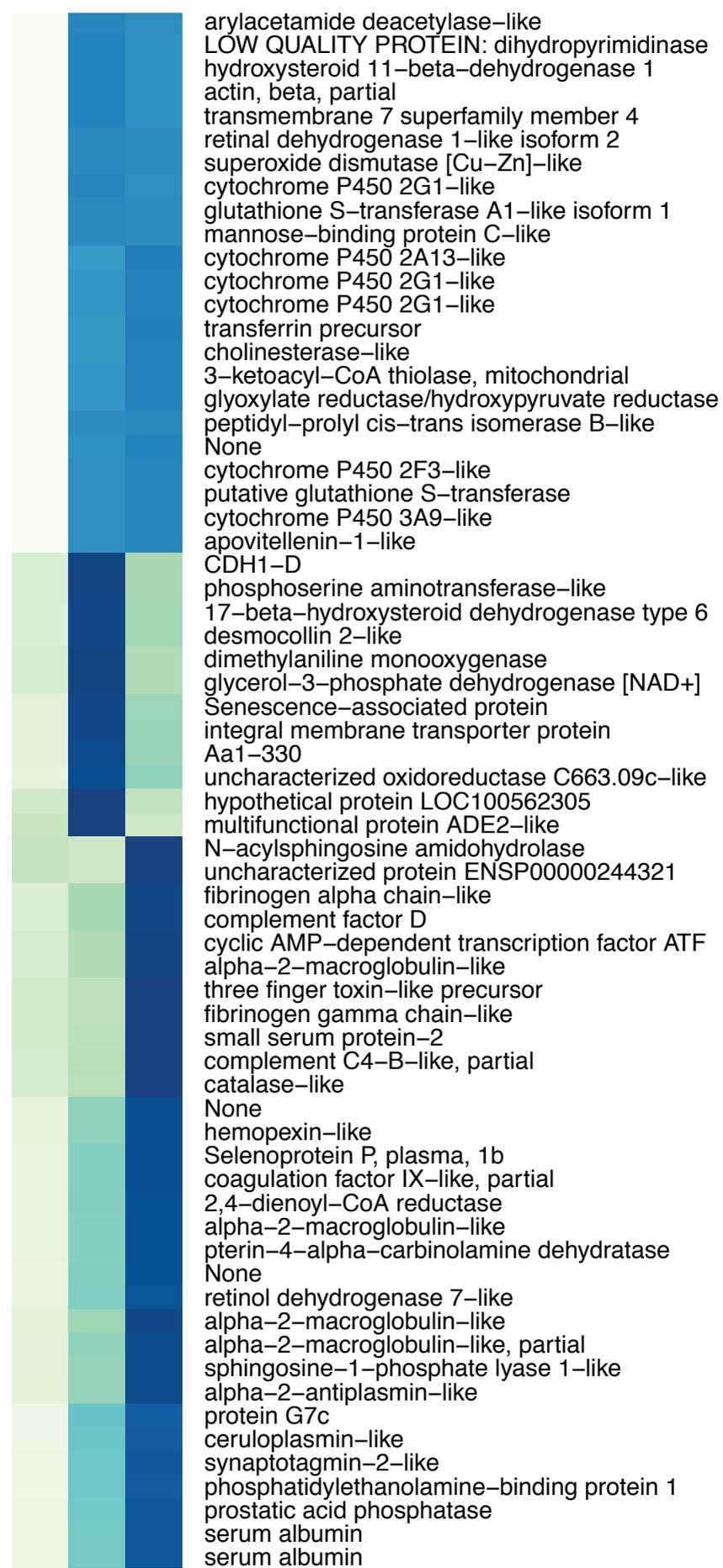
Heart





0 1 2

days after feeding

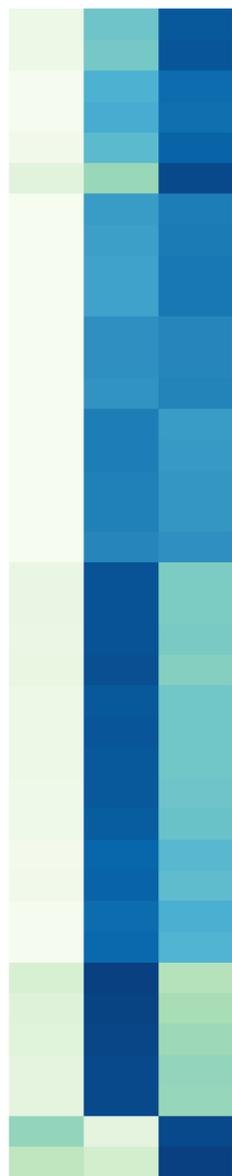


0 1 2

days after feeding

Pancreas

-1 0 1



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



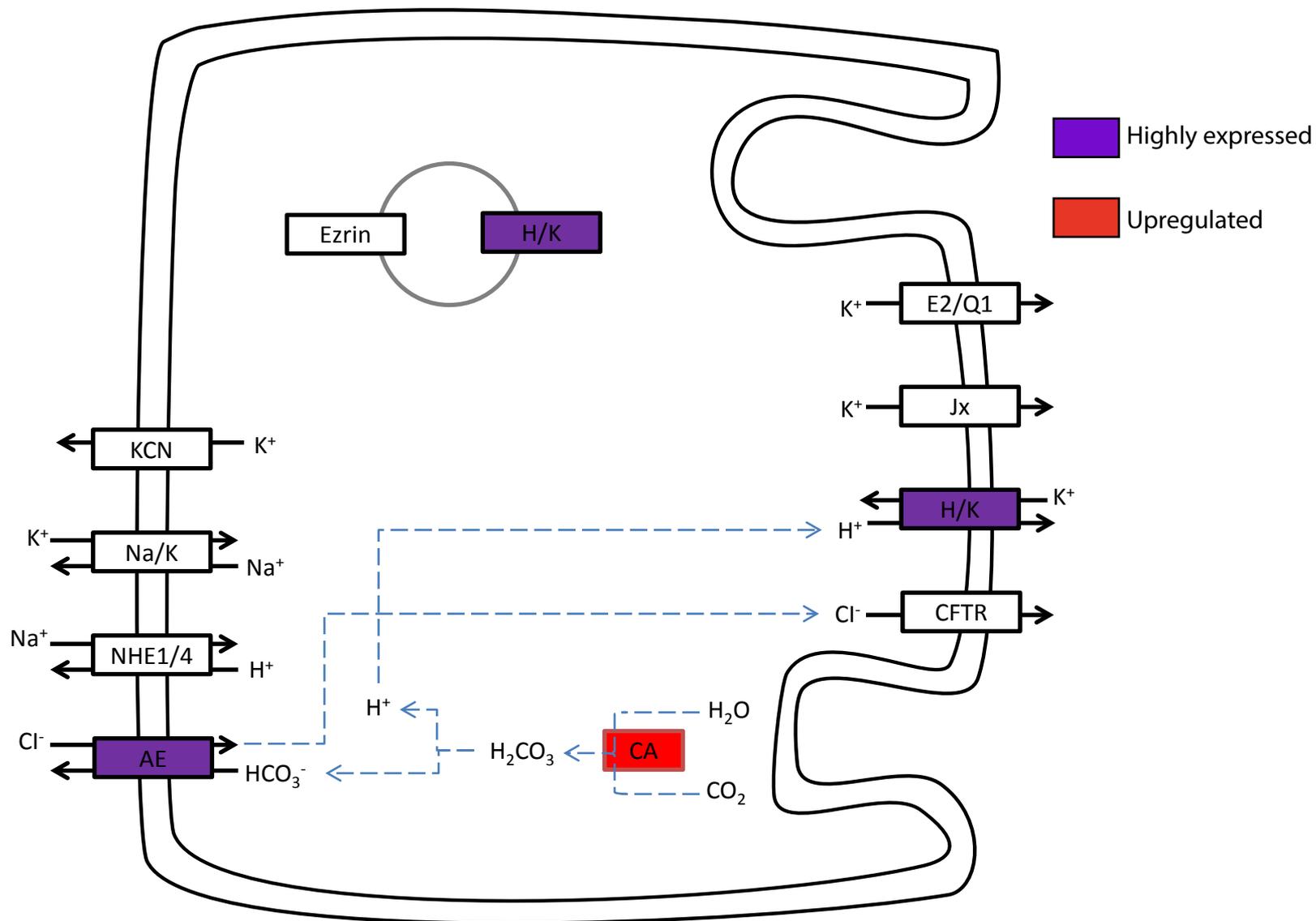
3. LC-MS/MS analyses



4. Protein identification



5. Identification of the python stomach secretome







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





Click here to access/download
Supplementary Material
FigureS4.pdf



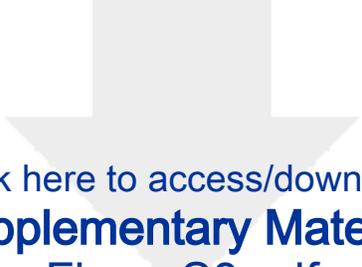
Click here to access/download
Supplementary Material
FigureS5.pdf



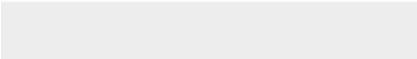
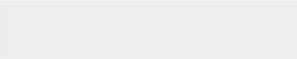
Click here to access/download
Supplementary Material
FigureS6.pdf



Click here to access/download
Supplementary Material
FigureS7.pdf



Click here to access/download
Supplementary Material
FigureS8.pdf

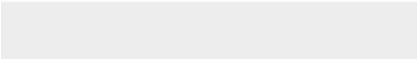
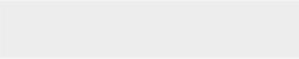


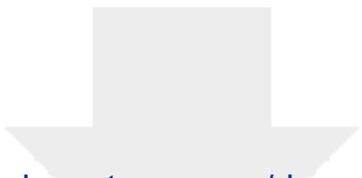


Click here to access/download
Supplementary Material
FigureS9.pdf

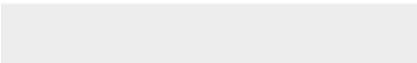


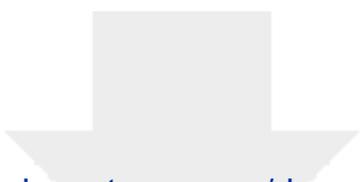
Click here to access/download
Supplementary Material
TableS3.xlsx



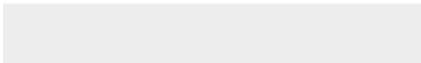


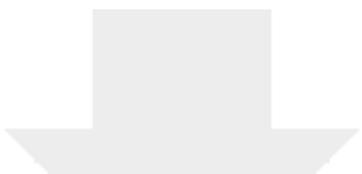
Click here to access/download
Supplementary Material
TableS5.xlsx





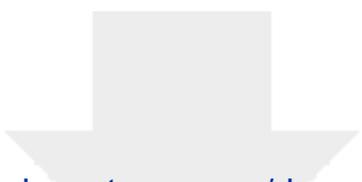
Click here to access/download
Supplementary Material
TableS6.xlsx





Click here to access/download
Supplementary Material
TableS7.xlsx



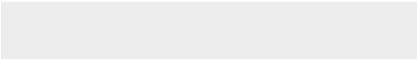
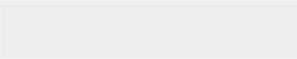


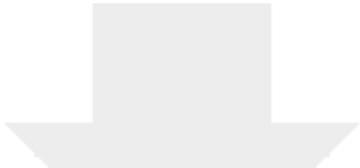
Click here to access/download
Supplementary Material
TableS8.xlsx





Click here to access/download
Supplementary Material
TableS10.xlsx



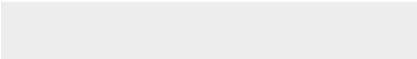
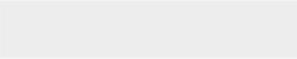


Click here to access/download
Supplementary Material
TableS12.xlsx



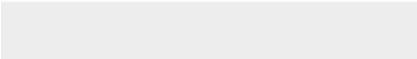
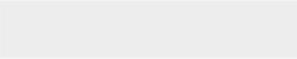


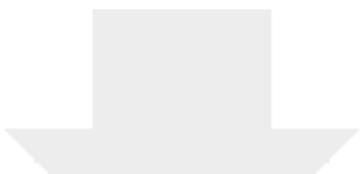
Click here to access/download
Supplementary Material
TableS13.xlsx



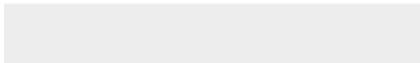
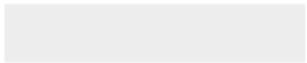


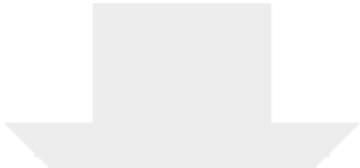
Click here to access/download
Supplementary Material
TableS14.xlsx



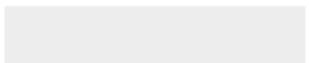


Click here to access/download
Supplementary Material
TableS9.xlsx





Click here to access/download
Supplementary Material
TableS11.xlsx



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: | <p>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.</p> <p>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.</p> <p>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.</p> |
| Response to Reviewers: | <p>Dear Editor,</p> <p>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.</p> <p>Sincerely yours Jinjie Duan (on behalf of all the authors)</p> |
| Additional Information: | |
| Question | Response |
| Are you submitting this manuscript to a special series or article collection? | No |
| Experimental design and statistics | Yes |
| Full details of the experimental design and statistical methods used should be given | |

| | |
|---|------------|
| <p>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.</p> <p>Have you included all the information requested in your manuscript?</p> | |
| <p>Resources</p> <p>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.</p> <p>Have you included the information requested as detailed in our Minimum Standards Reporting Checklist?</p> | <p>Yes</p> |
| <p>Availability of data and materials</p> <p>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.</p> <p>Have you have met the above requirement as detailed in our Minimum Standards Reporting Checklist?</p> | <p>Yes</p> |

Transcriptome Analysis of the Response of Burmese Python to Digestion

Jinjie Duan¹, Kristian Wejse Sanggaard^{2,3}, Leif Schauser⁵, Sanne Enok Lauridsen⁴, Jan J.
Enghild^{2,3}, Mikkel Heide Schierup^{1,4} and Tobias Wang⁴

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

23 **Abstract**

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 
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

1
2
3
4
5 47 thereby providing  el insight into the digestion mechanism.
6
7

8 48 Keywords:
9

10
11 49 Burmese Python, transcriptome, tissue expression, digestion, pathway, proteome
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

50 Background

51 All animals exhibit dynamic changes in the size and functional capacities of bodily organs
52 and tissues to match energetic maintenance costs to the prevailing physiological demands [1].

53 This phenotypic flexibility is particularly pronounced for the digestive organs in animals that
54 naturally experience prolonged periods of fasting,  capable of ingesting large prey items at
55 irregular intervals. The Burmese python is an iconic example of this extreme phenotype [1].

56 Many species of pythons easily endure months of fasting, but remain capable of subduing and
57 ingesting very large meals. In Burmese pythons, digestion is attended by a large and rapid
58 rise in mass and/or functional capacities of the intestine, stomach, liver, heart and kidneys [2-
59 4] in combination with a stimulation of secretory processes and an activation of enzymes and
60 transporter proteins. These physiological responses are associated with a many-fold rise in
61 aerobic metabolism. Hence, the Burmese python is an excellent model to study the
62 mechanisms underlying extreme metabolic transitions and physiological remodelling in
63 response to altered demand [1, 3, 5-10].

64 The postprandial changes in the morphology and physiology of the intestine, heart and other
65 organs have been described in some detail in pythons [1, 5, 8, 9, 11], but only a few studies
66 [12-14] have addressed the underlying transcriptional changes of this interesting biological
67 response. Transcriptome sequencing technology now allows comprehensive surveys [15, 16],
68 and we therefore decided to use transcriptome sequencing of heart, liver, stomach, pancreas
69 and intestine in snakes that had fasted for one month and at 24 and 48h into the postprandial
70 period. These organs were chosen because a number of earlier studies reveal profound
71 phenotypic changes during the postprandial period [1-4, 17], and are therefore likely to
72 exhibit large changes in gene expression. Differential gene expression in some of these
73 organs e previously been reported [12-14], but we provide new data on 48h into the

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

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 approach. 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

105 nificant 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 re 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 ke
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.

128 Transcriptome annotation

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

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  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 \geq 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
1
2 155 expression patterns among tissues and time points within the digestive period. The first three
3
4
5 156 principal components (PCs) accounted for ~58% of the variation (Supplementary Fig. S3).
6
7 157 Despite the large overlap in expressed genes (Supplementary Fig. S2), the different tissues
8
9
10 158 exhibited distinct transcriptional signatures shown by the PCA in Figure 2, showing a
11
12 159 tendency for 24h to represent an intermediate position between fasting and 48h. Liver,
13
14 160 intestine and stomach displayed greater shifts in the PCA plots compared to heart and
15
16
17 161 pancreas, and the largest changes occurred between fasting and 24h in the stomach and
18
19 162 intestine. This fits well with the expectation that the stomach and intestine respond early in
20
21
22 163 digestion [3]. The dramatic changes in gene expression in the liver are also consistent with
23
24 164 previous observations on pythons [12].

28 165 Pattern of transcriptional responses to feeding

31
32 166 The postprandial response involves thousands of genes and large changes in gene expression.
33
34 167 To restrict the analysis of these many genes, we used a conservative approach where we
35
36 168 selected genes that are both highly and differentially expressed with two strict thresholds (see
37
38
39 169 methods). Application of these two thresholds yielded 43 genes for heart, 206 genes for liver,
40
41 170 114 genes for stomach, 89 genes for pancreas and 158 genes for intestine, respectively, that
42
43
44 171 were differentially expressed in response to digestion (Fig. 3). To illustrate in greater detail,
45
46 172 we enlarged the five sub-clusters with the most prominent increase in expression. These sub-
47
48
49 173 clusters, labelled a - e in Figure 3, are shown with full annotation in Figures 4-8. To unravel
50
51 174 the functional implications of these responses, we searched for genes encoding for proteins
52
53
54 175 involved in processes of tissue re-organization, cellular metabolism and digestion within
55
56 176 these sub-clusters for each organ.

60 177 GO enrichment analysis and colored KEGG pathway maps

178 To  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

202 We identified the secretome of the python stomach during digestion (Fig. 9). The resulting
1
2 203 mass spectrometry data (containing 122538 MS/MS spectra) was used to interrogate our
3
4 204 python transcriptome database, which includes transcriptome from stimulated stomach tissue.
5
6
7 205 In total, 549 python proteins were identified using this approach. Afterwards, all
8
9 206 identifications based on a single tryptic peptide were removed reducing the number of
10
11
12 207 identified python proteins to 314 (Supplementary Table S7).
13
14

15
16 208 Five classical types of pepsinogens exist, namely pepsinogen A, B, and F,
17
18 209 progastricsin (or pepsinogen C), and prochymosin [29]. Of these, our analyses
19
20 210 (Supplementary table S8 and S9) show that pythons primarily rely on progastricsin for
21
22 211 proteolytic digestion, as the five most abundant proteases identified in the gastric juice are
23
24
25 212 annotated as progastricsin-like. **Alignment of the sequences of the various transcripts for**
26 
27
28 213 **gastricsin-like proteins shows considerable differences in sequence**, which indicate the
29
30 214 presence of numerous different proteins with similar functions. This annotation is based on
31
32 215 accession XP_003220378.1 and XP_003220378.1 from *Anolis carolinensis*. Alignment of the
33
34 216 python sequences with the two anole sequences, as well as with the well-characterized human
35
36 217 gastricsin variant, shows that both the active site residues, as well as cysteine bridges, are
37
38 218 conserved. It demonstrates the similarity between these enzymes and suggests that the
39
40 219 identified python sequences indeed represent catalytically active proteolytic enzymes
41
42
43 220 (Supplementary Fig. S9). The last identified pepsinogen-like python sequence
44
45 221 (m.31615_Py95) was annotated based on the predicted embryonic pepsinogen-like sequence
46
47 222 (XP_003220239.1), also from *Anolis carolinensis*. Here, the annotation originates from an
48
49 223 embryonic pepsinogen identified in chicken [30]. This protease was identified in the python's
50
51
52 224 gastric juice with a lower emPAI value than the gastricsin sequences indicating a lower
53
54
55 225 concentration of this enzyme (Supplementary table S8), although the transcript displays the
56
57
58 226 highest concentration of the analysed pepsinogens in the post-prandial period (Supplementary
59
60
61
62
63
64
65

227 table S9). As the name icate it is exclusively expressed during the embryonic period [30,
1
2 228 31], and logenetic analysis of the sequence suggest that its closest homolog, among the
3
4 229 classical pepsinogens, is prochymosin [30]. Also, prochymosin displays a temporal
5
6
7 230 expression pattern and is, in mammals, mainly expressed in new-born species. However, the
8
9
10 231 identified python snake embryonic chicken pepsinogen homolog does not display a similar
11
12 232 development-related temporal expression pattern and is, as shown, used among adult species
13
14 233 for digestion. However, it does not exclude that the protease is expressed during the python's
15
16
17 234 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
24
25
26 237 proteins, such as collagen. To facilitate the downstream analyses of the python proteins, we
27
28
29 238 produced a list of background proteins related to the prey. Hence, interrogation of the mass
30
31
32 239 spectrometry data against the 16693 mouse protein sequences in the Swiss-Prot database
33
34 240 resulted in the identification of 212 mouse proteins, after removing hits based on single
35
36 241 peptides (Supplementary table S10). To produce a list of identified python proteins, most
37
38
39 242 likely present in the digestive fluid samples due to blood contaminations during collection,
40
41 243 we characterized the python plasma proteome. The most abundant plasma proteins are
42
43
44 244 produced by the liver. Consequently, our python transcriptome sequence database, which
45
46 245 encompasses liver transcriptomes, is expected to contain the protein sequences of the python
47
48
49 246 plasma proteins. Thus, our python plasma LC-MS/MS data was used to interrogate our
50
51 247 python sequence database. It provided an overview of the most abundant python plasma
52
53 248 proteins (Supplementary table S11). In total, 64 plasma proteins were identified with
54
55
56 249 minimum two tryptic peptides. The result supports the liver transcriptome data, since the
57
58 250 abundant (based on emPAI) plasma proteins relate with the transcripts that are detected at
59
60
61
62
63
64
65



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



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.*  hon 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

275 proteomics analysis. In total, 37 proteins constituted the putative python stomach secretome
1
2 276 (Supplementary table S8). These could be divided into 18 gastric mucosal-related proteins
3
4
5 277 (e.g. mucin homologous and gastrokine), seven proteolytic enzymes (mainly pepsin
6
7 278 homologous), four other hydrolytic enzymes (e.g. phospholipases), and eight other proteins
8
9
10  279 (e.g. gastric intrinsic factor) (Supplementary table S8). Previous gastric juice proteomics
11
12 280 analyses were performed on samples obtained from fasting humans, most likely to avoid the
13
14 281 complex prey-protein background. In our study, we identify, for the first time, stomach-
15
16
17 282 related proteins from a digesting individual and thereby demonstrate that the sensitivity of
18
19 283 modern LC-MS/MS equipment allows the identification of gastric juice proteins that are
20
21
22 284 present during digestion.
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

285 **Discussion**

1
2
3
4 286 A primary motivation for our description of the temporal changes in gene expression profiles
5
6 287 as the visceral organs of Burmese pythons made the transition from fasting to digestion was
7
8 288 to identify key regulatory genes and pathways responsible for the pronounced tissue
9
10
11 289 restructuring and the increased functional capacity during the postprandial period. An equally
12
13 290 important motivation was to  **press specific hypothesis** on the upregulation of certain
14
15
16 291 pathways known to be involved in the secretion of digestive juices and enzymes as well as
17
18 292 the absorption of the nutrients as digestion proceed. We achieved these goals by identifying
19
20
21 293 the biochemical and physiological roles of the highly expressed genes with increased
22
23 294 expression during digestion and by using KEGG analysis of specific pathways underlying
24
25 295 physiological responses known to be stimulated by digestion. We also present GO
26
27
28 296 enrichment analyses of both up-regulated genes and highly expressed genes in all organs
29
30 297 (Supplementary Figs. S4-S8), showing that “biological process” is the most common
31
32
33 298 enriched category.

34
35
36 299 The influence of digestion on gene expression profiles in heart, liver, kidney and small
37
38
39 300 intestine has been studied previously in pythons [12-14]. These earlier studies reported
40
41 301 thousands of genes being either up- or downregulated within the first day of digestion [12-
42
43
44 302 14], and we confirm these substantial changes in gene expression at 24h and 48h. However,
45
46 303 we merely identified hundreds of genes, probably because we selected a more stringent
47
48
49 304 threshold for calling the differential expression. Given the differences in the selection of
50
51 305 thresholds and analysis strategy for differential expression and differences in times of
52
53
54 306 sampling, it is difficult to make a direct comparison between our study and that of Castoe et
55
56 307 al (2013). Nevertheless, for heart, liver and small intestine, both studies have determined a
57
58 308 number of upregulated genes at 24h where we identified 15, 93 and 61 upregulated genes,
59
60
61
62
63
64
65

309 respectively. Comparing upregulated genes between two studies (see supplementary material
1
2 310 for detailed method and result), we found there was  **good overlap** in identifying upregulated
3
4 311 genes in the liver where more than half of the 93 genes identified in our study were identified
5
6
7 312 as upregulated genes by Castoe et al (2013). However, there was less overlap for heart and
8
9
10 313 the small intestine. These differences may be due to the use of different quantification
11
12 314 methods for gene expression in the various studies, but may also be a result of the limited
13
14  315 biological replicates in our study. **Nevertheless, genes identified as being upregulated in both**
16
17 **316 studies, are probably of high confidence.**

18
19
20 317 Physiological interpretation of the upregulated genes in the stomach

21
22
23
24 318 The considerable changes in gene expression in the stomach were reflected in a pronounced
25
26 319 rise in expression of ribosomal 40S and 60S proteins (Fig. 4) that is likely to have attended a
27
28
29 320 rise in protein synthesis required for the marked transition from a quiescent fasting state to
30
31 321 the activated digestive state. This is also supported by the presence of ribosomal functions in
32
33 322 the enriched GO analysis of the stomach of the highly-expressed genes (Supplementary Fig.
34
35 323 S4B). During fasting, gastric acid secretion and presumably also the secretion of digestive
36
37 324 enzymes and lysozymes, is halted, such that the gastric juice has a neutral pH, whilst
38
39 325 ingestion of prey is followed by an immediate activation of gastric acid secretion [36, 37].
40
41 326 The stimulation of the secretory actions of the stomach is attended by an increased mass of
42
43
44 327 the stomach, where particularly the mucosa expands already within the first 24h [38].
45
46
47
48
49

50 328 The KEGG analysis, however, shows that the genes encoding for the gastric H,K
51
52 329 ATPase, the active and ATP consuming ion-transporter responsible for gastric acid secretion,
53
54 330 are highly expressed in fasting animals, and not additionally elevated in the postprandial
55
56
57 331 period (Fig. 10). This strongly indicates that the enzymatic machinery for gastric acid
58
59 332 secretion is maintained during fasting, a trait that may enable fast activation of acid secretion,
60
61
62
63
64
65

333 at modest energetic expenditure, to kill bacteria and match gastric pH to the optimum value
1
2 334 for pepsin. This interpretation is consistent with a number of recent studies indicating a rather
3
4
5 335 modest contribution of gastric acid secretion to the specific dynamic action (SDA) response
6
7 336 in pythons [39, 40], but we also did observe a high prevalence of ATP synthase subunits (Fig.
8
9
10 337 4) amongst the highly upregulated genes, which does indicate a rise in aerobic metabolism
11
12 338 (see also supplementary Fig. S4). Furthermore, the upregulation of the gene encoding for
13
14 339 creatine kinase (Fig. 4) indicate increased capacity for aerobic respiration required costs of
15
16 340 acid secretion and the stimulation of the accompanying gastric functions. It has been
17
18
19 341 proposed that gastric processes account for more than half of the rise in total metabolism
20
21 342 during digestion [36], and aerobic metabolism of isolated gastric strips *in vitro* increased
22
23 343 during digestion [41]. However, while metabolism of the stomach certainly must increase
24
25
26 344 during the postprandial period, more recent studies indicate a considerably smaller
27
28
29 345 contribution of gastric acid secretion to the total SDA response  considerable lower than
30
31 346 50% [39, 40, 42].
32
33
34

35 347 Our KEGG analysis also showed a large rise in expression of the gene encoding for
36
37 348 carbonic anhydrase (Fig. 10), the enzyme that hydrates CO₂ and provide protons for gastric
38
39 349 acid secretion. Gastric acid secretion, therefore, does not appear  under transcriptional
40
41
42 350 regulation, but is likely to involve translocation of existing H,K ATPases in vesicles from
43
44 351 intracellular vacuoles to the apical membrane of the oxyntopeptic cells that are responsible
45
46
47 352 for both gastric acid secreting as well as the release of pepsinogen in reptiles [43]. An
48
49
50 353 activation of the processes involved in vesicle transport is further supported by increased
51
52 354 transcription of the gene encoding for CD63 (Fig. 4), which belongs to the tetraspanin family
53
54
55 355 and mediate signal transduction events.
56
57

58 356 In contrast to acid secretion, expression of several genes encoding for digestive
59
60
61
62
63
64
65

357 enzymes (embryonic pepsinogen-like, gastricsin precursor and gastricsin-like) (Fig. 4) were
1
2 358 upregulated, which is consistent with *de novo* synthesis of the enzymes responsible gastric
3
4
5 359 protein degradation. Also, there was good overlap between the upregulation of the relevant
6
7 360 genes encoding for the proteins identified in the stomach secretome, such as gastrokines,
8
9
10 361 pepsin homologous, phospholipases and gastric intrinsic factor (Supplementary table S8). In
11
12 362 this context, it is also interesting that mucin 6 (Fig. 4), the gene coding for the large
13
14
15 363 glycoprotein (gastric mucin) that protects the gastric mucosa from the acidic and
16
17 364 proteolytically active chyme in the stomach lumen was upregulated. Thus, as gastric acid
18
19 365 secretion is activated, probably in response to increased levels of the gastrin as well as
20
21
22 366 luminal factors, there is an accompanying activation of the protective mucus layer that
23
24 367 prevents auto-digestion of the gastric mucosa. It is also noteworthy that the genes for both
25
26
27 368 gastrokine 1 and 2 were upregulated during digestion (Fig. 4). Gastrokines are constitutively
28
29 369 produced proteins in the gastric mucosa in mammals and chickens, and while the
30
31
32 370 physiological function remains somewhat elusive, they appear to upregulated during mucosal
33
34 371 remodelling in response to inflammation (*e.g.* in connection with ulcers) and often
35
36
37 372 downregulated in cancers. Thus, it is likely that the gastrokines are involved in regulating the
38
39 373 restructuring of the mucosa during digestion in pythons.

42 374 In addition to analysing the gene expression profiles of the stomach, we also used a
43
44
45 375 proteomics approach, assisted by our python transcriptome sequence database, to identify the
46
47
48 376 hydrolytic enzymes in the gastric juice secreted during digestion. We identified python
49
50 377 proteins on a complex background of highly abundant mice proteins. Python's digested food
51
52 378 is, when it enters the duodenum, rally similar to digested food in *e.g.* humans. Thus, the
53
54
55 379 digestive enzymes secreted by the pancreas are probably functional similar to known
56
57 380 hydrolytic enzymes from other species. Consequently, the enzymes that facilitate the extreme
58
59
60 381 digestion process and ow for have to be present in the stomach's digestive fluid.

382 We hypothesized that relative aggressive proteolytic digestive enzymes in the gastric
1
2 383 juice facilitate digestion of large and un-masticated whole prey items [8]. In our analysis, six
3
4
5 384 out of the seven identified proteolytic enzymes were pepsinogen homologous (Peptidase
6
7 385 subfamily A1A), and these were also the most abundant hydrolytic enzymes in the gastric
8
9
10 386 juice according to the emPAI values (Supplementary table S8). Most likely other pepsinogen
11
12 387 isoforms exist in the gastric juice, as our approach predominantly target the most abundant
13
14 388 proteolytic enzymes. The importance of the proteomics-identified pepsinogens was also
15
16
17 389 substantiated by the transcriptomics data (Supplementary table S9). Here, we found that the
18
19 390 six different pepsinogens were upregulated between 2.2 and 22.2 fold from the fasting
20
21
22 391 animals to 48 hours after ingestion of mice. In average the pepsinogen transcripts were
23
24 392 upregulated 10.7 fold. It supports that these proteases play a substantial role in the aggressive
25
26
27 393 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
1
2 407 enterocytes [44], while the smooth muscle in the gut wall is much less responsive [45].
3
4
5 408 Earlier studies on gene expression profiles during digestion in the python intestine revealed
6
7 409 massive upregulation of more than one thousand genes, commencing within the first six
8
9
10 410 hours after ingestion [12, 13]. Importantly, this previous study [13] identified a number of
11
12 411 genes that are likely to be involved in the restructuring of the microvilli, cell division and
13
14 412 apoptosis, as well as brush-border transporter proteins. In line with these earlier findings, our
15
16
17 413 GO enrichment analysis also highlights functions pertaining to mitotic cell division, which
18
19 414 supports a contribution to growth by hyperplasia faster cell turnover (Supplementary Fig. S5).
20
21
22 415 The expansion of the individual enterocytes is accompanied by pronounced elongation of the
23
24 416 microvilli [46] and the resulting rise in surface area of the intestinal lining is accompanied by
25
26
27 417 an ten-fold increase in intestinal transport capacity for amino acids and other nutrients [1, 4,
28
29 418 47].
30
31

32
33 419 Earlier studies provided strong evidence for an upregulation of genes coding for nutrient
34
35 420 transporter proteins, such as D-glucose, L-proline and L-leucine [13]. In this context, it is
36
37
38 421 noteworthy that there were no nutrient transporters amongst the highly expressed and
39
40 422 upregulated genes in the intestine (Fig. 5), but our KEGG analysis nevertheless showed
41
42 423 increased expression of the serosal L-type amino acid transporter. Clearly, it would be
43
44
45 424 worthwhile to quantitatively analyse the extent to which *de novo* synthesis of the various
46
47 425 nutrient transporters, particularly those for amino acids, is increased during digestion and
48
49
50 426 how much such synthesis tribute to absorptive capacity. It would seem adaptive if many
51
52 427 of the transporters merely have to be activated, either by insertion within the luminal
53
54
55 428 membrane or exposed as the enterocytes expand, to allow for an energetically cheap manner
56
57 429 of matching intestinal performance to the sudden appearance of nutrients in the intestine after
58
59
60 430 a meal. The GO enrichment analysis also pointed to an enrichment of various metabolic
61
62
63
64
65

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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

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

479 capacity in postprandial period [53] and cytochrome oxidase activity is almost doubled
1
2 480 during digestion [54], and we confirm that transcription for heat shock proteins may be
3
4
5 481 increased [53], possibly to protect against oxidative damage as result of the increased
6
7 482 metabolism. As in earlier studies [53], our observation of increased ATP synthase
8
9
10 483 lipid-binding protein and fatty acid binding protein 3 (Fig. 6) provide evidence for increased
11
12 484 fatty acid metabolism, which may reflect the substantial rise in circulating fatty acids in the
13
14 485 plasma.

16
17
18 486 It was originally suggested that the postprandial rise in stroke volume could be
19
20
21 487 ascribed to an impressive and swift growth of the heart [10], possibly triggered lipid-
22
23 488 signalling [53]. However, a number of recent studies, primarily from our laboratory, have
24
25 489 shown that increased cardiac mass is not an obligatory postprandial response amongst
26
27
28 490 pythons [54-56], and that stroke volume may be increased in response to increased venous
29
30 491 return rather than cardiac hypertrophy [54]. It is nevertheless, noteworthy that our and the
31
32
33 492 previous studies show a clear increase in the expression of contractile proteins (e.g. myosin
34
35 493 and actin) as well as tubulin (Fig. 6), which may reflect increased protein-turnover in
36
37
38 494 response to increased myocardial workload rather than cell proliferation or hypertrophy. The
39
40 495 enriched GO analyses also point to major changes in the extracellular space as well as both
41
42 496 elastin and collagen, which may indicate some level of cardiac reorganization at the cellular
43
44
45 497 or subcellular level that may alter compliance of the myocardial wall and influence cardiac
46
47 498 filling (Supplementary Fig. S8). It is noteworthy that the increased expression of BNP may
48
49
50 499 serve a signalling function as described in response to the cardiac hypertrophy that attends
51
52 500 hypertension.

53
54
55
56 501 Physiological interpretation of the genes in the liver

57
58
59
60 502 The liver exhibited a diverse expression profile in response to digestion that is likely to
61
62
63
64
65

503 reflect its many metabolic functions in connection with metabolism, synthesis and
1
2 504 detoxification during the postprandial period. This pattern is also evident from the many
3
4
5 505 metabolic functions identified in the enriched GO analysis (Supplementary Fig. S7). There
6
7 506 were marked upregulations of the P450 system (Fig. 7), which stems well with a rise in
8
9
10 507 synthesis and breakdown of hormones and signalling molecules, cholesterol synthesis in
11
12 508 response to lipid absorption and possibly also an increased metabolism of potentially toxic
13
14 509 compounds in the prey. A rise in cholesterol metabolism was supported by increased
15
16
17 510 expression apolipoproteins (Fig. 7). The hepatic involvement in lipid metabolism was also
18
19 511 supported by the increased expression of genes for Alpha-2-macroglobulin and serum
20
21 512 albumin (Fig. 7). The increased expression of albumin obviously also fits nicely with the
22
23
24 513 proteomic analysis of plasma proteins and it is likely that the postprandial rise in plasma
25
26
27 514 albumin serves a functional role in the lipid transport between the intestine and the liver as
28
29 515 well as other metabolically active organs
30
31

32
33 516 It is also noteworthy that a number of genes associated with the protection of
34
35 517 oxidative stress, such as catalase, heat shock protein and glutathione transferase were
36
37
38 518 markedly upregulated (Fig. 7). It was recently argued that snakes digesting large meals
39
40 519 experience oxidative damage due to reactive oxygen metabolites requiring increased
41
42
43 520 antioxidant responses to protect cellular functions [57].
44
45

46 521 Physiological interpretation of the genes in the pancreas
47
48
49

50 522 We sampled the entire pancreas for our analysis of gene expression and our data therefore
51
52 523 reflect both endocrine and exocrine pancreatic functions.  the vast majority of the upregulated
53
54 524 genes concerned the exocrine pancreas, and we found ample evidence for upregulated
55
56
57 525 expression of genes associated with the digestive functions, such as lipases, trypsin,
58
59
60 526 chymotrypsin and elastase and other enzymes for digestion of protein and lipid (Fig. 8). This
61
62
63
64
65

527 general upregulation of secretory processes is likely to explain the prevalence of processes
1
2 528 associated with protein synthesis in the enriched GO analysis (Supplementary Fig. S6). There
3
4
5 529 was even an increased expression of amylase (Fig. 8) that breaks down polysaccharides. In
6
7 530 connection with this latter function, the increased expression of insulin (Fig. 8) from the
8
9
10 531 endocrine pancreas is likely to reflect increased cellular signalling for postprandial uptake of
11
12 532 both glucose and amino acids. As in the other organs, we found increased expression of
13
14 533 cytochrome oxidase (Fig. 8) indicative of increased metabolism during digestion, and the rise
15
16
17 534 in heat shock protein expression may reflect a response to formation of reactive oxygen-
18
19 535 species as metabolism is stimulated by increased secretion of the pancreas.
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

536 **Conclusions**

1
2
3
4 537 Our study confirms that the extensive physiological and anatomical reorganization of the
5
6 538 visceral organs of pythons during the postprandial period is driven by differential expression
7
8 539 of hundreds or even thousands of genes. Many of the upregulated functions pertain to energy
9
10 540 production to support the rise in aerobic metabolism associated with digestion and absorption
11
12
13 541 of the large meals. In terms of the gastrointestinal organs, the gene expression profiles also
14
15 542 support the view that many of the digestive functions, such as gastric acid secretion and
16
17
18 543 nutrient absorption, can be stimulated with little gene expression indicating that the proteins
19
20
21 544 involved in these processes are merely need to be activated during the postprandial period,
22
23 545 and thus avoiding the energy and time-consuming processes associated with *de novo*
24
25 546 synthesis. This digestive strategy may, at least in part, explain how intermittent feeders, such
26
27
28 547 as snakes, retain the capacity for fast and reliable upregulation of the digestive processes
29
30 548 immediately after ingestion.
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

549 **Methods**

1
2
3
4 550 Stimulation of the postprandial response, collection of tissue biopsies and purification of
5
6 551 RNA for mRNA-seq analyses
7
8

9
10 552 Six *Python molurus* (Tiger Python/Burmese Python) with a body mass ranging from 180 to
11
12 553 700 g (average 373 g) were obtained from a commercial supplier and housed in vivaria with a
13
14 554 heating system providing temperatures of 25-32 °C. The animals were fed rodents once a
15
16
17 555 week and fresh water was always available. The animals appeared healthy and all
18
19 556 experiments were performed according to Danish Federal Regulations. All six individuals
20
21
22 557 were fasted for one month and divided in three groups. Four animals were fed a rodent meal
23
24 558 of 25 % of body weight and euthanized with an intra-peritoneal injection of pentobarbital (50
25
26 559 mg kg⁻¹; Mebumal) at 24h (N = 2) or 48h after feeding (N = 2). The remaining two snakes
27
28
29 560 served as fasted controls. During deep anaesthesia, two biopsies were obtained from each
30
31 561 snake from each of the following tissues: The heart (ventricles), liver, stomach, intestine, and
32
33
34 562 pancreas. In regard to the stomach tissue samples, one sample was obtained from the
35
36 563 proximal part of the stomach and one sample was obtained from the distal part. In total, 60
37
38
39 564 biopsies were collected. The samples were taken from the same part of the different tissues in
40
41 565 all individuals. After sampling, the biopsies were weighted and immediately snap frozen in
42
43 566 liquid nitrogen; stomach and intestinal tissues were rinsed in sterile saline solution before
44
45
46 567 weighting to avoid contamination with rodent tissue from the ingested meal. Subsequently,
47
48
49 568 all 60 biopsies were homogenized in liquid nitrogen and the four biological replicates (two
50
51 569 biopsies from each individual) were pooled in a 1:1 manner based on mass. It resulted in 15
52
53 570 samples (five tissues X three time points). From these samples, total RNA was purified using
54
55
56 571 the Nucleospin RNA II kit (Machery-Nagel GmbH & Co.), as recommended by the
57
58 572 manufacturer. The RNA concentration and quality were assessed by Nanodrop ND 1000
59
60
61
62
63
64
65

1
2
3
4
5
6 573 Spectrophotometer (Thermo Scientific) analyses, agarose gel-electrophoreses, and Agilent
7
8
9
10 574 BioAnalyzer (Agilent) analyses.
11
12
13
14 575 Library production and sequencing
15
16
17 576 Poly-A transcripts were enriched and the transcripts broken in the presence of Zn²⁺.
18
19 577 Subsequently, double-stranded cDNA was synthesized using random primers and RNase H.
20
21 578 After end repair and purification, the fragments were ligated with bar-coded paired-end
22
23 579 adapters, and fragments with insert sizes of approximately 150-250 bp were isolated from an
24
25 580 agarose gel. Each of the 15 samples derived from five tissues (heart, liver, stomach, pancreas
26
27 581 and intestine) at the three time points (fasted for one month, 24h and 48h post-feeding) were
28
29 582 amplified by PCR to generate DNA colonies template libraries and the libraries were then
30
31 583 purified. In addition, to sample as broadly from transcriptome as possible, we also produced
32
33 584 normalized libraries for each tissue in order to capture the reads from lowly expressed, tissue-
34
35 585 specific genes. Here, a part of the samples, which originating from the same tissue, were
36
37 586 pooled before the PCR analyses, i.e. in total five pooled samples were generated. These five
38
39 587 samples were split in two and after PCR amplification and library purification they were
40
41 588 normalized using two different normalization protocols, i.e. in total 10 normalized libraries
42
43 589 were prepared. Library quality of all 25 samples was then assessed by a titration-run (1 x 50
44
45 590 bp) on an Illumina HiSeq 2000 instrument. Finally, the sequencing was performed on the
46
47 591 same instrument using paired-reads (2 × 101 bp). One channel was used for the 15 non-
48
49 592 normalized libraries and one channel was used for the 10 normalized libraries.
50
51
52 593 Data pre-processing and *de novo* transcriptome assembly
53
54
55
56 594 To reduce the amount of erroneous data, the raw paired reads were processed by i) removing
57
58 595 reads that contained the sequencing adaptor, ii) removing reads that contained ambiguous
59
60
61
62
63
64
65

596 characters (Ns), and iii) trimming bases that had  low average quality (Q<20) within a
1
2 597 sliding window of length 10.
3
4

5
6 598 To develop a comprehensive transcriptomics resource for the Burmese python, all
7
8 599 high-quality reads from 25 libraries were pooled together for *de novo* assembly. To determine
9
10
11 600 the optimal assembly, *de novo* assembly was performed using Velvet (version 1.2.03) [18]
12
13
14 601 and Oases (version 0.2.06) [58] with different k-mer parameters. The performance of these
15
16 602 assemblies was assessed according to number of transcripts, total length of transcripts, N50
17
18 603 length, mean length, proportion of mapped reads and number of transcripts which length is
19
20
21 604 larger than N50 (Supplementary Table S2).
22
23

24 605 Assessment of the transcriptome assembly

25
26
27
28 606 The transcriptome assembly was evaluated by rnaQUAST 1.4.0 with default parameters
29
30
31 607 supplying reference genome sequences and genome annotation of Burmese python (GenBank
32
33 608 assembly accession: GCA_000186305.2).
34
35

36
37 609 BUSCO_v2 [20] was used to test the completeness of transcriptome assembly with
38
39 610 dependencies NCBI BLAST+ 2.4.0 [59] and HMMER 3.1b2 [60]. The vertebrata lineage set
40
41
42 611 was used and accessed on 28 Nov 2016.
43
44

45 612 Transcriptome annotation

46
47
48 613 To assess the identity of the most closely related gene in other organisms, the assembled
49
50 614 transcripts were compared with the sequences in the National Center for Biotechnology
51
52
53 615 Information (NCBI) non-redundant protein (nr) database using blastx [61] with an e-value
54
55 616 cut-off of 0.01. The nr annotation term of each transcript was assigned with the first best hit,
56
57
58 617 which was not represented in uninformative description (e.g., 'hypothetical protein', 'novel
59
60 618 protein', 'unnamed protein product', 'predicted protein' or 'Uncharacterized protein')
61
62
63
64
65

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

641 Colored KEGG Pathway and GO enrichment analysis

1
2
3
4 642 For each tissue, all assembled genes were mapped to KEGG human pathway maps using
5
6 643 KOBAS 2.0 [64] with e-value $1e-50$. Then genes were colored by representing FPKM value
7
8 644 and trend of differential expression value (Table 2).

10
11
12 645 Blast2GO was used to implement GO enrichment analysis (Fisher's exact test) with
13
14 646 threshold of FDR 0.001. The reference set is the whole transcripts with GO slim annotation.
15
16
17 647 For each organ, the selected test set is either upregulated or highly expressed genes defined in
18
19 648 Table 2. Finally, we performed Blast2GO to reduce to most specific GO terms.

20
21
22
23 649 Isolation of samples for proteomics analyses

24
25
26
27 650 Two Burmese pythons (weighing 400 and 800 g, respectively) were fed a rodent meal
28
29 651 corresponding to approximately 25% of their body mass. Approximately 24 h into the
30
31
32 652 postprandial period the animals were euthanized with an overdose of pentobarbital (100 mg
33
34 653 kg^{-1} , i.m.). Immediately afterwards, an incision was made to expose the stomach, which was
35
36 654 then ligated at the lower oesophagus and the pylorus, before the intact stomach was excised
37
38
39 655 by a cleavage just below the two sutures resulting in the stomach being released from the rest
40
41 656 of the animal. All undigested mouse remains were manually removed by forceps and 25
42
43 657 ml/kg tris-buffered saline (TBS) was injected into the stomach. The stomach was then ligated
44
45
46 658 at the opened end, rinsed by gently shaking the tissue, and finally the digestive fluid-
47
48
49 659 containing solution was collected and stored on ice. To ensure collection of all gastric fluid,
50
51 660 the stomach was rinsed additional two-three times with 12 ml/kg TBS. Subsequently, the
52
53 661 samples were filtered and centrifuged, and the supernatant stored at $-80\text{ }^{\circ}\text{C}$. We also obtained
54
55
56 662 two samples of gastric juice from a third individual (200 g) that had been fed 4 g peptone
57
58 663 (Sigma Aldrich), suspended in water. Peptone is a mixture of small peptides and amino acids
59
60
61
62
63
64
65

664 and the solution was injected directly into the stomach and after three hours the snake was
1
2 665 euthanized by an overdose of pentobarbital. The stomach was removed, rinsed with TBS, and
3
4
5 666 a single sample collected and stored, as described above. We analysed two samples from each
6
7 667 of the three individuals, resulting in a total of six digestive fluid samples being analysed by
8
9
10 668 MS/MS. In addition, we obtained a single plasma sample from each snake by direct cardiac
11
12 669 puncture followed by centrifugation and storage for later analysis.
13
14

15 16 670 Sample preparation for mass spectrometry analyses 17

18
19 671 The proteins in the six obtained python digestive fluid samples were recovered by
20
21
22 672 trichloroacetic acid precipitation. The resulting pellets were resuspended in 8 M Urea, 5 mM
23
24 673 DTT, 0.1 M ammonium bicarbonate pH 8.0 and incubated for 30 minutes at room
25
26
27 674 temperature in order to denature and reduce the proteins. Subsequently, the proteins were
28
29 675 alkylated by the addition of iodoacetamide to a final concentration of 25 mM. The samples
30
31
32 676 were incubated for additional 20 minutes at room temperature and then diluted five times
33
34 677 with a 50 mM ammonium bicarbonate, pH 8.0 buffer before the addition of approximately 2
35
36 678 µg sequencing grade modified trypsin (Promega) per 50 µg protein in the sample.
37
38
39 679 Subsequently, the samples were incubated at 37 °C for approximately 16 h. The proteins in
40
41 680 the plasma sample were denatured, reduced, alkylated, and digested with trypsin, as described
42
43
44 681 for the digestive fluid samples. Finally, the resulting peptides in all samples were
45
46 682 micropurified and stored at -20 C until the LC-MS/MS analyses.
47
48

49 50 683 Liquid chromatography-tandem mass spectrometry analyses 51

52
53 684 Nano-liquid chromatography-tandem mass spectrometry (LC-MS/MS) analyses were
54
55
56 685 performed on a nanoflow HPLC system (Thermo Scientific, EASY-nLC II) connected to a
57
58 686 mass spectrometer (TripleTOF 5600, AB Sciex) equipped with an electrospray ionization
59
60
61
62
63
64
65

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

1
2 711 obtained using the emPAI-values given by the Mascot 2.5.0 software after analysis of the
3 712 MS/MS data [67].
4
5

6 713 To identify the proteins secreted into the python stomach, identified python plasma
7
8 714 proteins, as well as the mouse protein homologs were removed from the list of identified
9
10 715 python digestive fluid proteins. With regard to the removal of prey protein homologs, the
11
12 716 overall mouse protein names were used to search the list of python proteins (e.g. “collagen”
13
14 717 was used as search term, not “collagen alpha-1(I) chain”) and to identify python proteins that
15
16 718 were identified based on homology with mouse. These proteins were removed from the list of
17
18 719 stomach-secreted python proteins. For each identified protein remaining on the list, we
19
20 720 reassessed the annotation of the python sequence, i.e. sequence comparisons were performed
21
22 721 using blastp version 2.2.30, and in addition, UniProt and NCBI protein databases, as well as
23
24 722 PubMed and SignalP 4.1, were interrogated to identify functional properties and cellular
25
26 723 location of the identified proteins. Plasma proteins, remaining collagen homologous,
27
28 724 intracellular proteins, and membrane proteins were discarded from the list of identified
29
30 725 python stomach secretome proteins.
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

726 **List of abbreviations**

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

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**
1
2
3
4 732 Ethics approval and consent to participate
5
6
7 733 Not applicable
8
9
10
11 734
12
13
14 735 Consent for publication
15
16
17
18 736 Not applicable
19
20
21
22 737
23
24
25 738 Availability of data and material
26
27
28
29 739 The raw RNA-Seq sequencing data that support the findings of this study have been
30
31 740 deposited in the NCBI BioProject database (accession no. PRJNA343735).
32
33
34 741 <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA343735>
35
36
37
38 742
39
40
41 743 Competing interests
42
43
44
45 744 The authors declare that they have no competing interests.
46
47
48
49 745
50
51
52 746 Funding
53
54
55
56 747 This work was supported by a grant (Novenia) from the Danish Research Council for
57
58 748 Strategic Research (grant identification number: 09-067076).
59
60
61
62
63
64
65

749

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

750 Authors' contributions

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

756

757 Acknowledgements

758 We thank Tania A. Nielsen (Aarhus, Denmark) for valuable assistance with RNA purification
759 and Fasteris SA (Switzerland) for library preparation and Illumina sequencing.

760 **Figure and table legends**

1
2
3 761 **Fig. 1. The workflow of Python RNA-Seq data analysis.** The diagram shows the main
4
5
6 762 steps and bioinformatics tools used in the study.

7
8 763 **Fig. 2. PCA plots of FPKM of 1862 genes.** PC, principal component. PC1 represents 25%,
9
10 764 PC2 represents 18% and PC3 represents 16% of total variation in the data. The name of the
11
12 765 label consists two part, one capital letter plus one number. Letter H, S, I, L, P represent
13
14 766 heart, stomach, intestine, liver and pancreas respectively. Number 0, 1, 2 represent fasting
15
16
17 767 for one month, 24h/1d after feeding and 48h/2d after feeding respectively.

18
19 768 **Fig. 3. Heat maps from hierarchical clustering of DEGs in each tissue.** Heat maps
20
21 769 showing the hierarchically clustered Spearman correlation matrix resulting from comparing
22
23 770 the normalized FPKM value for each pair of genes. Heat map columns represent samples
24
25 771 and rows correspond to genes. Expression values (FPKM) are \log_2 -transformed and then
26
27 772 median-centered by gene. Relative levels of gene expression are represented by colors.
28
29 773 Pale colour is low expression and darker blue is high expression. Five sub-clusters labelled
30
31 774 a to e are shown with full annotation in Fig. 4-8.

32
33 775 **Fig. 4. The cluster of upregulated genes with NCBI nr annotation in stomach.** It shows
34
35 776 the cluster e in Fig. 3. Heat map columns represent samples and rows correspond to genes.
36
37 777 Expression values (FPKM) are \log_2 -transformed and then median-centered by gene.
38
39 778 Relative levels of gene expression are represented by colors. Pale colour is low expression
40
41 779 and darker blue is high expression.

42
43 780 **Fig. 5. The cluster of upregulated genes with NCBI nr annotation in intestine.** It shows
44
45 781 the cluster b in Fig. 3. Heat map columns represent samples and rows correspond to genes.
46
47 782 Expression values (FPKM) are \log_2 -transformed and then median-centered by gene.
48
49 783 Relative levels of gene expression are represented by colors. Pale colour is low expression
50
51 784 and darker blue is high expression.

52
53 785 **Fig. 6. The cluster of upregulated genes with NCBI nr annotation in heart.** It shows the
54
55 786 cluster a in Fig. 3. Heat map columns represent samples and rows correspond to genes.
56
57
58
59
60

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.**

814 **Table 2. Colour coding of genes in KEGG pathway maps.** Three criteria are used to
1
2 815 classify and colour genes. First, i) whether the maximum FPKM of the gene among fasting,
3
4 816 24h and 48h is over 10, then ii) whether the gene is differential expressed in at least one of
5
6 817 the pairwise comparison among fasting, 24h and 48h with FC over 4. Finally, iii) for those
7
8
9 818 genes expressed, but not differential expressed, whether it is highly expressed with
10
11 819 maximum FPKM among three time points over 200. The term expression trend indicates the
12
13 820 trend of gene expression across fasting, 24h and 48h. e.g. The trend up means the gene is
14
15 821 upregulated from either fasting to 24h, fasting to 48h or 24h to 48h. The trend up-then-down
16
17 822 means the gene is firstly upregulated from fasting to 24h, then downregulated from 24h to
18
19
20 823 48h.

21
22 824 **Table 3. The number of DEGs across fasting, 24h and 48h in each tissue.** The
23
24 825 expression trend is consistent with definition in Table 2.
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

- 1
- 2
- 3
- 4 827 1. Secor SM, Diamond J: **A vertebrate model of extreme physiological**
- 5 828 **regulation.** *Nature* 1998, **395**:659-662.
- 6 829 2. Cox CL, Secor SM: **Matched regulation of gastrointestinal performance in the**
- 7 830 **Burmese python, *Python molurus*.** *J Exp Biol* 2008, **211**:1131-1140.
- 8 831 3. Secor SM, Diamond J: **Adaptive responses to feeding in Burmese pythons:**
- 9 832 **pay before pumping.** *J Exp Biol* 1995, **198**:1313-1325.
- 10 833 4. Secor SM, Diamond JM: **Evolution of regulatory responses to feeding in**
- 11 834 **snakes.** *Physiological and Biochemical Zoology* 2000, **73**:123-141.
- 12 835 5. Aubret F, Shine R, Bonnet X: **Evolutionary biology: adaptive developmental**
- 13 836 **plasticity in snakes.** *Nature* 2004, **431**:261-262.
- 14 837 6. Cohn MJ, Tickle C: **Developmental basis of limblessness and axial patterning**
- 15 838 **in snakes.** *Nature* 1999, **399**:474-479.
- 16 839 7. Ott BD, Secor SM: **Adaptive regulation of digestive performance in the genus**
- 17 840 ***Python*.** *J Exp Biol* 2007, **210**:340-356.
- 18 841 8. Secor SM: **Digestive physiology of the Burmese python: broad regulation of**
- 19 842 **integrated performance.** *J Exp Biol* 2008, **211**:3767-3774.
- 20 843 9. Zaar M, Overgaard J, Gesser H, Wang T: **Contractile properties of the**
- 21 844 **functionally divided python heart: two sides of the same matter.** *Comp*
- 22 845 *Biochem Physiol A Mol Integr Physiol* 2007, **146**:163-173.
- 23 846 10. Andersen JB, Rourke BC, Caiozzo VJ, Bennett AF, Hicks JW: **Physiology:**
- 24 847 **postprandial cardiac hypertrophy in pythons.** *Nature* 2005, **434**:37-38.
- 25 848 11. Vidal N, Hedges SB: **Molecular evidence for a terrestrial origin of snakes.**
- 26 849 *Proc Biol Sci* 2004, **271 Suppl 4**:S226-229.
- 27 850 12. Castoe TA, de Koning AP, Hall KT, Card DC, Schield DR, Fujita MK, Ruggiero RP,
- 28 851 Degner JF, Daza JM, Gu W, et al: **The Burmese python genome reveals the**
- 29 852 **molecular basis for extreme adaptation in snakes.** *Proc Natl Acad Sci U S A*
- 30 853 2013, **110**:20645-20650.
- 31 854 13. Andrew AL, Card DC, Ruggiero RP, Schield DR, Adams RH, Pollock DD, Secor SM,
- 32 855 Castoe TA: **Rapid changes in gene expression direct rapid shifts in intestinal**
- 33 856 **form and function in the Burmese python after feeding.** *Physiol Genomics*
- 34 857 2015, **47**:147-157.
- 35 858 14. Castoe TA, Fox SE, Jason de Koning A, Poole AW, Daza JM, Smith EN, Mockler TC,
- 36 859 Secor SM, Pollock DD: **A multi-organ transcriptome resource for the**
- 37 860 **Burmese Python (*Python molurus bivittatus*).** *BMC Res Notes* 2011, **4**:310.
- 38 861 15. Wang Z, Gerstein M, Snyder M: **RNA-Seq: a revolutionary tool for**
- 39 862 **transcriptomics.** *Nat Rev Genet* 2009, **10**:57-63.
- 40 863 16. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B: **Mapping and**
- 41 864 **quantifying mammalian transcriptomes by RNA-Seq.** *Nat Methods* 2008,
- 42 865 **5**:621-628.
- 43 866 17. Wang T, Hung CC, Randall DJ: **The comparative physiology of food**
- 44 867 **deprivation: from feast to famine.** *Annu Rev Physiol* 2006, **68**:223-251.
- 45 868 18. Zerbino DR, Birney E: **Velvet: algorithms for de novo short read assembly**
- 46 869 **using de Bruijn graphs.** *Genome Res* 2008, **18**:821-829.
- 47 870 19. Bushmanova E, Antipov D, Lapidus A, Suvorov V, Prjibelski AD: **rnaQUAST: a**
- 48 871 **quality assessment tool for de novo transcriptome assemblies.**
- 49 872 *Bioinformatics* 2016, **32**:2210-2212.
- 50
- 51
- 52
- 53
- 54
- 55
- 56
- 57
- 58
- 59
- 60
- 61
- 62
- 63
- 64
- 65

- 873 20. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM: **BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs.** *Bioinformatics* 2015:btv351.
- 1 874
- 2 875
- 3 876 21. Pruitt KD, Tatusova T, Maglott DR: **NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins.** *Nucleic Acids Res* 2007, **35**:D61-65.
- 4 877
- 5 878
- 6 879 22. Benton MJ: **Phylogeny of the major tetrapod groups: morphological data and divergence dates.** *J Mol Evol* 1990, **30**:409-424.
- 7 879
- 8 880
- 9 881 23. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M: **Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research.** *Bioinformatics* 2005, **21**:3674-3676.
- 10 881
- 11 882
- 12 883
- 13 884 24. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, et al: **InterPro: the integrative protein signature database.** *Nucleic Acids Res* 2009, **37**:D211-215.
- 14 884
- 15 885
- 16 886
- 17 887 25. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L: **Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation.** *Nat Biotechnol* 2010, **28**:511-515.
- 18 888
- 19 889
- 20 890
- 21 891 26. Fan HP, Di Liao C, Fu BY, Lam LC, Tang NL: **Interindividual and interethnic variation in genomewide gene expression: insights into the biological variation of gene expression and clinical implications.** *Clin Chem* 2009, **55**:774-785.
- 22 891
- 23 892
- 24 893
- 25 894
- 26 895 27. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M: **KEGG for integration and interpretation of large-scale molecular data sets.** *Nucleic Acids Res* 2012, **40**:D109-114.
- 27 895
- 28 896
- 29 897
- 30 898 28. Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes.** *Nucleic Acids Res* 2000, **28**:27-30.
- 31 898
- 32 899
- 33 900 29. Kageyama T: **Pepsinogens, progastricsins, and prochymosins: structure, function, evolution, and development.** *Cell Mol Life Sci* 2002, **59**:288-306.
- 34 900
- 35 901
- 36 902 30. Hayashi K, Agata K, Mochii M, Yasugi S, Eguchi G, Mizuno T: **Molecular cloning and the nucleotide sequence of cDNA for embryonic chicken pepsinogen: phylogenetic relationship with prochymosin.** *J Biochem* 1988, **103**:290-296.
- 37 903
- 38 904
- 39 905 31. Watanuki K, Yasugi S: **Analysis of transcription regulatory regions of embryonic chicken pepsinogen (ECPg) gene.** *Dev Dyn* 2003, **228**:51-58.
- 40 905
- 41 906
- 42 907 32. Aoki N, Deshimaru M, Terada S: **Active fragments of the antihemorrhagic protein HSF from serum of habu (*Trimeresurus flavoviridis*).** *Toxicon* 2007, **49**:653-662.
- 43 908
- 44 909
- 45 910 33. Perales J, Neves-Ferreira AG, Valente RH, Domont GB: **Natural inhibitors of snake venom hemorrhagic metalloproteinases.** *Toxicon* 2005, **45**:1013-1020.
- 46 910
- 47 911
- 48 912 34. Tomihara Y, Yonaha K, Nozaki M, Yamakawa M, Kawamura Y, Kamura T, Toyama S: **Purification of an antihemorrhagic factor from the serum of the non-venomous snake *Dinodon semicarinatus*.** *Toxicon* 1988, **26**:420-423.
- 49 913
- 50 914
- 51 915 35. Lee HJ, Sorimachi H, Jeong SY, Ishiura S, Suzuki K: **Molecular cloning and characterization of a novel tissue-specific calpain predominantly expressed in the digestive tract.** *Biol Chem* 1998, **379**:175-183.
- 52 915
- 53 916
- 54 917
- 55 918 36. Secor SM: **Gastric function and its contribution to the postprandial metabolic response of the Burmese python *Python molurus*.** *Journal of Experimental Biology* 2003, **206**:1621-1630.
- 56 918
- 57 919
- 58 920
- 59
- 60
- 61
- 62
- 63
- 64
- 65

- 921 37. Bessler SM, Secor SM: **Effects of feeding on luminal pH and morphology of**
1 922 **the gastroesophageal junction of snakes.** *Zoology (Jena)* 2012, **115**:319-329.
- 2 923 38. Helmstetter C, Pope RK, T'Flachebba M, Secor SM, Lignot J-H: **The effects of**
3 924 **feeding on cell morphology and proliferation of the gastrointestinal tract of**
4 925 **juvenile Burmese pythons (Python molurus).** *Canadian Journal of Zoology*
5 926 2009, **87**:1255-1267.
- 7 927 39. Enok S, Simonsen LS, Wang T: **The contribution of gastric digestion and**
8 928 **ingestion of amino acids on the postprandial rise in oxygen consumption,**
9 929 **heart rate and growth of visceral organs in pythons.** *Comp Biochem Physiol A*
10 930 *Mol Integr Physiol* 2013, **165**:46-53.
- 12 931 40. Nørsgaard S, Andreassen K, Malte CL, Enok S, Wang T: **Low cost of gastric acid**
13 932 **secretion during digestion in ball pythons.** *Comparative Biochemistry and*
14 933 *Physiology Part A: Molecular & Integrative Physiology* 2016, **194**:62-66.
- 16 934 41. Secor SM, Taylor JR, Grosell M: **Selected regulation of gastrointestinal acid-**
17 935 **base secretion and tissue metabolism for the diamondback water snake**
18 936 **and Burmese python.** *J Exp Biol* 2012, **215**:185-196.
- 20 937 42. Andrade DV, De Toledo LF, Abe AS, Wang T: **Ventilatory compensation of the**
21 938 **alkaline tide during digestion in the snake Boa constrictor.** *J Exp Biol* 2004,
22 939 **207**:1379-1385.
- 23 940 43. Koelz HR: **Gastric acid in vertebrates.** *Scand J Gastroenterol Suppl* 1992, **193**:2-
24 941 6.
- 26 942 44. Starck JM, Beese K: **Structural flexibility of the intestine of Burmese python**
27 943 **in response to feeding.** *Journal of Experimental Biology* 2001, **204**:325-335.
- 28 944 45. Holmberg A, Kaim J, Persson A, Jensen J, Wang T, Holmgren S: **Effects of**
29 945 **digestive status on the reptilian gut.** *Comparative Biochemistry and Physiology*
30 946 *a-Molecular and Integrative Physiology* 2002, **133**:499-518.
- 32 947 46. Lignot JH, Helmstetter C, Secor SM: **Postprandial morphological response of**
33 948 **the intestinal epithelium of the Burmese python (Python molurus).** *Comp*
34 949 *Biochem Physiol A Mol Integr Physiol* 2005, **141**:280-291.
- 36 950 47. Secor SM, Whang EE, Lane JS, Ashley SW, Diamond J: **Luminal and systemic**
37 951 **signals trigger intestinal adaptation in the juvenile python.** *American Journal*
38 952 *of Physiology-Gastrointestinal and Liver Physiology* 2000, **279**:G1177-G1187.
- 39 953 48. Overgaard J, Andersen JB, Wang T: **The effects of fasting duration on the**
40 954 **metabolic response to feeding in Python molurus: an evaluation of the**
41 955 **energetic costs associated with gastrointestinal growth and upregulation.**
42 956 *Physiol Biochem Zool* 2002, **75**:360-368.
- 44 957 49. Secor SM, White SE: **Prioritizing blood flow: cardiovascular performance in**
45 958 **response to the competing demands of locomotion and digestion for the**
46 959 **Burmese python, Python molurus.** *J Exp Biol* 2010, **213**:78-88.
- 48 960 50. Secor SM, Hicks JW, Bennett AF: **Ventilatory and cardiovascular responses of**
49 961 **a python (Python molurus) to exercise and digestion.** *J Exp Biol* 2000,
50 962 **203**:2447-2454.
- 52 963 51. Skovgaard N, Møller K, Gesser H, Wang T: **Histamine induces postprandial**
53 964 **tachycardia through a direct effect on cardiac H2-receptors in pythons.**
54 965 *American Journal of Physiology - Regulatory, Integrative and Comparative*
55 966 *Physiology* 2009, **296**:R774-R785.
- 57 967 52. Enok S, Simonsen LS, Pedersen SV, Wang T, Skovgaard N: **Humoral regulation**
58 968 **of heart rate during digestion in pythons (Python molurus and Python**
59 969 **regius).** *Am J Physiol Regul Integr Comp Physiol* 2012, **302**:R1176-1183.

- 970 53. Riquelme CA, Magida JA, Harrison BC, Wall CE, Marr TG, Secor SM, Leinwand LA:
1 971 **Fatty acids identified in the Burmese python promote beneficial cardiac**
2 972 **growth.** *Science* 2011, **334**:528-531.
- 3 973 54. Enok S, Leite G, Leite C, Gesser H, Hedrick MS, Wang T: **Improved cardiac filling**
4 974 **facilitates the postprandial elevation of stroke volume in Python regius.** *J*
5 975 *Exp Biol* 2016.
- 6 976 55. Slay CE, Enok S, Hicks JW, Wang T: **Reduction of blood oxygen levels enhances**
7 977 **postprandial cardiac hypertrophy in Burmese python (Python bivittatus).** *J*
8 978 *Exp Biol* 2014, **217**:1784-1789.
- 9 979 56. Jensen B, Larsen CK, Nielsen JM, Simonsen LS, Wang T: **Change of cardiac**
10 980 **function, but not form, in postprandial pythons.** *Comp Biochem Physiol A Mol*
11 981 *Integr Physiol* 2011, **160**:35-42.
- 12 982 57. Butler MW, Lutz TJ, Fokidis HB, Stahlschmidt ZR: **Eating increases oxidative**
13 983 **damage in a reptile.** *J Exp Biol* 2016, **219**:1969-1973.
- 14 984 58. Schulz MH, Zerbino DR, Vingron M, Birney E: **Oases: robust de novo RNA-seq**
15 985 **assembly across the dynamic range of expression levels.** *Bioinformatics*
16 986 2012, **28**:1086-1092.
- 17 987 59. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL:
18 988 **BLAST+: architecture and applications.** *BMC Bioinformatics* 2009, **10**:421.
- 19 989 60. Eddy SR: **A probabilistic model of local sequence alignment that simplifies**
20 990 **statistical significance estimation.** *PLoS Comput Biol* 2008, **4**:e1000069.
- 21 991 61. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment**
22 992 **search tool.** *J Mol Biol* 1990, **215**:403-410.
- 23 993 62. Langmead B, Salzberg SL: **Fast gapped-read alignment with Bowtie 2.** *Nat*
24 994 *Methods* 2012, **9**:357-359.
- 25 995 63. Li B, Dewey CN: **RSEM: accurate transcript quantification from RNA-Seq data**
26 996 **with or without a reference genome.** *BMC Bioinformatics* 2011, **12**:323.
- 27 997 64. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, Kong L, Gao G, Li CY, Wei L: **KOBAS**
28 998 **2.0: a web server for annotation and identification of enriched pathways**
29 999 **and diseases.** *Nucleic Acids Res* 2011, **39**:W316-322.
- 30 1000 65. Perkins DN, Pappin DJC, Creasy DM, Cottrell JS: **Probability-based protein**
31 1001 **identification by searching sequence databases using mass spectrometry**
32 1002 **data.** *Electrophoresis* 1999, **20**:3551-3567.
- 33 1003 66. Dyrlund TF, Poulsen ET, Scavenius C, Sanggaard KW, Enghild JJ: **MS Data Miner:**
34 1004 **A web-based software tool to analyze, compare, and share mass**
35 1005 **spectrometry protein identifications.** *Proteomics* 2012, **12**:2792-2796.
- 36 1006 67. Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, Mann M:
37 1007 **Exponentially modified protein abundance index (emPAI) for estimation of**
38 1008 **absolute protein amount in proteomics by the number of sequenced**
39 1009 **peptides per protein.** *Mol Cell Proteomics* 2005, **4**:1265-1272.
- 40 1010

| Parameter | <i>De novo</i> 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 transcript 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 |

| 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 |

| 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%) |

Figure 1

[Click here to download Figure Figure1.pdf](#)

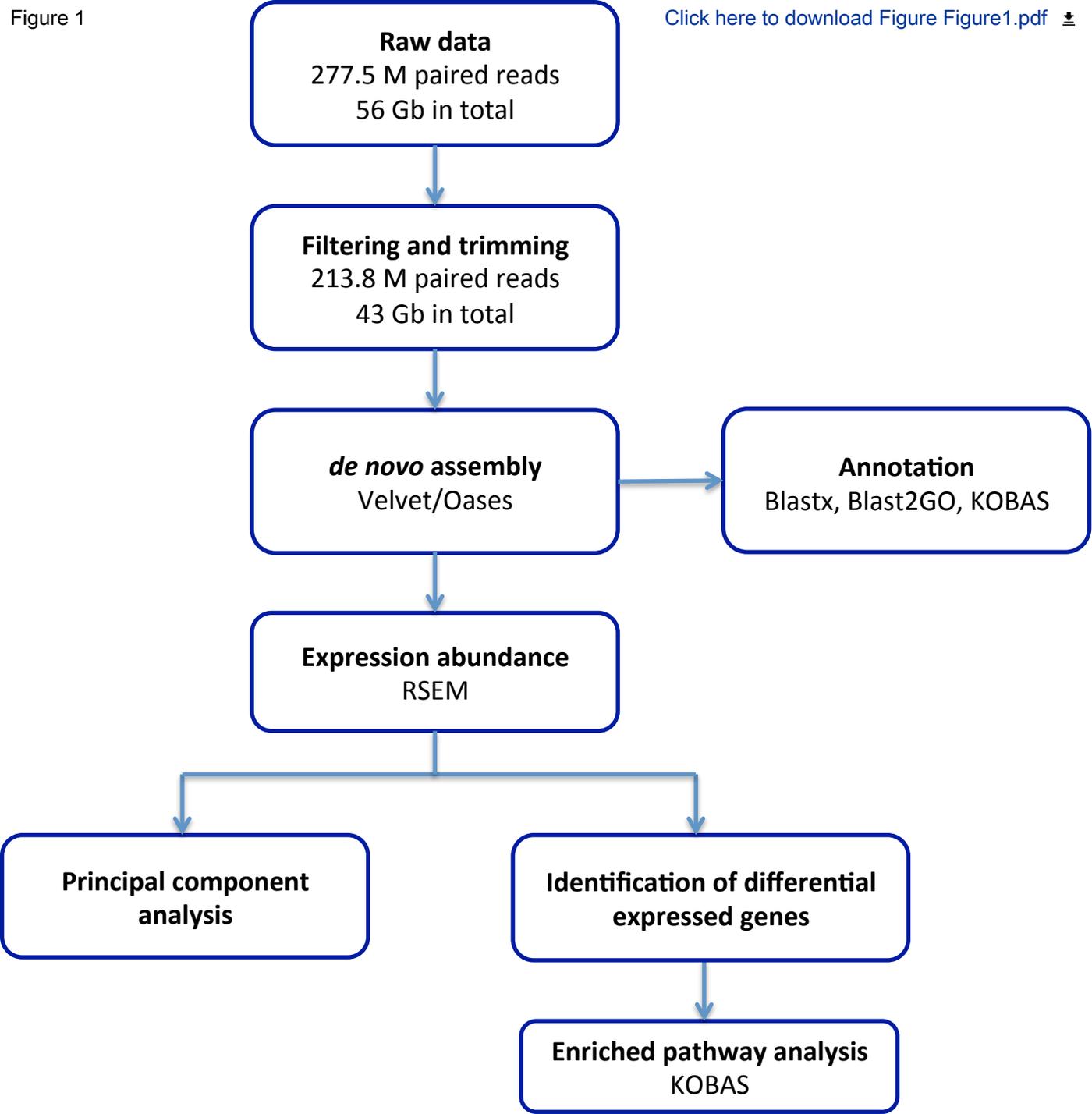


Figure 2

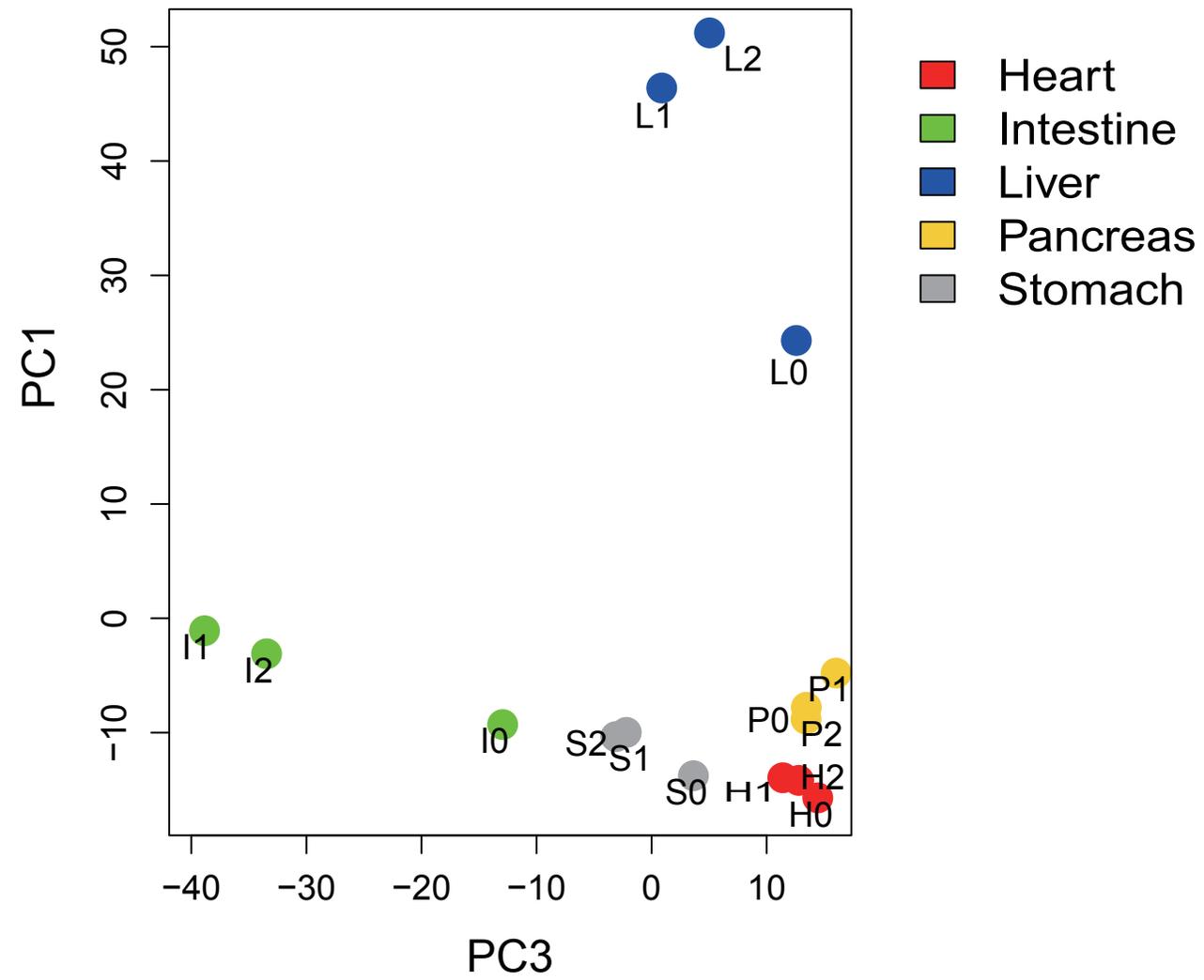
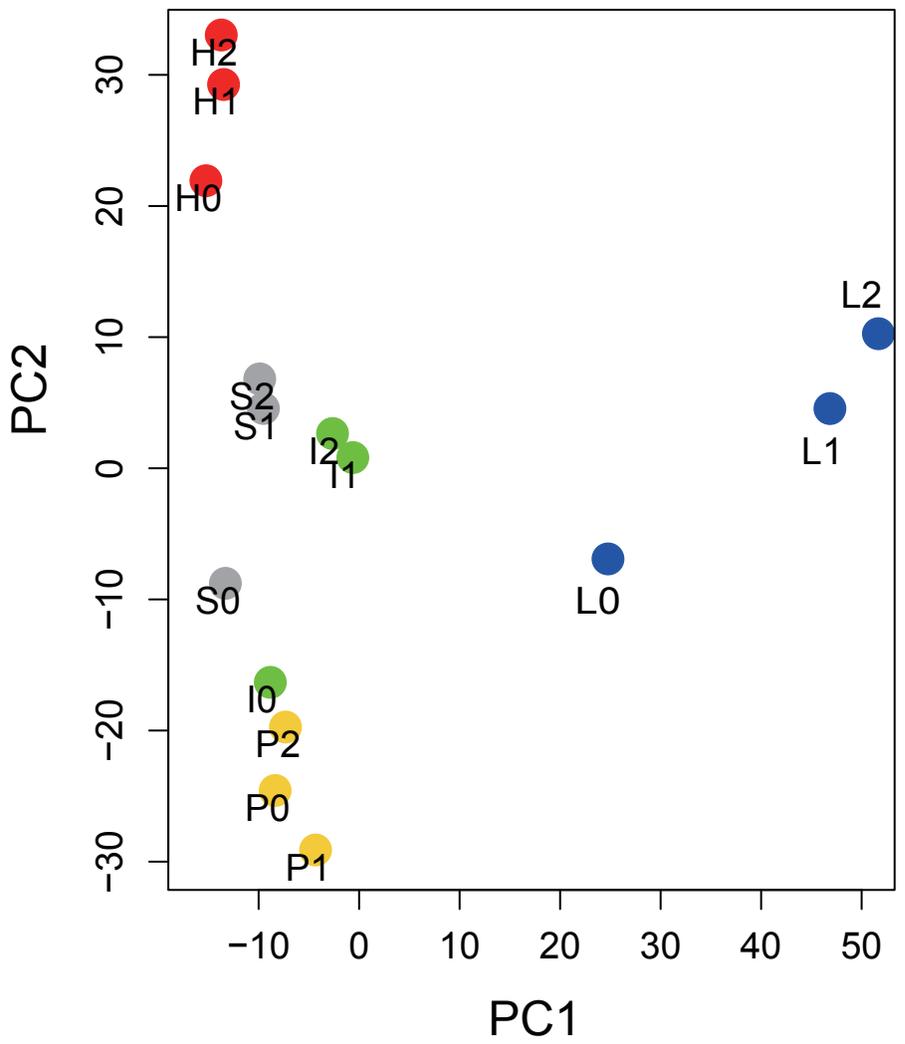


Figure 3

[Click here to download Figure Figure3.pdf](#)

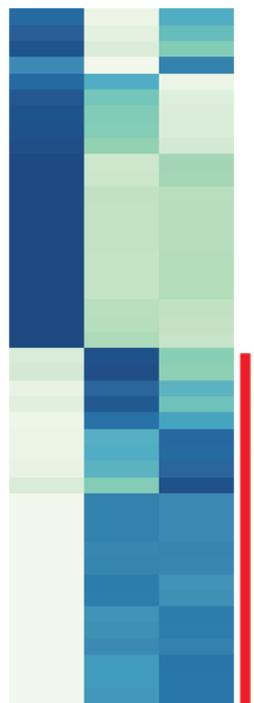
Heart

Intestine

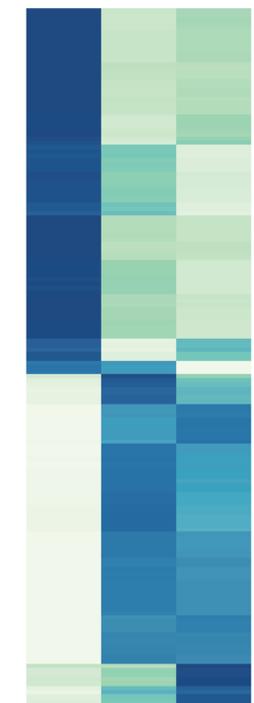
Liver

Pancreas

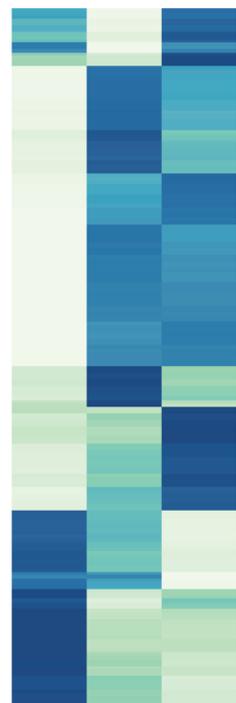
Stomach



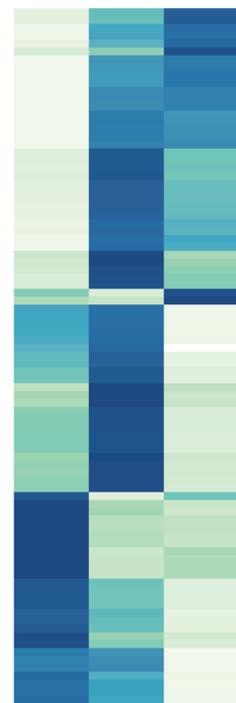
a



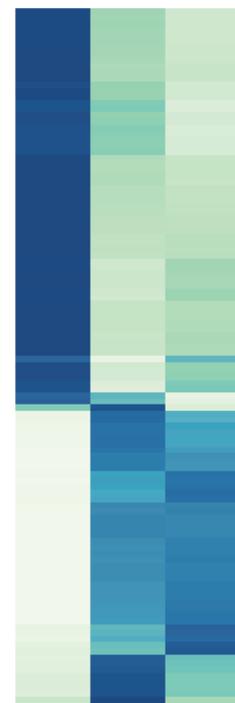
b



c

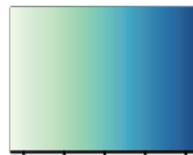


d



e

Color Key



-1 0 1

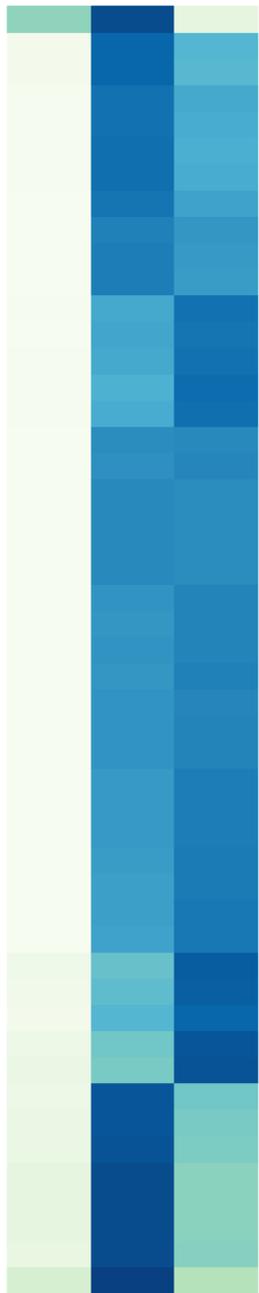
0 1 2

days after feeding

Figure 4

[Click here to download Figure4.pdf](#)

Stomach



- 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

[Click here to download Figure5.pdf](#)

Intestine

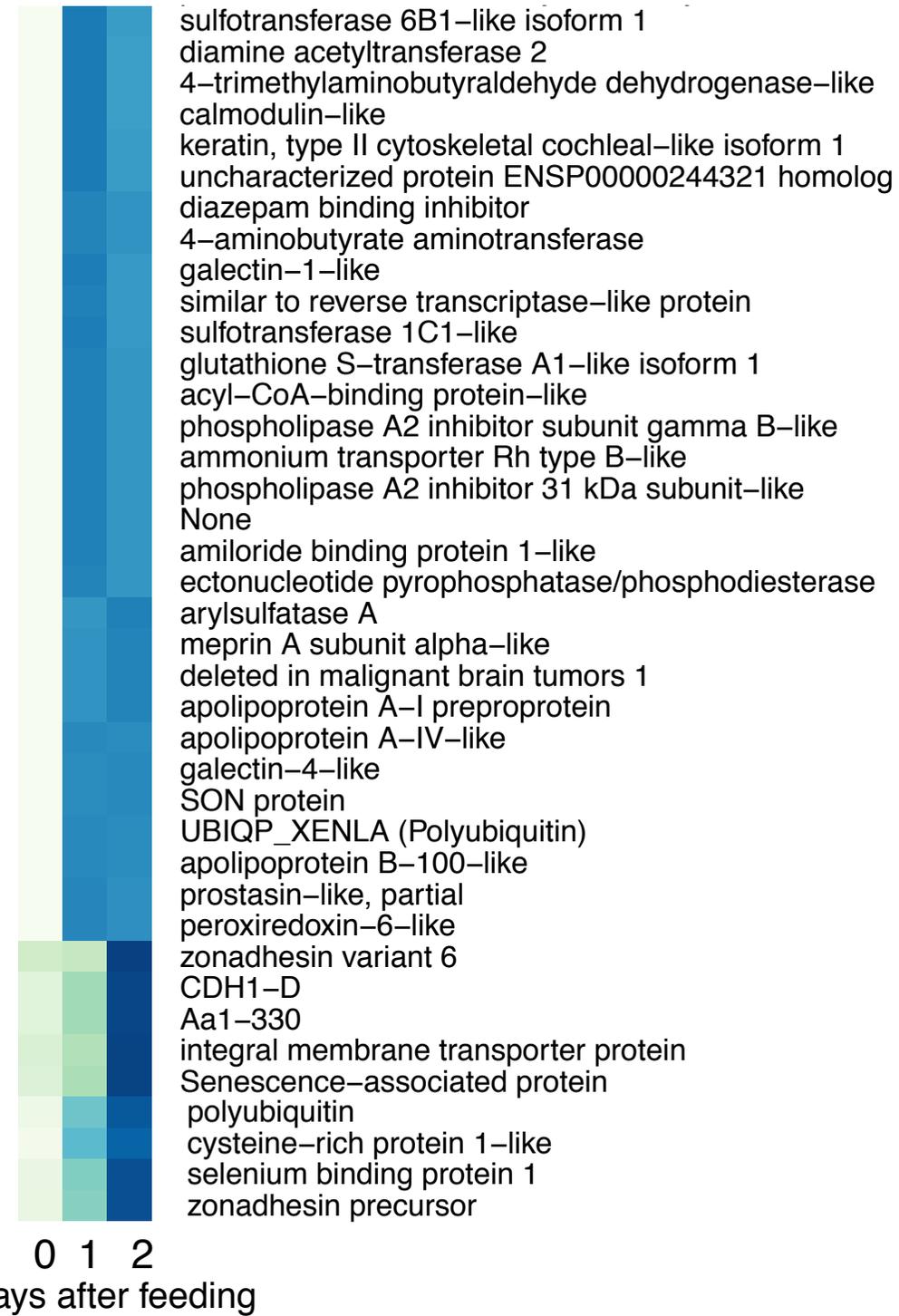
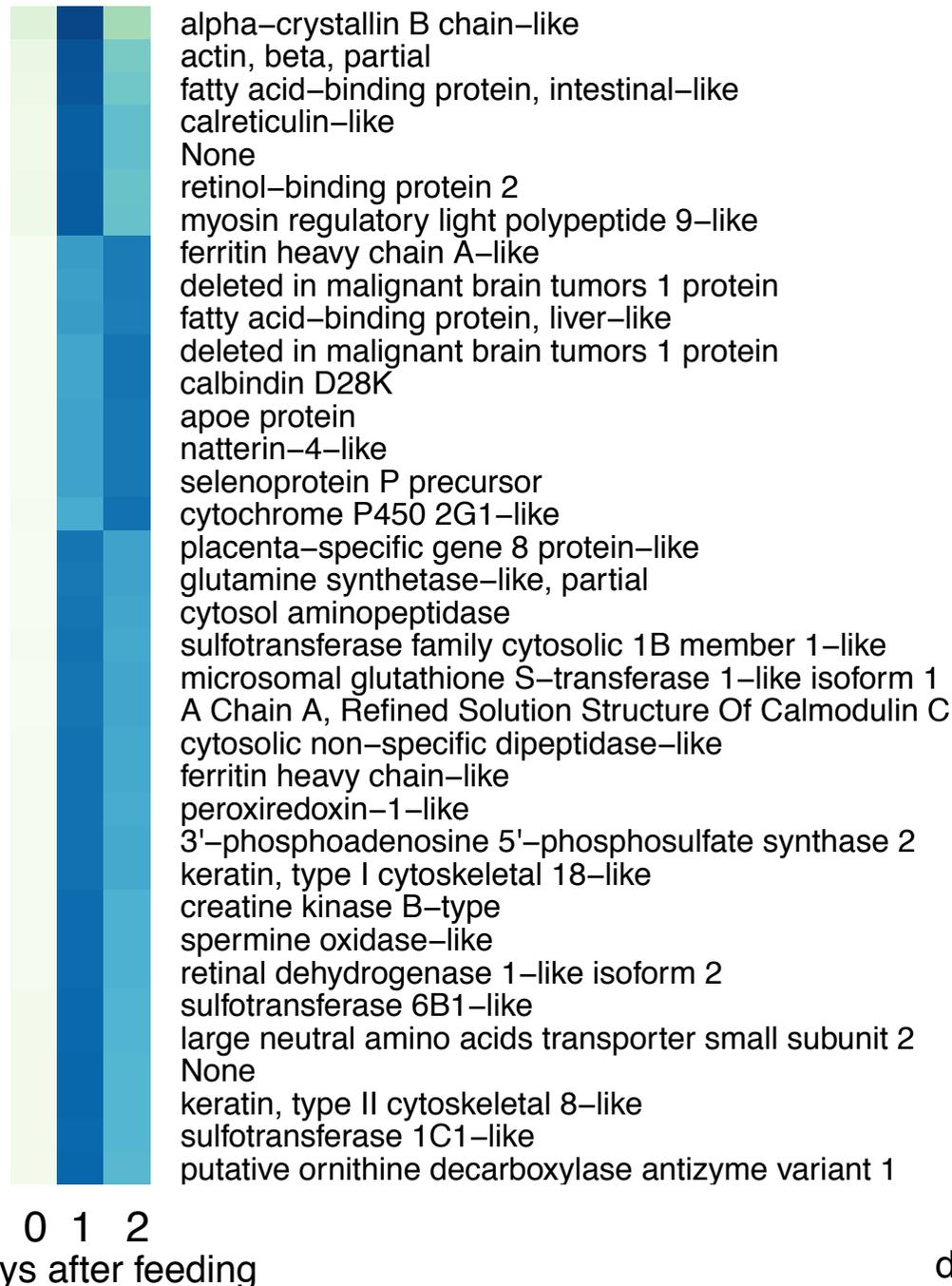
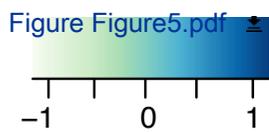
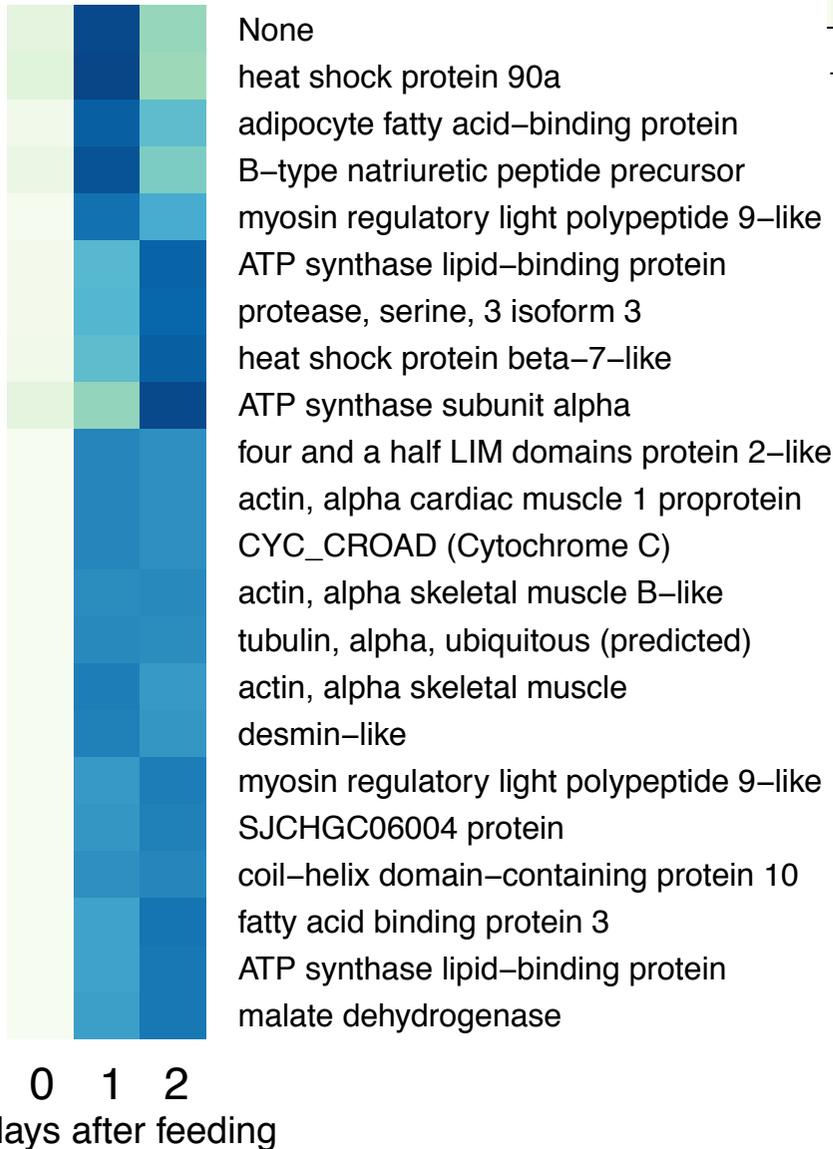
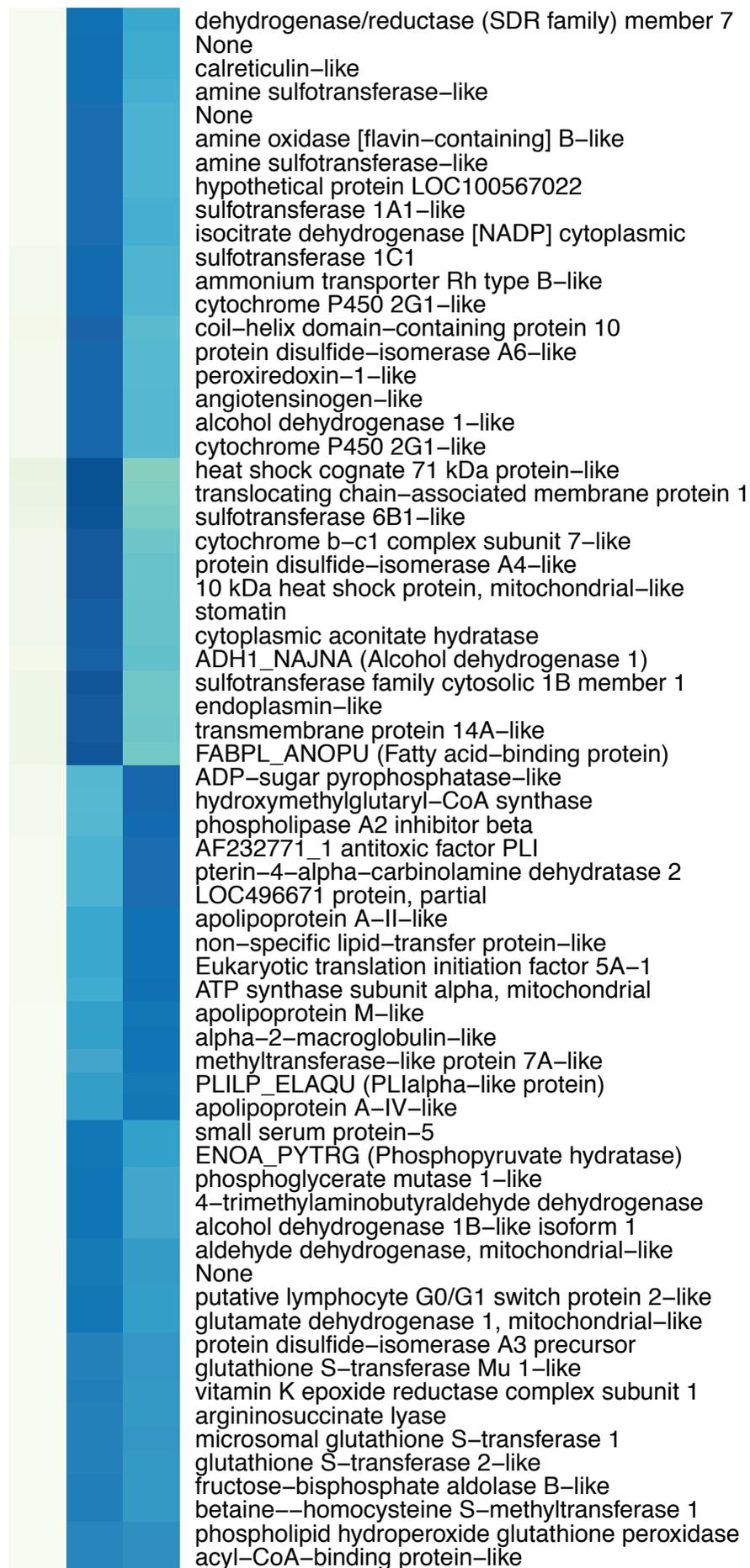
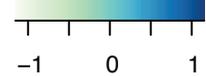


Figure 6

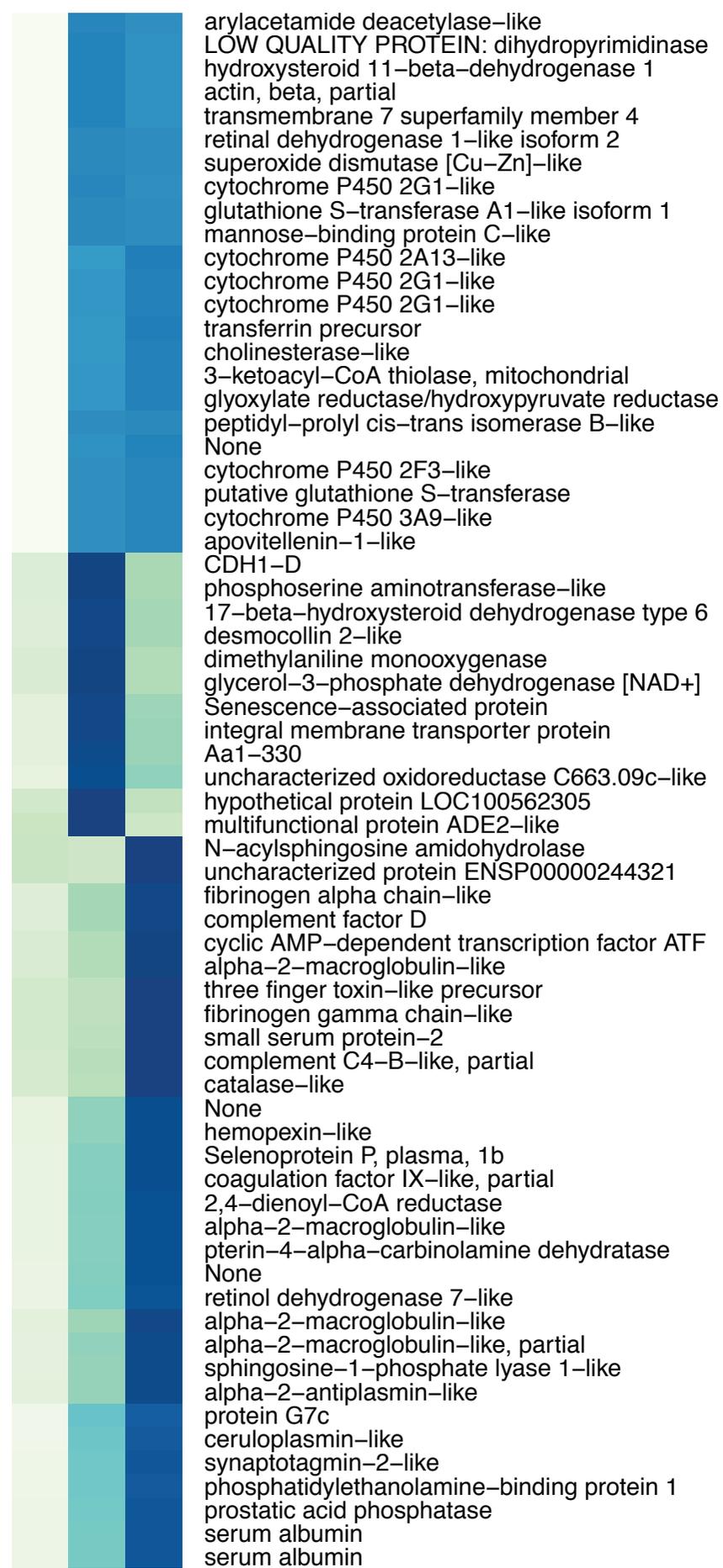
Heart





0 1 2

days after feeding

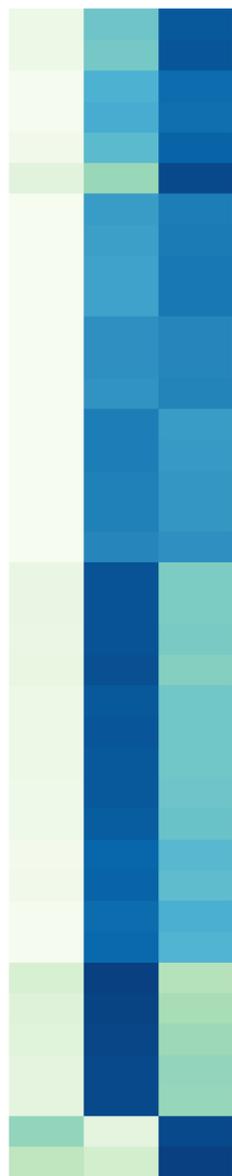


0 1 2

days after feeding

Pancreas

-1 0 1



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



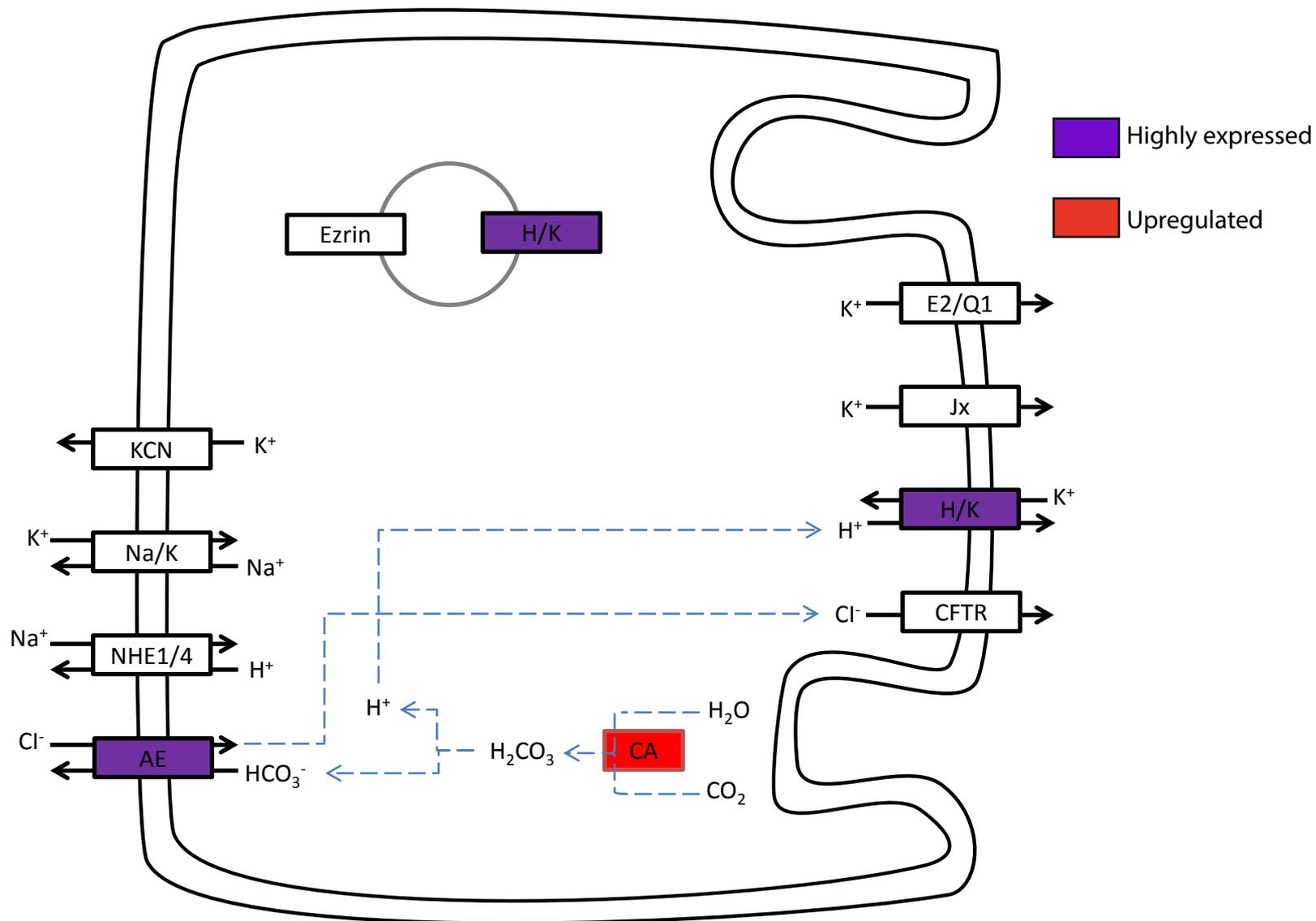
3. LC-MS/MS analyses



4. Protein identification

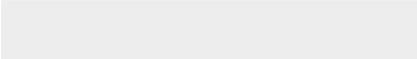
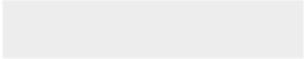


5. Identification of the python stomach secretome





Click here to access/download
Supplementary Material
FigureS4.pdf





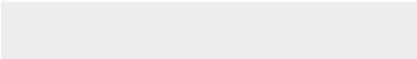
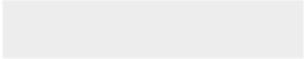
Click here to access/download
Supplementary Material
FigureS5.pdf



Click here to access/download
Supplementary Material
FigureS6.pdf



Click here to access/download
Supplementary Material
FigureS7.pdf

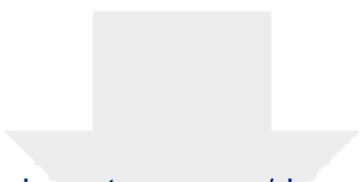




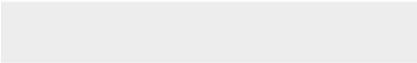
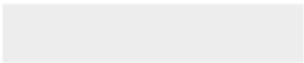
Click here to access/download
Supplementary Material
FigureS8.pdf

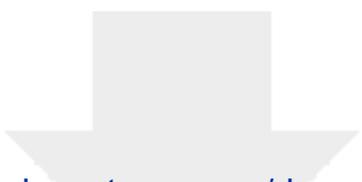


Click here to access/download
Supplementary Material
FigureS9.pdf

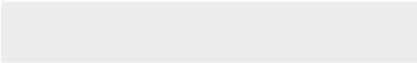
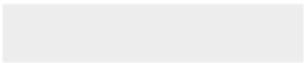


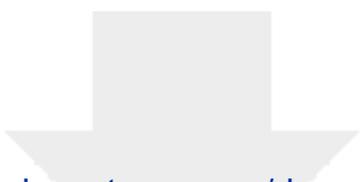
Click here to access/download
Supplementary Material
TableS3.xlsx





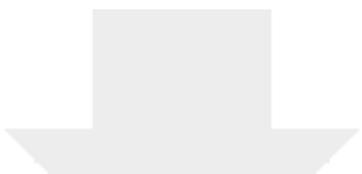
Click here to access/download
Supplementary Material
TableS5.xlsx



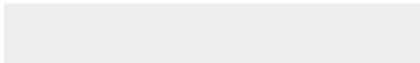


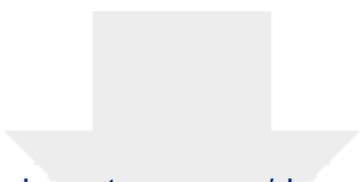
Click here to access/download
Supplementary Material
TableS6.xlsx





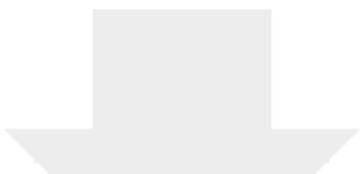
Click here to access/download
Supplementary Material
TableS7.xlsx



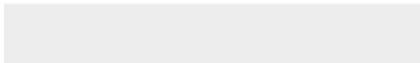


Click here to access/download
Supplementary Material
TableS8.xlsx



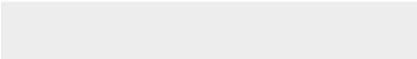
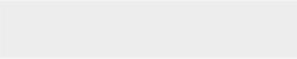


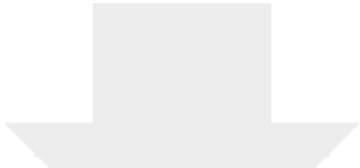
Click here to access/download
Supplementary Material
TableS9.xlsx



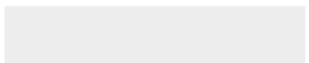


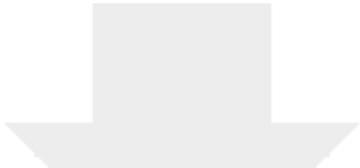
Click here to access/download
Supplementary Material
TableS10.xlsx





Click here to access/download
Supplementary Material
TableS11.xlsx



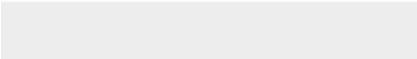
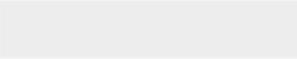


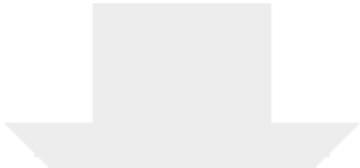
Click here to access/download
Supplementary Material
TableS12.xlsx



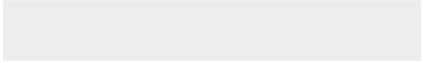
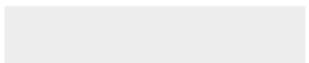


Click here to access/download
Supplementary Material
TableS13.xlsx





Click here to access/download
Supplementary Material
TableS14.xlsx





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





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

discordance between transcriptome and proteome in our study, which may be due to delayed protein synthesis and degradation. However, the imbalance didn't affect albumin-like transcripts because these six albumin-like transcripts were the six most highly expressed albumin-like transcripts in liver.

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.

We apologize for not clarifying the method in where we originally described. We have added a description to explain the method used on protein quantitation (line 710): "Semi-quantitative proteomics data was obtained using the emPAI-values given by the Mascot 2.5.0 software after analysis of the MS/MS data [67]."

Lines 201, 208: carolinensis should not be capitalized

Thank you for spotting this mistake. It has now been corrected.

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.

For the gene expressions in the intestine, heart and liver where previous data exist (Castoe et al., and Andrew et al), we have added a paragraph in the discussion (second paragraph of the new discussion) describing the overlap of upregulated genes in our and the previous studies. Information on the methods of comparison are now described in the supplementary material. It is noteworthy that the data from the liver was rather similar between studies, whereas the heart and small intestine revealed rather large differences between the studies.

Figures and Tables

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

We have added the missing information on the color-coding in the KEGG pathway that illustrates the gastric acid secretion (figure 10).

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-venomation.

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.