De novo transcriptome assembly of the African bullfrog *Pyxicephalus adspersus* for molecular analysis of aestivation

Naoki Yoshida[1*], Chikara Kaito[1,2]

[1]Laboratory of Immunology and Microbiology, Graduate School of Pharmaceutical Sciences, The University of Tokyo

[2]Graduate School of Medicine, Dentistry, and Pharmaceutical Sciences, Okayama University

Running Head: *De novo* transcriptome assembly of the African bullfrog

*Corresponding author

E-mail: kamerou2005@yahoo.co.jp

Abstract

The molecular mechanisms of aestivation, a state of dormancy that occurs under dry conditions at ordinary temperature, are not yet clarified. Here, we report the first *de novo* transcriptome assembly of the African bullfrog *Pyxicephalus adspersus*, which aestivates for 6-10 months during the hot and dry season. Polyadenylated RNA from tissues was sequenced to 75,320,390 paired-end reads, and the *de novo* assembly generated 101,682 transcripts. Of these, 100,093 transcripts had open reading frames encoding more than 25 amino acids. BLASTx analysis against the Uniprot *Xenopus tropicalis* protein database revealed 64,963 transcripts having little homology with an E value higher than 1E-5 and 8147 transcripts having no homology, indicating that the African bullfrog has many novel genes that are absent in *X. tropicalis*. The other 28,570 transcripts had homology with an E value lower than 1E-5 for which molecular functions were estimated by gene ontology (GO) analysis and found to contain the aestivation-related genes conserved among other aestivating organisms, including the African lungfish. This study is the first to identify a comprehensive set of genes expressed in the African bullfrog, thus providing basic information for molecular level analysis of aestivation.

Introduction

Dormancy is a phenomenon by which organisms temporarily decrease their physiologic activity under conditions such as low temperature, drying, and depletion of energy sources, and is observed in many organisms from microorganisms to mammals. In vertebrates, hibernation and aestivation are two common types of dormancies performed under either a low temperature condition or a drying and ordinary temperature condition, respectively. Molecular biologic and physiologic analyses of vertebrate hibernation have been performed for the Siberian chipmunk and Northern leopard frog. In Siberian chipmunks, hibernation is initiated by two cues: a seasonal change in the photoperiod depending on the melatonin level, and a hibernation-specific protein complex induced by cold temperature [1,2]. Energy-related genes are regulated to decrease body temperature and to suppress the contractile activity of cardiac myocytes. In the Northern leopard frog, glucose and glycerol concentrations in the body are increased to confer freeze tolerance, and the mammalian target of rapamycin (mTOR) is activated to prevent muscle atrophy during hibernation [3,4,5]. These molecular and physiologic aspects of hibernation have attracted attention as the potential basis for novel therapeutics against malignant tumors or tissue necrosis.

Aestivation in vertebrates is observed in many amphibian and fish species. Aestivation is similar to hibernation in terms of the reduced metabolic activity, but it is clearly distinguished from hibernation in terms of the environmental conditions; aestivation is performed under dry conditions at ordinary temperatures, whereas hibernation is performed under aqueous conditions at low temperatures [6]. Therefore, the mechanisms of aestivation are assumed to differ from those of hibernation, such as

56 suppression of energy metabolism at an ordinary temperature or tolerance against
57 dehydration. Molecular level analyses of aestivation are important to uncover
58 mechanisms of dormancy that have not been clarified in studies of hibernation, but such
59 studies are scarce at present.

60 So far, three species that undergo aestivation – the Green striped burrowing frog
61 (*Cyclorana alboguttata*), Couch's spadefoot toad (*Scaphiopus couchii*), and African
62 lungfish (*Protopterus annectens*) – have been analyzed at the transcriptional level to
63 identify genes whose expression is induced or suppressed during aestivation [7,8,9].
64 Very little is known about the molecular mechanisms of aestivation, however, including
65 how energy metabolism is suppressed or how tissue damage is minimized in the
66 aestivating organisms.

67 In the present study, we aimed to identify genes expressed in the African bullfrog,
68 *P. adspersus* using *de novo* transcriptome assembly. The African bullfrog, a species
69 belonging to the family of *Pyxicephalidae*, is a large frog – body size of males is larger
70 than 20 cm and that of females is ~14 cm. This frog inhabits a savanna area of east
71 Africa, south Africa, and the southern part of central Africa, where the air temperature
72 ranges from 20-30˚C throughout the year, and there are pronounced dry and rainy
73 seasons. During the dry season, the frogs burrow underground and form a tough cocoon
74 to reduce evaporative water loss and decrease the respiration rate to less than 10% that
75 during the active stage [10,11]. The aestivation of the frog continues for 6-10 months.
76 Because aestivation of the frog can be artificially induced and breeding methods are
77 established for breeding the frogs as pets, the African bullfrog is a suitable research
78 model animal to investigate the molecular mechanisms of aestivation. Furthermore, the
79 African bullfrog is phylogenetically and geographically distant from the other two

aestivating frogs: the Green striped burrowing frog and Couch's spadefoot toad, which inhabit Australia and North America, respectively. Therefore, studies of the aestivation of the African bullfrog will help to elucidate the molecular mechanisms of aestivation that are conserved among different aestivating frog species. In this study, we identified and characterized the genes of the African bullfrog to provide basic information for investigating the aestivation mechanism.

Materials and Methods

Ethics statement

All protocols and procedures conformed to the Regulations for Animal Care and Use of the University of Tokyo. No formal approval is required by the University of Tokyo to perform experiments using amphibians.

Experimental animal

A captive bred African bullfrog was purchased from a specialty reptile and amphibian store (Hachurui Club, Nakano, Japan). The purchased frog was maintained in a plastic container with coarse sand (5-7 mm diameter) and water. The frog was fed every day with house crickets (*Acheta domestica*) or wax worms (the larvae of the *Galleria mellonella*) for the first 3 weeks, and then fed every day with artificial diets (Samuraijapan, Ibaraki, Japan) for 2 months.

102 ## Isolation of genomic DNA and total RNA

103 The young adult frog (15 g body mass) was kept without feeding for 36 h. The frog was
104 anesthetized by placing it into crushed ice for 10 min and then dissected on ice. The
105 intestines were separated, the intestinal contents removed in phosphate buffered saline
106 (pH7.4), and the intestines frozen in liquid nitrogen. The other tissues, including inner
107 organs, muscle, and skin, were quickly excised to 3 - 5 mm^3 (0.1 - 0.3 g) and frozen in
108 liquid nitrogen. The frozen samples were maintained at -80˚C.

109 Genomic DNA was isolated by a standard phenol/chloroform extraction method
110 [12]. Frozen muscle tissues (5 mm^3, 0.2 g) were submerged in a mixture of 5 ml of
111 phenol/chloroform/isoamyl alcohol (25:24:1) and 5 ml of TE Buffer (10 mM Tris-HCl
112 [pH 8.0], 1mM EDTA). Then, the tissues were homogenized at 10,000 rpm three times
113 for 30 s each with a Polytron benchtop homogenizer (Kinematica, Switzerland). The
114 sample was centrifuged at room temperature at 2300 *g* for 10 min, and the upper
115 aqueous phase (1.0 ml) was transferred to a new tube. The sample was centrifuged at
116 room temperature at 20,400 *g* for 5 min to completely remove the organic solvents. The
117 upper aqueous phase (0.7 ml) was transferred to a new tube, vortexed with an
118 equivalent volume of chloroform, and centrifuged at room temperature for 5 min at
119 20,400 *g*. The upper aqueous phase (0.5 ml) was transferred to a new tube, mixed with a
120 2.5-fold amount of ethanol and a 0.1-fold amount of 3M sodium acetate, and
121 centrifuged at 20,400 *g* for 15 min at room temperature. The precipitate was washed
122 with 70% ethanol and resolved in 50 μl of TE buffer.

123 Total RNAs were extracted from the inner organs, intestines, muscles, and skin
124 using the chaotropic extraction protocol for mouse pancreatic RNA, described by

125 Robert [13]. Frozen tissues (3 - 5 mm^3, 0.1 - 0.3 g) were submerged in 10 ml of
126 TRIZOL Reagent (Life Technologies, Carlsbad, CA, USA), an amount three times
127 larger than that recommended by the supplier. The tissues were then homogenized at
128 14,000 rpm three times for 30 s each using the Polytron homogenizer. After incubation
129 for 5 min at room temperature, 2 ml of chloroform was added and the mixture was
130 vortexed for 15 s. The mixture was incubated for 3 min at room temperature and
131 centrifuged at 12,000 *g* for 10 min at 4˚C. The upper aqueous phase (4 ml) was
132 transferred to a fresh 50-ml tube and mixed with an equivalent amount of isopropyl
133 alcohol. The sample was centrifuged at 12,000 *g* for 10 min and an RNA pellet was
134 obtained. The RNA pellet was vortexed with 10 ml of 75% ethanol and centrifuged at
135 7500 *g* for 5 min at 4˚C. The RNA pellet was air-dried and dissolved in 200 µl of
136 RNase-free water by incubating for 10 min at 55˚C. Further purification to remove
137 contaminated genomic DNA was performed using a Monarch Total RNA Miniprep Kit
138 (New England Biolabs, MA, USA), according to the supplier's protocol. The RNA was
139 eluted from a column by 100 µl of RNase-free water and kept at -80˚C. The
140 concentration and purity of the isolated RNA was determined using a spectrophotometer
141 and denaturing agarose gel electrophoresis. The concentrations of RNA isolated from
142 11 tissue segments of inner organs, intestines, muscle, skin, and head ranged from
143 0.52-3.34 µg/µl or 0.38-1.60 µg/mg tissue.
144

145 Determination of mitochondrial 16S/12S rRNA sequence

146 Primers (16SA-L-frog: 5'-CGCCTGTTTACCAAAAACAT-3', 16SB-H-frog:
147 5'-CCGGTCTGAACTCAGATCACGT-3', 12SB-F:

148 5'-AAACTGGGATTAGATACCCCACTAT-3', 12SB-H:
149 5'-GAGGGTGACGGGCGGTGTGT-3') used to amplify mitochondrial 16S or 12S
150 rRNA partial regions were identical to those used in phylogenetic classification of
151 Anura, reported by Hirota [14]. Polymerase chain reaction (PCR) was performed in 25
152 µl of 1x Q5 reaction buffer (0.5 units of Q5 High-Fidelity DNA Polymerase (New
153 England Biolabs), 200 µM dNTPs, 0.5 µM forward primer, 0.5 µM reverse primer and
154 6.5 ng genomic DNA) using the thermal cycle program (30 s at 98˚C, 30 cycles [30 s,
155 98˚C; 30 s, 62˚C; 30 s, 72˚C], and 2 min at 72˚C). PCR products were examined by
156 agarose gel electrophoresis and sequenced using the same primers.
157

158 mRNA library preparation and Illumina next-generation 159 sequencing

160 An equivalent amount of total RNAs (0.52 - 3.34 µg/µl) isolated from 11 tissue
161 segments of inner organs (3 segments), intestines (2 segments), muscles (2 segments),
162 skin (3 segments), and head (1 segment) were mixed to obtain 100 ng/µl total RNA.
163 The mixed total RNA was analyzed by Tape Station (RNA Screen tape, Agilent
164 Technologies Ltd., USA) and determined to RIN (RNA integrity number) = 9.0. The
165 poly (A)+ fraction was isolated from the total RNA, followed by its fragmentation. The
166 fragmented mRNA was then reverse-transcribed to double-stranded cDNA using
167 random hexamers. The double-stranded cDNA fragments were processed for adaptor
168 ligation, size selection for 200-bp inserts, and amplification of strand-specific cDNA
169 libraries. Prepared libraries were subjected to paired-end 2 x 100 bp sequencing on the
170 HiSeq 2500 platform with v4 chemistry.

171

172 ## *De novo* transcriptome assembly and bioinformatic analysis

173 All analyses were performed mainly using the CLC Genomics Workbench 11.0
174 (QIAGEN Bioinformatics, Germany). The 2 x 100 bp paired-end reads were checked in
175 terms of the sequencing quality and trimmed (removal of adaptor and duplication) with
176 quality score limit of 0.05 and a maximum number of two ambiguous nucleotides. The
177 clean reads were then assembled and aligned with a word size of 20, bubble size of 50,
178 and minimum transcript length of 200, based on the algorithm of the De Brujin graph.
179 The open reading frame (ORF) in the assembled *P. adspersus* transcripts was predicted
180 under the following conditions: start codon as ATG, search as Both Strand, open-ended
181 sequence, minimum length (codons) as 25, and genetic code as standard. The transcripts
182 were homology searched using local BLASTx (National Center for Biotechnology
183 Information, NCBI) against the Uniprot database (https://www.uniprot.org/)
184 (all-proteins or *Xenopus tropicalis*). Homologous proteins found in the Uniprot database
185 (*X. tropicalis*) with an E-value lower than 1E-5 were subjected to gene ontology (GO)
186 analysis (https://www.uniprot.org) [15] to assign the GO terms of biologic processes,
187 molecular functions, and cellular components.

188

189 # Results

190 ## Confirmation of animal species used in this study

191 To confirm that the experimental animal used in this study was in fact the African
192 bullfrog, a captive bred frog was raised for 5 years and the appearance of the adult frog
193 was observed. The frog was approximately 20 cm in length and had olive-green skin

194 with longitudinal skin folds (**Fig 1A**). These morphologic characteristics were
195 consistent with those of the African bullfrog reported by Channing [16] and Bishop
196 [17], and clearly differed from the characteristics of *P. edulis*, a species closely related
197 to the African bullfrog that is 12 cm in length and has olive green-yellow skin with
198 spotted folds. Furthermore, we determined the nucleotide sequence of mitochondrial
199 16S and 12S rRNA of the frog used for subsequent *de novo* assembly analysis. NCBI
200 BLAST analysis revealed that the mitochondrial 16S rRNA (583 bp, GenBank:
201 LC440402) was 100% identical to the mitochondrial 16S rRNA of *Pyxicephalus cf.*
202 *adspersus* FB-2006 isolate 0954 (551 bp, GenBank: DQ347304) (**Fig 1B**). The
203 mitochondrial 12S rRNA (424 bp, GenBank: LC440403) was 99% identical to the
204 mitochondrial 12S rRNA of *P. cf. adspersus* FB-2006 isolate 0954 (1,364 bp, GenBank:
205 DQ347013) (**S1 Fig**). These morphologic characteristics and the results of the
206 homology analysis of rRNA sequences confirmed that the frog used in this study was *P.*
207 *adspersus,* the African bullfrog.

208 The two types of frog, *P. adspersus* and *P. edulis*, are sometimes confused with
209 each other [16,18]. Top hits against the 16S and 12S rRNA sequences of the frog used
210 in this study are found in both species in no particular order. Furthermore, the
211 phylogenetic tree analysis of the top hit 16S and 12S rRNA sequences revealed that the
212 two species were not clustered (**Fig 1B and S1 Fig**). We speculate that the
213 nomenclatures of these rRNA sequences were mistakenly registered in the database.

214 Figure 1C shows the phylogenetic tree of mitochondrial 16S rRNA partial
215 sequences of the African bullfrog used in this study and of representative frog species
216 belonging to the major family of Anura. The two aestivating frogs, *C. alboguttata* and
217 *S. cauchii*, are phylogenetically distant from *P. adspersus*, as suggested by Pyron's

218 report [19].

219

220 RNA sequencing and *de novo* assembly

221 RNA sequencing and *de novo* assembly data are summarized in Table 1. RNA
222 sequencing produced 75,604,146 raw reads. *De novo* transcriptome assembly of the
223 trimmed 75,320,390 reads generated 101,682 transcripts. The transcript size ranged
224 from 236 to 35,997 bases with a mean length of 866 bases and an N50 length of 1701
225 bases. The number of transcripts ranging from 1000 to 1999 bases was highest among
226 that of transcripts ranging from 500 to 5000 bases (**Fig 2**).

227 *In silico* analysis to identify ORF-encoding proteins larger than 25 amino acids
228 revealed that 100,093 transcripts (98.4% of total transcripts) were protein-coding genes
229 (**Table 2**). The longest ORF-encoded protein was 11,679 amino acids long. The average
230 ORF-encoded protein was 140.7 amino acids long. In contrast, 1590 transcripts (1.6%
231 of total transcripts) had no ORF, suggesting that these transcripts were non-coding
232 RNAs. When the ORF detection conditions were changed to 30, 50 or 100 amino acids
233 long, the number of transcripts carrying the ORF was decreased to 97,122 (95.5%),
234 71,305 (70.1%) or 29,296 (28.8%), respectively (**Table 2**).

235 Within these *de novo* assembled transcripts, we identified that transcript #275 had
236 high similarity with cytochrome *b* on the basis of the BLASTx search. The DNA
237 sequence of transcript #275 (GenBank: LC440404) was 100% identical to *P. adspersus*
238 isolate Brak01 cytochrome *b* gene (GenBank: FJ613662) (**S2 Fig**). This result
239 confirmed that that the frog used in this study was *P. adspersus*.

240

241 **Table 1. Summary of the RNA sequence reads and the *de novo* transcriptome**
242 **assembly for the African bullfrog**

Sequencing results	Length of raw reads (bp)	100
	Total number of raw reads	75,604,146
	Total number of clean reads	75,320,390
***De novo* transcriptome assembly results**	Total number of transcripts	101,682
	Total length of all transcripts (bases)	88,093,202
	GC contents of all transcripts (%)	40.8
	N50 length of all transcripts (bases)	1,701
	Mean length	866

243

244

245 **Table 2. Detection of ORF in the African bullfrog transcripts**

	Minimum ORF length			
	25 aa	**30 aa**	**50 aa**	**100 aa**
Total transcripts				101,683
Detected ORF	729,832	562,570	219,579	45,086
Transcripts coding ORF	100,093 (98.4%)	97,122 (95.5%)	71,305 (70.1%)	29,296 (28.8%)
Maximum ORF length	11,679	11,679	11,679	11,679
Minimum ORF length	25	30	50	100
Average ORF length (aa)	140.7	144.1	181.4	341.2
Transcripts of non-coding RNA	1,590 (1.6%)	4,561 (4.6%)	30,378 (29.9%)	72,387 (71.2%)

246

247

248 ## Transcriptome annotation

249 We performed a BLASTx homology search of the African bullfrog transcripts against
250 the Uniprot protein database (all-proteins or *Xenopus tropicalis* proteins). In the
251 all-proteins database, 27,286 transcripts (26.8%) had an E-value lower than 1E-5,
252 53,064 transcripts (52.2%) had an E-value higher than 1E-5, and 21,332 transcripts
253 (21.0%) showed no hits with the database (**Fig 3A**, All proteins). In the *X. tropicalis*
254 protein database, 28,572 transcripts (28.1%) had an E-value lower than 1E-5, 64,963
255 transcripts (63.9%) had an E-value higher than 1E-5, and 8147 transcripts (8.0%)
256 showed no hits with the database (**Fig 3A**, *X. tropicalis*). Figure 3B shows the
257 distribution of transcripts with an E-value lower than 1E-5 according to the amino acid
258 identity with the protein database. 22,670 transcripts (83.1%) and 25,407 transcripts
259 (88.9%) had an amino acid identity higher than 50% in the all-proteins database or in
260 the *X. tropicalis* protein database, respectively.

261

262 ## Gene ontology analysis

263 The 28,572 transcripts with an E-value lower than 1E-5 in Uniprot *X. tropicalis* protein
264 database were subjected to gene ontology (GO) analysis and classified according to
265 three major GO categories: cellular components, molecular function, and biologic
266 processes (**Fig 4**). In the cellular component category, 11,634 (33.4%) and 7689
267 (22.0%) genes were associated with the cellular components and membrane
268 components, respectively. In the molecular function category, 11,142 (45.6%) and 6995
269 (28.6%) genes were associated with binding and catalytic activity, respectively. In the
270 biologic process category, 9765 (30.5%), 6213 (19.4%), and 5553 (17.3%) genes were

271 associated with cellular processes, biologic regulation, and metabolic processes,
272 respectively.
273

274 Comparison with the aestivation-related genes reported in 275 other organisms

276 To determine whether the African bullfrog transcripts contain genes homologous to the
277 aestivation-related genes of the two aestivating frogs and the African lungfish, we
278 performed a homology search of aestivation-related genes against the African bullfrog
279 transcripts (**Table 3**) [8,9,10]. The #2083 transcript in the African bullfrog had high
280 identity with riboflavin-binding protein precursor (RfBP) in Couch's spadefoot toad,
281 which has increased expression during aestivation and might be involved in egg
282 maturation in the frog during the aestivating stage. The African bullfrog genes had high
283 identity with genes of the Green striped burrowing frog, which are highly upregulated
284 during aestivation, such as adaptor related protein complex 3 (beta 2 subunit) that
285 functions in vesicle budding from the Golgi apparatus; an immunoglobulin heavy chain
286 that recognizes foreign antigens; and a serine/threonine protein kinase, Chk1, that
287 activates DNA repair in response to DNA damage. In addition, the African bullfrog
288 genes had high identity with African lungfish genes that are expressed during
289 aestivation, such as argininosuccinate synthetase 1 and carbamoylphosphate synthetase
290 III, which are key enzymes in the ornithine-urea cycle and function to detoxify
291 ammonia to urea; rhesus family C glycoprotein that maintains muscle function; and
292 superoxide dismutase 1, an anti-oxidant. These findings suggest that African bullfrog
293 has a set of genes that are highly expressed in other vertebrates during aestivation.

294

295 **Table 3. Presence of aestivation-related genes in the African Bullfrog**

Species	Gene	Predicted Function	Top Hit Contig of African Bullfrog	E-value	% Identity
Scaphiopus couchii (Couch's spadefoot toad)	Riboflavin binding protein precursor (AF102545)	Maturation of eggs in preparation for the explosive breeding	#2083 (ICLD012083)	8.63E-114	67
Cyclorana alboguttata (Green-striped burrowing frog)	Adaptor related protein complex 3, beta 2 subunit (contig #8496)	Budding of vesicles from the golgi membrane	#4968 (ICLD014968)	0	100
	Solute carrier family 28 member 3-like (contig #7653)	Homeostasis of endogenous nucleosides	#2599 (ICLD012599)	1.99E-45	100
	Nuclease HARBI1-like (contig #14756)	Nuclease activity	#59155 (ICLD0159155)	1.31E-122	88
	Ribosome biogenesis protein NSA2-like (contig #40)	Biogenesis of the 60S ribosomal subunit	#4109 (ICLD014109)	5.97E-103	99
	Coiled coil domain containing protein 81 (contig #15484)	Positive regulator of T-cell maturation and inflammatory function	#18527 (ICLD0118527)	0	75
	Pentraxin precursor (contig #27476)	Mediating uptake of synaptic material	#3473 (ICLD013473)	7.25E-171	90
	Immunoglobulin heavy prechain (contig #36672)	Recognization of foreign antigens and initiation of immune responses	#3055 (ICLD013055)	1.01E-118	82
	Receptor type tyrosine protein phosphatase T-like (contig #7294)	Signal transduction and cellular adhesion	#10027 (ICLD0110027)	0	99
	Immunoglobulin J chain precursor (contig #48432)	Serving to link two monomer units of either IgM or IgA	#1190 (ICLD011190)	1.82E-15	76
	UHRF1 binding protein 1-like (contig #1661)	Negative regulator of cell growth	#8912 (ICLD018912)	0	95
	Inositol 3 phosphate synthase 1 (contig #6523)	Key enzyme in myo-inositol biosynthesis pathway that catalyzes the conversion of glucose 6-phosphate to 1-myo-inositol 1-phosphate	#13219 (ICLD0113219)	0	93

	Serine/threonine protein kinase Chk1 (contig #13899)	Activation of DNA repair in response to the presence of DNA damage or unreplicated DNA	#24942 (ICLD0124942)	0	85
Protopterus annectens (African Lungfish)	Rhesus family C glycoprotein (KX583647)	Glycolysis and muscle development	#10697 (ICLD0110697)	4.05E-162	79
	Argininosuccinate synthetase 1 (JZ575533)	Ornithine-urea cycle capacity to detoxify ammonia to urea	#297 (ICLD01297)	1.72E-56	81
	Carbamoylphosphate synthetase III (JZ575539)	Ornithine-urea cycle capacity to detoxify ammonia to urea	#265 (ICLD01265)	3.70E-108	91
	Betain homocyctein-S-transferase 1 (JZ575536)	Prevent the accumulation of hepatic homocyctein	#243 (ICLD01243)	5.19E-86	94
	Superoxide dismutase 1 (JZ575606)	Anti ROX	#1087 (ICLD011087)	2.73E-62	100

tBLASTx search was performed to identify the African bullfrog contig having high identity to the genes that are up-regulated during the aestivation in the three organisms. The ID numbers in parenthesis indicate Genbank accession numbers or contig numbers (*C. alboguttata*).

299 # Discussion

300 The African bullfrog is well known as an aestivating frog and is bred all over the world.
301 Molecular biologic analysis of this frog, however, has not been performed and no
302 genetic information has been available. In this study, we performed the first *de novo*
303 transcriptome assembly and identified 101,682 transcripts. Of the 101,682 transcripts
304 identified, 64,963 had an E-value higher than 1E-5 and 8147 transcripts had no hits in a
305 BLASTx search in the *X. tropicalis* database. These results suggest that 73,110 genes,
306 71.9% of total transcripts, are the novel genes in the African bullfrog and are not present
307 in *X. tropicalis*. The high proportion of the novel genes in total transcripts is also
308 observed in other organisms whose *de novo* transcriptome assembly was recently
309 performed, such as the Caecilian (*Caecilia tentaculate*), the Sodom apple (*Calotropis*
310 *gigantea*), and the Black ant (*Formica fusca*) [20, 21, 22], and confirms that *de novo*
311 transcriptome assembly is a powerful tool to identify novel genes. The novel genes of
312 the African bullfrog might encode proteins without known domains, or non-coding
313 RNAs. With regard to the very limited genomic and transcriptomic information in
314 amphibians, the novel genes of the African bullfrog are the significant biologic
315 resources to perform molecular biologic analysis of the physiologic functions of the
316 African bullfrog, including its aestivation ability.

317 In contrast, 28,572 transcripts (29.1%) had an E-value lower than 1E-5 in the *X.*
318 *tropicalis* database, suggesting that these genes are conserved in *X. tropicalis*. These
319 genes were subjected to gene ontology analysis and their molecular functions were
320 estimated. They are assumed to have functions conserved between *P. adspersus* and *X.*
321 *tropicalis*. Furthermore, we revealed that the African bullfrog has a set of

aestivation-related genes, which has been reported in two aestivating frogs and the African lungfish. These genes presumably function in the aestivation mechanism conserved between different aestivating vertebrates. The aestivation-related genes are not specific to the aestivating organisms, but are conserved in other non-aestivating organisms, and must therefore be transcriptionally controlled by factor(s) specific to the aestivating environment or the aestivating organisms. How the conserved genes are regulated during aestivation remains an important unanswered question.

The genetic information of the African bullfrog obtained in this study enables future comparative transcriptome analysis between the active stage and the aestivating stage of the frog and contributes to uncover the molecular mechanisms of aestivation.

Data Availability

Mitochondrial 16S rRNA of the African bullfrog (583 bp): GenBank: LC440402

Mitochondrial 12S rRNA of the African bullfrog (424 bp): GenBank: LC440403

Cytochrome *b* of the African bullfrog (transcripts #275): GenBank: LC440404

Raw reads of 2 x 100 bp RNA sequence of the African bullfrog: DDBJ sequencing read archive (SRA): DRR164258, BioProject: PRJDB7457, BioSample: SAMD00143251, Experiment: DRX154877

Transcript sequences of the African bullfrog: GenBank: ICLD01000001-ICLD01101682

Supporting Information

S1 Fig. The phylogenetic tree analysis of 12S rRNA sequence between the frog used in this study and several *P. adspersus* and *P. edulis* species deposited in database. Sequences were aligned using MUSCLE algorithm implemented in MEGA X. Phylogenetic tree was constructed using neighbor-joining algorithm with bootstrap test of 1000 replicates.

S2 Fig. The phylogenetic tree analysis of cytochrome b partial sequence between the frog used in this study and several frogs deposited in database. Sequence of cytochrome *b* of the African bullfrog (GenBank: LC440404) was derived from *de novo* assembled transcript #275. Sequences were aligned using MUSCLE algorithm implemented in MEGA X. Phylogenetic tree was constructed using neighbor-joining algorithm with bootstrap test of 1000 replicates.

REFERENCES

1. Michael H, Hastings Francis JPE. Hibernation Proteins: Preparing for Life in the Freezer. Cell. 2006 125(1): 21-23. doi: 10.1016/j.cell.2006.03.019. PMID: 16615885
2. Kondo N, Sekijima T, Kondo J, Takamatsu N, Tohya K, Ohtsu T. Circannual Control of Hibernation by HP Complex in the Brain. Cell. 2006 125(1):161-72. doi: 10.1016/j.cell.2006.03.017. PMID: 16615897
3. Costanzo JP, Amaral MC, Rosendale AJ, Lee RE. Hibernation physiology, freezing adaptation and extreme freeze tolerance in a northern population of the wood frog. J Exp Biol. 2013; 216(18): 3461-73. doi: 10.1242/jeb.089342. PMID: 23966588
4. Christenson MK, Trease AJ, Potluri LP, Jezewski AJ, Davis VM, Knight LA, et al. De

369 novo Assembly and Analysis of the Northern Leopard Frog *Rana pipiens* Transcriptome. J
370 Genomics. 2014; 2:141-9. doi: 10.7150/jgen.9760. PMID: 25371763
371 5. Hadj-Moussa H, Storey KB. Micromanaging Freeze Tolerance: The Biogenesis and
372 Regulation of microRNAs in Frozen Frogs. Cell Mol Life Sci. 2018; 75(19): 3635-3647.
373 doi: 10.1007/s00018-018-2821-0, PMID: 29681008
374 6. Storey KB. Turning Down The Fires Of Life: Metabolic Regulation Of Hibernation And
375 Estivation. Comp Biochem Physiol B Biochem Mol Biol. 2001; 126. doi:
376 10.1016/S0305-0491(00)80178-9
377 7. Reilly BD, Schlipalius DI, Cramp RL, Ebert PR, Franklin CE. Frogs and estivation:
378 transcriptional insights into metabolism and cell survival in a natural model of extended
379 muscle disuse. Physiol Genomics. 2013; 45(10): 377-88. doi:
380 10.1152/physiolgenomics.00163.2012. PMID: 23548685
381 8. Storey KB, Dent ME, Storey JM. Gene expression during estivation in spadefoot toads,
382 *Scaphiopus couchii*: upregulation of riboflavin binding protein in liver. J Exp Zool. 1999;
383 284(3): 325-33. PMID: 10404124
384 9. Hiong KC, Ip YK, Wong WP, Chew SF. Differential gene expression in the liver of the
385 African lungfish, *Protopterus annectens*, after 6 months of aestivation in air or 1 day of
386 arousal from 6 months of aestivation. PLoS One. 2015; 10(3): e0121224. doi:
387 10.1371/journal.pone.0121224. PMID: 25822522
388 10. Loveridge JP, Withers PC. Metabolism and Water Balance of Active and Cocooned
389 African Bullfrogs *Pyxicephalus adspersus*. Physiol Biochem Zoo. 1981; 54(2). doi:
390 210.1086/physzool.54.2.30155821
391 11. Secor SM. Physiological responses to feeding, fasting and estivation for anurans. J Exp
392 Biol. 2005; 208(13): 2595-608. doi: 10.1242/jeb.01659. PMID: 15961745
393 12. Strauss WM. Preparation of Genomic DNA from Mammalian Tissue. Curr Protoc
394 Immunol. 2001; Chapter 10 (10.2). doi: 10.1002/0471142735.im1002s08. PMID:

395 18432684
396 13. De Lisle RC. Isolation of Pancreatic RNA. Pancreapedia: Exocrine Pancreas Knowledge
397 Base. doi: 10.3998/panc.2014.9
398 14. Sumida M, Kondo Y, Kanamori Y, Nishioka M. Inter-and intraspecific evolutionary
399 relationships of the rice frog *Rana limnocharis* and the allied species *R. cancrivora* inferred
400 from crossing experiments and mitochondorial DNA sequences of the 12S and 16S rRNA
401 genes. Mol Phylogenet Evol. 2002; 25(2): 293-305. doi: 10.1016/S1055-7903(02)00243-9.
402 PMID: 12414311
403 15. Birol I, Behsaz B, Hammond SA, Kucuk E, Veldhoen N, Helbing CC. *De novo*
404 Transcriptome Assemblies of *Rana (Lithobates) catesbeiana* and *Xenopus laevis* Tadpole
405 Livers for Comparative Genomics without Reference Genomes. PLoS One. 2015; 10(6):
406 e0130720. doi: 10.1371/journal.pone.0130720. PMID: 26121473
407 16. Channing A, DU Preez L, Passmore N. Status, vocalization and breeding biology of two
408 species of African bullfrogs (Ranidae: *Pyxicephalus*). J Zoo. 1994; 234(1): 141-148. doi:
409 10.1111/j.1469-7998
410 17. Bishop P. Announcing AmphibiaWeb (University of California Berkeley, CA, USA). 2004
411 Oct 19 [Edited by Janzen P, Tunstall T. 2005 Sep 21]. Available from:
412 https://amphibiaweb.org/species/4966
413 18. Engelbrecht D, Mashao M, Halajian A. Notes on the breeding behavior and ecology of
414 Edible Bullfrogs *Pyxicephalus edulis* Peters, 1854 in the Limpopo Province, South Africa.
415 Herpetology Notes. 2015; 8: 365-369.
416 19. Pyron RA, Wiens JJ. A large -scale phylogeny of Amphibia including over 2,800 species,
417 and a revised classification of extant frogs, salamanders, and caecilians. Mol Phylogenet
418 Evol. 2011; 61(2): 543-583. doi: 10.1016/j.ympev.2011.06.012. PMID: 21723399
419 20. Torres M, Creevey CJ, Kornobis E, Gower DJ, Wilkinson M, San D. Multi-tissue
420 transcriptomes of caecilian amphibians highlight incomplete knowledge of vertebrate gene

421 families. DNA Res. 2018; 0(0): 1-8. doi: 10.1093/dnares/dsy034. PMID: 30351380

422 21. Muriira N, Xu W, Muchugi A, Xu J, Liu A. *De novo* sequencing and assembly analysis of
423 transcriptome in the Sodom apple (*Calotropis gigantea*). BMC Genomics. 2015 (16):723.
424 doi: 10.1186/s12864-015-1908-3. PMID: 26395839

425 22. Morandin C, Pulliainen U, Bos N, Schultner E. *De novo* transcriptome assembly and its
426 annotation for the black ant *Formica fusca* at the larval stage. Scientific Data. 2018.
427 180282. doi: 10.1038/sdata.2018.282. PMID 30561435

428

429 # Figure legend

430 **Fig. 1.** The African bullfrog used in this study.

431 A) The adult male African bullfrog used in this study is shown.

432 B) Phylogenetic tree analysis of the 16S rRNA sequence between the frog used in this
433 study and several *P. adspersus* and *P. edulis* species deposited in the database.
434 Sequences were aligned using MUSCLE algorithm in MEGA X. Phylogenetic tree was
435 constructed using neighbor-joining algorithm and the node support values were assessed
436 by bootstrap test of 1000 replicates.

437 C) The phylogenetic tree using the 16S rRNA sequence between the representative frog
438 species belonging to the major family of Anura. *P. adspersus* used in this study is
439 shown in red and that of the other two aestivating frogs are shown in blue. Phylogenetic
440 tree was constructed in the same manner in Fig 1B.

441

442 **Fig. 2.** Distribution of the length of the African bullfrog transcripts.

443 The number of transcripts was count and the transcript length was determined.

444

445 **Fig. 3**. Similarity of the African bullfrog transcripts against the Uniprot protein
446 databases.

447 A) BLASTx search of the African bullfrog transcripts against the Uniprot database
448 (all-proteins or *X. tropicalis* proteins), and the number of transcripts were counted
449 according to E-values.

450 B) Distribution of the identity of the African bullfrog transcripts with an E-value lower
451 than 1E-05 against the *X. tropicalis* protein database.

452

453 **Fig. 4.** Gene ontology terms assigned to the African bullfrog genes.

454 The 28,570 genes with high identity (E-value lower than 1E-5) in a BLASTx search of

455 the *X. tropicalis* protein database were subjected to gene ontology analysis.

Figure 1

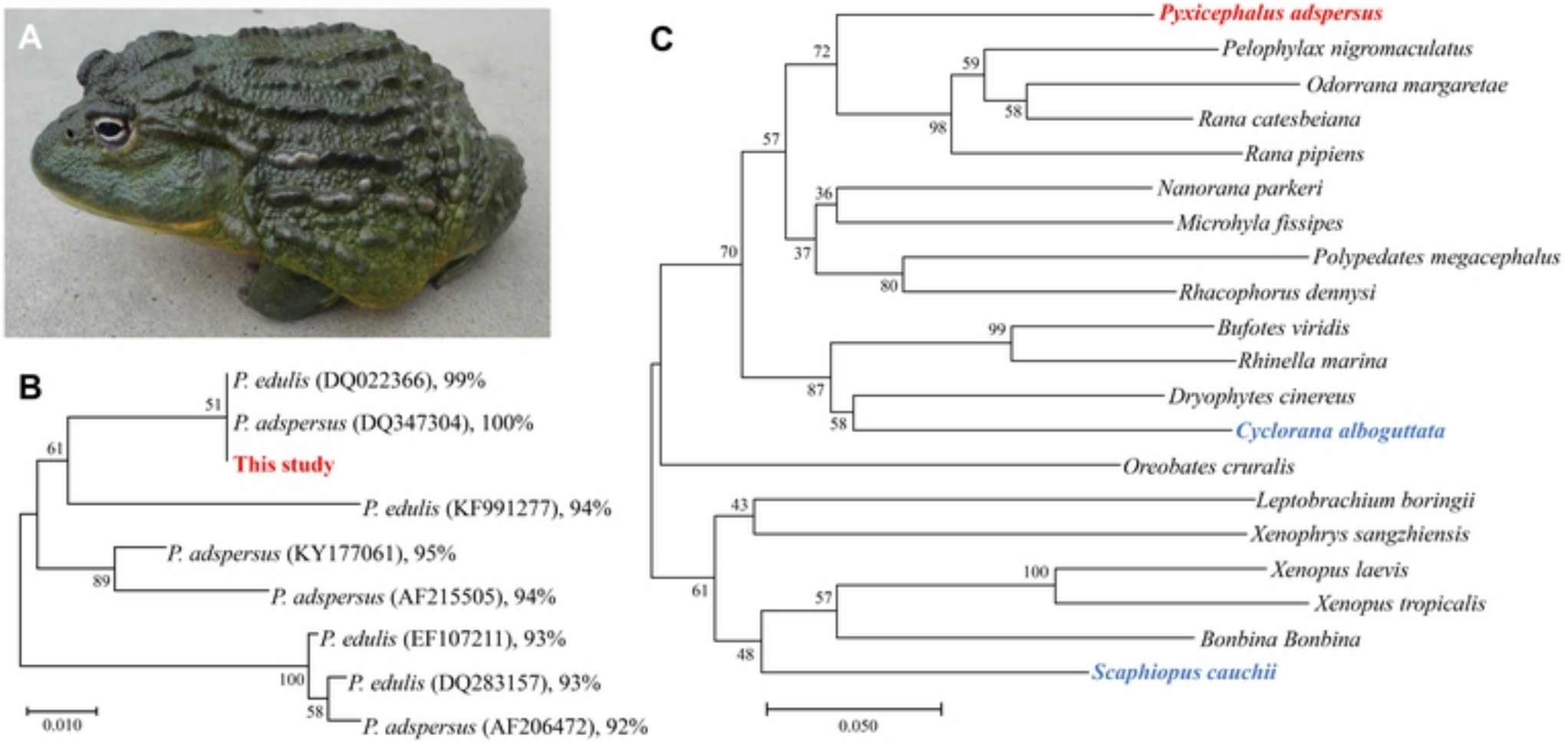

Figure 1

Figure 2

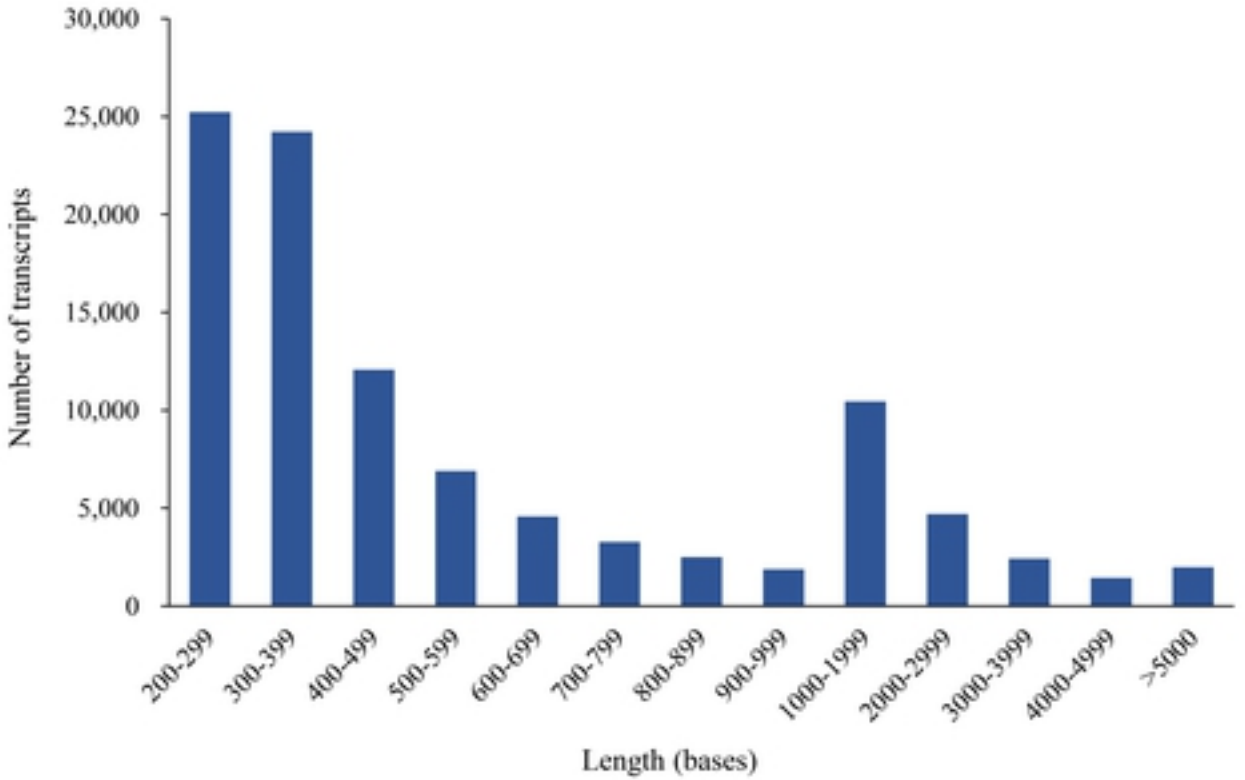

Figure 2

Figure 3

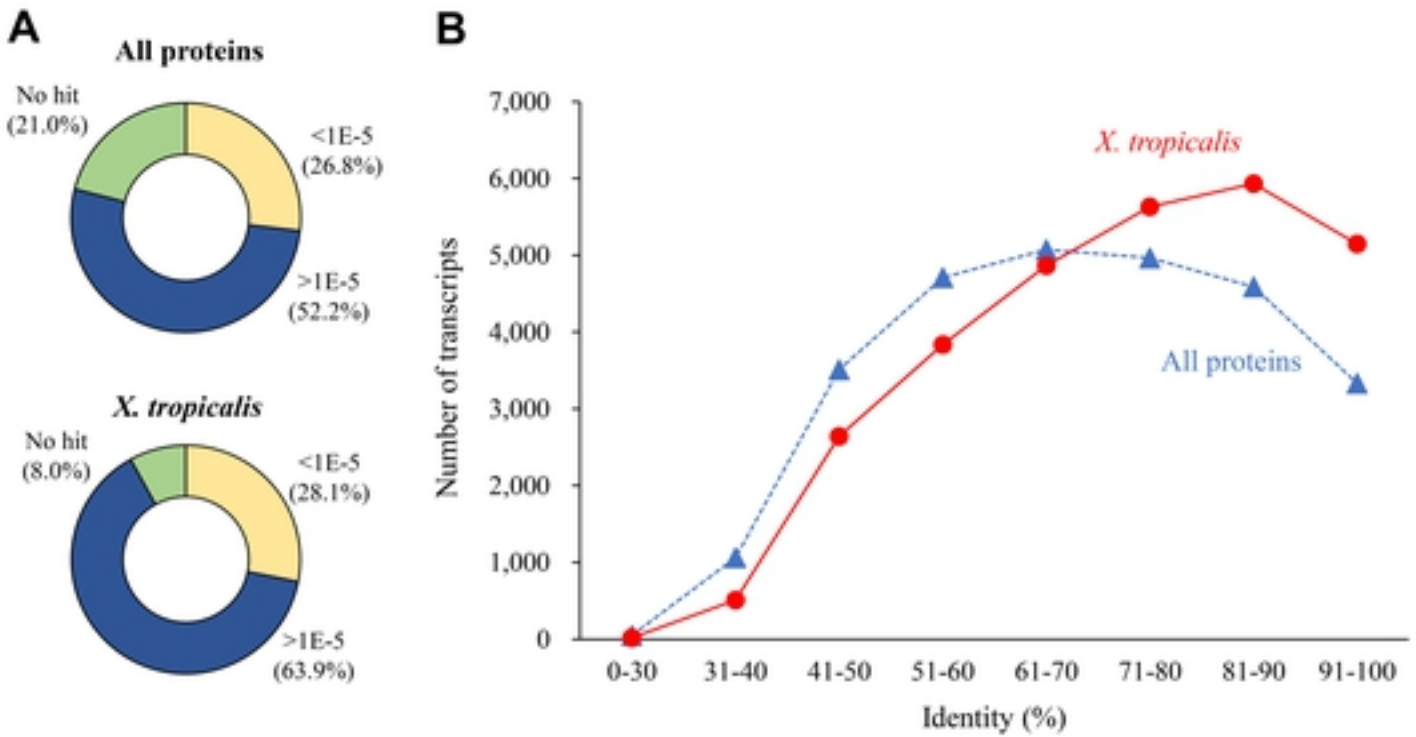

Figure 3

Figure 4

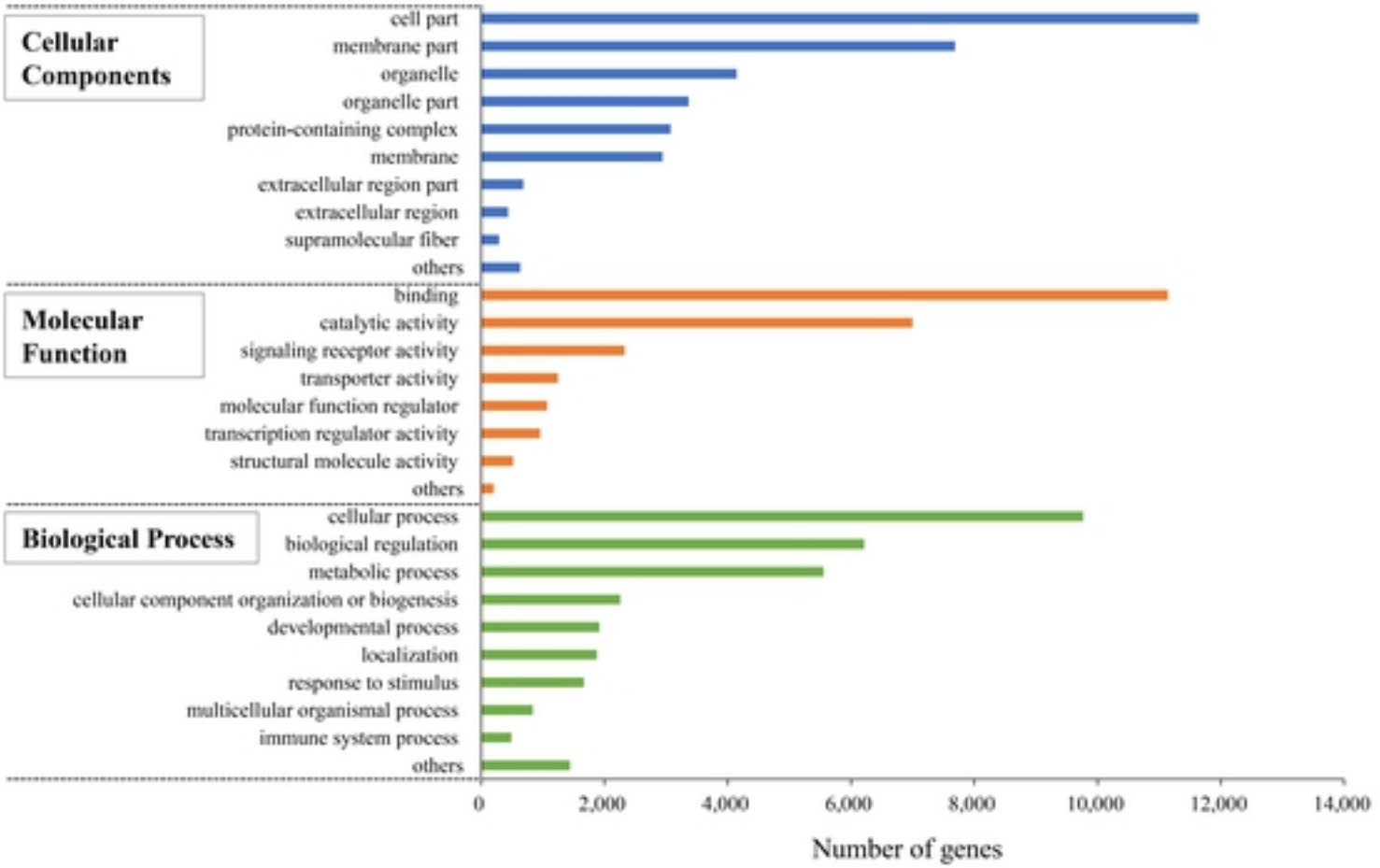

Figure 4