

<b>Manuscript Number:</b>	GIGA-D-16-00040R4	
<b>Full Title:</b>	The genome of the Antarctic-endemic copepod, <i>Tigriopus kingsejongensis</i>	
<b>Article Type:</b>	Data Note	
<b>Funding Information:</b>	Korea Polar Research Institute (KR) (PE16070)	Dr. Hyun Park
	Korea Polar Research Institute (PE14260)	Dr Sanghee Kim
<b>Abstract:</b>	<p>Background: The Antarctic intertidal zone is continuously subjected to extremely fluctuating biotic and abiotic stressors. The West Antarctic Peninsula is the most rapidly warming region on Earth. Organisms living in Antarctic intertidal pools are therefore interesting for research into evolutionary adaptation to extreme environments and the effects of climate change.</p> <p>Findings: We report the whole genome sequence of the Antarctic-endemic harpacticoid copepod <i>Tigriopus kingsejongensis</i>. The 37 Gb raw DNA sequence was generated using the Illumina Miseq platform. Libraries were prepared with 65-fold coverage and a total length of 295 Mb. The final assembly consists of 48 368 contigs with an N50 contig length of 17.5 kb, and 27 823 scaffolds with an N50 contig length of 159.2 kb. A total of 12 772 coding genes were inferred using the MAKER annotation pipeline. Comparative genome analysis revealed that <i>T. kingsejongensis</i>-specific genes are enriched in transport and metabolism processes. Furthermore, rapidly evolving genes related to energy metabolism showed positive selection signatures.</p> <p>Conclusions: The <i>T. kingsejongensis</i> genome provides an interesting example of an evolutionary strategy for Antarctic cold adaptation, and offers new genetic insights into Antarctic intertidal biota.</p>	
<b>Corresponding Author:</b>	Hyun Park KOREA, REPUBLIC OF	
<b>Corresponding Author Secondary Information:</b>		
<b>Corresponding Author's Institution:</b>		
<b>Corresponding Author's Secondary Institution:</b>		
<b>First Author:</b>	Seunghyun Kang	
<b>First Author Secondary Information:</b>		
<b>Order of Authors:</b>	Seunghyun Kang	
	Do-Hwan Ahn	
	Jun Hyuck Lee	
	Sung Gu Lee	
	Seung Chul Shin	
	Jungeun Lee	
	Gi-Sik Min	
	Hyoungseok Lee	
	Hyun-Woo Kim	
	Sanghee Kim	
	Hyun Park	

<b>Order of Authors Secondary Information:</b>	
<b>Response to Reviewers:</b>	<p>Dear Hans,</p> <p>We have tried to address all the comments and we have uploaded both clean and comment version of manuscripts.</p> <p>Thanks and best regards.</p>
<b>Additional Information:</b>	
<b>Question</b>	<b>Response</b>
Are you submitting this manuscript to a special series or article collection?	No
<p><b>Experimental design and statistics</b></p> <p>Full details of the experimental design and statistical methods used should be given in the Methods section, as detailed in our <a href="#">Minimum Standards Reporting Checklist</a>. 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>	Yes
<p><b>Resources</b></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 <a href="#">Research Resource Identifiers</a> (RRIDs) for antibodies, model organisms and tools, where possible.</p> <p>Have you included the information requested as detailed in our <a href="#">Minimum Standards Reporting Checklist</a>?</p>	Yes
<p><b>Availability of data and materials</b></p> <p>All datasets and code on which the conclusions of the paper rely must be either included in your submission or deposited in <a href="#">publicly available repositories</a> (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>	Yes

Have you have met the above requirement as detailed in our [Minimum Standards Reporting Checklist?](#)

1 **The genome of the Antarctic-endemic copepod, *Tigriopus***  
2  
3  
4 ***kingsejongensis***  
5  
6  
7  
8  
9

10 **Seunghyun Kang<sup>1¶</sup>, Do-Hwan Ahn<sup>1¶</sup>, Jun Hyuck Lee<sup>1,2</sup>, Sung Gu Lee<sup>1,2</sup>, Seung Chul**  
11  
12 **Shin<sup>1</sup>, Jungeun Lee<sup>1</sup>, Gi-Sik Min<sup>3</sup>, Hyoungseok Lee<sup>1</sup>, Hyun-Woo Kim<sup>4\*</sup>, Sanghee Kim<sup>5\*</sup>**  
13  
14 **& Hyun Park<sup>1,2\*</sup>**  
15  
16  
17  
18  
19

20 <sup>1</sup>Unit of Polar Genomics, Korea Polar Research Institute, Yeonsu-gu, Incheon, South Korea  
21

22 <sup>2</sup>Polar Sciences, University of Science & Technology, Yuseong-gu, Daejeon, South Korea  
23

24 <sup>3</sup>Department of Biological Sciences, Inha University, Incheon, South Korea  
25  
26

27 <sup>4</sup>Department of Marine Biology, Pukyong National University, Busan, South Korea  
28

29 <sup>5</sup>Division of Polar Life Sciences, Korea Polar Research Institute, Yeonsu-gu, Incheon, South  
30  
31 Korea  
32  
33  
34  
35  
36

37 \*Corresponding authors  
38

39 Hyun-Woo Kim  
40

41 Department of Marine Biology  
42

43 Pukyong National University  
44

45 Busan  
46  
47

48 South Korea  
49  
50

51 Email: [kimhw@pknu.ac.kr](mailto:kimhw@pknu.ac.kr)  
52

53 Phone: +82-51-629-5926  
54  
55  
56  
57

58 Sanghee Kim  
59  
60  
61  
62  
63  
64  
65

1 Division of Polar Life Sciences

2  
3 Korea Polar Research Institute

4  
5 Yeonsu-gu

6  
7  
8 Incheon

9  
10 South Korea

11  
12  
13 Email: [sangheekim@kopri.re.kr](mailto:sangheekim@kopri.re.kr)

14  
15  
16 Phone: +82-32-760-5515

17  
18  
19  
20 Hyun Park

21  
22 Unit of Polar Genomics

23  
24  
25 Korea Polar Research Institute

26  
27 Yeonsu-gu

28  
29  
30 Incheon

31  
32 South Korea

33  
34  
35 Email: [hpark@kopri.re.kr](mailto:hpark@kopri.re.kr)

36  
37  
38 Phone: +82-32-760-5570

39  
40  
41  
42 ¶These authors contributed equally to this work.

43  
44  
45  
46 Seunghyun Kang

47  
48  
49 [skang@kopri.re.kr](mailto:skang@kopri.re.kr)

50  
51  
52  
53  
54 Do-Hwan Ahn

55  
56  
57 [ahndh@kopri.re.kr](mailto:ahndh@kopri.re.kr)

58  
59  
60  
61  
62  
63  
64  
65

1 Jun Hyuck Lee

2  
3 [junhyucklee@kopri.re.kr](mailto:junhyucklee@kopri.re.kr)

4  
5  
6  
7  
8 Sung Gu Lee

9  
10 [holynine@kopri.re.kr](mailto:holynine@kopri.re.kr)

11  
12  
13  
14  
15 Seung Chul Shin

16  
17 [ssc@kopri.re.kr](mailto:ssc@kopri.re.kr)

18  
19  
20  
21  
22 Jungeun Lee

23  
24 [jelee@kopri.re.kr](mailto:jelee@kopri.re.kr)

25  
26  
27  
28  
29 Gi-Sik Min

30  
31 [mingisik@inha.ac.kr](mailto:mingisik@inha.ac.kr)

32  
33  
34  
35  
36  
37 Hyoungseok Lee

38  
39 [soulaid@kopri.re.kr](mailto:soulaid@kopri.re.kr)

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 **Abstract**  
2  
3

4 **Background:** The Antarctic intertidal zone is continuously subjected to extremely fluctuating  
5 biotic and abiotic stressors. The West Antarctic Peninsula is the most rapidly warming region  
6  
7 on Earth. Organisms living in Antarctic intertidal pools are therefore interesting for research  
8  
9 into evolutionary adaptation to extreme environments and the effects of climate change.  
10  
11  
12

13  
14 **Findings:** We report the whole genome sequence of the Antarctic-endemic harpacticoid  
15 copepod *Tigriopus kingsejongensi*. The 37 Gb raw DNA sequence was generated using the  
16  
17 Illumina Miseq platform. Libraries were prepared with 65-fold coverage and a total length of  
18  
19 295 Mb. The final assembly consists of 48 368 contigs with an N50 contig length of 17.5 kb,  
20  
21 and 27 823 scaffolds with an N50 contig length of 159.2 kb. A total of 12 772 coding genes  
22  
23 were inferred using the MAKER annotation pipeline. Comparative genome analysis revealed  
24  
25 that *T. kingsejongensis*-specific genes are enriched in transport and metabolism processes.  
26  
27 Furthermore, rapidly evolving genes related to energy metabolism showed positive selection  
28  
29 signatures.  
30  
31  
32  
33  
34  
35  
36

37 **Conclusions:** The *T. kingsejongensis* genome provides an interesting example of an  
38  
39 evolutionary strategy for Antarctic cold adaptation, and offers new genetic insights into  
40  
41 Antarctic intertidal biota.  
42  
43  
44

45 **Keywords:** Copepoda; Genome; Antarctic; Adaptation; *Tigriopus*  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

## Data description

Approximately 12 000 species have been described in the diverse copepod subclass [1, 2]. These species dominate the zooplankton community, contributing about 70% of total zooplankton biomass [3], and are an important link between phytoplankton and higher trophic levels in the marine meiobenthic food web [4]. Harpacticoid copepods of the genus *Tigriopus* Norman 1868 are dominant members of shallow supratidal rock pools, distributed worldwide among habitats that vary widely in salinity, temperature, desiccation risk, and UV radiation. They are a model system in investigations of osmoregulation [5], temperature adaptation [6, 7] and environmental toxicology [8]. With publically available copepod genome resources (e.g., *Tigriopus californicus* [9], *T. japonicus* [10], *Eurytemora affinis* [11] and salmon louse *Lepeophtheirus salmonis* [12]), it is now possible to explore their fundamental biological processes and physiological responses to diverse environments.

Antarctica is not only an extreme habitat for extant organisms, but also a model for research on evolutionary adaptations to cold environments [13, 14]. The Antarctic intertidal zone, particularly in the Western Antarctic Peninsula region, is one of the most extreme, yet fastest warming environments on Earth. Thus, it is a potential barometer for global climate change [15]. Antarctic intertidal species that have evolved stenothermal phenotypes through adaptation to year-round extreme cold may now face extinction by global warming. The response of these species to further warming in Western Antarctica is of serious concern; however, to date, few studies have focused on Antarctic intertidal zone species.

First described in 2014, *T. kingsejongensis* was recognized as a new species endemic to a rock pool in the Antarctic Peninsula. It is extremely cold-tolerant and can survive in frozen sea water [16]. Compared to the congener *T. japonicus*, which is found in coastal areas of the Yellow Sea, morphological differences of this species include increased numbers of caudal



1 setae in nauplii, an optimal growth temperature of approximately 8°C, and differing  
2 developmental characteristics. *Tigriopus kingsejongensis* has evolved to overcome the unique  
3 environmental constraints of Antarctica, therefore providing an ideal experimental model for  
4 extreme habitat research. This species may represent a case of rapid speciation, since the  
5 intertidal zone on King George Island and the surrounding areas did not exist 10 000 years  
6 ago [17]. *Tigriopus kingsejongensis* likely evolved as a distinct species within this relatively  
7 short time period. Thus, interspecies and intraspecies comparative analyses of Antarctic  
8 *Tigriopus* species will help to define the trajectory of adaptation to the Antarctic  
9 environment, and also provide insights into the genetic basis of *Tigriopus* divergence and  
10 evolution.  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29

### 30 **Library construction and sequencing**

31 *Tigriopus kingsejongensis* specimens were collected using hand-nets from tidal pools in  
32 Potter Cove, near King Sejong Station, on the northern Antarctic Peninsula (62°14'S,  
33 58°47'W) (Fig. 1 and Fig. S1) in January 2013. The water temperature was  $1.6 \pm 0.8^\circ\text{C}$  during  
34 this month. High molecular weight genomic DNA from pooled *T. kingsejongensis* was  
35 extracted using the DNeasy Blood and Tissue Kit (Qiagen, Venlo, The Netherlands). For  
36 Illumina Miseq sequencing, four library types were constructed with 350, 400, 450, and  
37 500 bp for paired-end libraries, and 3 kb and 8 kb for mate-pair libraries, prepared using the  
38 standard Illumina sample preparation methods (Table 1). All sequencing processes were  
39 performed according to the manufacturer's instructions (Illumina, Carlsbad, USA).  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54

55 RNA was prepared from pooled *T. kingsejongensis* and *T. japonicus* specimens from two  
56 different temperature experiments (4°C and 15°C) using the RNeasy Mini Kit (Qiagen). For  
57 Illumina Miseq sequencing, subsequent experiments were carried out according to the  
58  
59  
60  
61  
62  
63  
64  
65

1 manufacturer's instructions (Illumina). The *de novo* transcriptome assembly was performed  
2  
3 with CLC Genomics Workbench (Qiagen), setting the minimum allowed contig length to 200  
4  
5 nucleotides. The assembly process generated 40 172 contigs with a maximum length of  
6  
7 23 942 bp and an N50 value of 1093 bp. Generated contigs were used as reference sequences  
8  
9 to map trimmed reads, and fold-changes in expression for each gene were calculated with a  
10  
11 significance threshold of  $P \leq 0.05$  using the CLC Genomics Workbench (Table 2 and 3).  
12  
13  
14  
15  
16  
17  
18  
19

## 20 **Genome assembly**

21  
22 First, k-mer analysis was conducted using jellyfish 2.2.5 [18] to estimate the genome size  
23  
24 from DNA paired-end libraries. The estimated genome size was 298 Mb, with the main peak  
25  
26 at a depth of  $\sim 39\times$  (Fig. 2). Then, assemblies were performed using a Celera Assembler with  
27  
28 Illumina short reads [19]. Prior to assembly, Illumina reads were trimmed using the FASTX-  
29  
30 Toolkit [20] with parameters  $-t$  20,  $-l$  70 and  $-Q$  33, after which a paired sequence from  
31  
32 trimmed Illumina reads was selected. Finally, trimmed Illumina reads with 65-fold coverage  
33  
34 (insert sizes 350, 400, 450, and 500 bp) were obtained and converted to the FRG file format  
35  
36 (required by the Celera Assembler) using FastqToCA. Assembly was performed on a 96-  
37  
38 processor workstation with Intel Xeon X7460 2.66 GHz processors and 1 Tb random access  
39  
40 memory (RAM) with the following parameters: `overlapper = ovl`, `unitigger = bogart`,  
41  
42 `utgGraphErrorRate = 0.03`, `utgGraphErrorLimit = 2.5`, `utgMergeErrorRate = 0.030`,  
43  
44 `utgMergeErrorLimit = 3.25`, `ovlErrorRate = 0.1`, `cnsErrorRate = 0.1`, `cgwErrorRate = 0.1`,  
45  
46 `merSize = 22`, and `doOverlapBasedTrimming = 1`. The initial Celera assembly was 305 Mb,  
47  
48 had an N50 contig size of 17 566 bp, and a maximum contig size of 349.5 kb. Scaffolding  
49  
50 was completed using the SSPACE 2.0 scaffolder using mate-paired data [21]. Subsequently,  
51  
52 we closed gaps using Gapfiller version 1.9 with  $65\times$  trimmed Illumina reads with default  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 settings [22]. *De novo* assembly of 203 million reads from paired-end and mate-paired  
2  
3 libraries yielded a draft assembly (65-fold coverage) with a total length of 295 Mb, and  
4  
5 contig and scaffold N50 sizes of 17.6 kb and 159.2 kb, respectively (Table 4 and Fig. 3).  
6  
7  
8  
9

## 10 11 12 **Annotation**

13  
14  
15 MAKER, a portable and easily configurable genome annotation pipeline, was used to  
16  
17 annotate the genome [23]. Repetitive elements were identified using RepeatMasker [24]. This  
18  
19 masked genome sequence was used with SNAP software [25] for *ab initio* gene prediction,  
20  
21 after which alignment of expressed sequence tags (ESTs) with BLASTn [26] and protein  
22  
23 information from tBLASTx [26] were included. The *de novo* repeat library of *T.*  
24  
25 *kingsejongensis* from RepeatModeler was used for RepeatMasker; proteins from five species  
26  
27 with data from *Drosophila melanogaster*, *Daphnia pulex*, *T. japonicus*, and *T. californicus*  
28  
29 were included in the analysis. RNA-seq-based gene prediction, data were aligned against the  
30  
31 assembled genome using TopHat [27], and Cufflinks [28] was used to predict cDNAs from  
32  
33 the resultant data. Next, MAKER polished the alignments using the program Exonerate [29],  
34  
35 which provided integrated information to synthesize SNAP annotation. Considering all  
36  
37 information, MAKER then selected and revised the final gene model. A total of 12 772 genes  
38  
39 were predicted in *T. kingsejongensis* using MAKER. Annotated genes contained an average  
40  
41 of 4.6 exons, with an average mRNA length of 1090 bp. Additionally, 12 562 of 12 772  
42  
43 genes were assigned preliminary functions based on automated annotation using Blast2GO  
44  
45 (Ver. 2.6.0) [30] (Fig. S2 and S3) with homology sequences from the SwissProt[31],  
46  
47 TrEMBL, National Center for Biotechnology Information (NCBI) non-redundant protein  
48  
49 databases [32] and REVIGO software was used to cluster related GO terms according to *P*-  
50  
51 value [33]. Infernal version 1.1 [34] and covariance models (CMs) from the Rfam database  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 [35] were used to identify other non-coding RNAs in the *T. kingsejongensis* scaffold. Putative  
2  
3 tRNA genes were identified using tRNAscan-SE [36] (Table S1), which uses a CM that  
4  
5 scores candidates based on their sequence and predicted secondary structures.  
6  
7

8 Non-gap sequences occupied 284.8 Mb (96.5%), and simple sequence repeats (SSRs)  
9  
10 amounted to 1.2 Mb (0.4%) (Table S2). Transposable elements (TEs) comprised 6.5 Mb;  
11  
12 roughly 2.3% of the assembled genome (Table S2 and S3). On the basis of homology and *ab*  
13  
14 *initio* gene prediction, the *T. kingsejongensis* genome contained 12 772 protein-coding genes  
15  
16 (Table 5). By assessing the quality of the 12 772 annotated gene models, 11 686 protein-  
17  
18 coding genes (91.5%) were supported by RNA-seq data, of which 7325 (63%) were similar to  
19  
20 proteins from other species. To estimate genome assembly and annotation completeness,  
21  
22 Core Eukaryotic Genes Mapping Approach (CEGMA) [37] and Benchmarking Universal  
23  
24 Single-Copy Orthologs (BUSCO) [38] analysis was used (Table 6). The CEGMA report  
25  
26 revealed that 193 of 248 CEGMA score genes were fully annotated (77.8% completeness),  
27  
28 and 206 of 248 genes were partially annotated (83% completeness). BUSCO, a similar  
29  
30 approach used for lineage-specific profile libraries such as eukaryotes, metazoans, and  
31  
32 arthropods, revealed 71% complete and 6% partial Metazoan orthologous gene sets in our  
33  
34 assembly; using an arthropod gene set, only 61.1% complete and 10.7% partial genes were  
35  
36 assigned. CEGMA and BUSCO gene sets largely comprised insects; other non-insect  
37  
38 arthropod genomes obtained similarly low assignment scores. Overall, the *T.*  
39  
40 *kingsejongensis* genome was moderately complete in non-dipteran arthropod genomes.  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54

## 55 **Gene families**

56

57 Orthologous groups were identified from 11 species (*T. kingsejongensis*, *Aedes aegypti*, *D.*  
58  
59 *melanogaster*, *Ixodes scapularis*, *Mesobuthus martensii*, *Strigamia maritima*, *Tetranychus*  
60  
61  
62  
63  
64  
65

1 *urticae*, *D. pulex*, *Homo sapiens*, *Ciona intestinalis*, and *Caenorhabditis elegans*) (Table 7)  
2  
3 using OrthoMCL [40] with standard parameters and options; transcript variants other than the  
4  
5 longest translation forms were removed. For *T. kingsejongensis*, the coding sequence from  
6  
7 the MAKER annotation pipeline was used. The 1:1:1 single-copy orthologous genes were  
8  
9 subjected to phylogenetic construction and divergence time estimation. Protein-coding genes  
10  
11 were aligned using the Probabilistic Alignment Kit (PRANK) with the codon alignment  
12  
13 option [41], and poorly aligned sequences with gaps were removed using Gblock under the  
14  
15 codon model [42]. A maximum likelihood phylogenetic tree was constructed using RAxML  
16  
17 with 1000 bootstrap values [43] and calibrated the divergence time between species with  
18  
19 TimeTree [44]. Finally, the average gene gain/loss rate along the given phylogeny was  
20  
21 identified using CAFÉ 3.1 [45].  
22  
23  
24  
25  
26

27 Orthologous gene clusters were constructed using four arthropod species (Antarctic copepod,  
28  
29 *T. kingsejongensis*; scorpion, *M. martensii*; fruit fly, *D. melanogaster*, and water flea, *D.*  
30  
31 *pulex*) to compare genomic features and adaptive divergence. In total, 2063 gene families are  
32  
33 shared by all four species, and 1028 genes are *T. kingsejongensis*-specific. *T. kingsejongensis*  
34  
35 shares 4559 (73.5%) gene families with *D. pulex*, which belongs to the same crustacean  
36  
37 lineage, Vericrustacea; 3531 (56.9%) with *D. melanogaster*; and 3231 (52.1%) with *M.*  
38  
39 *martensii* (Fig. 4A). Gene Ontology (GO) analysis revealed the 1028 *T. kingsejongensis*-  
40  
41 specific genes are enriched in transport (single-organism transport, GO:0044765;  
42  
43 transmembrane transport, GO:0055085; ion transport, GO:0006811; cation transport,  
44  
45 GO:0006812) and single-organism metabolic processes (GO:0044710) (Table S4 and S5).  
46  
47  
48  
49  
50

51 Subsequently, gene gain-and-loss was analyzed in 11 representative species: *T.*  
52  
53 *kingsejongensis* gained 735 and lost 4401 gene families (Fig. 4B). This species exhibits a  
54  
55 gene family turnover of 5136, the largest value among the eight arthropods. The second  
56  
57 largest value was obtained from *T. uticae* and the third from *M. martensii*. Non-insect  
58  
59  
60  
61  
62  
63  
64  
65

1 arthropod genomes were relatively poorly assigned with CEGMA or BUSCO sets (Table 6).  
2  
3 The assignment reports of these largely insect-based gene sets tend to have low assignment  
4 scores in non-insect or non-dipteran genomes [38, 46, 47]. This implies that careful  
5 examination of gene family turnover is needed in non-insect arthropod genomes, as well as  
6 globally approved arthropod orthologous gene sets.  
7  
8  
9

10  
11 Analysis of gene family expansion and contraction in *T. kingsejongensis* (Tables S6–S9)  
12 revealed 232 significantly expanded gene families, which are significantly overrepresented in  
13 amino acid and carbohydrate metabolism pathways, according to the Kyoto Encyclopedia of  
14 Genes and Genomes (KEGG) [48].  
15  
16  
17  
18  
19  
20  
21

## 22 23 24 25 26 27 **Genome evolution**

28  
29 Adaptive functional divergence caused by natural selection is commonly estimated based on  
30 the ratio of nonsynonymous ( $dN$ ) to synonymous ( $dS$ ) mutations. To estimate  $dN$ ,  $dS$ , the  
31 average  $dN/dS$  ratio ( $w$ ), and lineage-specific positively selected genes (PSGs) in *T.*  
32 *kingsejongensis* and *T. japonicus*, protein-coding genes from *T. japonicus* were added to  
33 define orthologous gene families among four species (*T. kingsejongensis*, *T. japonicus*, *D.*  
34 *pulex*, and *D. melanogaster*) using the program OrthoMCL with the same conditions  
35 previously described. We identified 2937 orthologous groups shared by all four species;  
36 single-copy gene families were used to construct a phylogenetic tree and estimate the time  
37 since divergence using the methods described above. Each of the identified orthologous  
38 genes was aligned using PRANK, and poorly aligned sequences with gaps were removed  
39 using Gblock. Alignments with less than 40% identity and genes shorter than 150 bp were  
40 eliminated in subsequent procedures. The values of  $dN$ ,  $dS$  and  $w$  were estimated from each  
41 gene using the Codeml program implemented in the Phylogenetic Analysis by Maximum  
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 Likelihood (PAML) package with the free-ratio model [49] under F3X4 codon frequencies;  
2  
3 orthologs with  $w \leq 5$  and  $dS \leq 3$  were retained [50]. To examine the accelerated  
4  
5 nonsynonymous divergence in either the *T. kingsejongensis* or *T. japonicus* lineages, a  
6  
7 binomial test [51] was used to determine GO categories with at least 20 orthologous genes.  
8  
9 To define PSGs in *T. kingsejongensis* and *T. japonicus*, basic and branch-site models were  
10  
11 applied, and Likelihood Ratio Tests (LRTs) were used to remove genes under relaxation of  
12  
13 selective pressure. To investigate the functional categories and pathways enriched in PSGs,  
14  
15 the Database for Annotation, Visualization and Integrated Discovery (DAVID) Functional  
16  
17 Annotation [52] was used with Fisher's exact test (cutoff:  $P \leq 0.05$ ).

18  
19 The average  $w$  value from 2937 co-orthologous genes of *T. kingsejongensis* (0.0027) is  
20  
21 higher than that of *T. japonicus* (0.0022). GO categories that show evidence of accelerated  
22  
23 evolution in *T. kingsejongensis* are: energy metabolism (generation of precursor metabolites  
24  
25 and energy, GO:0006091; cellular respiration, GO:0045333) and carbohydrate metabolism  
26  
27 (monosaccharide metabolic process, GO:0005996; hexose metabolic process, GO:0019318)  
28  
29 (Fig. 5A, Table S10). Branch-site model analysis showed that genes belonging to these  
30  
31 functional categories have undergone a significant positive selection process by putative  
32  
33 functional divergence in certain lineages. There are 74 and 79 PSGs in *T. kingsejongensis*  
34  
35 (Table S11) and *T. japonicus* (Table S12), respectively.

36  
37 The functional categories enriched in *T. kingsejongensis*, when compared to *T. japonicus*,  
38  
39 support the idea that functional divergence in *T. kingsejongensis* is strongly related to energy  
40  
41 metabolism (oxidative phosphorylation, GO:0006119; energy-coupled proton transport down  
42  
43 electrochemical gradient, GO:0015985; ATP synthesis-coupled proton transport, GO  
44  
45 0015986; generation of precursor metabolites and energy, GO:0006091) (Fig. 5B, Table S13  
46  
47 and S14). In particular, three of the identified genes are involved in the oxidative  
48  
49 phosphorylation (OxPhos) pathway, which provides the primary cellular energy source in the  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 form of adenosine triphosphate (ATP). These three genes are nuclear-encoded mitochondrial  
2  
3 genes: the catalytic F1 ATP synthase subunit alpha (*ATP5A*) (Fig. S4), a regulatory subunit  
4  
5 acting as an electron transport chain such as ubiquinol-cytochrome *c* reductase core protein  
6  
7 (*UQCRC1*) (Fig. S5), and an electron transfer flavoprotein alpha subunit (*ETFFA*) (Fig. S6).  
8  
9

## 10 11 12 **Availability of supporting data**

13  
14  
15  
16 *T. kingsejongensis* genome and transcriptome data are deposited in the Sequence Read  
17  
18 Archive (SRA) as BioProjects PRJNA307207 and PRJNA307513, respectively. Other  
19  
20 supporting data is available in the *GigaScience* repository, GigaDB [53].  
21  
22  
23  
24  
25  
26

## 27 **List of abbreviations**

28  
29  
30 ATP: Adenosine triphosphate; BUSCO: Benchmarking Universal Single-Copy Orthologs;  
31  
32 CEGMA: Core Eukaryotic Genes Mapping Approach; CM: Covariance model; DAVID:  
33  
34 Database for Annotation, Visualization and Integrated Discovery; *dN*: Nonsynonymous  
35  
36 mutations; *dS*: Synonymous mutations; EST: Expressed sequence tag; GO: Gene Ontology;  
37  
38 KEGG: Kyoto Encyclopedia of Genes and Genomes; LRT: Likelihood Ratio Test; OxPhos:  
39  
40 Oxidative phosphorylation; PAML: Phylogenetic Analysis by Maximum Likelihood;  
41  
42 PRANK: Probabilistic Alignment Kit; PSG: Positively selected gene; RAM: Random access  
43  
44 memory; SRA: Sequence Read Archive; SSR: Simple sequence repeat; TE: Transposable  
45  
46 element; *w*: *dN/dS* ratio  
47  
48  
49  
50  
51  
52  
53  
54  
55

## 56 **Competing interests**

57  
58  
59 The authors declare no competing interests.  
60  
61  
62  
63  
64  
65



1  
2  
3  
4 **Funding**  
5  
6

7 This work was supported by the Korea Polar Research Institute-funded the grant ‘Antarctic  
8 organisms: cold-adaptation mechanism and its application’ (PE16070), and basic research  
9 program (PE14260).  
10  
11  
12  
13  
14  
15  
16  
17

18 **Authors’ contributions**  
19  
20

21 HP, S Kim and HWK conceived and designed experiments and analyses; S Kang, DHA,  
22 SGL, SCS, JL, GSM and HL performed experiments and conducted bioinformatics.  
23  
24 Seunghyun Kang, HWK, S Kim and HP. wrote the paper. All authors read and approved the  
25  
26  
27  
28  
29 final manuscript.  
30  
31  
32  
33  
34

35 **Acknowledgements**  
36  
37

38 We would like to thank Joseph A. Covi for comments and discussion.  
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

## References

1. Huys R, Boxshall GA: *Copepod evolution*. Ray Society; 1991.
2. Humes AG: How many copepods? *Hydrobiologia* 1994, 292:1-7.
3. Wells P, Persoone G, Jaspers E, C. C: *Marine ecotoxicological tests with zooplankton*. In: Persoone, G., Jaspers, E., Claus, C. (Eds.), *Ecotoxicological Testing for the Marine Environment*. Inst. Mar. Sci. Res., Bredene; 1984.
4. Ruppert E, Fox R, Barnes R: *Invertebrate Zoology, A Functional Evolutionary Approach*. Brooks/Cole-Thomson Learning. Belmont, CA 2003.
5. Goolish E, Burton R: Energetics of osmoregulation in an intertidal copepod: Effects of anoxia and lipid reserves on the pattern of free amino accumulation. *Funct Ecol* 1989:81-89.
6. Lazzaretto I, Libertini A: Karyological comparison among different Mediterranean populations of the genus *Tigriopus* (Copepoda Harpacticoida). *Boll Zool* 2009, 53:197-201.
7. Davenport J, Barnett P, McAllen R: Environmental tolerances of three species of the harpacticoid copepod genus *Tigriopus*. *J Mar Biol Assoc UK* 1997, 77:3-16.
8. Raisuddin S, Kwok KW, Leung KM, Schlenk D, Lee J-S: The copepod *Tigriopus*: A promising marine model organism for ecotoxicology and environmental genomics. *Aquat Toxicol* 2007, 83:161-173.
9. Whole Genome Assembly of *Tigriopus californicus* provided by the Weizhong Li lab, UCSD Calit2 [[http://i5k.nal.usda.gov/Tigriopus\\_californicus](http://i5k.nal.usda.gov/Tigriopus_californicus)]
10. Lee J-S, Rhee J-S, Kim R-O, Hwang D-S, Han J, Choi B-S, Park GS, Kim I-C, Park HG, Lee Y-M: The copepod *Tigriopus japonicus* genomic DNA information (574Mb) and molecular anatomy. *Mar Environ Res* 2010, 69:S21-S23.
11. Whole genome assembly of *Eurytemora affinis*

- 1 [http://i5k.nal.usda.gov/Eurytemora\_affinis ]  
2  
3  
4 12. The Salmon Louse Genome Project [http://sealouse.imr.no/ ]  
5  
6  
7 13. Thorne MAS, Kagoshima H, Clark MS, Marshall CJ, Wharton DA: Molecular  
8  
9 analysis of the cold tolerant Antarctic Nematode, *Panagrolaimus davidi*. *PLOS one*  
10  
11 2014, 9:e104526.  
12  
13  
14 14. Everatta MJ, Worlandb MR, Balea JS, Conveyb P, Hayward SAL: Pre-adapted to the  
15  
16 maritime Antarctic? – Rapid cold hardening of the midge, *Eretmoptera murphyi*. *J*  
17  
18 *Insect Physiol* 2012, 58:1104–1111.  
19  
20  
21 15. Bromwich DH, Nicolas JP, Monaghan AJ, Lazzara MA, Keller LM, Weidner GA,  
22  
23 Wilson AB: Central West Antarctica among the most rapidly warming regions on  
24  
25 Earth. *Nature Geoscience* 2013, 6:139-145.  
26  
27  
28 16. Park E-O, Lee S, Cho M, Yoon SH, Lee Y, Lee W: A new species of the genus  
29  
30 *Tigriopus* (Copepoda: Harpacticoida: Harpacticidae) from Antarctica. *Proc Biol Soc*  
31  
32 *Wash* 2014, 127:138-154.  
33  
34  
35  
36 17. Birkenmajer K: Geology of Admiralty Bay, King George Island (South Shetland  
37  
38 Islands). An outline. *Pol Polar Res* 1980, 1:29-54.  
39  
40  
41 18. Marçais G, Kingsford C: A fast, lock-free approach for efficient parallel counting of  
42  
43 occurrences of k-mers. *Bioinformatics* 2011, 27:764-770.  
44  
45  
46 19. Myers EW, Sutton GG, Delcher AL, Dew IM, Fasulo DP, Flanigan MJ, Kravitz SA,  
47  
48 Mobarry CM, Reinert KH, Remington KA, et al: A whole-genome assembly of  
49  
50 *Drosophila*. *Science* 2000, 287:2196-2204.  
51  
52  
53 20. Gordon A, Hannon G: Fastx-toolkit. *FASTQ/A short-reads preprocessing tools*  
54  
55 *(unpublished)* [http://hannonlab.cshl.edu/fastx\\_toolkit](http://hannonlab.cshl.edu/fastx_toolkit) 2010.  
56  
57  
58 21. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W: Scaffolding pre-assembled  
59  
60 contigs using SSPACE. *Bioinformatics* 2011, 27:578-579.  
61  
62  
63  
64  
65

- 1 22. Nadalin F, Vezzi F, Policriti A: GapFiller: a *de novo* assembly approach to fill the gap  
2 within paired reads. *BMC Bioinformatics* 2012, 13:S8.  
3  
4
- 5 23. Holt C, Yandell M: MAKER2: an annotation pipeline and genome-database  
6 management tool for second-generation genome projects. *BMC Bioinformatics* 2011,  
7 12:491.  
8  
9
- 10 24. Smit AFA HR, Green, P.: RepeatMasker Open-3.0. 1996-2004  
11 (<http://www.RepeatMasker.org>).  
12  
13
- 14 25. Korf I: Gene finding in novel genomes. *BMC Bioinformatics* 2004, 5:59.26.  
15  
16
- 17 26. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ:  
18 Gapped BLAST and PSI-BLAST: a new generation of protein database search  
19 programs. *Nucleic Acids Res* 1997, 25:3389-3402.  
20  
21
- 22 27. Trapnell C, Pachter L, Salzberg SL: TopHat: discovering splice junctions with RNA-  
23 Seq. *Bioinformatics* 2009, 25:1105-1111.  
24  
25
- 26 28. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg  
27 SL, Wold BJ, Pachter L: Transcript assembly and quantification by RNA-Seq reveals  
28 unannotated transcripts and isoform switching during cell differentiation. *Nat Biotech*  
29 2010, 28:511-515.  
30  
31
- 32 29. Slater GS, Birney E: Automated generation of heuristics for biological sequence  
33 comparison. *BMC Bioinformatics* 2005, 6:31.  
34  
35
- 36 30. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M: Blast2GO: a  
37 universal tool for annotation, visualization and analysis in functional genomics  
38 research. *Bioinformatics* 2005, 21:3674-3676.  
39  
40
- 41 31. Boeckmann B, Bairoch A, Apweiler R, Blatter M-C, Estreicher A, Gasteiger E,  
42 Martin MJ, Michoud K, O'Donovan C, Phan I: The SWISS-PROT protein  
43 knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res* 2003,  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

- 1 31:365-370.  
2  
3  
4 32. Acland A, Agarwala R, Barrett T, Beck J, Benson DA, Bollin C, Bolton E, Bryant  
5  
6 SH, Canese K, Church DM: Database resources of the national center for  
7  
8 biotechnology information. *Nucleic Acids Res* 2014, 42:D7.  
9  
10  
11 33. Supek F, Bošnjak M, Škunca N, Šmuc T: REVIGO summarizes and visualizes long  
12  
13 lists of gene ontology terms. *PloS one* 2011, 6:e21800.  
14  
15  
16 34. Nawrocki EP, Kolbe DL, Eddy SR: Infernal 1.0: inference of RNA alignments.  
17  
18 *Bioinformatics* 2009, 25:1335-1337.  
19  
20  
21 35. Gardner PP, Daub J, Tate J, Moore BL, Osuch IH, Griffiths-Jones S, Finn RD,  
22  
23 Nawrocki EP, Kolbe DL, Eddy SR, Bateman A: Rfam: Wikipedia, clans and the  
24  
25 "decimal" release. *Nucleic Acids Res* 2011, 39:D141-145.  
26  
27  
28 36. Lowe TM, Eddy SR: tRNAscan-SE: a program for improved detection of transfer  
29  
30 RNA genes in genomic sequence. *Nucleic Acids Res* 1997, 25:955-964.  
31  
32  
33 37. Parra G, Bradnam K, Korf I: CEGMA: a pipeline to accurately annotate core genes in  
34  
35 eukaryotic genomes. *Bioinformatics* 2007, 23:1061-1067.  
36  
37  
38 38. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM: BUSCO:  
39  
40 assessing genome assembly and annotation completeness with single-copy orthologs.  
41  
42 *Bioinformatics* 2015:btv351.  
43  
44  
45 39. Chipman AD, Ferrier DE, Brena C, Qu J, Hughes DS, Schröder R, Torres-Oliva M,  
46  
47 Znassi N, Jiang H, Almeida FC: The first myriapod genome sequence reveals  
48  
49 conservative arthropod gene content and genome organisation in the centipede  
50  
51 *Strigamia maritima*. *PLoS Biol* 2014, 12:e1002005.  
52  
53  
54  
55 40. Li L, Stoeckert CJ, Roos DS: OrthoMCL: identification of ortholog groups for  
56  
57 eukaryotic genomes. *Genome Res* 2003, 13:2178-2189.  
58  
59  
60 41. Löytynoja A, Goldman N: An algorithm for progressive multiple alignment of  
61  
62  
63  
64  
65

- 1 sequences with insertions. *Proc Natl Acad Sci U S A* 2005, 102:10557-10562.
- 2
- 3
- 4 42. Castresana J: Selection of conserved blocks from multiple alignments for their use in
- 5 phylogenetic analysis. *Mol Biol Evol* 2000, 17:540-552.
- 6
- 7
- 8 43. Stamatakis A: RAxML version 8: a tool for phylogenetic analysis and post-analysis of
- 9 large phylogenies. *Bioinformatics* 2014, 30:1312-1313.
- 10
- 11
- 12
- 13 44. Hedges SB, Dudley J, Kumar S: TimeTree: a public knowledge-base of divergence
- 14 times among organisms. *Bioinformatics* 2006, 22:2971-2972.
- 15
- 16
- 17
- 18 45. Han MV, Thomas GW, Lugo-Martinez J, Hahn MW: Estimating gene gain and loss
- 19 rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol*
- 20 *Biol Evol* 2013, 30:1987-1997.
- 21
- 22
- 23
- 24
- 25 46. Rider SD, Morgan MS, Arlian LG: Draft genome of the scabies mite. *Parasites &*
- 26 *Vectors* 2015, 8:585.
- 27
- 28
- 29
- 30 47. Hoy M, Waterhouse R, Wu K, Estep A, Ioannidis P, Palmer W, Pomerantz A, Simão
- 31 F, Thomas J, Jiggins F: Genome sequencing of the phytoseiid predatory mite
- 32 *Metaseiulus occidentalis* reveals completely atomised Hox genes and super-dynamic
- 33 intron evolution. *Genome biology and evolution* 2016, 8:1762-1775.
- 34
- 35
- 36
- 37
- 38
- 39
- 40 48. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M: KEGG as a reference
- 41 resource for gene and protein annotation. *Nucleic Acids Res* 2015:D457–D462.
- 42
- 43
- 44
- 45 49. Yang Z: PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*
- 46 2007, 24:1586-1591.
- 47
- 48
- 49
- 50 50. Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, Storz JF, Antunes A, Greenwold MJ,
- 51 Meredith RW: Comparative genomics reveals insights into avian genome evolution
- 52 and adaptation. *Science* 2014, 346:1311-1320.
- 53
- 54
- 55
- 56
- 57 51. Consortium TCSaA: Initial sequence of the chimpanzee genome and comparison with
- 58 the human genome. *Nature* 2005, 437:69-87.
- 59
- 60
- 61
- 62
- 63
- 64
- 65

- 1 52. Huang DW, Sherman BT, Lempicki RA: Systematic and integrative analysis of large  
2 gene lists using DAVID bioinformatics resources. *Nature protocols* 2008, 4:44-57.  
3  
4  
5  
6 53. Kang S; Ahn D; Lee JH; Lee SG; Shin SC; Lee J; Min G; Lee H; Kim H; Kim S; Park  
7 H. Supporting data for "The genome of the Antarctic-endemic copepod, *Tigriopus*  
8 *kingsejongensis*". GigaScience Database. 2016. <http://dx.doi.org/10.5524/100249>  
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 **Figure legends**  
2  
3

4 **Figure 1. Photograph of an adult *Tigriopus kingsejongensis* specimen (scale bar = 200  $\mu$**   
5 **m).**  
6  
7

8  
9  
10  
11 **Figure 2. Estimation of the *Tigriopus kingsejongensis* genome size based on 33-mer**  
12 **analysis.**  
13

14 X-axis represents the depth (peak at 39 $\times$ ) and the y-axis represents the proportion. Genome  
15 size was estimated to be 298 Mb (total k-mer number/volume peak).  
16  
17  
18  
19  
20  
21  
22

23 **Figure 3. Scaffold and contig size distributions of *Tigriopus kingsejongensis*.**  
24

25 The percentage of the assembly included (y-axis) in contigs or scaffolds of a minimum size  
26 (x-axis, log scale) is shown for the contig (red) and scaffold (blue).  
27  
28  
29  
30  
31  
32

33 **Figure 4. Comparative genome analyses of the *T. kingsejongensis* genome.**  
34

35 A. Venn diagram of orthologous gene clusters between four arthropod lineages.  
36  
37

38 B. Gene family gain-and-loss analysis. The number of gained gene families (red), lost gene  
39 families (blue) and orphan gene families (black) are indicated for each species. Time lines  
40 specify divergence times between the lineages.  
41  
42  
43  
44  
45  
46  
47

48 **Figure 5. *Tigriopus kingsejongensis*-specific adaptive evolution.**  
49

50 A. Global mean  $w$  (ratio of nonsynonymous ( $dN$ ) to synonymous mutations ( $dS$ )) distribution  
51 by GO categories of *T. kingsejongensis* and *T. japonicus*. GO categories showing supposedly  
52 accelerated nonsynonymous divergence (binomial test, test statistic  $< 0.05$ ) in *T.*  
53 *kingsejongensis* and *T. japonicus* are colored in red and blue, respectively.  
54  
55  
56  
57  
58  
59

60 B. A total of seven enzyme-coding genes were positively selected genes (PSGs) involved in  
61  
62  
63  
64  
65



1 the four metabolic pathways (oval frame) of *T. kingsejongensis*: energy (purple), nucleotide  
2  
3 (red), lipid (green), and carbohydrate (blue) metabolic pathways. The three genes belonging  
4  
5 to the oxidative phosphorylation pathway (KEGG pathway map00190) (rectangular frame)  
6  
7 are presented below the enzymes involved. Solid lines indicate direct processes and dashed  
8  
9 lines indicate that more than one step is involved in a process.  
10  
11  
12  
13  
14  
15

## 16 **Table legends**

17  
18  
19 **Table 1. DNA library statistics.**  
20  
21

22 **Table 2. Transcriptome sequencing and assembly analysis for *Tigriopus japonicus*.**  
23  
24

25 **Table 3. RNA-seq statistics analysis for *Tigriopus kingsejongensis*.**  
26  
27

28 **Table 4. Genome assembly statistics.**  
29  
30

31 **Table 5. *Tigriopus kingsejongensis* genes: general statistics.**  
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

Library		Reads (n)	Average length	Sequences (bp) (n)	Reads (trimmed) (n)	Average length	Sequences (trimmed) (n)
Paired-end	Sum	99 710 266		29 271 916 613	65 644 374		14 668 956 871
	350S1	6 661 392	300	2 005 078 992	4 446 394	233	1 034 231 244
	350S2	4 933 058	265	1 311 700 122	4 618 711	211	975 471 763
	400S1	65 668 598	300	19 766 247 998	36 863 154	228	8 397 426 481
	450S1	3 418 988	300	1 029 115 388	2 812 455	230	646 302 159
	450S2	8 009 162	245	1 968 652 020	7 660 814	199	1 527 566 312
	500S1	11 019 068	289	3 191 122 093	9 242 846	226	2 087 958 911
Mate-Paired	Sum	103 373 998		7 753 049 850	73 515 391		5 169 006 268
	3KS1	8 374 238	75	628 067 850	6 745 546	73	493 099 413
	3KS2	9 250 994	75	693 824 550	5 281 513	65	344 618 723
	3KS3	51 349 594	75	3 851 219 550	39 147 167	72	2 816 638 666
	3KS4	3 063 232	75	229 742 400	1 740 986	65	112 554 745
	8KS1	9 847 636	75	738 572 700	7 887 612	73	572 246 251
	8KS2	16 322 038	75	1 224 152 850	9 653 293	65	630 842 698
	8KS3	5 166 266	75	387 469 950	3 059 274	65	199 005 774
Total		203 084 264		37 024 966 463	139 159 765		19 837 963 139
Coverage (folds)				120.7			64.7

**Table 1.**

1  
2  
3  
4 **Table 2.**  
5  
6

<b>Sequencing</b>	
Total reads (n)	37 956 160
Total bases (n)	7 714 415 316
Trimmed reads (n)	35 577 636
Trimmed bases (n)	5 989 188 343
<b>Assembly</b>	
Contigs (n)	40 172
Total contig length (bases)	28 850 726
N50 contig length (bases)	1093
Max scaffold length (bases)	23 942
<b>Annotation</b>	
With BLAST results	20 392
Without BLAST hits	7090
With mapping results	8172
Annotated sequences	4518

29  
30  
31 **Table 3.**  
32  
33

<b>Temperature</b>		
	<b>4°C</b>	<b>15°C</b>
Total reads (n)	15 786 118	16 417 072
Total bases (n)	3 567 662 668	3 763 295 032
Trimmed reads (n)	14 845 103	15 388 513
Trimmed bases (n)	2 761 189 158	2 833 805 442

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 **Table 4.**  
5

Type	Parameter	Assembly size according to Celera Assembler
Scaffold	Total scaffold length (bases)	295 233 602
	Gap size (bases)	10 474 460
	Scaffolds (n)	11 558
	N50 scaffold length (bases)	159 218
	Max scaffold length (bases)	3 401 446
Contig	Total contig length (bases)	305 712 242
	Contigs (n)	48 368
	N50 contig length (bases)	17 566
	Max contig length (bases)	349 507

27  
28  
29  
30  
31 **Table 5.**  
32

Genes (n)	12 772
Gene length sum (bp)	82 293 116
Exons per genes (n)	4.6
mRNA length sum (bp)	43 306 342
Average mRNA length (bp)	1090
Number of tRNA	1393
Number of rRNA	215

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

**Table 6.**

Species	<i>Tigriopus kingsejongensis</i>	<i>Daphnia pulex</i>	<i>Ixodes scapularis</i>	<i>Mesobuthus martensii</i>	<i>Strigamia maritima</i>	<i>Tetranychus urticae</i>	<i>Drosophila melanogaster</i>	<i>Aedes aegypti</i>
Assembly	This study	GCA_000187875.1	GCA_000208615.1	GCA_000484575.1	Smar1.22	GCA_000239435.1	Dmel_r5.55	AaegL3
Sample type	genome	genome	genome	genome	genome	genome	genome	genome
CEGMA <sup>a</sup>	83/77.8	99.2/98.8	79.8/41.9 <sup>g</sup>	57.3/24.2 <sup>g</sup>	95.1 <sup>f</sup>	98.0/95.2 <sup>g</sup>	100/100	99.2/83.5
BUSCO <sup>b</sup>	61.1 [10.5], 10.7, 28.1	83 [3.9], 11, 5.1 <sup>e</sup>	68.9 [2.4], 21.0, 10.1 <sup>g</sup>	34.4 [4.0], 23.0, 42.7 <sup>g</sup>	84 [5.9], 12, 3.2 <sup>e</sup>	68.8 [5.8], 9.9, 21.3 <sup>g</sup>	98 [6.4], 0.6, 0.3 <sup>e</sup>	86 [13], 10, 3.2 <sup>e</sup>
BUSCO <sup>c</sup>	70.9 [13.6], 6.0, 23.0							
BUSCO <sup>d</sup>	67.1 [16.8], 5.1, 27.7							

<sup>a</sup>248 CEGMA genes found/complete

<sup>b</sup>BUSCO Arthropods complete [duplicated], fragmented, missing

<sup>c</sup>BUSCO Metazoa complete [duplicated], fragmented, missing

<sup>d</sup>BUSCO Eukaryotes complete [duplicated], fragmented, missing

<sup>e</sup> [38]

<sup>f</sup> [39]

<sup>g</sup> [47]

1 **Table 7.**

2  
3  
4

5 Species	6 Source of data	7 No. of coding genes	8 No. of gene families	9 No. of genes in gene families	10 No. of orphan genes	11 No. of unique gene families	12 Average No. of genes in gene families
13 <i>Aedes aegypti</i>	14 Ensembl genome 25	15 15 797	16 7958	17 12 792	18 7839	19 854	2.10
20 <i>Caenorhabditis elegans</i>	21 Ensembl gene 78	22 20 447	23 6536	24 13 737	25 13 911	26 1528	1.29
27 <i>Ciona intestinalis</i>	28 Ensembl gene 78	29 16 671	30 7017	31 9058	32 9654	33 503	1.25
34 <i>Daphnia pulex</i>	35 Ensembl genome 25	36 30 590	37 6710	38 8362	39 7208	40 368	2.27
41 <i>Drosophila melanogaster</i>	42 Ensembl gene 78	43 13 918	44 9673	45 21 917	46 20 917	47 2408	1.98
48 <i>Homo sapiens</i>	49 Ensembl gene 78	50 20 300	51 8696	52 17 186	53 11 604	54 1065	1.39
55 <i>Ixodes scapularis</i>	56 Ensembl genome 25	57 20 486	58 8097	59 11 277	60 12 389	61 873	2.38
62 <i>Mesobuthus martensii</i>	63 <a href="http://lifecenter.sgst.cn/main/en/scorpion.jsp">http://lifecenter.sgst.cn/main/en/scorpion.jsp</a>	64 32 016	65 8389	66 19 961	67 23 627	68 2276	1.43
69 <i>Strigamia maritima</i>	70 Ensembl genome 25	71 14 992	72 7727	73 11 012	74 7265	75 583	1.79
76 <i>Tetranychus urticae</i>	77 Ensembl genome 25	78 18 224	79 6602	80 11 788	81 11 622	82 939	1.42
83 <i>Tigriopus kingsejongensis</i>	84 this study	85 12 772	86 6205	87 8813	88 6567	89 649	

90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100

## Supplementary figures

Figure S1. Map showing location of the *Tigriopus kingsejongensis* sampling site.

Figure S2. BLAST top-hit species distribution of *Tigriopus kingsejongensis*.

Data obtained using BLASTx against the National Center for Biotechnology Information's (NCBI) non-redundant protein database with an E value cutoff of  $1e^{-5}$ .

Figure S3. Gene Ontology distribution of annotated genes.

Gene Ontology (GO) annotation of predicted *Tigriopus kingsejongensis* genes was conducted using the GO annotation. The figure illustrates the number of genes from major GO modules of molecular function (MF), biological process (BP), and cellular component (CC).

1 **Figure S4. *Tigriopus kingsejongensis*-specific amino acid changes in ATP synthase**  
2  
3 **subunit alpha.**

4  
5  
6 **A.** Clustal X alignment of the amino acid sequences between four species. *Tigriopus*  
7  
8 *kingsejongensis*-specific amino acid changes representing positive selections are presented  
9  
10 with red boxes.

11  
12 **B.** Cartoon of the protein crystal structure of the ATP synthase (PDB ID: 1BMF).

13  
14  
15 **C.** The specific amino acid change Ala166 is colored in red (in stick form) and positioned  
16  
17 within the external loop region of nucleotide-binding domain. The three domains of the ATP  
18  
19 synthase subunit alpha illustrated in cartoon form are colored accordingly (blue, beta-barrel  
20  
21 domain; green, nucleotide-binding domain; purple: C terminal domain).  
22  
23  
24  
25  
26

27  
28 **Figure S5. *Tigriopus kingsejongensis*-specific amino acid changes in ubiquinol-**  
29 **cytochrome c reductase core protein I.**

30  
31  
32 **A.** Clustal X alignment of the amino acid sequences between four species. *Tigriopus*  
33  
34 *kingsejongensis*-specific amino acid changes representing positive selections are presented  
35  
36 with red boxes.  
37  
38

39  
40 **B.** Cartoon of the protein crystal structure of ubiquinol-cytochrome c reductase (PDB ID:  
41  
42 1QCR).  
43

44  
45 **C.** Positions of the specific amino acid changes in ubiquinol-cytochrome c reductase core  
46  
47 protein I are colored red (stick form). The insulinase domain is yellow and the peptidase M16  
48  
49 domain is green.  
50  
51  
52  
53

54  
55 **Figure S6. *Tigriopus kingsejongensis*-specific amino acid changes in electron-**  
56 **transferring flavoprotein.**

57  
58  
59 **A.** Clustal X alignment of the amino acid sequences between four species. *Tigriopus*  
60  
61  
62  
63  
64  
65

1 *kingsejongensis*-specific amino acid changes representing positive selections are presented  
2  
3 with red boxes. Among the ten amino acid changes, the five sites are located within the N-  
4  
5 terminal domain and the other five are positioned within the FAD binding domain.  
6  
7

8 **B.** Cartoon of the protein crystal structure of the electron-transferring flavoprotein (PDB ID:  
9  
10 1EFV). The five amino acid sites within the FAD binding domain are colored in red (stick  
11  
12 form). Electron-transferring flavoprotein alpha subunit is green; FAD-binding domain is  
13  
14 represented by color-coded electrostatic surface (blue, positive charge; red, negative charge;  
15  
16 grey, neutral charge); FAD is orange (stick form). Notably, the Asp463 residue makes a salt  
17  
18 bridge with Arg437 in the homology model structure of electron-transferring flavoprotein  
19  
20 from *T. kingsejongensis*. In addition, Gln454 is located near the bound FAD co-factor and  
21  
22 may form a hydrogen bond with the N7A atom of FAD in the model structure of electron-  
23  
24 transferring flavoprotein from *T. kingsejongensis*.  
25  
26  
27  
28  
29  
30  
31  
32

### 33 **Supplementary tables**

34  
35 **Table S1. Number of tRNA in the *Tigriopus kingsejongensis* genome.**  
36  
37  
38  
39

40 **Table S2. Known repetitive and transposable elements in the *Tigriopus kingsejongensis***  
41  
42 **genome.**  
43  
44  
45

46  
47 **Table S3. Transposable elements in the *Tigriopus kingsejongensis* genome.**  
48  
49

50 **Table S4. Gene Ontology (GO) of lineage-specific gene families in the *Tigriopus***  
51  
52 ***kingsejongensis* genome.**  
53  
54

55 REVIGO software was used to cluster related GO terms (in bold letters) according to *P*-  
56  
57 value.  
58  
59  
60  
61  
62  
63  
64  
65



1 **Table S5. Annotated domains of lineage-specific gene families in the *Tigriopus***  
2  
3 ***kingsejongensis* genome.**  
4

5  
6  
7  
8 **Table S6. Gene Ontology (GO) of expanded gene families in the *Tigriopus***  
9 ***kingsejongensis* genome.**  
10

11  
12  
13 REVIGO software was used to cluster related GO terms (in bold letters) according to *p*-value.  
14  
15

16  
17  
18 **Table S7. Gene annotation of the expanded genes in the *Tigriopus kingsejongensis***  
19 **genome.**  
20

21  
22  
23  
24  
25 **Table S8. Gene Ontology (GO) of contracted genes in the *Tigriopus kingsejongensis***  
26 **genome.**  
27

28  
29  
30 REVIGO software was used to cluster related GO terms (in bold letters) according to *P*-  
31  
32 value.  
33  
34

35  
36  
37 **Table S9. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway of expanded**  
38 **genes in the *Tigriopus kingsejongensis* genome.**  
39

40  
41  
42  
43  
44  
45 **Table S10. Gene Ontology (GO) categories displaying *w* (ratio of nonsynonymous (dN)**  
46 **to synonymous mutations (dS)) in the genomes of *Tigriopus kingsejongensis* and *T.***  
47 ***japonicus*.**  
48  
49

50  
51  
52  
53  
54  
55 **Table S11. Lists and annotations of positively selected genes in the *Tigriopus***  
56 ***kingsejongensis* genome.**  
57  
58

59  
60  
61  
62  
63  
64  
65

1 **Table S12. Lists and annotations of positively selected genes in the *Tigriopus japonicus***  
2  
3  
4 **genome.**

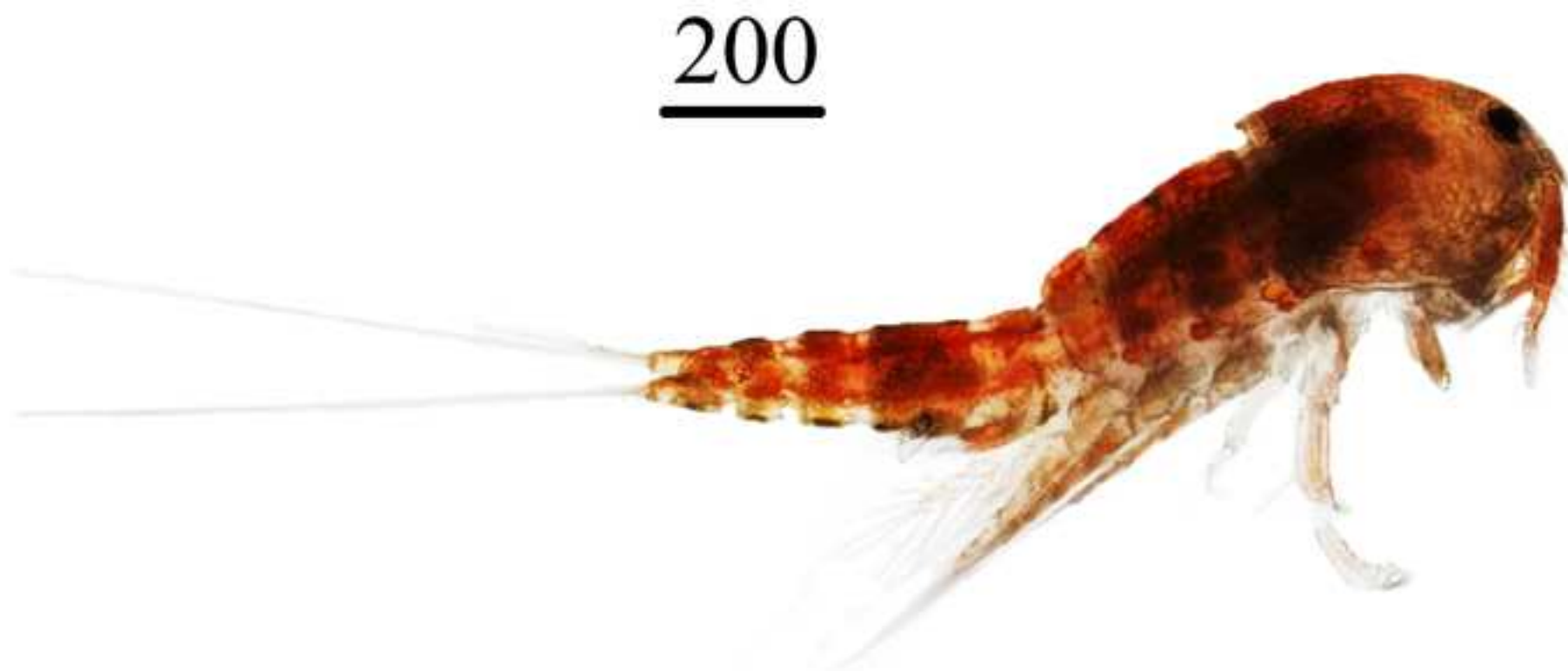
5  
6  
7  
8  
9 **Table S13. Enriched Gene Ontology (GO) categories identified by positively selected**  
10 **genes from the *Tigriopus kingsejongensis* genome.**

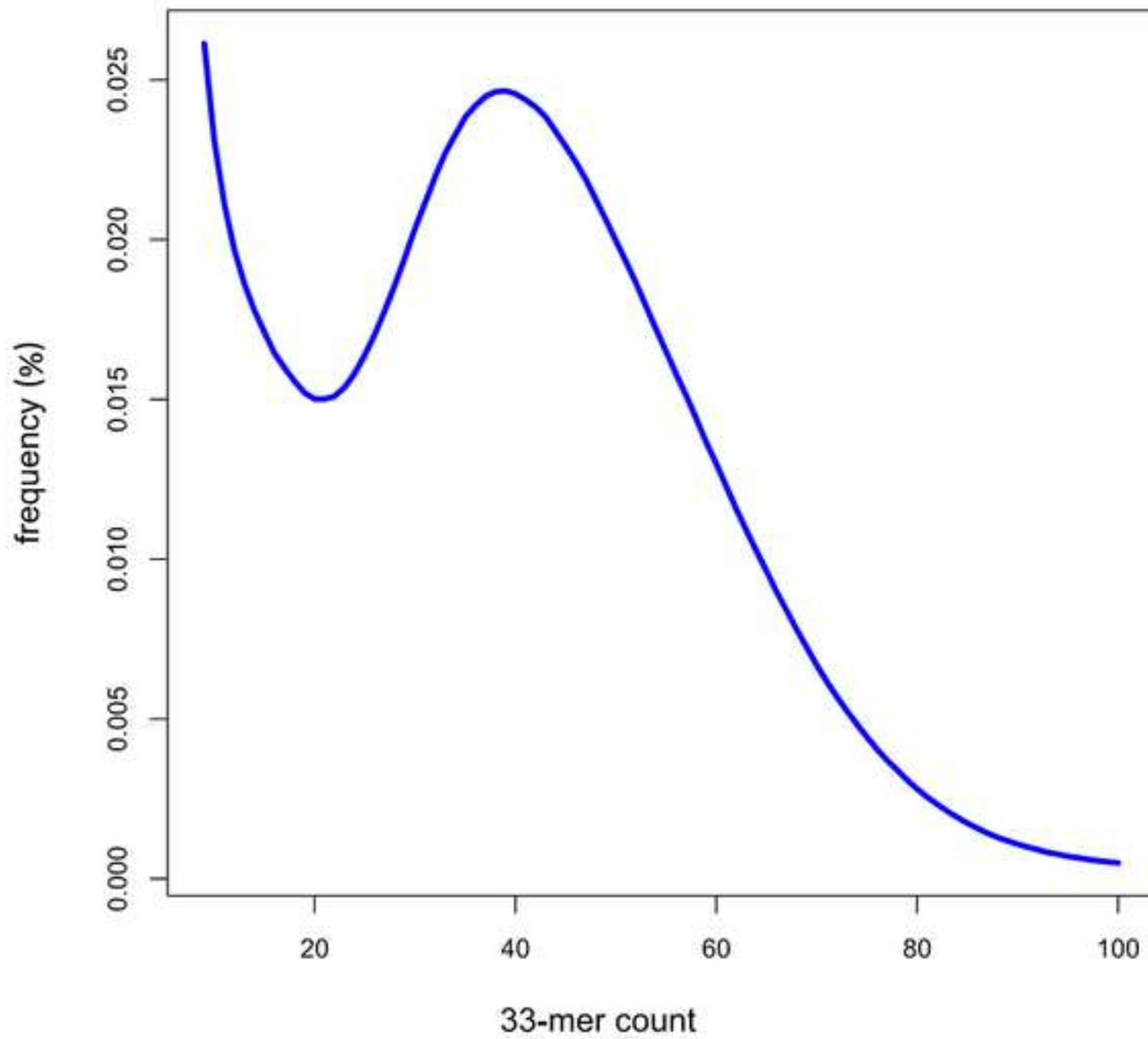
11  
12  
13 REVIGO software was used to cluster related GO terms (in bold letters) according to *P*-  
14  
15  
16 value.

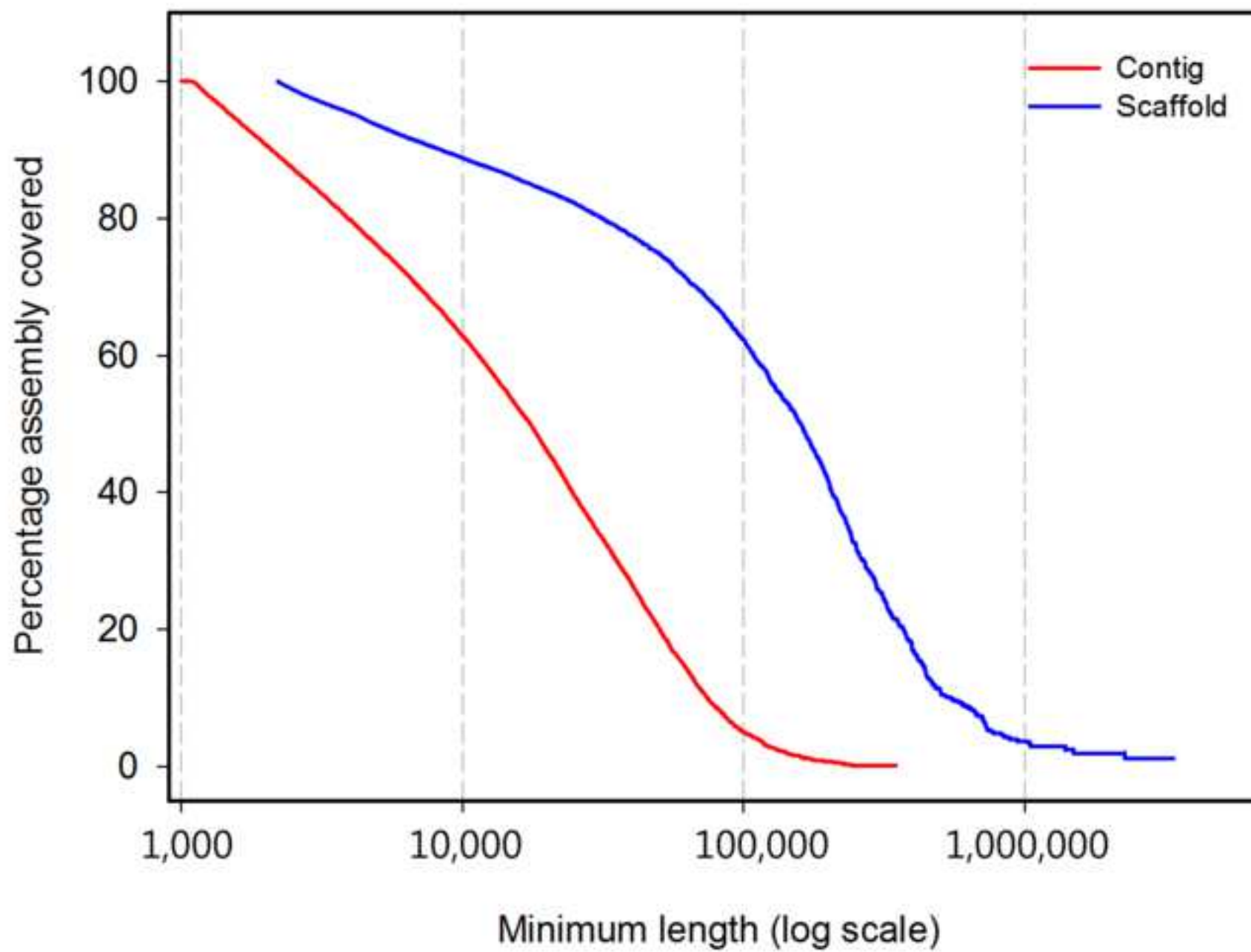
17  
18  
19  
20  
21 **Table S14. Enriched Gene Ontology (GO) categories identified by positively selected**  
22 **genes from the *Tigriopus japonicus* genome.**

23  
24  
25  
26 REVIGO software was used to cluster related GO terms (in bold letters) according to *P*-  
27  
28  
29 value.

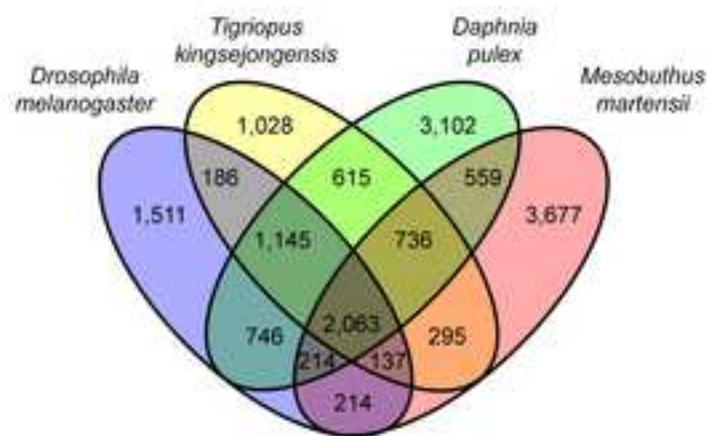
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65



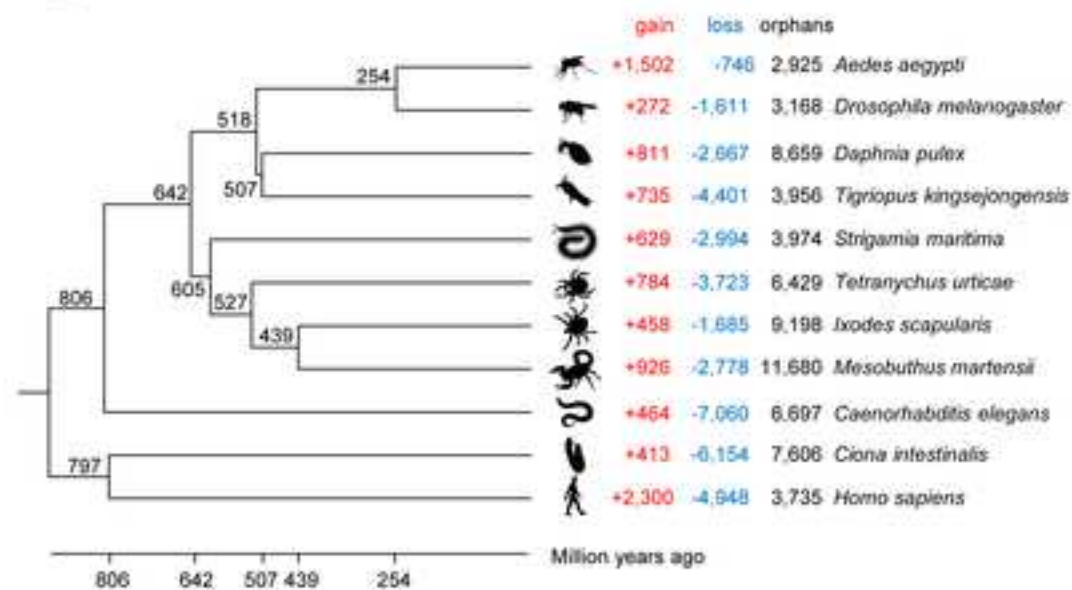




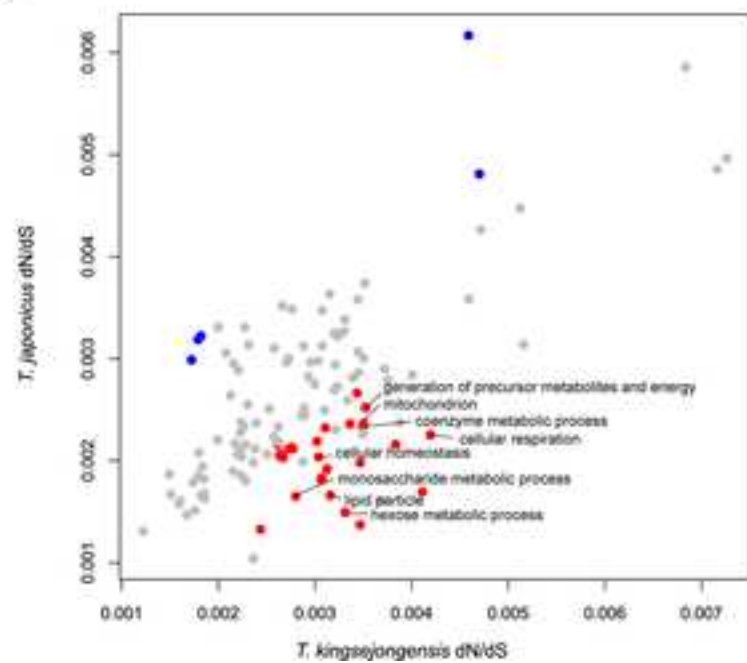
A



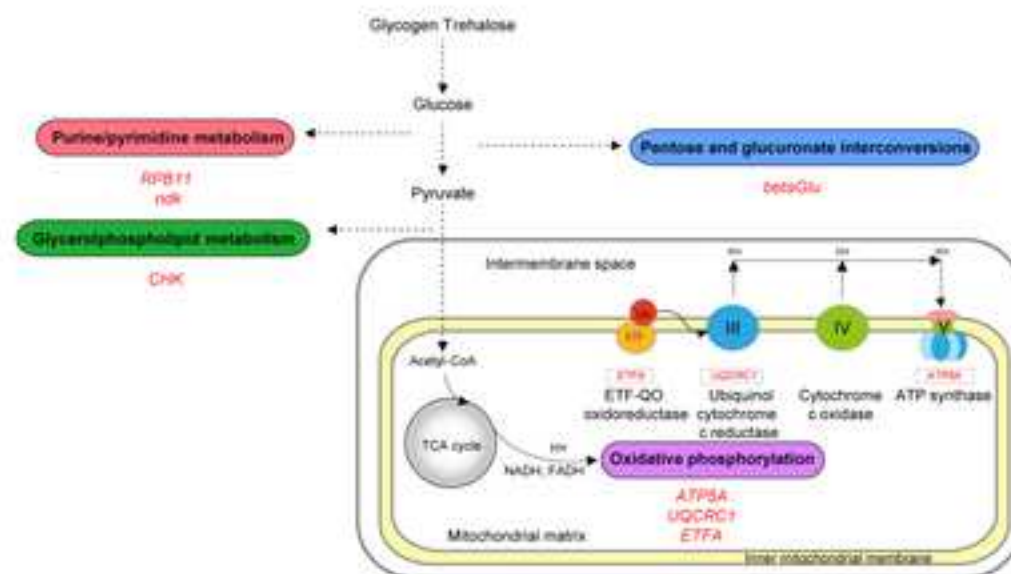
B

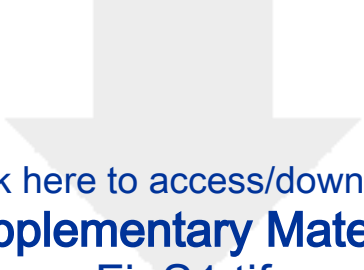


A

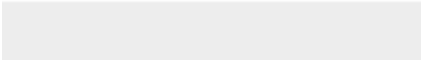



B

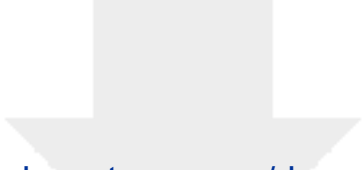




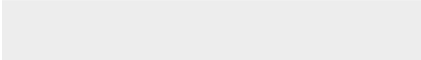

Click here to access/download  
**Supplementary Material**  
FigS1.tif

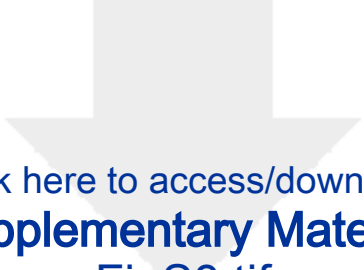




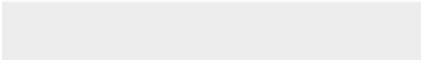



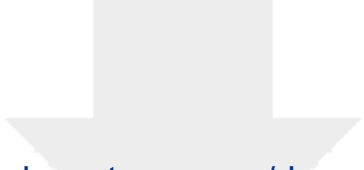
Click here to access/download  
**Supplementary Material**  
FigS2.tif



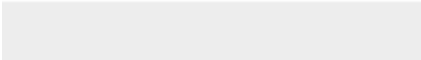



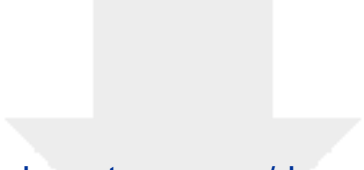
Click here to access/download  
**Supplementary Material**  
FigS3.tif



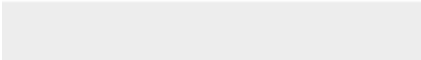



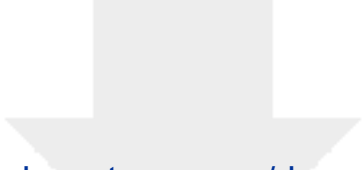
Click here to access/download  
**Supplementary Material**  
FigS4.tif






Click here to access/download  
**Supplementary Material**  
FigS5.tif





Click here to access/download  
**Supplementary Material**  
FigS6.tif





Click here to access/download  
**Supplementary Material**  
TableS1.docx




Click here to access/download  
**Supplementary Material**  
TableS2.docx






Click here to access/download  
**Supplementary Material**  
TableS3.docx






Click here to access/download  
**Supplementary Material**  
TableS4.docx



Click here to access/download  
**Supplementary Material**  
TableS5.docx



Click here to access/download  
**Supplementary Material**  
TableS6.docx



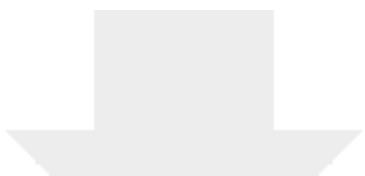
Click here to access/download  
**Supplementary Material**  
TableS7.docx



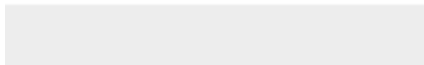

Click here to access/download  
**Supplementary Material**  
TableS8.docx



Click here to access/download  
**Supplementary Material**  
TableS9.docx



Click here to access/download  
**Supplementary Material**  
TableS10.docx

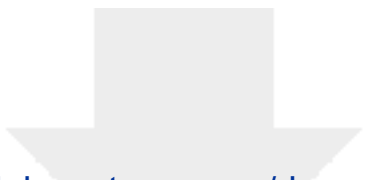




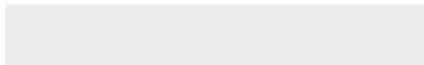

Click here to access/download  
**Supplementary Material**  
TableS11.docx

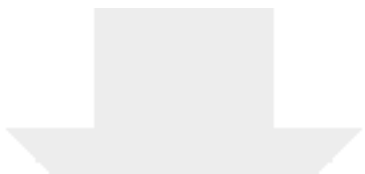




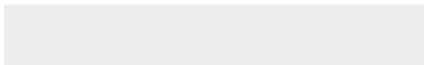



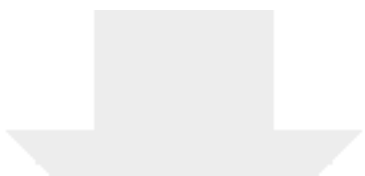
Click here to access/download  
**Supplementary Material**  
TableS12.docx





Click here to access/download  
**Supplementary Material**  
TableS13.docx





Click here to access/download  
**Supplementary Material**  
TableS14.docx

