

1 **Mitochondrial genome annotation and phylogenetic placement of *Oreochromis***  
2 ***andersonii* and *O. macrochir* among the cichlids of southern Africa**

3

4 Ian Bbole<sup>1,#\*</sup>, Jin-Liang Zhao<sup>2,3,4</sup>, Shou-Jie Tang<sup>2</sup>, Cyprian Katongo<sup>5</sup>

5 <sup>1</sup> Department of Fisheries, Mansa, Zambia

6 <sup>2,#</sup> Key Laboratory of Freshwater Aquatic Genetic Resources, Ministry of Agriculture,

7 Shanghai Ocean University, Shanghai 201306, China

8 <sup>3</sup>Centre for Research on Environmental Ecology and Fish Nutrition (CREEFN), Ministry of

9 Agriculture, Shanghai Ocean University, Shanghai 201306, China

10 <sup>4</sup>National Demonstration centre for Experimental Fisheries Science Education, Shanghai

11 Ocean University, Shanghai 201306, China

12 <sup>5</sup> Biological Sciences Department, University of Zambia, Lusaka, Zambia.

13

14 \* Corresponding author:

15 Email: [ibbole@gmail.com](mailto:ibbole@gmail.com) (IB)

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

## 33 **Abstract**

34 Genetic characterization of southern African cichlids has not received much attention. Here,  
35 we describe the mitogenome sequences and phylogenetic positioning of *Oreochromis andersonii* and  
36 *O. macrochir* among the cichlids of southern Africa. The complete mitochondrial DNA sequences  
37 were determined for *O. andersonii* and *O. macrochir*, two important aquaculture and fisheries species  
38 endemic to southern Africa. The complete mitogenome sequence lengths were 16642 bp and 16644  
39 bp for *O. andersonii* and *O. macrochir* respectively. The general structural organization follows that  
40 of other teleost species with 13 protein-coding genes, 2 *rRNAs*, 22 *tRNAs* and a non-coding control  
41 region. Phylogenetic placement of the two species among other African cichlids was performed using  
42 Maximum Likelihood (ML) and Bayesian Markov-Chain-Monte-Carlo (MCMC). The consensus  
43 trees confirmed the relative positions of the two cichlid species with *O. andersonii* being very closely  
44 related to *O. mossambicus* and *O. macrochir* showing a close relation to both species. Among the 13  
45 mitochondrial DNA protein coding genes *ND6* may have evolved more rapidly and *COIII* was the  
46 most conserved. There are signs that *ND6* may have been subjected to positive selection in order for  
47 these cichlid lineages to diversity and adapt to new environments. More work is needed to  
48 characterize the southern Africa cichlids as they are important species for capture fisheries,  
49 aquaculture development and understanding biogeographic history of African cichlids. Bio-  
50 conservation of some endangered cichlids is also essential due to the threat by invasive species.

## 51 **Introduction**

52 Africa is the origin centre for cichlid diversity with well over 2000 species having diverse  
53 morphology, behaviour and ecology [1]. In Southern Africa *Oreochromis andersonii* (Castelnau 1861)  
54 and *Oreochromis macrochir* (Boulenger 1912) are two important mouth brooding endemic cichlid  
55 species in this region [2]. The former occurs in the upper Zambezi, Middle Zambezi, Kafue,  
56 Okavango and Cunene Rivers while the latter is distributed in the Upper Zambezi, Kafue, and Congo  
57 River systems [2-4]. *Oreochromis macrochir* has further been introduced to the Hawaiian Islands,  
58 Okavango and Ngami region and Cunene River basin [4]. *Oreochromis andersonii* (Three-spot tilapia)

59 and *O. macrochir* (Green head tilapia) are important for both capture fisheries and aquaculture in  
60 Southern Africa [5]. However, due to the increase in fishing pressure as a result of an ever growing  
61 human population in this region as well as the introduction of Nile tilapia (*O. niloticus*) in almost all  
62 river systems where the native species occur, populations of these native species has greatly dwindled  
63 to vulnerable levels [6,7]. Nile tilapia hybridization with these native species and consequent decline  
64 in their population may as well make these species critically endangered in some southern African  
65 rivers such as the Kafue River [8-10]. Despite Nile tilapia dominating aquaculture production in  
66 Southern Africa, efforts are been made to domesticate native species. This is partly as a result of the  
67 growing concern of ecosystem changes in most river systems due to Nile tilapia invasion. A number  
68 of studies in Zambia have shown potential for aquaculture of some native tilapia species [11-14]. In-  
69 fact, the Zambian Department of Fisheries have adopted *O. andersonii* as a candidate species for  
70 aquaculture development [15].

71 Molecular genetic studies on cichlids in Africa have been biased towards phylogenetics of East  
72 African Lakes [16-18]. Nile tilapia has also received global attention because of its importance as an  
73 edible fish [19]. A few peer reviewed scientific papers have reported the use of molecular genetics in  
74 the study of southern African cichlids. Phylogenetic relationships have been inferred among and  
75 between cichlid species based on selected mitochondrial DNA regions and some allozymes [20-24].  
76 However, none of these phylogenetic analyses were based on complete mitogenome sequences.  
77 Complete mitogenome sequences have been employed in phylogenetic analyses of different fish  
78 species [25-33]. Among the important cichlids for aquaculture in Southern Africa (*O. andersonii*, *O.*  
79 *macrochir*, *Tilapia rendalli*, *O. mossambicus* and *O. niloticus*) complete mitogenome sequence have  
80 only been determined for *O. mossambicus* and *O. niloticus*.

81 In this study we describe the complete mitogenome sequences of *O. andersonii* and *O. macrochir*  
82 deposited in the GenBank with accession numbers MG603674 and MG603675 respectively. Based on  
83 the mitogenome sequences of these two species and other cichlids we confirm the position of these  
84 species among the cichlid species of Africa and analyse the evolutionary rates of the protein coding  
85 genes.

## 86 **Materials and methods**

### 87 **Sample collection and DNA extraction**

88 Tissue samples (fin clips) of *O. andersonii* and *O. macrochir* were collected from Upper  
89 Zambezi River (16.10 S, 23.295 E) and Lake Bangweulu (11.35 S, 29.58 E) in Zambia respectively  
90 with approval from the Zambian Department of Fisheries. These areas had no report of *O. niloticus*  
91 presence at the time of collecting the tissues. DNA was extracted from fin clips using a TIANamp  
92 Marine Animals DNA Kit using the manufacturer's instructions (Tiangen, China). All applicable  
93 international guidelines for the care and use of animals were followed.

### 94 **PCR amplification and sequencing**

95 Thirty primers were designed for both *O. andersonii* and *O. macrochir* and 2 other primers  
96 for each species (S1Table). The primers were designed using aligned complete mitogenomes of *O.*  
97 *mossambicus* (Accession number: AY597335.1) and *O. niloticus* (Accession number: GU370126.1)  
98 [30]. PCR was performed using an Eppendorf Thermal Cycler (Eppendorf, Germany). The total  
99 reaction mixture of 25.0 µl containing 17.5 µl distilled water, 4.5 µl PCR mix (Tiangen, China), 1.0 µl  
100 forward primer, 1.0 µl reverse primer, and 1.0 µl template DNA (50 ng/ µl) was used. The reaction  
101 was denatured at 94°C for 5 minutes, followed by 35 cycles of denaturation at 94 °C for 30 seconds,  
102 annealing at 48-53°C for 30 seconds and elongation at 72°C for 45 seconds; the last extension step  
103 was carried out at 72°C for 7 minutes. Agarose (1.0%) electrophoresis was performed to visualize the  
104 PCR products. The PCR products were purified using the 3S Spin PCR product Purification Kit  
105 (Biocolor Inc., Shanghai, China). The purified DNA was sequenced on an ABI 3730 xl capillary  
106 sequencer employing the same primers as for PCR. The various regions of mitochondrial DNA were  
107 sequenced and projected as electronic outputs by the computer connected to the sequencer.

### 108 **Sequence editing, alignment and annotation**

109 The raw sequences of the various regions of mitochondrial DNA of *O. andersonii* and *O.*  
110 *macrochir* were edited and assembled using BioEdit Version 7.2.6 [34]. They were further edited and

111 aligned using complete mitogenomes of *O. niloticus* and *O. aureus* [30], *O. variabilis* [32] and *O.*  
 112 *mossambicus* (Assession number: AY597335.1). Blast database searches on NCBI site were  
 113 performed to verify the target sequences amplified. Transfer RNA (*tRNA*) genes and their secondary  
 114 structures were identified using tRNAScan-SE 2.0 [35]. MEGA Version 7.0.26 [36] was used to  
 115 calculate the composition of amino acids, nucleotides and codon usage in the sequences. Annotation  
 116 of the sequences was performed with DOGMA [37], MITOS [38], and MitoAnnotator [39]. Further  
 117 verification was done with *O. niloticus* and *O. mossambicus* genome organization. The nucleotide  
 118 composition skewness was measured following the formulas: AT skew  $[(A - T)/(A + T)]$  and GC  
 119 skew  $[(G - C)/(G + C)]$  [40]. MEGA Version 7.0.26 was used to test for mode of selection in protein  
 120 coding genes acting at non-synonymous sites using the ration  $dN/dS$ .

## 121 Phylogenetic analysis

122 To infer phylogeny, 29 other cichlids and another 18 non-cichlid species from 7 different  
 123 families mainly found in southern Africa were obtained from the NCBI site (Table 1). The protein-  
 124 coding genes were used for phylogenetic analysis except *ND6* which is encoded by the opposite  
 125 strand and considered to possess a distinct heterogeneous base composition than the other 12 protein-  
 126 coding genes [41]. The concatenated protein sequences were aligned using MUSCLE with default  
 127 settings [42].

128

129 **Table 1. List of species with their accession numbers used in phylogenetic analysis in this study.**

Species	Accession No.	Reference	Species	Accession No.	Reference
<i>Oreochromis mossambicus</i>	AY597335.1	-	<i>Buccochromis nototaenia</i>	NC_031416.1	-
<i>Oreochromis niloticus</i>	GU370126.1	[30]	<i>Placidochromis longimanus</i>	NC_028156.1	-
<i>Oreochromis niloticus</i>	GU477624.1	-	<i>Lethrinops lethrinus</i>	NC_031419.1	-
<i>GIFT</i>			<i>Maylandia zebra</i>	NC_027944.1	-
<i>Oreochromis variabilis</i>	NC_026109	[32]	<i>Hemilapia oxyrhyncha</i>	NC_031415.1	-
<i>Oreochromis esculentus</i>	NC_025555	-	<i>Clarias gariepinus</i>	NC_027661.1	[52]
<i>Oreochromis aureus</i>	NC_013750	[30]	<i>Clarias fuscus</i>	NC_023941	-
<i>Oreochromis</i> sp ‘red tilapia’	NC_014060	-	<i>Lepomis macrochirus</i>	NC_015984	-
<i>Sarotherodon melanotheron</i>	NC_015611.1	[43]	<i>Micropterus salmoides</i>	NC_008106.1	[53]
<i>Coptodon zillii</i>	NC_026110	[32]			

<i>Serranochromis robustus</i>	NC_031418.1	-	<i>Ambloplites rupestris</i>	NC_035659.1	-
<i>Astatotilapia calliptera</i>	NC_018560.1	[44]	<i>Anabas testudineus</i>	NC_024752	-
<i>Tylochromis polylepis</i>	NC_011171	[45]	<i>Macropodus ocellatus</i>	NC_024753.1	-
<i>Haplochromis burtoni</i>	NC_027289	[46]	<i>Cyprinus carpio</i>	KJ511883.1	-
<i>Pundamilia nyererei</i>	NC_028011	[47]	<i>Labeo altivelis</i>	NC_029444.1	-
<i>Copadichromis virginalis</i>	NC_029761	-	<i>Labeobarbus intermedius</i>	NC_031531.1	-
<i>Fossorochromis rostratus</i>	NC_028089	[48]	<i>Barbus fasciolatus</i>	NC_031616.1	-
<i>Aulonocara stuartgranti</i>	NC_029380	[49]	<i>Petrocephalus microphthalmus</i>	NC_015098	-
<i>Tropheus duboisi</i>	NC_009063	[50]	<i>Mormyrops anguilloides</i>	AP011576.1	[54]
<i>Petrochromis trewavasae</i>	NC_018814.1	[51]	<i>Protopterus annectens</i>	NC_018822.1	-
<i>Neolamprologus brichardi</i>	NC_009062.1	[50]	<i>Protopterus aethiopicus</i>	NC_014764.2	[55]
<i>Cynotilapia afra</i>	NC_018564	[44]	<i>Mugil cephalus</i>	NC_003182.1	[56]
<i>Nimbochromis linni</i>	NC_018558	[44]	<i>Liza macrolepis</i>	NC_027239.1	-
<i>Pseudotropheus crabro</i>	NC_018559.1	[44]	<i>Myxus capensis</i>	NC_017892.1	-
<i>Petrotilapia nigra</i>	NC_018557.1	[44]			

130

131 Phylogenetic analysis was performed using Maximum Likelihood (ML) in MEGA Version 7.0.26 [36]

132 and Bayesian Markov-Chain-Monte-Carlo (MCMC) method in MrBayes (Version 3.2.6) [57].

133 Maximum Likelihood method was based on the General Time Reversible model. A discrete Gamma

134 distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter

135 = 0.4073)) and it allowed for some sites to be evolutionarily invariable ([+I], 15.08% sites). The

136 bootstrap consensus tree was inferred from 1000 replicates taken to represent the evolutionary history

137 of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap

138 replicates were collapsed. The Bayesian posterior probabilities were estimated using 1,000,000

139 generations, sampling every 100 generations at which point the standard deviation of split frequencies

140 of the two independent runs was below 0.01. A consensus tree was constructed from the saved trees

141 after discarding the first 25% trees as burn-in.

## 142 Results and Discussion

### 143 Mitogenome annotation

144 Annotation results of *O. andersonii* and *O. macrochir* using three methods; DOGMA, MTOS  
 145 and MitoAnnotator [37-39] revealed some differences in sizes of the protein coding genes and the two  
 146 *rRNAs*. However, annotation results using the three methods were similar for all *tRNAs* (S2 Table).  
 147 The largest difference was observed in gene *ND5* where DOGMA and MITOS differed from  
 148 MitoAnnotator in the final position with 495 bp gene size. The gene sizes for *O. andersonii* and *O.*  
 149 *macrochir* were similar in all the three annotation methods used except *ND2* final position with a  
 150 difference of 7 bp and *ND5* initial position with a difference of 3 bp (S2 Table). Comparing the results  
 151 obtained, annotation from Mitoannotator was selected for submission to the Genbank (Table 2).

152

153 **Table 2. Complete automated annotated sequences of *O. andersonii* (16642 bp) and *O. macrochir***  
 154 **(16644 bp) generated in MitoAnnotator**

155

Protein	Position		Size (bp)	Codon		Intergenic nucleotide <sup>b</sup>	Strand <sup>c</sup>
	Start	Stop		Start	Stop <sup>a</sup>		
tRNA <sup>Phe</sup>	1	69	69				H
12S rRNA	70	1013	944			0	H
tRNA <sup>Val</sup>	1014	1085	72			0	H
16S rRNA	1086	2779	1694			0	H
tRNA <sup>Leu</sup>	2780	2853	74			0	H
ND1	2854	3828	975	ATG	TAG	0	H
tRNA <sup>Ile</sup>	3832	3901	70			3	H
tRNA <sup>Gln</sup>	3901	3971	71			-1	L
tRNA <sup>Met</sup>	3971	4039	69			0	H
ND2	4040	5094	1055	ATG	TAA	0	H
tRNA <sup>Trp</sup>	5095	5166	72			0	H
tRNA <sup>Ala</sup>	5168	5236	69			1	L
tRNA <sup>Asn</sup>	5238	5310	73			1	L
tRNA <sup>Cys</sup>	5344	5409	66			33	L
tRNA <sup>Tyr</sup>	5410	5479	70			0	L
COI	5481	7082	1602	GTG	TAA	1	H
tRNA <sup>Ser</sup>	7083	7153	71			0	L
tRNA <sup>Asp</sup>	7157	7229	73			3	H
COII	7235	7925	691	ATG	T++	5	H
tRNA <sup>Lys</sup>	7926	7999	74			0	H
ATPase 8	8001	8168	168	ATG	TAA	1	H
ATPase 6	8159	8832	674	ATG	TAA	-10	H
COIII	8833	9616	784	ATG	TAA	0	H

tRNA <sup>Gly</sup>	9617	9688	72			0	H
ND3	9689	10037	349	ATG	TAG	0	H
tRNA <sup>Arg</sup>	10038	10106	69			0	H
ND4L	10107	10403	297	ATG	TAA	0	H
ND4	10397	11786	1390	ATG	T++	-7	H
tRNA <sup>His</sup>	11787	11855	69			0	H
tRNA <sup>Ser</sup>	11856	11922	67			0	H
tRNA <sup>Leu</sup>	11927	11999	73			4	H
ND5	12000	14300	2301	ATG	TAA	0	H
ND6	13840	14361	522	ATG	TAA	-461	L
tRNA <sup>Glu</sup>	14362	14430	69			0	L
Cyt b	14435	15575	1141	ATG	T++	4	H
tRNA <sup>Thr</sup>	15576	15647	72			0	H
tRNA <sup>Pro</sup>	15648	15717	70			0	L
CR	15718	16642/16644	925/927			0	-

156 The forward slashes (/) denote the values of *O. andersonii*/*O. macrochir*, if none the values for both species are identical.

157 <sup>a</sup> TA+ and T++ represent incomplete stop codon

158 <sup>b</sup> The positive numbers indicate nucleotides separating two adjacent genes while the negative numbers represent nucleotide overlap.

159 <sup>c</sup> H=heavy and L=light strands.

160

161

## 162 Mitogenome organization

163 The complete mitogenome sequences of *O. andersonii* and *O. macrochir* were 16642 bp  
164 (accession #: MG603674) and 16644 bp (accession #: MG603675) respectively. This was within the  
165 range of published mitogenomes sequences of *Oreochromis* species i.e *O. niloticus* with 16625 bp and  
166 *O. aureus* with 16628 bp [30]; *Oreochromis variabilis* with 16626 bp [32] and GenBank deposited  
167 sequence of *O. mossambicus* (AY597335.1). As anticipated, the general structural organization  
168 follows those of other teleost species with 13 protein coding genes, 2 *rRNAs*, 22 *tRNAs*, and a non-  
169 coding control region (Table 2 and figure 1). Both species were observed to exhibit the Heavy (H) and  
170 Light (L) strand coding pattern previously observed in other teleosts. The L-strand was observed in  
171 one protein coding gene (*NADH* dehydrogenase subunit 6 (*ND6*)) and eight *tRNAs* (*tRNA<sup>Gln</sup>*, *tRNA<sup>Ala</sup>*,  
172 *tRNA<sup>Asn</sup>*, *tRNA<sup>Cys</sup>*, *tRNA<sup>Tyr</sup>*, *tRNA<sup>Ser</sup>*, *tRNA<sup>Glu</sup>*, *tRNA<sup>Pro</sup>*). The start codons for protein coding genes for  
173 both *O. andersonii* and *O. macrochir* were ATG except for gene *COI* which was GTG. The stop  
174 codons on the other hand were either TAG or TAA except for gene *COII*, *ND4*, and *Cyt b* which had  
175 incomplete stop codons of T++. The incomplete codon stops observed for the two species is a  
176 common feature in most vertebrates including some fish species [27, 28, 33].



177 The overall nucleotide base composition was very similar between the two species. For *O. andersonii*;  
 178 T=26.3%, C=30.1%, A=28.0%, G=15.6 % whereas *O. macrochir* T=26.2%, C=30.2%, A=28.1%,  
 179 G=15.5%. Further, the GC% and AT% were the same for both species at 45.7 and 54.3 respectively  
 180 (Table 3). This base composition is very similar to what has been reported for other *Oreochromis*  
 181 species [30,32] and other species as well [31, 32]. The lower value for GC compared to AT is another  
 182 common feature which has been observed in most vertebrate mitogenome resulting from anti-bias  
 183 against G in the third codon position. The species in this study had lower values of G in all the three  
 184 codon positions of protein coding genes and non-coding control region (Table 3). Furthermore, the  
 185 GC-skew and AT-skew which describe the overall patterns of nucleotide composition in DNA  
 186 sequences [40], were -0.317 and 0.031 for *O. andersonii* and -0.322 and 0.035 for *O. macrochir*  
 187 respectively. This result shows an excess of G over C and excess of A over T (Table 3).

188

189 **Table 3. Base percent composition of complete mitogenome sequences of *O. andersonii* (16642**  
 190 **bp) and *O. macrochir* (16644 bp) generated in MEGA 7.0.26**

Codon position	% Nucleotide composition	Complete mitogenome		Protein coding genes		ND6		tRNAs		rRNAs		Control Region	
		* <i>O.A</i>	* <i>O.M</i>	<i>O.A</i>	<i>O.M</i>	<i>O.A</i>	<i>O.M</i>	<i>O.A</i>	<i>O.M</i>	<i>O.A</i>	<i>O.M</i>	<i>O.A</i>	<i>O.M</i>
1	T	27.0	27.0	31.0	31.0	35.0	35.0	26.0	26.0	19.0	19.0	32.0	31.0
	C	31.3	31.3	30.1	30.1	8.0	8.0	22.8	23.0	26.6	26.7	20.4	19.1
	A	27.7	27.9	25.3	25.2	6.9	6.9	27.2	27.2	32.4	32.6	32.0	34.3
	G	13.8	13.8	13.9	14.0	50.0	50.0	23.6	23.4	21.7	21.7	15.5	15.5
2	T	24.0	24.0	27.0	27.0	43.0	43.0	31.0	31.0	21.0	22.0	31.0	30.0
	C	29.8	30.0	36.2	36.2	21.8	22.4	20.7	20.7	25.7	25.6	20.5	23.6
	A	29.3	29.3	26.6	26.8	10.9	10.9	29.0	29.0	32.5	32.7	32.5	32.7
	G	16.5	16.4	10.0	9.9	24.1	24.1	19.9	19.9	20.4	20.3	15.6	13.6
3	T	27.0	27.0	24.0	24.0	37.0	37.0	24.0	24.0	21.0	21.0	35.0	38.0
	C	29.3	29.2	30.8	31.0	1.1	1.1	21.2	21.0	28.0	28.2	20.8	19.4
	A	27.0	26.9	26.9	26.8	20.5	19.4	28.4	28.6	31.9	31.9	31.2	29.4
	G	16.3	16.3	17.9	17.9	41.4	42.5	26.4	26.3	19.5	19.2	13.0	13.3
Total composition	T(U)	26.3	26.2	27.5	27.4	38.3	38.3	27.0	27.0	20.5	20.4	32.9	33.0
	C	30.1	30.2	32.4	32.5	10.3	10.5	21.6	21.6	26.8	26.8	20.5	20.7
	A	28.0	28.1	26.2	26.2	12.8	12.3	28.2	28.2	32.3	32.4	31.9	32.1
	G	15.6	15.5	13.9	13.9	38.5	38.9	23.2	23.3	20.5	20.4	14.7	14.1
	% GC	45.7	45.7	46.3	46.4	48.9	49.4	44.7	44.9	47.3	47.2	35.2	34.8
	% AT	54.3	54.3	53.7	53.6	51.1	50.6	55.3	55.1	52.7	52.8	64.8	65.2
	GC Skew	-0.317	-0.322	-0.400	-0.401	0.578	0.575	0.036	0.038	-0.133	-0.136	-0.165	-0.190
	AT Skew	0.031	0.035	-0.024	-0.022	-0.499	-0.514	0.022	0.022	0.223	0.227	-0.015	-0.014
	Total length (bp)	16642	16644	11427	11427	522	522	1554	1554	2638	2638	925	927

191 \**O.A*= *Oreochromis andersonii*

\**O.M*= *Oreochromis macrochir*

192 Intergenic overlaps were observed in 4 genes and spacers in 10 in the mitogenome of *O. andersonii*  
193 and *Oreochromis macrochir*. The most significant overlap of 461 nucleotides was observed between  
194 *ND5* and *ND6*. This large overlap is not a common feature among most reported teleosts  
195 mitogenomes. Overlap of 10 nucleotides between *ATPase 8* and *ATPase 6* was also observed for both  
196 species followed by *ND4L* and *ND4* gene with overlap of 7 nucleotides. Intergenic spacers totalled 56  
197 bp in 10 regions for both *O. andersonii* and *O. macrochir* (see Table 2). The most significant  
198 intergenic spacers in both species were between *tRNA<sup>Asn</sup>* and *tRNA<sup>Cys</sup>* (33 nucleotides) followed by  
199 *tRNA<sup>Asp</sup>* and *COII* (5 nucleotides). Again intergenic overlaps and spacers follow what has been  
200 reported for most vertebrate mitogenomes inclusive of different fish species except for the overlap  
201 between *ND5* and *ND6*.

## 202 **Protein coding genes**

203 The nucleotide composition of the protein coding genes was very similar between *O.*  
204 *andersonii* and *O. macrochir* (Table 3). The total GC% composition was 46.3 and AT% was 53.7 for  
205 *O. andersonii* and 46.4 and 53.6 for *O. macrochir*. The GC-skew for *O. andersonii* and *O. macrochir*  
206 were -0.400 and -0.401 and the AT-skew values were -0.024 and -0.022 respectively. However, the L-  
207 strand gene *ND6* had a positive GC-skew value for both species (Table 3). Again it showed overall  
208 anti-G bias supporting earlier findings in fish mitogenomes [26, 27, 33]. However, unlike many  
209 authors who report strong anti-G bias on third codon positions, both *O. andersonii* and *O. macrochir*  
210 were observed to have the strong anti-G bias on the second codon position (*O. andersonii* = 10.0, *O.*  
211 *macrochir* = 9.9).

212 The adaptive radiation of African cichlids is unparalleled so far among the vertebrates. To understand  
213 the role of mitogenome protein coding genes in the evolution of these cichlid species we analysed the  
214 rate of non synonymous (*dN*) and synonymous (*dS*) nucleotide substitutions in 13 protein coding  
215 genes of 20 species. These included; Riverine, Lake Victoria, Lake Tanganyika and Lake Malawi  
216 species (*Sarotherodon melanotheron*, *Coptodon zillii*, *Oreochromis aureus*, *O. niloticus*, *O. niloticus*  
217 GIFT, *O. mossambicus*, *O. variabilis*, *O. andersonii*, *O. macrochir*, *O. esculentus*, *Oreochromis* sp  
218 ‘red tilapia’, *Petrochromis trewavasae*, *Tropheus duboisi*, *Astatotilapia calliptera*, *Cynotilapia afra*,

219 *Maylandia zebra*, *Pundamilia nyererei*, *Tylochromis polylepis*, *Haplochromis burtoni*, *Lethrinops*  
220 *lethrinus*). The mean value of  $dN$  and  $dN/dS$  was highest in *ND6* gene and lowest in *COIII* gene for all  
221 the 13 protein coding genes (Fig 2). Our findings indicate that all protein coding genes evolved under  
222 purifying selection except for *ND6* gene which had an elevated rate of  $dN/dS$  indicative of evolution  
223 under positive selection. We can conclude that *ND6* gene may have evolved more rapidly than any  
224 other protein coding gene among these African cichlids.

225 The overall  $p$ -genetic distance was used to measure the conservation of the protein coding genes in  
226 the mitogenome of these 20 cichlid species. Calculation was performed on the 1<sup>st</sup> and 2<sup>nd</sup>, 3<sup>rd</sup> and  
227 whole sequence codon positions. On the 1<sup>st</sup> and 2<sup>nd</sup> codon positions, the highest overall mean  $p$ -  
228 distance was on gene *ND6* (0.1581) followed by *ATPase 6* (0.1268) and least was gene *COIII*  
229 (0.0193). For the whole sequence, the highest overall  $p$ -distance was recorded in *ATPase 6* (0.1389)  
230 followed by *ND6* (0.1288) and *ATPase 8* (0.0990) had the least value (Fig 3). Based on these results  
231 *ND6* likely may have the highest evolutionary rate and *COIII* been the most conserved gene among  
232 the mitogenome protein coding genes of these cichlid species which could have radiated into many  
233 species due to geographical isolation [1].

## 234 **Ribosomal and Transfer RNAs**

235 The *tRNA* genes for both *O. andersonii* and *O. macrochir* possessed anti-codons that match  
236 the mitochondrial vertebrate code. The length of the *tRNAs* for both species ranged between 66 -74 bp  
237 while the total length was 1554 bp for both *O. andersonii* and *O. macrochir*. For both species all the  
238 22 *tRNAs* with exception of *tRNA<sup>Ser(GCT)</sup>* inferred secondary structures folded into classic cloverleaf .  
239 The common features of the secondary structures were: a 7 bp aminoacyl stem and anticodon loop, 5  
240 bp TΨC and anticodon arms, and 4 bp DHU arm (S1 Fig). Similar results have been reported on  
241 South American catfish and Asian arowana [26, 33]. However, non-complementary pairing and size  
242 variations in the secondary structures were observed in both species.

243 The two *rRNAs* for both species had a total length of 944 bp and 1694 bp which was within the range  
244 reported for vertebrate mitogenomes. The nucleotides percent compositions of *rRNAs* of *O.*

245 *andersonii* (T=20.5, C= 26.8, A=32.3, G=20.5) and *O. macrochir* (T= 20.4, C= 26.8, A=32.4, G=20.4)

246 were very similar. They also showed a higher percentage of AT than GC pairs (52.7, 47.3 for *O.*  
247 *andersonii*; 52.8, 47.2 for *O. macrochir*) (Table 3)

## 248 **Non-coding region**

249 The mitochondrial non-coding region or control region of *O. andersonii* and *O. macrochir*  
250 was located between *tRNA<sup>pro</sup>* and *tRNA<sup>phe</sup>* typical for vertebrate mitogenomes. Its length was 925 bp  
251 for *O. andersonii* and 927 bp for *O. macrochir*. The nucleotide base composition was similar between  
252 the two species although more variable compared to other regions of the mitogenome (Table 3). The  
253 AT% (35.2:34.8, *O. andersonii*: *O. macrochir*) composition to GC% (64.8:65.2, *O. andersonii*: *O.*  
254 *macrochir*) was also more highly skewed towards AT compared to other regions of the mitogenome  
255 due to anti-G bias on the third codon position.

256 *Oreochromis andersonii* and *O. macrochir* control regions were aligned with four other species from  
257 the same genus to characterize the control region. Domain I consisted of a hypervariable region with  
258 a length of 281 bp for both species including a Termination -associated Sequence (TAS) with motif-  
259 ATGCAT similar to a putative TAS of *O. aureus* [30]. Domain II or central conserved region was  
260 identified with three conservative sequence blocks (CSB) which are involved in heavy-H strand  
261 replication [58]. The first was CSB-F (ATGTAGTAAGAGCCCACC) followed by CSB-E  
262 (AAGGACAGTACTTGTGGGGGT) and then CBS-D (TATTCCTGGCATCTGGTTCCT) to  
263 complete domain II for both species. The third domain at the 3' end of the control region consisted of  
264 CBS-1 (*O. andersonii* - ATTACATAACTGATATCAAGAGCATA; *O. macrochir*-  
265 ACCACATAACTGATATCTAGAGCATA) and CBS-2 (AAACCCCCCTACCCCC). The last  
266 conservative block observed was CBS-3 (TGCAAACCCCCCGGAAACAG) (Fig 4)

## 267 **Amino acid composition**

268 For both *O. andersonii* and *O. macrochir* leucine, proline, serine, threonine, asparagine were  
269 the most frequently translated amino acids from the mitochondrial genome (Table 5 and Fig 5). This  
270 is similar to the findings on other fish species [26, 33]. The relative synonymous codon usage (RSCU)  
271 indicated that the most frequently used codon in the mitogenome of the two species was GCC for

272 alanine with *O. andersonii* having an RSCU value of 1.74 while *O. macrochir* had 1.67. The second  
273 highest codon usage values were for serine (UCU) for *O. andersonii* having RSCU=1.54 and Arginine  
274 (CGC) for *O. macrochir* having an RSCU=1.58. RSCU values of 0.40 and 0.43 were observed to be  
275 the least for alanine (GCG) for *O. andersonii* and *macrochir* respectively.

## 276 **Phylogenetic analysis**

277 Phylogenetic analysis using 12 protein coding genes of 29 cichlid species and 18 other  
278 species belonging to 7 different families of mainly Southern African fish were used to examine the  
279 phylogenetic placement of *O. andersonii* and *O. macrochir* among the cichlids of Africa. Most of the  
280 families were monophyletic. The consensus tree forms a clear clade consisting of genera *Oreochromis*  
281 (maternal mouthbrooders), *Sarotherodon* (biparental and paternal mouthbrooders) and *Coptodon*  
282 (substrate spawners) agreeing with the classification of Trewavas of these species [2]. This  
283 classification seems to indicate that the development of the mouthbrooding reproductive behaviour  
284 emerged later after the substrate and biparental. This phylogenetic relationship was also observed by  
285 Nagal [59]. The close relatedness of *Sarotherodon melanotheron* to *Oreochromis aureus* rather  
286 forming a monophyletic clade may suggest a need to redefine the genus [32, 59].

287 The three *Oreochromis* species (*O. macrochir*, *O. andersonii* and *O. mossambicus*) from southern  
288 Africa in this classification seem to have evolved later compared to the west and east African species.  
289 Data from the consensus tree in this study placed *O. andersonii* closely related to *O. mossambicus*  
290 (Fig 6). This is in agreement with other traditional classifications [20, 60]. The close relatedness of  
291 these two species though not sharing the same habitat may indicate that they may have separated most  
292 recently. On the other hand, *O. macrochir* clusters with both *O. mossambicus* and *O. andersonii*  
293 indicative of a closer phylogenetic relationship among these species of Southern Africa. The  
294 consensus tree for the other Lake Malawi cichlids is similar to what has been reported by [44]. The  
295 families Mugilidae, Centrarchidae and Anabantidae were the most closely related to the cichlids (Fig  
296 6).

## 297 **Conclusion**

298           This study has determined the complete mitochondrial DNA sequences of *O. andersonii*  
299 (16623 bp) and *O. macrochir* (16624 bp) and further characterization has confirmed their similarity to  
300 teleost vertebrate mitochondrial. Using Bayesian posterior probabilities and Maximum Likelihood  
301 analysis of concatenated mitochondrial genome of 12 protein coding genes, consensus trees has  
302 revealed a close phylogenetic relationship between *O. andersonii* and *O. mossambicus*. Further, *O.*  
303 *macrochir* was found to be closely related to both *O. andersonii* and *O. mossambicus*. The protein  
304 coding genes evolved under purifying selection except for *ND6* which indicated evolution under  
305 positive selection. Gene *ND6* likely may have the highest evolutionary rate and *COIII* been the most  
306 conserved gene. More work is needed to characterize the southern Africa cichlids as they are  
307 important species especially for capture fisheries, aquaculture development and understanding  
308 biogeographic history of African cichlids.

## 309 **Acknowledgments**

310           We thank the Department of Fisheries – Zambia staff for support during sample collection.

## 311 **References**

312

- 313           1. Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, Fan S, et al. The genomic substrate for  
314           adaptive radiation in African cichlid fish. *Nature*. 2014; 513:375-381.  
315           doi:10.1038/nature13726.
- 316           2. Trewavas E. Tilapiine fishes of the Genera *Sarotherodon*, *Oreochromis*, and *Danakilia*.  
317           London: British Museum (Natural History); (1983).
- 318           3. Skelton P. Freshwater Fishes of Southern Africa. Second edition. South Africa: Struik  
319           Publishers (Pty) Ltd; 2001.
- 320           4. Froese R, Pauly D; 2016 [cited 2018 Jan 6]. Database: FishBase [Internet]. Available from:  
321           <http://www.fishbase.org>

- 322 5. Lind CE, Brummett RE, Ponzoni RW. Exploitation and conservation of fish genetic resources  
323 in Africa: issues and priorities for aquaculture development and research. Reviews in  
324 Aquaculture. 2012; 4:125–141.
- 325 6. Tweddle D, Marshall BE; 2007. [cited 2017 Oct 23]. Database: *Oreochromis andersonii*. The  
326 IUCN Red List of Threatened Species [Internet]. Available from:  
327 <http://dx.doi.org/10.2305/IUCN.UK.2007.RLTS.T60623A12385801.en>
- 328 7. Tweddle D, Peel RA, Weyl OLF. Challenges in fisheries management in the Zambezi, one of  
329 the great rivers of Africa. Fisheries Management and Ecology. 2015; 22:99-111.
- 330 8. Canonico GC, Arthington A, McCrary JF, Thiem ML. The effects of introduced tilapias on  
331 native biodiversity. Aquatic Conserv: Mar Freshw Ecosyst. 2005; 15:463–483.  
332 doi:10.1002/aqc.699.
- 333 9. Deines AM, Bbole I, Katongo C, Feder JL, and Lodge DM. Hybridisation between native  
334 *Oreochromis* species and introduced Nile tilapia *O. niloticus* in the Kafue River, Zambia. Afr  
335 J Aquat Sci. 2014; 39:23-34.
- 336 10. Bbole I, Katongo C, Deines AM, Shumba O, Lodge DM. Hybridization between non-  
337 indigenous *Oreochromis niloticus* and native *Oreochromis* species in the lower Kafue River  
338 and its potential impacts on fishery. J. Ecol. Nat. Environ. 2015; 6(6):215-225.
- 339 11. Crayon-Thomas E. Background history on the research and use of *Oreochromis niloticus* in  
340 Zambia and their use on a private fish farm (Kalimba Farm) between 1985 and 1993.  
341 Technical Consultation on Species for Small Reservoir Fisheries and Aquaculture in Southern  
342 Africa. Livingstone, Zambia, 7–11 November 1994
- 343 12. Kefi AS, Kang’ombe J, Kassam D, Katongo C. Growth, reproduction and sex ratios in  
344 *Oreochromis andersonii* (Castelnau, 1861) fed with varying Levels of 17 $\alpha$ -methyl  
345 testosterone. J Aquac Res Dev. 2012; 3:130–137.



- 346 13. Musuka CG, Musonda FF. Current and future prospects of commercial fish farming in  
347 Zambia. *Int J Bioflux Soc.* 2012 5:79–87.
- 348 14. Nsonga A. Indigenous fish species a panacea for cage aquaculture in Zambia: A case for  
349 *Oreochromis macrochir* (Boulenger, 1912) at Kambashi out-grower scheme. *Int. j. fish.*  
350 *Aquat.* 2014; 2(1):102-105.
- 351 15. Kefi AS, Kang'ombe J, Kassam D, Katongo C. Optimal Dietary Plant Based Lipid on Growth  
352 of *Oreochromis andersonii* (Castelnau, 1861). *Turk J Fish Aquat Sci.* 2013; 13:503-508.
- 353 16. Salzburger W, Meyer A, Baric S, Verheyen E, Sturmbauer C. Phylogeny of the Lake  
354 Tanganyika Cichlid Species Flock and Its Relationship to the Central and East African  
355 Haplochromine Cichlid Fish Faunas. *Syst Biol.* 2002; 51(1):113-135.
- 356 17. Salzburger W, Meyer A. The species flocks of East African cichlid fishes: recent advances in  
357 molecular phylogenetics and population genetics. *Naturwissenschaften.* 2004; 91:277–290.
- 358 18. Meyer BS, Matschiner M, Salzburger W. A tribal level phylogeny of Lake Tanganyika  
359 cichlid fishes based on a genomic multi-marker approach. *Mol Phylogenet Evol.* 2015;  
360 83:56–71.
- 361 19. Kocher TD, Lee W, Sobolewska H, Penman D, Andrew B. A genetic linkage map of a cichlid  
362 fish the tilapia (*Oreochromis niloticus*). *Genetics.* 1998; 148:1225–1232.
- 363 20. Sodsuk PK, McAndrew BJ, Turner GF (1995) Evolutionary relationships of the Lake Malawi  
364 *Oreochromis* species: evidence from allozymes. *J Fish Biol.* 1995; 47:321–333.
- 365 21. Feresu-Shonhiwa F, Howard JH. Electrophoretic identification and phylogenetic relationships  
366 of indigenous tilapiine species of Zimbabwe. *J Fish Biol.* 1998; 53:1178–1206.
- 367 22. Katongo C, Koblmüller S, Duftner N, Makasa L, Sturmbauer C. Phylogeography and  
368 speciation in the *Pseudocrenilabrus philander* species complex in Zambian Rivers.  
369 *Hydrobiologia.* 2005; 542:221–233. doi: 10.1007/s10750-004-1389-x



- 370 23. D'Amato ME, Esterhuysen MM, Van der Waal BCW, Brink D, Volckaert FAM. Hybridization  
371 and Phylogeography of the Mozambique tilapia *Oreochromis mossambicus* in Southern  
372 Africa evidenced by mitochondrial and microsatellite DNA genotyping. *Conserv. Genet.* 2007;  
373 8: 475-488.
- 374 24. Katongo C, Koblmüller, S, Duftner N, Mumba L, Sturmbauer C. Evolutionary history and  
375 biogeographic affinities of the Serranochromine cichlids in Zambian rivers. *Mol.*  
376 *Phylogenetics Evol.* 2007; 45:326–338. doi:10.1016/j.ympev.2007.02.011
- 377 25. Peng Z, Wang J, He S. The complete mitochondrial genome of the helmet catfish  
378 *Cranoglanis boudierius* (Siluriformes: Cranoglanididae) and the phylogeny of otophysan  
379 fishes. *Gene.* 2006; 376:290–297.
- 380 26. Yue GH, Liew WC, Orban L. The complete mitochondrial genome of a basal teleost, the  
381 Asian arowana (*Scleropages formosus*, Osteoglossidae). *BMC Genomics.* 2006; 7:242.
- 382 27. Liu Y, Cui Z. The complete mitochondrial genome sequence of the cutlassfish *Trichiurus*  
383 *japonicas* (Perciformes:Trichiuridae): Genome characterization and phylogenetic  
384 considerations. *Mar Genomics.* 2009; 2:133-142.
- 385 28. Zhang X, Yue B, Jiang W, Song Z. The complete mitochondrial genome of rock carp  
386 *Procypris rabaudi* (Cypriniformes: Cyprinidae) and phylogenetic implications. *Mol Biol Rep.*  
387 2009; 36:981-991.
- 388 29. Cheng YZ, Wang RX, Xu TJ. The mitochondrial genome of the spinyhead croaker *Collenchys*  
389 *lucida*: Genome organization and phylogenetic consideration. *Mar Genomics.* 2011; 4:17–23.
- 390 30. He A, Luo Y, Yang H, Liu L, Li S, Wang C. Complete mitochondrial DNA sequences of the  
391 Nile tilapia (*Oreochromis niloticus*) and Blue tilapia (*Oreochromis aureus*): genome  
392 characterization and phylogeny applications. *Mol Biol Rep.* 2011; 38:2015–2021. DOI  
393 10.1007/s11033-010-0324-7

- 394 31. Wang C, Wang J, Yang J, Song X, Chen Q, Xu J, Yang Q, Li S. Complete mitogenome  
395 sequence of black carp (*Mylopharyngodon piceus*) and its use for molecular phylogeny of  
396 leuciscine fishes. *Mol Biol Rep.* 2012; 39:6337-6342. doi: 10.1007/s11033-012-1455-9
- 397 32. Kinaro ZO, Xue L, Volatiana JA. Complete mitochondrial DNA sequences of the Victoria  
398 tilapia (*Oreochromis variabilis*) and Redbelly Tilapia (*Tilapia zilli*): genome characterization  
399 and phylogeny analysis. *Mitochondrial DNA Part A.* 2015; 27(4): 2455–2457.
- 400 33. Villela LC, Alves AL, Varela ES, Yamagishi ME, Giachetto PF, da Silva NM, et al.  
401 Complete mitochondrial genome from South American catfish *Pseudoplatystoma reticulatum*  
402 (Eigenmann & Eigenmann) and its impact in Siluriformes phylogenetic tree. *Genetica.* 2017.  
403 145:51–66. doi: 10.1007/s10709-016-9945-7.
- 404 34. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program  
405 for Windows 95/98/NT. *Nucl Acids Symp Ser.* 1999; 41:95-98.
- 406 35. Lowe TM, Eddy SR. tRNAscan-SE: A program for improved detection of transfer RNA  
407 genes in genomic sequence. *Nucl Acids Res.* 1997; 25:955-964.
- 408 36. Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis  
409 Version 7.0 for Bigger Datasets. *Mol Biol Evol.* 2016; 33(7):1870–1874.  
410 <https://doi.org/10.1093/molbev/msw054>
- 411 37. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with  
412 DOGMA. *Bioinformatics.* 2004; 20(17):3252-5. doi: 10.1093/bioinformatics/bth352
- 413 38. Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritsch G, et al. MITOS: improved  
414 de novo metazoan mitochondrial genome annotation. *Mol Phylogenet Evol.* 2013; 69(2):313-  
415 9. doi: 10.1016/j.ympev.2012.08.023
- 416 39. Iwasaki W, Fukunaga T, Isagozawa R, Yamada K, Maeda Y, Satoh TP, et al. MitoFish and  
417 MitoAnnotator: a mitochondrial genome database of fish with an accurate and automatic  
418 annotation pipeline. *Mol Biol Evol.* 2013; 30:2531–2540.

- 419 40. Perna NT, Kocher TD. Patterns of nucleotide composition at fourfold degenerate sites of  
420 animal mitochondrial genomes. *J Mol Evol.* 1995; 41:353–358. doi: 10.1007/BF01215182.
- 421 41. Miya M, Nishida M. The mitogenomic contributions to molecular phylogenetics and  
422 evolution of fishes: a 15- year retrospect. *Ichthyol Res.* 2015; 62:29-71. doi: 10.1007/s10228-  
423 014-0440-9.
- 424 42. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput.  
425 *Nucleic Acids Res.* 2004; 32:1792-1797.
- 426 43. He A-Y, Tang S-J, Jiang Y-T, Li S-F, Wang C-H. Complete mitochondrial genome of  
427 blackchin tilapia *Sarotherodon melanotheron* (Perciformes, Cichlidae). *Mitochondrial DNA.*  
428 2011; (5-6):171-3. doi: 10.3109/19401736.2011.636439
- 429 44. Hulsey CD, Keck BP, Alamillo H, O’Meara BC. Mitochondrial genome primers for Lake  
430 Malawi cichlids. *Molec Ecol Res.* 2013b; 13:347–353.
- 431 45. Azuma Y, Kumazawa Y, Miya M, Mabuchi K, Nishida M. Mitogenomic evaluation of the  
432 historical biogeography of cichlids toward reliable dating of teleostean divergences. *BMC*  
433 *Evol Biol.* 2008; 8:215. doi: 10.1186/1471-2148-8-215
- 434 46. Hu XX, Lu J, He SY, Li XF, Liu HJ, Han LJ. The complete mitochondrial genome of  
435 *Astatotilapia burtoni*. *Mitochondrial DNA.* 2016; 27(4):2379-80.  
436 doi:10.3109/19401736.2015.1028039
- 437 47. Chen L, Song X, Chen X, Dang X, Wang W. The complete mitochondrial genome of the  
438 *Pundamilia nyererei* (Perciformes, Cichlidae). *Mitochondrial DNA.* 2016; 27(5):3567-8.  
439 doi:10.3109/19401736.2015.1074221
- 440 48. Qi DS, Zhang SC, Jiang J, Zhang LQ, Fu YY, Du ZM, et al. The complete mitochondrial  
441 genome of the *Fossorochromis rostratus*. *Mitochondrial DNA.* 2016; 27(6):4284-4285.

- 442 49. Zhao J, Gao J. Complete mitochondrial genome of *Aulonocara stuartgranti* (Flavescent  
443 peacock cichlid). *Mitochondrial DNA A DNA Mapp Seq Anal.* 2017; 28(2):279-280.  
444 doi:10.3109/19401736.2015.1118079
- 445 50. Mabuchi K, Miya M, Azuma Y, Nishida M. Independent evolution of the specialized  
446 pharyngeal jaw apparatus in cichlid and labrid fishes. *BMC Evol Biol.* 2007; 7:10. doi:  
447 10.1186/1471-2148-7-10.
- 448 51. Fischer C, Koblmuller S, Gully C, Schlotterer C, Sturmbauer C, Thallinger GG. Complete  
449 mitochondrial DNA sequences of the threadfin cichlid (*Petrochromis trewavasae*) and the  
450 blunthead cichlid (*Tropheus moorii*) and patterns of mitochondrial genome evolution in  
451 cichlid fishes. *PLoS One.* 2013; 8(6):e67048. doi: 10.1371/journal.pone.0067048.
- 452 52. Han C, Li Q, Xu J, Li X, Huang J. Characterization of *Clarias gariepinus* mitochondrial  
453 genome sequence and a comparative analysis with other catfishes. *Biologia.* 2016;  
454 70(9):1245-1253.
- 455 53. Broughton RE, Reneau PC. Spatial covariation of mutation and nonsynonymous substitution  
456 rates in vertebrate mitochondrial genomes. *Mol Biol Evol.* 2006; 23 (8):1516-1524
- 457 54. Lavoue S, Miya M, Arnegard ME, Sullivan JP, Hopkins CD, Nishida M. Comparable ages for  
458 the independent origins of electrogenesis in African and South American weakly electric  
459 fishes. *PLoS ONE.* 2012; 7(5): E36287.
- 460 55. Inoue JG, Miya M, Lam K, Tay BH, Danks JA, Bell J, et al. Evolutionary origin and  
461 phylogeny of the modern holocephalans (Chondrichthyes: Chimaeriformes): a mitogenomic  
462 perspective. *Mol Biol Evol.* 2010; 27(11):2576-2586. doi: 10.1093/molbev/msq147
- 463 56. Miya M, Kawaguchi A, Nishida M. Mitogenomic exploration of higher teleostean  
464 phylogenies: a case study for moderate-scale evolutionary genomics with 38 newly  
465 determined complete mitochondrial DNA sequences. *Mol Biol Evol.* 2001; 18(11):1993-2009.

- 466 57. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed  
467 models. *Bioinformatics*. 2003; 19:1572–1574.
- 468 58. Faber JE, Stepien CA. Tandemly repeated sequences in mitochondrial DNA control region  
469 and phylogeography of the Pike-Perches *Stizostedion*. *Mol. Phylogenetics Evol.* 1998;  
470 10(3):310-322
- 471 59. Nagl S, Tichy H, Mayer WE, Samonte IE, McAndrew BJ, Klein J. Classification and  
472 phylogenetic relationships of African tilapiine fishes inferred from mitochondrial DNA  
473 sequences. *Mol. Phylogenetics Evol.* 2001; 20(3):361-374.
- 474 60. McAndrew BJ, Majumdar K C. Evolutionary relationships within three Tilapiine genera  
475 (Pisces: Cichlidae). *Zool J Linnean Soc.* 1984; 80:421–435.

## 476 **Supporting information**

477

478 **S1 Fig. Secondary structures of 22 *tRNAs* of mitochondrial genome of *O. andersonii* (A)  
479 and *O. macrochir* (B) generated by tRNAScan-SE 2.0**

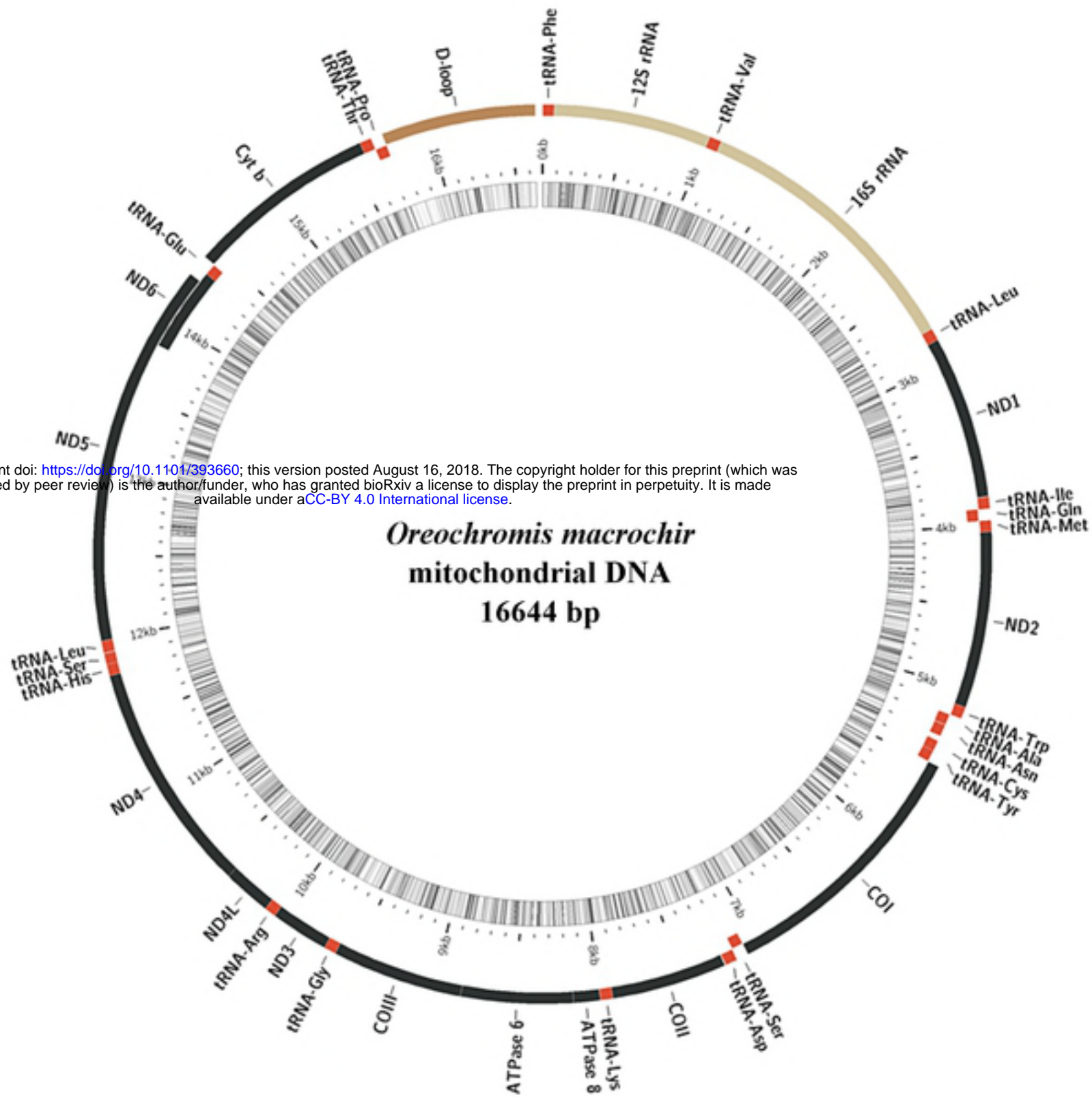
480 **S1 Table. Designed primers used for amplification and sequencing of complete  
481 mitogenome of *O. andersonii* and *O. macrochir***

482 **S2 Table. Complete automated annotation of *O. andersonii* and *O. macrochir*  
483 mitochondrial genome.**

484

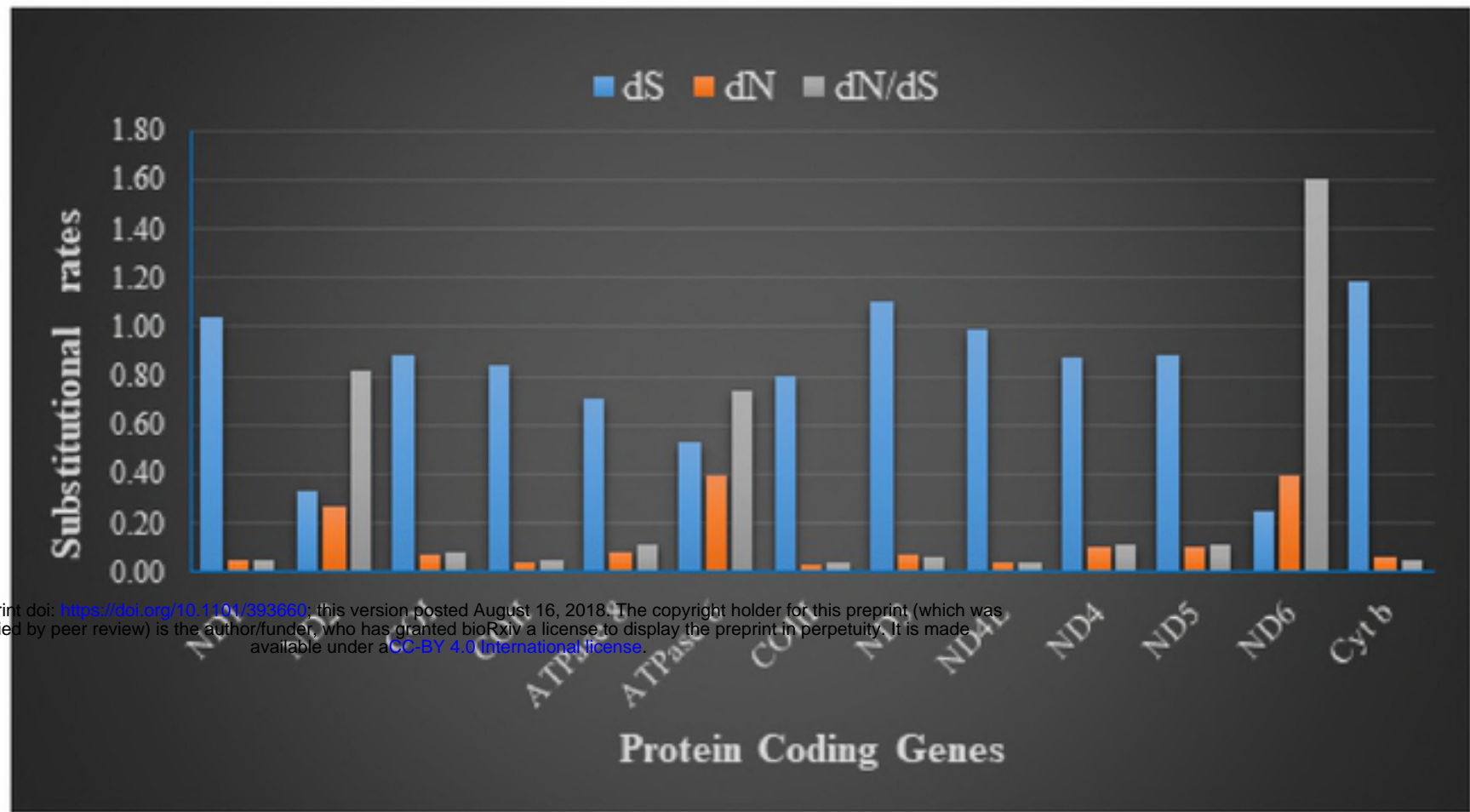
485

bioRxiv preprint doi: <https://doi.org/10.1101/393660>; this version posted August 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

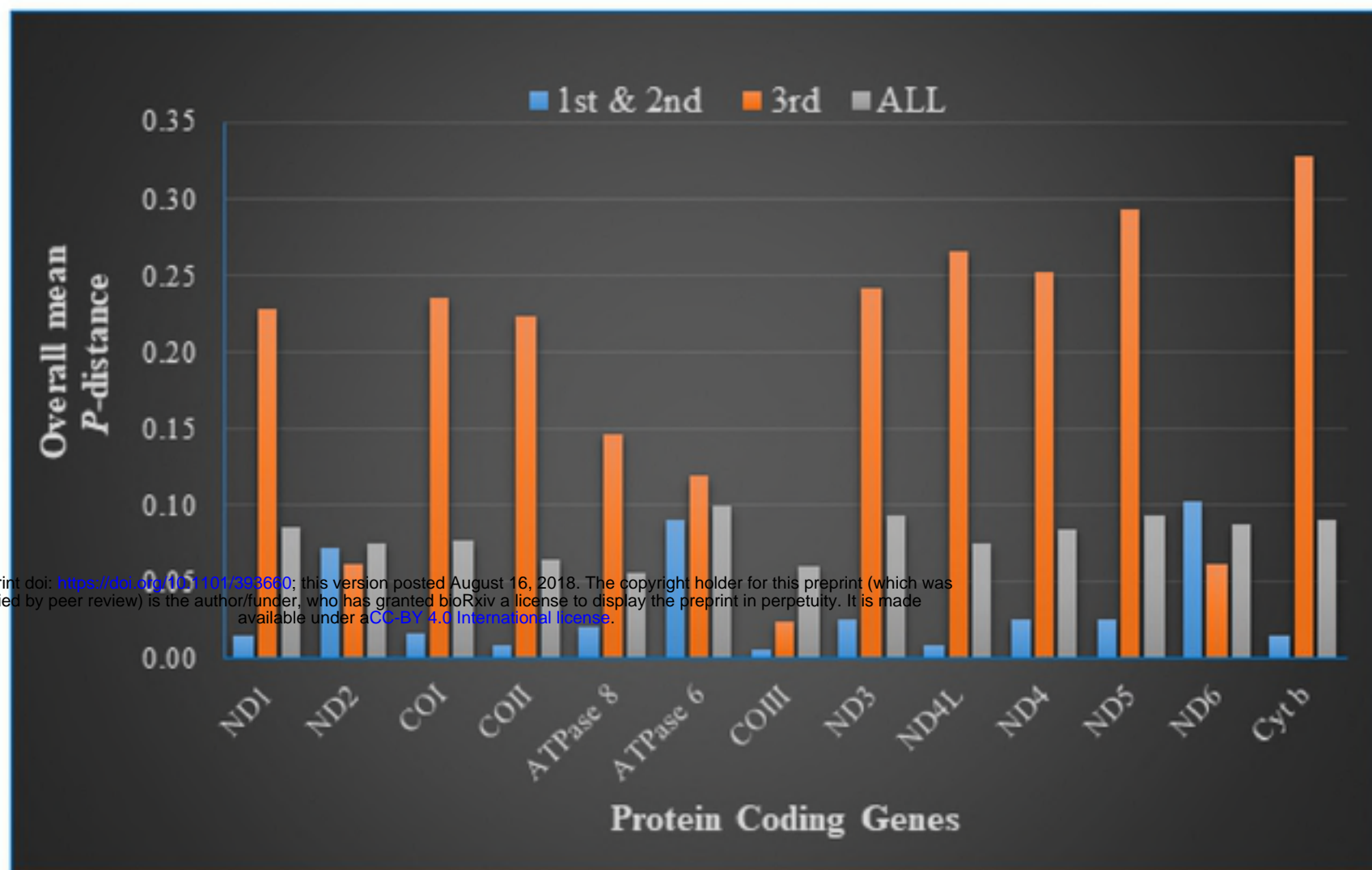


**Fig. 1.** Mitogenome organisation of *O. macrochir* generated by MitoAnnotator. The genes on the outer side of the circle are coded on the H-strand while those on the inner circle are coded with L-strand. *O. andersonii* showed similar organization.





**Fig 2.** The rate of non-synonymous substitution ( $dS$ ), rate of synonymous substitution ( $dN$ ) and the ratio of  $dS$  and  $dN$  for each protein coding gene.



**Fig. 3** Estimates of average evolutionary divergence over all sequence pairs of 20 cichlid species for each of the 13 protein coding genes calculated based on codon positions 1<sup>st</sup> + 2<sup>nd</sup>, 3<sup>rd</sup> and full sequence



*O. andersonii*

CCGAGCTCTGCCTTTATGTAAAATGCAATGCAT<sup>1</sup>ATATGTATTATCACCATTATTTTATATCAAACATATCCTATATATAAAT  
 ACATACAACCTTTAAAAACATACATTGTTTTCCACATATTTGTTATCAACATTTACAACCTAAGAAAAACATAAACCAATAA  
 ATGAAATCTTCCAAAACATCCCAAACCCTGAACGATAGTTAAGACCGAACACAACCTCTCATAAGTCAAGATATACCAAG  
 TACCCAACATTCTATTAACCTCTGAGTTATTTAATGTAGTAAGAGCCCACC<sup>2</sup>ATCAGTTGATTTCCTTAATGTCAACGGTTCTT  
 GAAGGTCAAGGACAGTACTTGTGGGGGT<sup>3</sup>TTCACACTTGAATTATTCCTGGCATCTGGTTCCT<sup>4</sup>ATTTCAAGTCCAATAATTG  
 TGATAGCTCCCATTCTTTCATTGACGCTTGCATAAGTTAATGGTGTAAATACATACTCCTCGTTACCCACCATGCCGGGCGT  
 TCTTCCAGGGTGTGGGGGGTTCTCTTTTTTTTTTTTCTTTCACCTTGACATTTTCAGAGTGCATACAGAAACGACAGACAAGGT  
 TGAACATTTTCTTGCTTGAACGGAAATAGTATGAGTGATGGTAAGATATTAATAGAAGAATTACATAACTGATATCAAGAG  
CATA<sup>5</sup>AAGTTAATCAAATTTTAAATTTTCTCCTAATTTTCTATCAACCTTCGGTTTTTGC GCGTTAAACCCCCCTACCC  
C<sup>6</sup>CCAAAACCTCTAAGATCTCTAATACTCCTGCAAACCCCCGGAAACAG<sup>7</sup>GAAAAGCTCTAGAAGTGACTTTTAGCGCTTTAA  
 TATGTGCATAAAATTAAGCTAATGCTGCTTATGTAGTACTGTTTACTGACCGGTCATGTATCCAATGTGTGTATATTATAC  
 TATTATAATATTGCACAT

bioRxiv preprint doi: <https://doi.org/10.1101/383660>; this version posted August 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

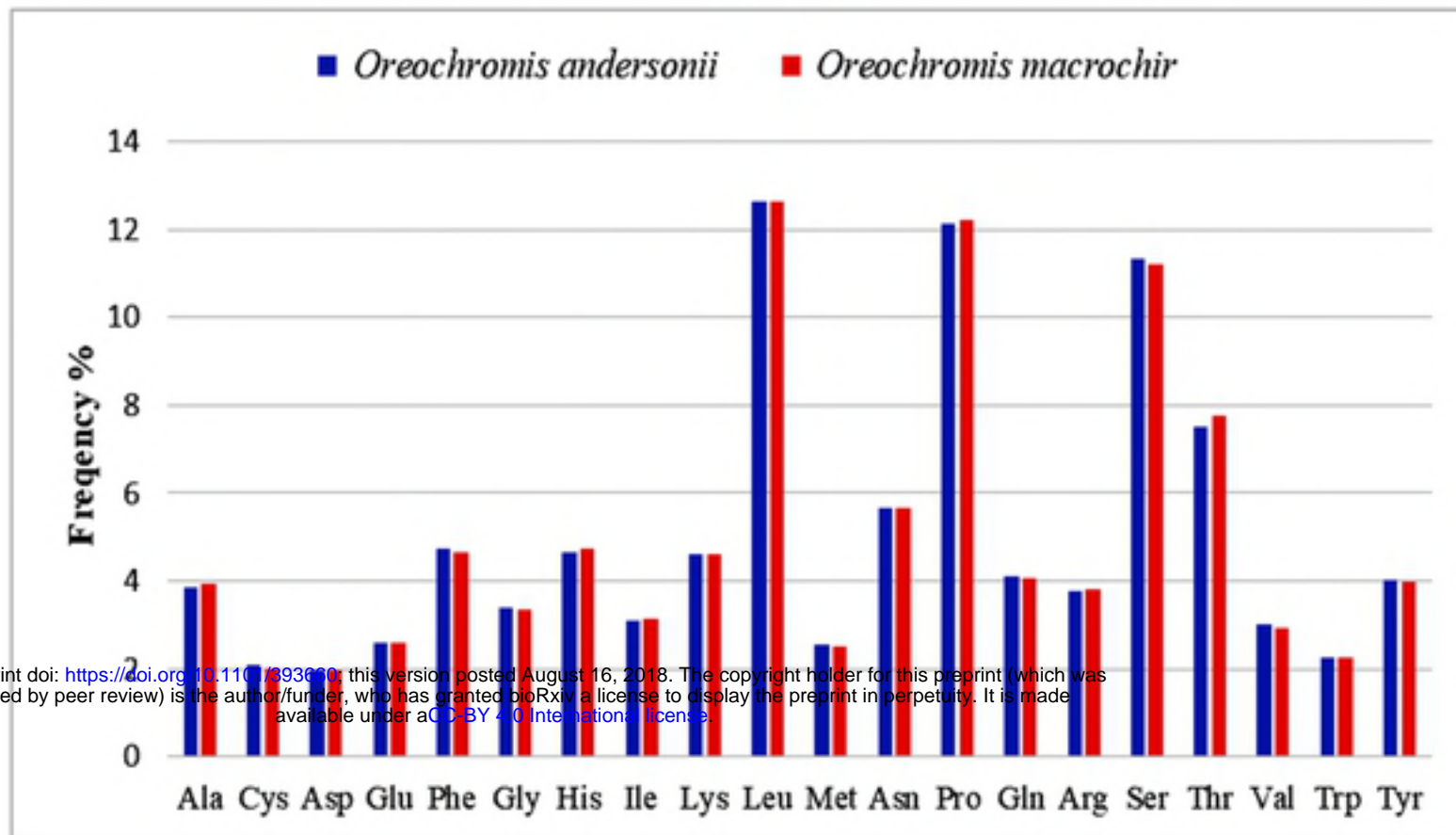
*O. macrochir*

CCGAGCTCTGCCTTTATGTAAAATACAATGCAT<sup>1</sup>ATATGTATTATCACCATTATTTTATATCAAACATATCCTATATATAAATA  
 CATAACAACCTTTAAAAACATACATTGTTTTCCACATATTTGTCATCAACATTTACAACCTAAGAAAAACATAAACCAATAAA  
 TGAAATTTTCCAAAACATTTCAAACCCTGAACGATAGTTAAGACCGAACACAACCTCTCATAAGTTAAGATATACCAAGT  
 ACCCAACATCCTATTAACCTCTCGAATTATTTAATGTAGTAAGAGCCCACC<sup>2</sup>ATCAGTTGATTTCCTTAATGTCAACGGTTCTTG  
 ATGGTCAAGGACAGTACTTGTGGGGGT<sup>3</sup>TTCACACTTGAATTATTCCTGGCATCTGGTTCCT<sup>4</sup>ATTTCAAGTCCAATAATTGT  
 TATAATCCCCATTCTTTCATTGACGCTTGCATAAGTTAATGGTGTAAATACATACTCCTCGTTACCCACCATGCCGGGCGTT  
 CTTTCCAGGGTGTGGGGGGTTCTCTTTTTTTTTTTTCTTTCACCTTGACATTTTCAGAGTGCATACAGAAACGACAGACAAGGT  
 GAACATTTTCTTGCTTGAACGGAAATAGTATGAATGGTGATAAGATATTAATAGAAGAACCACATAACTGATATCTAGAG  
CATA<sup>5</sup>AAGTTAATTAATCTTTCAATTTTCTCCTAATTTCTCTATCAACCTTCGGTTTTTGC GCGTTAAACCCCCCTACC  
CCC<sup>6</sup>CCAAAACCTCTAAGATCTCTAATACTCCTGCAAACCCCCGGAAACAG<sup>7</sup>GAAAAGCTCTAGAAGTGACTTTTAGCGCTTT  
 AATATGTGCATAAAATATTACGTAATGTGTGTATATGTAGTACTATCAGTGCACGAGTCATGTATCCAATGTGTGTATATTAT  
 ACTATTATAATATTGCACAT

Superscript numbers represent the sequences of the different parts of the control region as follows:

- <sup>1</sup> TAS – Domain I
- <sup>2</sup> CSB-F, <sup>3</sup> CSB-E, <sup>4</sup> CSB- D – Domain II
- <sup>5</sup> CSB- 1, <sup>6</sup> CSB- 2, <sup>7</sup> CSB- 3 – Domain III

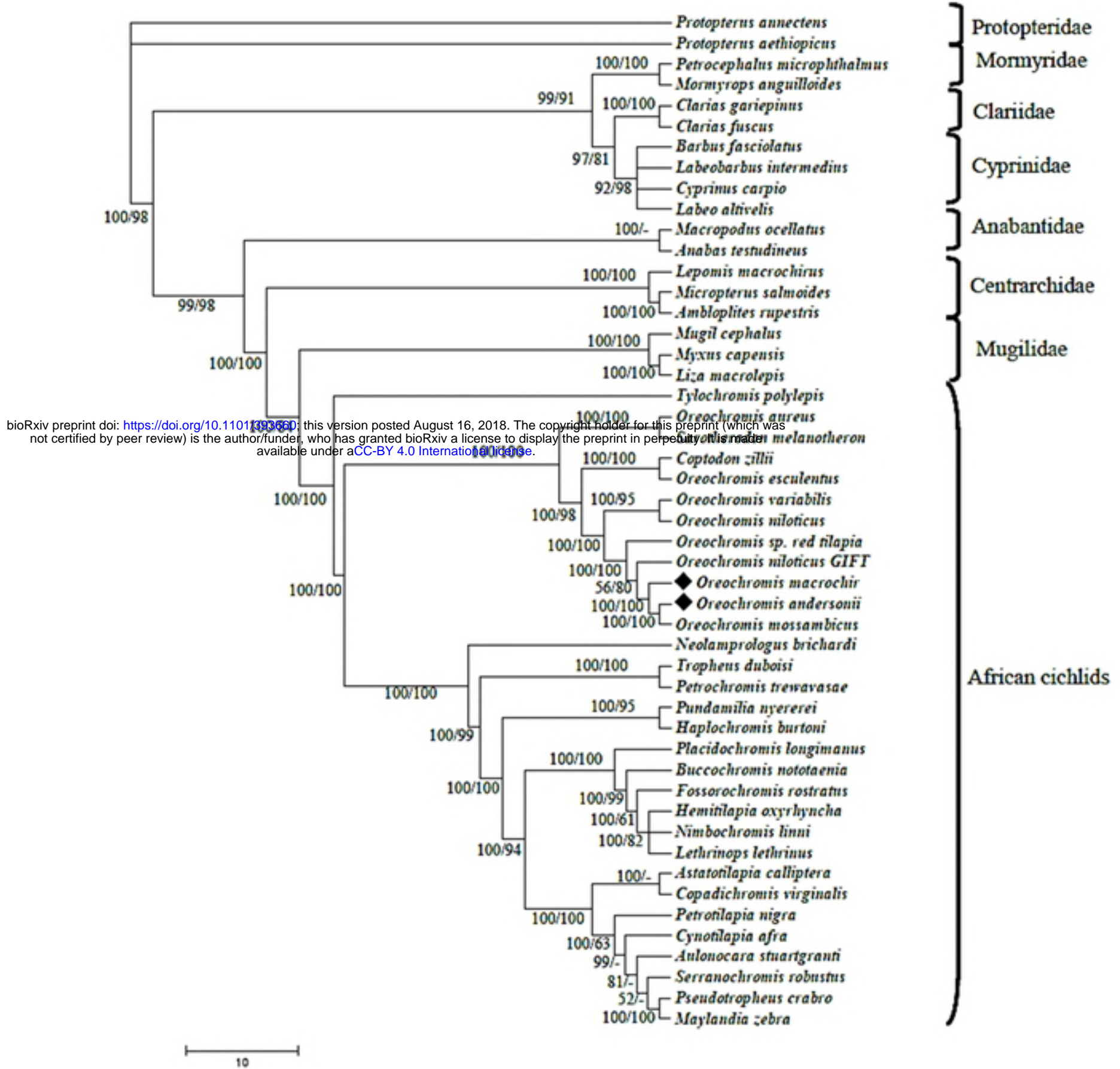
**Fig 4.** Different parts of the control region of *O. andersonii* and *O. macrochir*



bioRxiv preprint doi: <https://doi.org/10.1101/393600>; this version posted August 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

**Fig 5.** Amino acid frequency in the complete mitogenome of *O. andersonii* and *O. macrochir*





**Fig 6.** The phylogenetic tree showing the position of *O. andersonii* and *O. macrochir* with other 47 species based on Bayesian posterior probabilities and bootstrap values of ML (numbers in branches) of 12 concatenated protein genes (without *ND6*).