

DOI: 10.3724/SP.J.1118.2011.00790

奥利亚罗非鱼线粒体基因组全序列测定与系统进化分析

蒋宗良^{1,2}, 张明², 林勇¹, 朱佳杰¹, 黎明星², 罗永巨¹, 甘西¹

1. 广西水产研究所, 广西 南宁 530004;

2. 广西大学动物繁殖研究所, 广西 南宁 530004

摘要: 采用 LA-PCR(long amplification polymerase chain reaction)扩增方法获得奥利亚罗非鱼(*Oreochromis aureus*)线粒体基因组全序列。分析表明, 序列全长 16 632 bp, 包括 13 个蛋白质编码基因、22 个 tRNA 基因、2 个 rRNA 基因和 1 个长度为 931 bp 的主非编码区。A、T、G、C 碱基的组成分别为 27.89%、25.50%、15.62%、30.99%。基因排列与罗非鱼属的其他物种一致。13 个蛋白质基因除 COX1 用 TGT 做起始密码子外, 其他均以 ATG 为起始, 终止密码子除 COX2、Cyt b、ND4 为不完整的 T, 其余基因均以典型的 TAA 做终止密码子。22 个 tRNA 的二级结构都具有典型三叶草结构。系统发育分析显示, 罗非鱼属与丽鱼科(Cichlidae)其他物种分开, 聚为单系群, 其中奥利亚罗非鱼与尼罗罗非鱼亲缘关系较近, 莫桑比克罗非鱼和罗非鱼 KM-2006 亲缘关系较近。本研究旨在为进一步利用线粒体基因组分析罗非鱼群体遗传多样性、亲缘关系以及丽鱼科系统进化提供基础依据。

关键词: 奥利亚罗非鱼; 线粒体基因组; 序列分析; 系统进化

中图分类号: Q959

文献标志码: A

文章编号: 1005-8737-(2011)04-0790-11

奥利亚罗非鱼(*Oreochromis aureus*)属于鲈形目(Perciformes)、丽鱼科(Cichlidae), 具有繁殖周期短、生长速度快、抗逆性强等特点, 是热带和亚热带地区重要的淡水经济鱼种。动物线粒体基因组是核外遗传物质, 具有分子量小、结构简单、母系遗传等特点, 成为生物多样性、群体遗传学与系统进化研究中的重要分子标记^[1-3]。后生动物 mtDNA 为典型的环状双链分子结构, 长度大多在 14~18 kb 之间, 编码 37 个基因, 包括 13 个蛋白质基因(ND1-ND6、ND4L、COX1-COX3、ATP6、ATP8 和 Cyt b)、22 个 tRNA 以及 12S rRNA 和 16S rRNA^[4]。

目前, 线粒体基因已作为种类鉴定和系统发育分析的重要分子标记, 并且线粒体 ND2、16S

rRNA 与 Cyt b 基因、D-loop 序列已被应用于罗非鱼遗传多样性及系统发育关系分析^[5-8]。迄今, 很多丽鱼科鱼类的线粒体基因组序列被测定, 包括罗非鱼属的莫桑比克罗非鱼(*O. mossambicus*)、罗非鱼 KM-2006 (*Oreochromis* sp. KM-2006)^[9]、尼罗罗非鱼(*Oreochromis niloticus*)^[10], 见表 1。但奥利亚罗非鱼线粒体基因组全序列尚未见报道。本研究测定了奥利亚罗非鱼线粒体基因组序列, 分析了其基因组的主要结构信息, 并与同属的莫桑比克罗非鱼、罗非鱼 KM-2006 和尼罗罗非鱼进行了比较分析, 构建了丽鱼科分子系统发生树, 初步探讨了奥利亚罗非鱼与丽鱼科其他鱼类的进化关系。本研究结果为进一步利用线粒体基因组分析罗非鱼群体遗传多样性、亲缘关系以及丽鱼

收稿日期: 2010-08-02; 修订日期: 2010-11-29.

基金项目: 农业部公益性行业科研项目专项(Nyhyzx07-044-04)资助; 广西科学基金资助项目(桂科自 09910082Z); 现代农业产业技术体系(nycytx-48-2).

作者简介: 蒋宗良(1985-), 男, 博士研究生, 专业方向: 动物遗传育种与生物技术. E-mail: jiangzlg@126.com

通信作者: 甘西(1957-), 男, 研究员, 研究方向: 水产生物技术. E-mail: ganxien@126.com

科系统进化提供了重要的基础信息。

1 材料与方法

1.1 样品采集

本研究所用的奥利亚罗非鱼(美国品系)取自广西水产研究所国家级罗非鱼良种场。

1.2 总 DNA 的提取

取奥利亚罗非鱼活鱼新鲜尾鳍, 采用 Qiagen Tissue Extraction Kit 试剂盒提取其总 DNA。总 DNA 用 1%琼脂糖凝胶电泳检测, -20 保存。

1.3 长 PCR 扩增和 DNA 测序

参考莫桑比克罗非鱼(*Oreochromis mossambicus*)线粒体基因组全序列(GenBank 登录号: NC_007231)设计 4 对引物扩增奥利亚罗非鱼全线粒体基因组(表 1)。PCR 产物通过步移法测序。部

分序列通过设计特异性引物扩增, 然后克隆到 pMT-18 载体上进行测序。

PCR 反应在 Eppendorf 梯度 PCR 仪上进行。PCR 扩增反应体积为 50 μ L, 含 2 μ L DNA 模板(50 ng/ μ L)、5 μ L 10 $\times$ LA PCR 缓冲液(Mg²⁺ free, Takara)、5 μ L MgCl₂ (2.5 mmol/L)、8 μ L dNTP (2.5 mmol/L), 正反向引物各 2 μ L (10 μ mol/L)、0.5 μ L LA Taq DNA 聚合酶 (5 U/ μ L, Takara), 加 25.5 μ L 去离子灭菌水补足。反应条件为: 循环起始设置为 94 预变性 2 min, 94 变性 30 s, 56~62 退火 1 min, 72 延伸 5 min(根据扩增产物的长度), 反应 35 个循环后。72 补齐 10 min。PCR 产物经 1.5%琼脂糖凝胶电泳后切胶, 用 DNA 凝胶回收试剂盒(Takara)纯化后, 委托上海杰李生物有限公司进行测序(ABI 3730x1 DNA 测序仪测序)。

表 1 奥利亚罗非鱼线粒体基因组扩增引物
Tab.1 Primers used to amplify the mitochondrial genome of *Oreochromis aureus*

编号 No.	扩增区域 amplifying area	上游引物(5'-3') upstream primer(5'-3')	下游引物(5'-3') downstream primer(5'-3')	退火温度/ anneal temperature
P1	475-4999	TGCGAAACAACTGGGATTAGATAC	GGGAAAGTTAGTTGGGTGGAAGGT	59
P2	4892-10170	GCAGCCCTTTCAGCCCTCCTTAGCC	GCGGAAAGAAGGTGGGTTCGGTGTA	62
P3	9835-14240	GTCTGCCCTTCTCCCTTCGCTTCT	TGTGGTAGCAGGAATAGGATGTGGG	56
P4	13937-1141	ACCTCCACCTGCCGAATACATCAA	GGTTTGTTGTTGGGAGGTGGTGT	62

1.4 基因注释和序列分析

原始测序峰图文件首先利用拼接软件 Phred 处理, 使每个碱基的质量值在 20 以上, 然后在默认参数下条件下用软件 Phrap 拼接^[11-12]。拼接结果和序列碱基质量通过 Consed 软件检查, 手工去除错误拼接^[13]。蛋白质编码基因和 rRNA 基因通过软件 DOGMA^[14]和 BLAST 比对(<http://www.ncbi.nlm.nih.gov/BLAST>)识别。每个基因的起始和终止的确切位置, 参考莫桑比克罗非鱼全线粒体基因组的注释并稍有修改。tRNA 基因通过 tRNAscan-SE 1.21^[15]软件预测。奥利亚罗非鱼线粒体基因组全序列已经提交 GenBank 数据库, 序列的登录号是 GU477630。

1.5 系统发生分析

从 GenBank 下载了 12 个丽鱼科品质的线粒体蛋白质编码基因的核苷酸序列和氨基酸序列,

外加本次测序的奥利亚罗非鱼, 以斑马鱼作为外群, 将 14 个个体的蛋白质编码基因的核苷酸序列氨基酸序列串联起来, 在默认参数设置条件下, 用软件 ClustalW 进行序列的多重比对。使用 2 种不同的系统发生分析方法构建系统发生树。1 种方法是最大似然法(ML), 使用的软件为 PHYML 2.4^[16], 另外一种方法是邻接法(NJ), 使用的软件为 MEGA 4.0^[17]。构建 14 个个体的类的系统进化关系, 系统树各分支的可靠性通过重抽样分析评价, 自展检验(bootstrap analysis) 设为 100 000 次。

2 结果与分析

2.1 罗非鱼线粒体基因组的特征

奥利亚罗非鱼线粒体基因组全长 16 632 bp。包括 13 个蛋白质编码基因、22 个 tRNA、2 个 rRNA 和 1 个 D-loop。其结构紧凑, 除控制区外, 有的

相邻基因之间还有少量的不编码的核苷酸, 有的相邻基因之间还有碱基重叠现象。与其他脊椎动物线粒体基因结构特征一样, 除了 ND6 和 8 个 tRNA 在 L 链上编码, 其他的所有基因都在 H 链

上编码。奥利亚罗非鱼线粒体基因组组成见表 2。

2.2 核苷酸组成

奥利亚罗非鱼线粒体基因组的碱基含量由高到低依次为 C(27.89%)、A(25.50%)、T(30.99%)、G

表 2 奥利亚罗非鱼线粒体基因组基因排布顺序表
Tab. 2 Profile of the mitochondrial genome of *Oreochromis aureus*

基因名称 gene name	位置 location	大小/bp size		密码子 codon		基因间隔大小 intergenic nucleotides	编码链 strand
		核苷酸 nucleotide	氨基酸 amino acid	起始密码 start codon	终止密码 stop codon		
tRNA-Phe	1-69	69					H
12S rRNA	70-1014	945					H
tRNA-Val	1015-1086	72					H
16S rRNA	1087-2780	1694					H
tRNA-Leu	2781-2854	74					H
ND1	2855-3829	975	324	ATG	TAG		H
tRNA-Ile	3833-3902	70				+3	H
tRNA-Gln	3902-3972	71				-1	L
tRNA-Met	3972-4040	69				-1	H
ND2	4041-5087	1047	348	ATG	TAA		H
tRNA-Trp	5087-5158	71				-1	H
tRNA-Ala	5160-5228	69				+1	L
tRNA-Asn	5230-5302	73				+1	L
tRNA-Cys	5338-5403	66				+35	L
tRNA-Tyr	5404-5473	70					L
COX1	5475-7070	1596	531	TGT	TAA	+1	H
tRNA-Ser	7071-7141	71					L
tRNA-Asp	7145-7217	73				+3	H
COX2	7223-7913	691	230	ATG	T-	+5	H
tRNA-Lys	7914-7987	74					H
ATP8	7989-8156	168	55	ATG	TAA	+1	H
ATP6	8147-8830	684	227	ATG	TAA	-10	H
COX3	8830-9669	840	279	ATG	TAA	-1	H
tRNA-Gly	9614-9685	72				-56	H
ND3	9686-10036	351	116	ATG	TAG		H
tRNA-Arg	10035-10103	69				-2	H
ND4L	10104-10400	297	98	ATG	TAA		H
ND4	10394-11774	1381	460	ATG	T-	-7	H
tRNA-His	11775-11843	69					H
tRNA-Ser	11844-11910	67					H
tRNA-Leu	11915-11987	73				+4	H
ND5	11988-13826	1839	612	ATG	TAA		H
ND6	13823-14344	522	173	ATG	TAA	-4	L
tRNA-Glu	14345-14413	69					L
Cyt b	14418-15552	1135	378	ATG	T-	+4	H
tRNA-Thr	15559-15630	72				+6	H
tRNA-Pro	15631-15700	70					L
D-loop	15701-16632	931					H

(15.62%), C 的含量最高, 而 G 最低。53.39%(A+T)>46.61(C+G), 表现出适度的 A+T 偏好性, 这与脊椎动物 A 或者 T 的密码子偏好性是一致的^[18], 其 A+T 含量略低于莫桑比克罗非鱼(54.39%)、SP(54.3%)和尼罗罗非鱼(54.18%)。4 种罗非鱼的线粒体基因组的长度相差不大, 其中莫桑比克最长(16 641bp); 其次为奥利亚罗非鱼(16 632 bp), 尼罗罗非鱼(16 627 bp)和 O.SP(16 626)最短(表 3)。

2.3 氨基酸和密码子的偏向性

奥利亚罗非鱼所有蛋白质编码基因编码 3831 个氨基酸。所有蛋白质编码基因序列的 AT 含量为 52.66%, 其中, 密码子第一位的 AT 含量为 52.43%, 第二位和第三位 AT 含量分别为 51.68 %和 53.89%。线粒体基因组一个明显的特点是密码子第三位的 AT 含量明显高于密码子的第一位和第二位。表 3 列出了 4 种罗非鱼线粒体基因组数据。

在 3 831 个氨基酸中, 亮氨酸(Leu)是使用最多的氨基酸, 占氨基酸总数的 17.1%; 丙氨酸(Ala)是第二常用的氨基酸, 占氨基酸总数的 9.1%; 而半胱氨酸(Cys)是使用最少的氨基酸。密码子 CUC(亮氨酸)是最常用的密码子, 占密码子总数的 5.4%; 密码子 AUU (异亮氨酸)是第二常用的密码子, 占密码子总数的 4.7%; 密码子 AGA、AGG 没有使用。线粒体基因组的一个普遍特征是序列中核苷酸 AT 的出现的高频率, 导致以 AT 编码的相应氨基酸呈现一定偏好性。在奥利亚罗非鱼中, AT 含量与莫桑比克罗非鱼、尼罗罗非鱼和罗非鱼 KM-2006 的 AT 含量相似(表 4)。

2.4 蛋白质编码基因

与典型线粒体基因组编码 13 个蛋白基因一样, 奥利亚罗非鱼线粒体也有 13 个蛋白质编码基因, 编码基因核苷酸总长是 11 526 bp, 占整个基因组的 69.35%。在奥利亚罗非鱼蛋白质编码基因序列中, 存在开放阅读框的重叠: ATP6 和 ATP8 之间有 10 个核苷酸重叠, 这与大多数脊椎动物一致^[19]; 在 ND4L 与 ND4 之间有 7 个核苷酸重叠, 这在所有脊椎动物中都是恒定的^[19]。此特征与尼罗罗非鱼也一致^[10]。除了 COX1 基因使用 TGT 作为启动子外, 所有其他的蛋白质编码基因

都使用 ATG 作为启动子, 这与莫桑比克罗非鱼、尼罗罗非鱼和罗非鱼 KM-2006 的 COX1 基因均使用 GTG 作为启动子不一致。ND2、COX1、ATP8、ATP6、COX3、ND4L、ND5、ND6 使用 TAA 做终止密码子, ND1、ND3 使用 TAG 作为终止密码子, 而 COX2、ND4、Cyt b 使用不完整的终止密码子, 其他罗非鱼中也存在不完整终止密码的情况, 不完全终止密码子在脊椎动物线粒体 DNA 中较为常见^[19], 转录后可以变成完整的终止密码。一般认为, 转录的 mRNA 可能利用其 PolyA 尾补全终止密码子从而终止翻译^[20]。

2.5 转运 RNA(tRNA)和核糖体 RNA(rRNA)

奥利亚罗非鱼线粒体基因组具有线粒体标准的 22 个 tRNA 基因, 长度在 66~74 bp 之间, 这些 tRNA 都能够折叠成典型的三叶草二级结构(图 1)。在接受臂和反密码子的颈部有一些错配碱基。tRNA 在 H 链和 L 链上都有编码, 其中, tRNA-Gln、tRNA-Ala、tRNA-Asn、tRNA-Cys、tRNA-Tyr、tRNA-Ser、tRNA-Glu、tRNA-Pro 在 L 链上编码, 其他 14 个在 H 链上编码。

尽管已经推测出 2 种 rRNA 基因的边界, 但 rRNA 基因精确的边界需要有转录图谱才能确定。奥利亚罗非鱼线粒体基因组 16S rRNA 位于 tRNA-Val 和 tRNA-Leu 间, 而 12S rRNA 位于 tRNA-Phe 和 tRNA-Val 间, 长度分别是 1 694 bp 和 945 bp, AT 含量分别是 54.72%和 49.9%; 奥利亚罗非鱼 16S rRNA 和 12S rRNA 的 AT 含量与其他 3 种的相差不大(表 3)。

2.6 非编码区

与其他大多数脊椎动物一样, 奥利亚罗非鱼线粒体基因组包含大量非编码的核苷酸。在奥利亚罗非鱼全序列中, 共有 13 处非编码区, 大小在 1~923 bp 之间。最大的非编码区位于基因 tRNA-Pro 和 tRNA-Phe 之间, 占非编码核苷酸的 92.86%。这一区域的 AT 含量高达 64.46%, 高于全基因组和蛋白编码区域 AT 含量的平均值(54.41%和 54.02%)。

后生动物线粒体基因组中通常包含多个大小不一的非编码区^[21-22], 最大的非编码区通常认为包含着线粒体基因的复制和转录的信号, 因此通

表 3 奥利亚罗非鱼线粒体基因组特征

Tab. 3 Genomic characteristics of *Oreochromis aureus* mitochondrial genome

物种名称 species	GenBank 登录号 GenBank Acc. No.	正链 light-strand		蛋白质编码基因 protein-coding					16S rRNA		12S rRNA		D-loop	
		长度/bp length	(A+T)/%	氨基酸 数目 No. of amino-acid	数目 No.	(A+T)/%			长度 /bp length	(A+T) /%	长度 /bp length	(A+T) /%	长度/bp length	(A+T) /%
						密码子(第一、二、三位) Codon positions								
						— — —								
						first	second	third						
<i>O.aurea</i>	GU_477630	16632	53.39	3832	52.66	52.43	51.68	53.89	1694	54.72	944	49.9	931	65.09
<i>O.mossambicus</i>	NC_007231	16641	54.39	3839	53.91	52.45	52.85	56.43	1697	54.04	941	50.69	923	65.33
<i>O. sp.</i>	NC_009057	16626	54.3	3816	53.67	52.89	55.49	52.62	1695	53.98	944	50.43	927	65.59
<i>O.niloticus</i>	NC_013663	16627	54.19	3816	53.72	51.41	57.52	52.21	1699	53.56	944	50.42	924	64.07

表 4 奥利亚罗非鱼线粒体 12 个蛋白质编码基因密码子使用情况

Tab. 4 Codon usage in 13 protein-coding genes of *Oreochromis aureus*

氨基酸 amino-acid	密码子 codon	出现频率/% frequency	氨基酸 amino-acid	密码子 codon	出现频率/% frequency	氨基酸 amino-acid	密码子 codon	出现频率/% frequency	氨基酸 amino-acid	密码子 codon	出现频率/% frequency
Phe	UUU	90	Ser	UCU	44	Tyr	UAU	41	Cys	UGU	8
	UUC	152		UCC	94		UAC	72		UGC	19
Leu	UUA	62		UCA	55		UAA	0	Trp	UGA	107
	UUG	27		UCG	7	His	UAG	0		UGG	14
	CUU	147	Pro	CCU	50		CAU	26	Arg	CGU	11
	CUC	208		CCC	120		CAC	82		CGC	18
Ile	CUA	180		CCA	45	Gln	CAA	90		CGA	40
	CUG	25		CCG	7		CAG	10		CGG	6
	AUU	164	Thr	ACU	53	Asn	AAU	27	Ser	AGU	13
	AUC	126		ACC	133		AAC	85		AGC	43
Met	AUA	93		ACA	118	Lys	AAA	68		AGA	0
	AUG	61	Ala	ACG	4		AAG	11		AGG	0
Val	GUU	58		GCU	68	Asp	GAU	15	Gly	GGU	35
	GUC	64		GCC	165		GAC	62		GGC	96
	GUA	57		GCA	99	Glu	GAA	87		GGA	71
	GUG	22		GCG	17		GAG	16		GGG	43

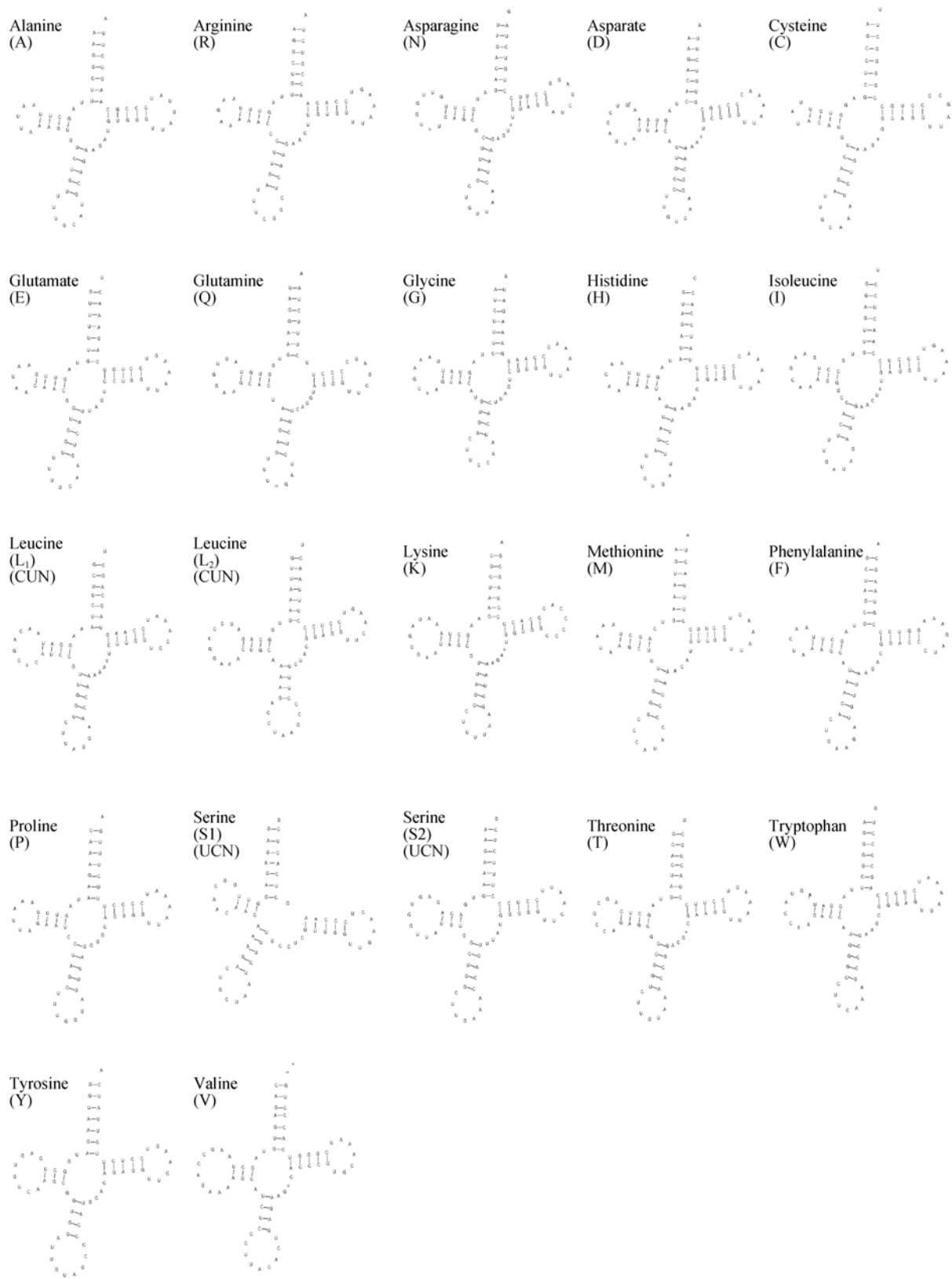


图 1 奥利亚罗非鱼 22 个 tRNA 的二级结构

Fig. 1 Putative secondary structures for the 22 tRNA genes of the *Oreochromis aureus* mitochondrial genome

常被称为控制区^[23]。非编码区序列较高的 AT 含量,是它的一个典型的特征,用来鉴定线粒体复制的起始。串连重复序列在动物线粒体控制区比较常见^[24],它经常形成稳定的二级结构,在线粒体基因的复制和转录的早期阶段发挥着重要作用^[25-26]。在奥利亚罗非鱼线粒体基因组中也存在着一些大小为 10~20 bp 重复单位组成的串连重复序列。

2.7 系统进化分析

使用 NJ、ML 法对丽鱼科已测序 13 个物种的蛋白质编码基因的核苷酸序列和氨基酸序列构建系统进化树,以斑马鱼作为外群,结构都显示了相似的进化拓扑结构,图 2-图 4 是基于表 5 所列 14 个物种的 13 个蛋白质编码的核苷酸序列和氨基酸序列采用 ML 和 NJ 法构建的系统发生树。

罗非鱼属的 4 种罗非鱼中,根据 ML 法构建

表 5 系统发生分析中使用的物种分类列表
Tab.5 Category list of species using in the phylogenetic analysis

物种 species	分类地位 classification	GenBank 登录号 GenBank accession No.	参考文献 reference
<i>Tropheus duboisi</i>	Tropheini; Tropheus	NC_009063	Mabuchi,K et al, 2007
<i>Neolamprologus brichardi</i>	Lamprologini; Neolamprologus	NC_009062	Mabuchi,K et al, 2007
<i>Astronotus ocellatus</i>	Astronotinae Astronotini; Astronotus	NC_009058	Mabuchi,K et al, 2007
<i>Hypselecara temporalis</i>	Cichlasomatinae; Heroini;Hypselecara	NC_011168	Azuma,Y et al, 2008
<i>Ptychochromoides katria</i>	Ptychochrominae; Ptychochromoides	NC_011169	Azuma,Y et al, 2008
<i>Paratilapia polleni</i>	Etroplinae; Paratilapia	NC_011170	Azuma,Y et al, 2008
<i>Tylochromis polylepis</i>	Tylochromini; Tylochromis	NC_011171	Azuma,Y et al, 2008
<i>Paretroplus maculatus</i>	Etroplinae; Paretroplus	NC_011177	Azuma,Y et al, 2008
<i>Etroplus maculatus</i>	Etroplinae; Etroplus	NC_011179	Azuma,Y et al, 2008
<i>Oreochromis</i> sp. KM-2006	Tilapiiini; Oreochromis	NC_009057	Mabuchi,K et al, 2007
<i>Oreochromis mossambicus</i>	Tilapiiini; Oreochromis	NC_007231	Chen,I.-S et al
<i>Oreochromis niloticus</i>	Tilapiiini; Oreochromis	NC_013663	Yang,L et al, 2010
<i>Oreochromis aureus</i>	Tilapiiini;Oreochromis	GU477624	本研究
<i>Danio rerio</i>	Cyprinidae; Danio	AC024175	Milam,J.E et al

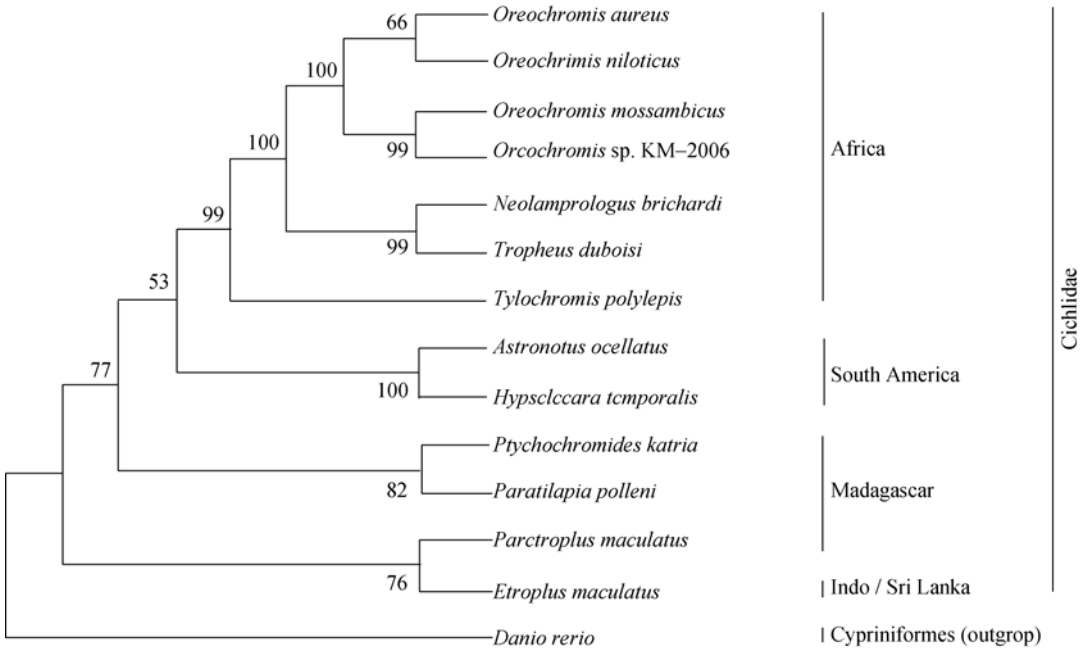


图 2 基于蛋白质编码基因的氨基酸序列采用邻接法构建的系统发生树

Fig. 2 Topology derived from NJ analysis of 13 concatenated mitochondrial nucleic acid data from 14 mitochondrial genomes

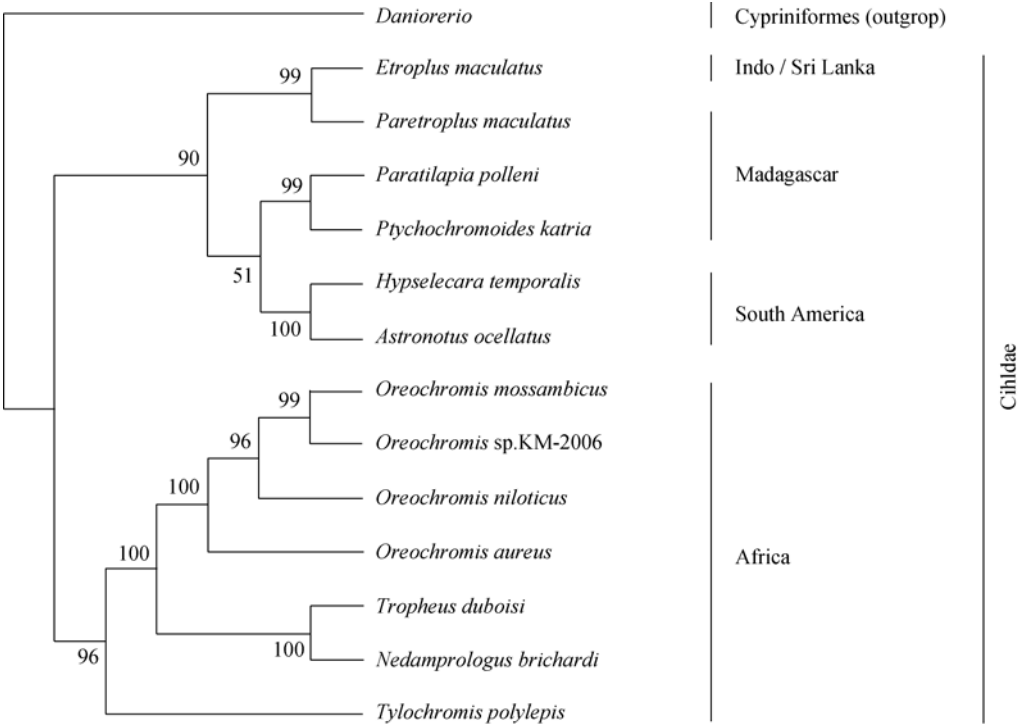


图 3 基于蛋白质编码基因的核苷酸序列采用最大似然法(ML)构建的系统发生树

Fig. 3 Topology derived from ML analysis of 13 concatenated mitochondrial nucleotide data from 14 mitochondrial protein-coding genes.

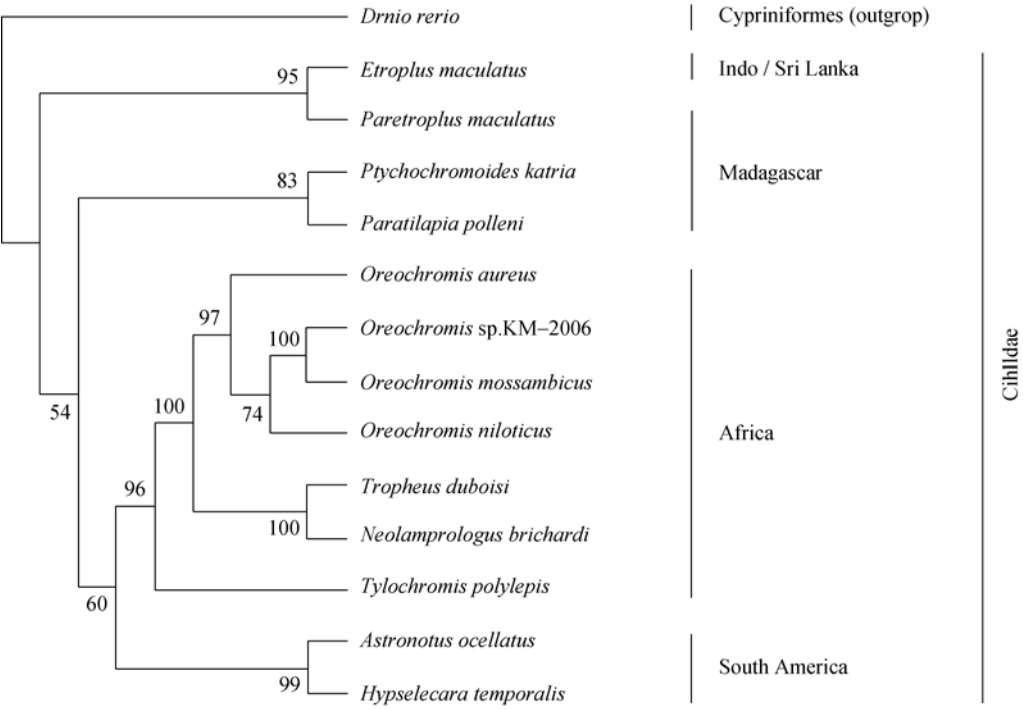


图 4 基于蛋白质编码基因的氨基酸序列采用邻接法构建的系统发生树

Fig. 4 Topology derived from ML analysis of 13 concatenated mitochondrial nucleic acid data from 14 mitochondrial genomes

的核苷酸和氨基酸树分析,莫桑比克罗非鱼与罗非鱼 KM-2006 首先聚成一支,然后与尼罗罗非鱼聚类,最后与奥利亚罗非鱼聚成一支后再跟其他丽鱼科的鱼类聚类。而根据 NJ 构建的氨基酸树分析,奥利亚罗非鱼与尼罗罗非鱼具成一支,莫桑比克罗非鱼与罗非鱼 KM-2006 聚成一支,然后这两支再聚在一起后再与其他丽鱼科的鱼类聚类。系统树清楚的显示出奥利亚罗非鱼与尼罗罗非鱼以及莫桑比克罗非鱼与罗非鱼 KM-2006 有非常近的亲缘关系,这与传统的形态学分类结果一致,说明线粒体基因组适用于丽鱼科系统发育分析。此系统进化树是在卢迈新^[10]建树基础上,增加了本研究所测序的奥利亚罗非鱼,结果与其一致。郭奕惠等^[7]对线粒体 DNA 16S rRNA 的研究结果表明莫桑比克罗非鱼是红罗非鱼的母本。本研究得出的基于全基因组系统分析表明,莫桑比克罗非鱼与罗非鱼 KM-2006 聚成一支,罗非鱼 KM-2006 即为红罗非鱼种,因此本研究支持了以上观点,这也符合线粒体母系遗传的特点。

3 讨论

罗非鱼类为热带鱼类,现已成为世界性的主要养殖鱼类之一。目前制约罗非鱼养殖业发展的突出问题是种质混杂和退化,罗非鱼的进化历史非常短^[27],亲缘关系很近,其遗传差异的区分十分不容易。开展罗非鱼的分子系统发生研究可了解罗非鱼进化的历史和进化机制,探讨罗非鱼种间亲缘关系,为罗非鱼保种育种提供理论依据。

本研究测定了奥利亚罗非鱼线粒体基因组全序列,其基因组成与其他硬骨鱼类类似。奥利亚罗非鱼线粒体基因组大小(16 632 bp)略小于莫桑比克罗非鱼线粒体基因组(16 641bp)、略大于尼罗罗非鱼(16 627 bp)和罗非鱼 KM-2006(16 626 bp),线粒体基因组结构特征跟其他罗非鱼线粒体基因结构特征一样,表现出适度的 A+T 偏好性;大部分基因在 H 链编码。编码基因起始密码子除 COX1 用 TGT 做起始密码子外,其他均以 ATG 为起始,终止密码子除 COX2、Cyt b、ND4 为不完整的 T,其余基因均为典型的 TAA。这与莫桑比

克罗非鱼、尼罗罗非鱼和罗非鱼 KM-2006 的 COX1 基因均使用 GTG 作为启动子不一致。总之,奥利亚罗非鱼线粒体的结构与其他 3 种罗非鱼均差别不大,这说明线粒体非常保守,适合于罗非鱼的系统进化发生分析。

目前,鱼类线粒体基因组的研究迅速发展,广泛应用于研究鱼类种内种间系统发生关系、物种起源及遗传分化、种内多态及地理分布关系、种质评估和种群遗传结构研究等方面。目前 GenBank 上已公布的鱼类线粒体基因组全序列已超过 400 种,通常为 16 kb 左右长度,呈双链闭合环状结构,一般包括 1 段 D-loop, 22 个 tRNA 基因, 13 个编码疏水性蛋白质亚基基因(包括 ND1-ND6、COI-COIII、Cyt b 等)和 2 个 rRNA (16S rRNA 和 12SrRNA)。对于近缘物种间的系统发育关系,一般选用进化速度比较快的区域如 D-loop、Cyt b、COX1、16S rDNA 等;而对于远缘物种,则选用相对保守的区域如 12S rDNA、ND1-ND5 等。目前,线粒体 ND2、16S rRNA、Cyt b 与 D-loop 基因序列已被应用于罗非鱼遗传多样性及系统发育关系分析^[5-8]。但是,较短的线粒体基因序列在分析高级分类阶元系统发育关系上表现出很大的局限性,应用全线粒体基因组序列的信息可以分析、重建一些处于高级分类阶元的系统发育关系。目前,采用线粒体全基因组进行罗非鱼系统分析的报道比较少,国内只有卢迈新^[10]测定了尼罗罗非鱼线粒体全序列,并采用线粒体全基因组构建了莫桑比克罗非鱼、尼罗罗非鱼和罗非鱼 KM-2006 的进化树。本研究系统发育分析显示,罗非鱼属与丽鱼科其他物种分开,聚为单系群,其中奥利亚罗非鱼与尼罗罗非鱼聚类后再与莫桑比克罗非鱼和罗非鱼 KM-2006 聚类。但要更深入的分析罗非鱼属各种间的进化关系,还需要更多的罗非鱼品种的线粒体基因组数据来揭示罗非鱼属系统发生关系。

参考文献:

- [1] Mindell DP, Sorenson MD, Dimcheff DEM, et al. Interordinal relationships of birds and other reptiles based on whole mi-

- tochondrial genomes[J]. Syst Biol, 1999, 48(1): 138–152.
- [2] Kumazawa Y, Ota H, Nishida M, et al. The complete nucleotide sequence of a snake (*Dinodon semicarinatus*) mitochondrial genome with two identical control regions[J]. Genetics, 1998, 150(1): 313–329.
- [3] Caccone A, Gentile G, Burns CE, et al. Extreme difference in rate of mitochondrial and nuclear DNA evolution in a large ectotherm *Galapagos tortoises* [J]. Mol Phylogenet Evol, 2004, 31(2): 794–798.
- [4] Wolstenholme D R. Animal mitochondrial DNA: structure and evolution [J]. Int Rev Cytol, 1992, 141: 173–216.
- [5] Klett V, Meyer A. What, if anything, is a tilapia? — mitochondrial ND2 phylogeny of tilapiines and the evolution of parental care systems in the African Cichlid fishes[J]. Mol Biol Evol, 2002, 19(6): 865–883.
- [6] 杨玲, 李娴, 王成武, 等. 5 个罗非鱼品种(系) 线粒体 DNA 16S rRNA 和 Cyt b 基因片段序列分析[J]. 长江大学学报: 自然科学版, 2008, 5(3): 35–40.
- [7] 郭奕惠, 黄桂菊, 喻达辉, 等. 我国主要养殖罗非鱼的 16S rRNA 序列特征分析[J]. 上海水产大学学报, 2007, 16(5): 490–494.
- [8] Maria Rowena R, Romana-Eguia, Minoru Ikeda, et al. Genetic diversity in farmed Asian Nile and redhybrid tilapia stocks evaluated from microsatellite and mitochondrial DNA analysis[J]. Aquaculture, 2004, 236: 131–150.
- [9] Mabuchi K, Miya M, Azuma Y, et al. Independent evolution of the specialized pharyngeal jaw apparatus in cichlid and labrid fishes[J]. J BMC Evol Biol, 2007, 7 (10): 2037–2046
- [10] 杨丽萍, 卢迈新, 叶星, 等. 尼罗罗非鱼线粒体基因组全序列测定与系统进化分析[J]. 中国生物化学与分子生物学报, 2010, 26(5): 484–490.
- [11] Ewing B, Green P. Basecalling of automated sequencer traces using phred II. Error probabilities[J]. Genome Res. 1998, 8: 186–194.
- [12] Ewing B, Hillier L, Wendl M, et al. Basecalling of automated sequencer traces using phred. I. Accuracy assessment[J]. Genome Res, 1998, 8: 175–185.
- [13] Gordon D, Abajian C, Green P. Consed: a graphical tool for sequence finishing[J]. Genome Res, 1998, 8: 195–202.
- [14] Wyman S K, Jansen R K, Boore J L. Automatic annotation of organellar genomes with DOGMA[J]. Bioinformatics, 2004, 20: 3 252–3 255.
- [15] Lowe T M, Eddy S R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence[J]. Nucleic Acids Res, 1997, 25: 955–964.
- [16] Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood[J]. Syst Biol, 2003, 52: 696–704.
- [17] MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Tamura K, Dudley J, Nei M, Kumar S. Mol Biol Evol. 2007 Aug; 24(8): 1596–9. Epub 2007 May 7.
- [18] Brown W M. The mitochondrial genome of animals [J]. Mol Evolut Gen, 1985, 21: 95–100.
- [19] Liu Y, Cui Z. The complete mitochondrial genome sequence of the cutlassfish *Trichiurus japonicus* (Perciformes: Trichiuridae): genome characterization and phylogenetic considerations[J]. Mar Genomics, 2009, 2(2): 133–142.
- [20] 党江鹏, 刘念, 叶伟, 等. 云斑车蝗线粒体基因组全序列测定与分析[J]. 昆虫学报, 2008, 51(7): 671–680.
- [21] Jacobs HT, Elliott DJ, Math VB, et al. Nucleotide sequence and gene organization of sea urchin mitochondrial DNA[J]. J Mol Biol, 1998, 202: 185–217.
- [22] Boyce T M, Zwick M E, Aquadro C F. Mitochondrial DNA in the bark weevils: sizes, structure and heteroplasmy[J]. Genetics, 1989, 123: 825–836.
- [23] Wolstenholme D R. Animal mitochondrial DNA: structure and evolution[J]. Intl Rev Cytology, 1992, 141: 73–216.
- [24] Lunt D H, Whipple L E, Hyman B C. mitochondrial DNA variable number tandem repeats (VNTRs): utility and problems in molecular ecology[J]. Mol Ecol, 1998, 7: 1441–1455.
- [25] Wilkinson G S, Chapman A M. Length and sequence variation in evening bat D-loop mtDNA[J]. Genetics, 1991, 128: 607–617.
- [26] Arnason E, Rand D M. Heteroplasmy of short tandem repeats in mitochondrial DNA of Atlantic cod, *Gadus morhua*[J]. Genetics, 1992, 132: 211–220.
- [27] Albertson R C, Markert J A, Danley P D, et al. Phylogeny of a rapidly evolving clade: the cichlid fishes of Lake Malawi, East Africa [J]. Proc Natl Acad Sci USA, 1999, 96: 5107–5110.

Complete mitochondrial genome sequence and phylogenetic analysis of *Oreochromis aureus*

JIANG Zongliang^{1,2}, ZHANG Ming², LIN Yong¹, ZHU Jiajie¹, LI Mingxing², LUO Yongju¹, GAN Xi¹

1. Guangxi Institute of Fisheries, Nanning 530021, China;

2. Animal Reproduction Institute, Guangxi University, Nanning 530004, China

Abstract: The complete mitochondrial genome of *Oreochromis aureus* was obtained using long amplification polymerase chain reaction (LA-PCR) and sequenced. The total length of the sequence was 16 632 bp, and includes 13 protein-coding genes, two ribosomal RNA genes, 22 transfer RNA genes, and a 931 bp major non-coding region. The nucleotide composition was 27.89% (A), 25.50% (T), 15.62% (C), and 30.99% (G). The arrangement of genes in *Oreochromis aureus* was identical to that of other Tilapiini. The COX1 gene uses TGT as an initiation codon, with the remaining 12 using ATG. Nine genes have complete stop codons (TAA), and three (COX2, CYTB, and ND4) have an incomplete T stop codon. The 22 tRNAs all have typical cloverleaf structures. Phylogenetic analysis indicated that *Oreochromis* is divided from other species of cichlids and forms a monophyletic group, among which *Oreochromis aureus* is closely related to other *Oreochromis* species, and *Oreochromis mossambicus* is closely related to *Oreochromis* sp. KM-2006.

Key words: *Oreochromis aureus*; mitochondrial genome; sequence analysis; phylogenetic tree

Corresponding author: GAN Xi. E-mail: ganxicn@126.com