

鲂属鱼类线粒体基因组的比较及其系统发育分析

赖瑞芳¹, 张秀杰¹, 李艳和¹, 吴俊颖¹, 杨东辉², 王卫民^{1*}

(1. 华中农业大学水产学院, 农业动物遗传育种与繁育教育部重点实验室,
农业部淡水生物繁育重点实验室, 湖北 武汉 430070;
2. 信阳农林学院水产科学系, 河南 信阳 46400)

摘要: 基于 GenBank 中团头鲂线粒体基因组全序列和三角鲂、厚颌鲂、广东鲂的部分线粒体基因组序列, 设计引物扩增出三角鲂、厚颌鲂和广东鲂 3 种鱼线粒体基因组全序列, 同时对 4 种鲂属鱼类线粒体基因组全序列进行了比较分析。结果表明, 4 种鲂属鱼类线粒体基因组基因排列顺序完全相同, 排列紧密, 均包含 13 个蛋白质编码基因、22 个 tRNA、2 个 rRNA、1 个非编码控制区 (D-loop 区) 和 1 个轻链复制起始区 (OL 区)。除 ND6 和 8 个 tRNA 在 L 链上编码外, 其余的基因均在 H 链上编码。4 种鲂属线粒体基因组 13 个蛋白质编码基因中, 均呈现出较强的 A+T 偏向性和 C 碱基偏好。全序列比对结果显示, 共有 758 个变异位点, 其中非简约性信息位点有 691 个, 占总变异位点的 91.16%, 简约性信息位点有 67 个, 仅占总变异位点的 8.84%。22 个 tRNA 基因中只有 11 个存在种间变异, 共 23 个变异位点, 主要发生在 tRNA 三叶草结构的 TΨC 和 DHU 臂环上。13 个蛋白质编码基因中共检测出 626 个变异位点, 这些变异主要发生在密码子第三位, 占总变异位点的 82.59%, 其中变异位点数最多的是 Cyt b 基因, 达 84 个, 其次是 ND4 基因 (83 个)。因此, Cyt b 和 ND4 基因可作为备选的分子标记, 用于鲂属群体间的遗传学研究。基于 4 种鲂属鱼类线粒体基因组全序列构建的 ML 树和 BI 树均显示, 三角鲂与厚颌鲂的亲缘关系最近, 团头鲂与它们的亲缘关系相对较近, 而广东鲂与前述 3 种鲂属鱼类的亲缘关系均较远。

关键词: 鲂属; 线粒体基因组; 变异位点; 分子标记; 系统发育

中图分类号: Q 343.1; S 965

文献标志码: A

在脊椎动物中, 线粒体基因组为双链闭环状分子, 长度为 15~18 kb。与其他脊椎动物一样, 鱼类线粒体基因组核酸序列和组成比较保守, 且排列紧密。根据碱性氯化铯密度梯度离心 G+T 含量的高低将双链分为重链 (H 链) 和轻链 (L 链)^[1], 它们绝大多数由 13 个蛋白质编码基因 (ATP6、ATP8、COX1~3、Cyt b、ND1~6 和 ND4L)、2 个核糖体 RNA 基因 (rRNAs: 12S rRNA 和 16S rRNA)、22 个转运 RNA 基因 (tRNAs)、1 个非编码区 (D-loop 区) 和轻链复制起始区组成。其中, 除 ND6 基因和 8 个 tRNA 基因 (tRNA^{Gln}、

tRNA^{Ala}、tRNA^{Asn}、tRNA^{Cys}、tRNA^{Tyr}、tRNA^{Ser}、tRNA^{Glu}、tRNA^{Pro}) 在 L 链上编码外, 其余的基因均在 H 链上编码^[2-3]。线粒体基因组具有信息量丰富、分子量小、基因结构及排列简单、多个拷贝、物种内几乎不发生重排、无内含子、母系遗传、进化速率快等优点, 是一个相对独立的复制单位, 是生物学家研究系统发育进化的有力工具, 也是惟一可以提供在基因组水平上进行研究的分子标记。在近十年间, 线粒体基因组已经被广泛地应用于脊椎动物群体的系统发生学、比较和进化基因组学、群体遗传学、分子进化、保护生物学和系

收稿日期: 2013-08-05 修回日期: 2013-11-21

资助项目: 国家自然科学基金项目 (31201988); 现代农业产业技术体系建设专项 (CARS-46-05); 中央高校基本科研业务费专项 (2011PY023, 2011SC27, 2011ZC011, 2012YB08, 2011QC030)

通信作者: 王卫民, E-mail: wangwm@mail.hzau.edu.cn

统发育树重建等方面的研究^[4-5]。

鲂属 (*Megalobrama*) 隶属鲤形目 (Cypriniformes)、鲤科 (Cyprinidae)、鲃亚科 (Culterinae), 主要分布于我国以及越南和俄罗斯的一些地区, 是我国重要的养殖鱼类之一。目前分类研究认为鲂属鱼类包括 4 种: 团头鲂 (*Megalobrama amblycephala*)、三角鲂 (*M. terminalis*)、厚颌鲂 (*M. pellegrini*) 和广东鲂 (*M. hoffmanni*)^[6-7]。

目前, 关于鲂属鱼类线粒体基因组的研究仍停留在 D-loop、Cyt b、ND2、COX1 和 COX2 等不同区域水平, 其主要原因是得到 mtDNA 全序列比较耗时耗力, 而且还受到各种技术的限制^[8-9]。然而, 现在研究普遍认为较长的线粒体基因组序列能够得到较为准确的进化关系^[10-11], 而鲂属鱼类中只有团头鲂的线粒体基因组全序列已经被克隆出来。针对鲂属鱼类线粒体基因组全序列数据信息还不够完善的现状, 本研究克隆了三角鲂、厚颌鲂和广东鲂 3 种鱼的线粒体基因组全序列, 并结合从 GenBank 数据库中获得的团头鲂线粒体基因组全序列, 对 4 种鲂属鱼类线粒体基因组进行比较及其系统发育分析, 为鲂属鱼类群体的生物多样性的保护、遗传学研究及分子标记遗传育种等提供遗传信息, 也为揭示鱼类线粒体基因组进化规律提供参考。

1 材料与方法

1.1 数据来源

实验所用的三角鲂和广东鲂采集于珠江水系, 厚颌鲂采集于长江上游的龙溪河。采集后的鳍条组织样品用无水乙醇固定, 置于 $-20\text{ }^{\circ}\text{C}$ 的冰箱中存储备用。总基因组 DNA 参照 Li 等^[12] 的方法用醋酸铵沉淀法从组织样品中提取。基于 GenBank 中团头鲂线粒体基因组全序列和三角鲂、厚颌鲂、广东鲂的部分线粒体基因组序列, 用 Primer premier 5.0 软件设计扩增了三角鲂、厚颌鲂和广东鲂 3 种鱼线粒体基因组全序列的 16 对引物 (表 1), 其中, p-mt16F 和 p-mt16R 是参照黄志坚等^[13] 的方法所设计的简并引物。PCR 扩增产物用琼脂糖凝胶电泳检测, 纯化后 T-A 克隆到大肠杆菌中, 然后将菌液直接送往上海英骏公司进行测序。测得的序列在 NCBI 中用 BLAST 进行同源序列搜索, 然后用 ClustalX 1.83 软件^[14] 和 BioEdit 7.0 软件^[15] 对测出序列进行比对、校正, 人工拼接得到完整的线粒体基因组全序列。3 种鲂属鱼类线粒体基因组全序列已经提交 GenBank, 序列登录号分别为 JX242528、JX242529 和 JX242530。团头鲂线粒体基因组全序列从 GenBank 数据库中下载 (登录号 NC_010341.1)。

表 1 鲂属鱼类线粒体基因组引物设计序列

Tab. 1 Primers used for amplification of *Megalobrama* mitochondrial genome

| 上游 forward | 引物序列 (5'-3') primer sequence (5' to 3') | 下游 reverse | 引物序列 (5'-3') primer sequence (5' to 3') |
|---------------|--|---------------|--|
| p-mt1F | CATAGCACTGAAGATGTTAAGATGG | p-mt1R | ATGGCTAGAAGTGGTGAGGTT |
| p-mt2F | GTCATACGCTTCTAGGTGTCC | p-mt2R | GCTTGATGGTGCTGGCTAA |
| p-mt3F | GGTGTGGCTGAGTTAGTCAAG | p-mt3R | ATGTTCTGTGGCTTAGAGATGTC |
| p-mt4F | CGGAGCAATCCAGGTCAGT | p-mt4R | CATATTCGGCTAGGAAGAAGAGG |
| p-mt5F | GCCTTATCAAGCCTCGCAGTA | p-mt5R | GGTGGTGGTGTGTGCTATTAG |
| p-mt6F | CCACTGACTACTAGCCTGAATAG | p-mt6R | GCTCTTGGTCTGATGTTATCCTAA |
| p-mt7F | ACTTAGGATAACATCAGACCAAGAG | p-mt7R | CTCCGTGAAGTGTGGCTAATC |
| p-mt8F | CGGAATTAATGCTTCTTACAGAC | p-mt8R | GTGGTTATGTGACTGGCTTGAA |
| p-mt9F | CACCTCTGAGCCGTTCCAT | p-mt9R | AATTATTGCCACTGCGACTTCT |
| p-mt10F | TCCTRGCWGMTGYCTTCTTC | p-mt10R | GCAGTTGTGGTTGTTCACAGC |
| p-mt11F | ACCACAACCGCACATAGTCT | p-mt11R | GTAGCATCAGCCTCATATAAGTTCT |
| p-mt12F | GAACTTATATGAGGCTGATGCTACT | p-mt12R | CTGTTGCTGCGAGGATTAGTC |
| p-mt13F | AAGCATAGCATGGTTCGCAAT | p-mt13R | AGAGTTAGGAGGAAGATAGTTAGGT |
| p-mt14F | TAACAGCCACACCACTAACATC | p-mt14R | AGAAGCCACCTCAGATTCATTG |
| p-mt15F | CTTGAAGAACCACGTTGTAGT | p-mt15R | GAGGATGAGGAATAATGCGAAGTA |
| p-mt16F | CACCCTGRCTMCCAAGCCA | p-mt16R | RBTGCGGAGACTTGTCATGTGTAA |

1.2 碱基组成特征及密码子使用情况

用 MEGA 5.10 软件^[16]对 4 种鲂属鱼类线粒体基因组序列碱基组成特征及其密码子使用情况进行分析,并用 Perna 等^[17]的方法进行 AT 和 GC 偏移度计算[$AT-skew = (A - T)/(A + T)$ 和 $GC-skew = (G - C)/(G + C)$]。相对同义密码子的使用频率(relative synonymous codon usage, RSCU)是衡量密码子使用偏好性的重要指标,它能直观地反映密码子使用的偏好程度^[18]。它是用观察到的某一同义密码子的使用次数除以预期该密码子的出现次数,若 RSCU 的值小于 1 说明该密码子出现的次数比预期低,若 RSCU 的值大于 1 说明该密码子出现的次数比预期多^[19]。

1.3 变异位点分析

通过 ClustalX 1.83 软件^[14]对各基因进行多重序列比对。然后,通过 DnaSP 5.0 软件^[20]对 4 种鲂属鱼类线粒体基因组的主编码基因进行基因特征及变异位点分析。

1.4 tRNA 基因二级结构预测

用在线 tRNAscan-SE 1.21 软件和 RNAstructure 5.3 软件对 4 种鲂属鱼类发生变异的 tRNA 进行二级结构的预测,从而对它们的变异位点进行详细的定位分析。

1.5 系统发育树构建

为探讨 4 种鲂属鱼类的亲缘关系,以翘嘴鲌(*Culter alburnus*) (登录号 NC_013616.1) 和蒙古鲌(*Culter mongolicus*) (登录号 NC_008683.1) 为外群,基于线粒体基因组全序列,使用 PhyML 3.0 软件^[21]的最大似然法(maximum-likelihood, ML)和 MrBayes 3.12 软件^[22]的贝叶斯推论法(Bayesian inference, BI)对 4 种鲂属鱼类进行系统发育树的构建。

2 结果

2.1 鲂属鱼类线粒体基因组的基本特征

4 种鲂属鱼类线粒体基因组全序列长度为 16 621 ~ 16 623 bp (团头鲂 16 623 bp、三角鲂 16 621 bp、厚颌鲂 16 621 bp、广东鲂 16 622 bp), 均编码 37 个基因(13 个蛋白质编码基因、22 个转运 RNA 基因和 2 个核糖体 RNA 基因)以及 1 个非编码区(D-loop 区)和轻链复制起始区(OL 区)。除 ND6 和 8 个 tRNA (tRNA^{Gln}、tRNA^{Ala}、tRNA^{Asn}、tRNA^{Cys}、tRNA^{Tyr}、tRNA^{Ser}、tRNA^{Glu}、

tRNA^{Pro}) 在 L 链上编码外,其余的基因均在 H 链上编码。它们的基因排列顺序、长度、编码链、起始和终止密码子使用等方面非常相似(表 2)。13 个蛋白质编码基因和 22 个 tRNA 基因的排列顺序和长度完全相同。整个基因组排列紧密,仅仅只有少许的碱基间隔区(13 处 33 bp,间隔碱基数为 1 ~ 13 bp)和重叠区(6 处 22 bp,重叠碱基数为 1 ~ 7 bp)。在蛋白质编码基因序列中,有 4 对基因间存在开放阅读框的重叠,其中 ATP8/ATP6 基因和 ND4L/ND4 基因间都存在 7 bp 的重叠,ND5 和 ND6 基因间存在 4 bp 的重叠,ATP6/COX3 基因间存在 1 bp 的重叠。同时,相邻 tRNA 基因之间也存在有个别碱基的重叠现象,均有 2 处 3 bp 的重叠(tRNA^{Ile}/tRNA^{Gln} 2bp、tRNA^{Thr}/tRNA^{Pro} 1 bp)。另外,鲂属鱼类线粒体基因组的 13 处间隔区中,最大间隔区位于 tRNA^{Asp} 和 COX2 之间,为 13 bp。基因间既没有重复又没有间隔的紧密排列基因对共计 21 处。

在蛋白质编码基因的起始和终止密码子方面,广东鲂的 ATP8 基因的终止密码子是 TAG,而其他 3 种鲂属鱼类均以 TAA 作为终止密码子,除此之外的 12 个蛋白质编码基因的起始密码子(ATG 或 GTG)和终止密码子(TAA、TA - 或 T -)在 4 种鲂属鱼类中完全相同(表 2)。

2.2 鲂属鱼类线粒体基因组的碱基组成与偏好

对 4 种鲂属鱼类线粒体基因组及各组成部分碱基的平均组成及偏好情况进行了统计,结果显示,4 种鲂属鱼类线粒体基因组全序列 A + T 平均含量为 56.0%,表明鲂属线粒体基因组全序列存在一定 A + T 偏向性(表 3)。鲂属线粒体基因组全序列还呈现出较强的 A 碱基和 C 碱基偏好(AT-skew = 0.12 和 GC-skew = -0.27),而且,4 种鲂属鱼类线粒体基因组的碱基组成还具有一定的规律性,除 OL 区序列和蛋白质编码基因第三密码子外,鲂属鱼类线粒体基因组的全序列、蛋白质编码基因、tRNA、rRNA 和 D-loop 区序列都存在一定的 A + T 偏向性、A 碱基偏好和 C 碱基偏好,其中 D-loop 区的 A + T 偏向性最为显著为 64.2%,且明显高于线粒体基因组其他各组成部分的基因,因此 D-loop 区又称为 A + T 丰富区。

2.3 鲂属鱼类蛋白质编码基因密码子使用情况

对 4 种鲂属鱼类的 13 个蛋白质编码基因的密码子平均使用频率和相对同义密码子平均使用

频率进行了统计,结果见表4。表4中粗体字显示的是编码同种氨基酸使用频率最高的密码子,这些密码子的RSCU值均大于1,均为偏好密码子。结果显示,所有蛋白质编码基因的密码子使用都存在着强烈的偏好性,其中NNA密码子

RSCU基本上都大于1,表明第三位点为A的密码子使用频率较高,密码子使用的这种偏好性与蛋白质编码基因的密码子第三位点的A偏好性一致。

表2 4种鲂属鱼类线粒体基因组基因长度、起始和终止密码子使用情况
Tab.2 Gene length, start codon and stop codon usage of mitochondrial genomes from four species of the genus *Megalobrama*

| 基因 gene | 团头鲂 <i>M. amblycephala</i> | | | 厚颌鲂 <i>M. pellegrini</i> | | | 三角鲂 <i>M. terminalis</i> | | | 广东鲂 <i>M. skolkovii</i> | | |
|---------------------------|----------------------------|-----------------------------|----------------------------|--------------------------|-----------------------------|----------------------------|--------------------------|-----------------------------|----------------------------|-------------------------|-----------------------------|----------------------------|
| | 长度/bp length | 起始 密码子 start codon | 终止 密码子 stop codon | 长度/bp length | 起始 密码子 start codon | 终止 密码子 stop codon | 长度/bp length | 起始 密码子 start codon | 终止 密码子 stop codon | 长度/bp length | 起始 密码子 start codon | 终止 密码子 stop codon |
| tRNA^{Phe} | 69 | | | 69 | | | 69 | | | 69 | | |
| 12S rRNA | 962 | | | 961 | | | 961 | | | 965 | | |
| tRNA^{Val} | 72 | | | 72 | | | 72 | | | 72 | | |
| 16S rRNA | 1 692 | | | 1 691 | | | 1 691 | | | 1 689 | | |
| tRNA^{Leu} | 76 | | | 76 | | | 76 | | | 76 | | |
| ND1 | 975 | ATG | TAA | 975 | ATG | TAA | 975 | ATG | TAA | 975 | ATG | TAA |
| tRNA^{Ile} | 72 | | | 72 | | | 72 | | | 72 | | |
| tRNA ^{Gln} | 71 | | | 71 | | | 71 | | | 71 | | |
| tRNA^{Met} | 69 | | | 69 | | | 69 | | | 69 | | |
| ND2 | 1 045 | ATG | T | 1 045 | ATG | T | 1 045 | ATG | T | 1 045 | ATG | T |
| tRNA^{Trp} | 71 | | | 71 | | | 71 | | | 71 | | |
| tRNA ^{Ala} | 69 | | | 69 | | | 69 | | | 69 | | |
| tRNA ^{Asn} | 73 | | | 73 | | | 73 | | | 73 | | |
| O _L | 32 | | | 32 | | | 32 | | | 32 | | |
| tRNA ^{Cys} | 68 | | | 68 | | | 68 | | | 68 | | |
| tRNA ^{Tyr} | 71 | | | 71 | | | 71 | | | 71 | | |
| COX1 | 1 551 | GTG | TAA | 1 551 | GTG | TAA | 1 551 | GTG | TAA | 1 551 | GTG | TAA |
| tRNA ^{Ser} | 71 | | | 71 | | | 71 | | | 71 | | |
| tRNA^{Asp} | 74 | | | 74 | | | 74 | | | 74 | | |
| COX2 | 691 | ATG | T | 691 | ATG | T | 691 | ATG | T | 691 | ATG | T |
| tRNA^{Lys} | 76 | | | 76 | | | 76 | | | 76 | | |
| ATP8 | 165 | ATG | TAA | 165 | ATG | TAA | 165 | ATG | TAA | 165 | ATG | TAG |
| ATP6 | 684 | ATG | TAA | 684 | ATG | TAA | 684 | ATG | TAA | 684 | ATG | TAA |
| COX3 | 785 | ATG | TA | 785 | ATG | TA | 785 | ATG | TA | 785 | ATG | TA |
| tRNA^{Gly} | 72 | | | 72 | | | 72 | | | 72 | | |
| ND3 | 349 | ATG | T | 349 | ATG | T | 349 | ATG | T | 349 | ATG | T |
| tRNA^{Arg} | 70 | | | 70 | | | 70 | | | 70 | | |
| ND4L | 297 | ATG | TAA | 297 | ATG | TAA | 297 | ATG | TAA | 297 | ATG | TAA |
| ND4 | 1 382 | ATG | TA | 1 382 | ATG | TA | 1 382 | ATG | TA | 1 382 | ATG | TA |
| tRNA^{His} | 69 | | | 69 | | | 69 | | | 69 | | |
| tRNA^{Ser} | 69 | | | 69 | | | 69 | | | 69 | | |
| tRNA^{Leu} | 73 | | | 73 | | | 73 | | | 73 | | |
| ND5 | 1 836 | ATG | TAA | 1 836 | ATG | TAA | 1 836 | ATG | TAA | 1 836 | ATG | TAA |
| ND6 | 522 | ATG | TAA | 522 | ATG | TAA | 522 | ATG | TAA | 522 | ATG | TAA |
| tRNA ^{Glu} | 69 | | | 69 | | | 69 | | | 69 | | |
| Cytb | 1 141 | ATG | T | 1 141 | ATG | T | 1 141 | ATG | T | 1 141 | ATG | T |
| tRNA^{Thr} | 72 | | | 72 | | | 72 | | | 72 | | |
| tRNA ^{Pro} | 70 | | | 70 | | | 70 | | | 70 | | |
| D-loop | 937 | | | 937 | | | 937 | | | 936 | | |

注:在H链上编码的基因用粗体显示

Notes: The genes encoded in H chain in bold

表 3 鲂属鱼类线粒体基因组各组成部分的碱基平均组成统计

Tab. 3 Average base compositions of mitochondrial genomes of the genus *Megalobrama*

| | 碱基组成/% nucleotide composition | | | | | | AT 偏好 | GC 偏好 |
|----------------------|-------------------------------|------|------|------|-------|-------|---------|---------|
| | T(U) | C | A | G | A + T | G + C | AT-skew | GC-skew |
| complete genome | 24.7 | 27.9 | 31.2 | 16.2 | 56.0 | 44.0 | 0.12 | -0.27 |
| protein-coding genes | 26.7 | 28.6 | 29.1 | 15.6 | 55.8 | 44.2 | 0.04 | -0.30 |
| 1st | 25.1 | 27.8 | 29.0 | 18.2 | 54.0 | 46.0 | 0.07 | 0.21 |
| 2nd | 30.5 | 28.4 | 25.4 | 15.7 | 55.9 | 44.1 | 0.09 | -0.34 |
| 3rd | 24.6 | 29.6 | 32.9 | 12.9 | 57.5 | 42.5 | 0.14 | -0.39 |
| tRNA | 25.1 | 24.6 | 30.3 | 20.0 | 55.4 | 44.6 | 0.09 | -0.10 |
| rRNA | 19.8 | 24.5 | 34.5 | 21.3 | 54.2 | 45.8 | 0.27 | -0.07 |
| D-loop | 30.9 | 21.5 | 33.3 | 14.3 | 64.2 | 35.8 | 0.04 | -0.20 |
| OL | 25.0 | 28.1 | 19.5 | 27.3 | 44.5 | 55.5 | -0.12 | -0.01 |

表 4 4 种鲂属鱼类 13 个蛋白质编码基因密码子平均使用频率

Tab. 4 Total codon average usage in the thirteen protein-coding genes from four species of the genus *Megalobrama*

| 密码子 codon | 使用次数 count | 相对同义 密码子 使用频率 RSCU | 密码子 codon | 使用次数 count | 相对同义 密码子 使用频率 RSCU | 密码子 codon | 使用次数 count | 相对同义 密码子 使用频率 RSCU | 密码子 codon | 使用次数 count | 相对同义 密码子 使用频率 RSCU |
|---------------|---------------|-----------------------------|---------------|---------------|-----------------------------|---------------|---------------|-----------------------------|---------------|---------------|-----------------------------|
| UUU(F) | 60.5 | 0.76 | UCU(S) | 65.5 | 1.14 | UAU(Y) | 69.3 | 0.95 | UGU(C) | 23.5 | 0.82 |
| UUC(F) | 99.8 | 1.24 | UCC(S) | 88.8 | 1.55 | UAC(Y) | 77.3 | 1.05 | UGC(C) | 34 | 1.18 |
| UUA(L) | 85 | 1.06 | UCA(S) | 74.5 | 1.3 | UAA(*) | 55.8 | 1.3 | UGA(W) | 78.5 | 1.47 |
| UUG(L) | 38.8 | 0.48 | UCG(S) | 21.8 | 0.38 | UAG(*) | 53.5 | 1.24 | UGG(W) | 28.5 | 0.53 |
| CUU(L) | 98.3 | 1.22 | CCU(P) | 94.3 | 1.13 | CAU(H) | 59 | 0.8 | CGU(R) | 18 | 0.69 |
| CUC(L) | 70 | 0.87 | CCC(P) | 85 | 1.01 | CAC(H) | 89 | 1.2 | CGC(R) | 29.8 | 1.14 |
| CUA(L) | 150.8 | 1.87 | CCA(P) | 132 | 1.58 | CAA(Q) | 77.3 | 1.38 | CGA(R) | 36.5 | 1.4 |
| CUG(L) | 40.5 | 0.5 | CCG(P) | 23.8 | 0.28 | CAG(Q) | 34.8 | 0.62 | CGG(R) | 20.3 | 0.78 |
| AUU(I) | 122 | 1.08 | ACU(T) | 80.3 | 1.08 | AAU(N) | 90.5 | 1 | AGU(S) | 31.3 | 0.55 |
| AUC(I) | 103.3 | 0.92 | ACC(T) | 87 | 1.17 | AAC(N) | 91 | 1 | AGC(S) | 62 | 1.08 |
| AUA(M) | 87.8 | 1.28 | ACA(T) | 104.3 | 1.4 | AAA(K) | 82.8 | 1.59 | AGA(S) | 21.5 | 0.5 |
| AUG(M) | 49.8 | 0.72 | ACG(T) | 26.3 | 0.35 | AAG(K) | 21.3 | 0.41 | AGG(S) | 41.3 | 0.96 |
| GUU(V) | 32 | 0.82 | GCU(A) | 35.5 | 0.71 | GAU(D) | 24 | 0.6 | GGU(G) | 32.3 | 0.75 |
| GUC(V) | 35 | 0.89 | GCC(A) | 79 | 1.59 | GAC(D) | 56.3 | 1.4 | GGC(G) | 39.5 | 0.92 |
| GUA(V) | 69.3 | 1.77 | GCA(A) | 73.3 | 1.47 | GAA(E) | 57.8 | 1.37 | GGA(G) | 66 | 1.54 |
| GUG(V) | 20.3 | 0.52 | GCG(A) | 11.5 | 0.23 | GAG(E) | 26.5 | 0.63 | GGG(G) | 33.3 | 0.78 |

注: * 表示终止密码子;()内的字母表示各氨基酸的简写;偏好密码子用粗体字显示

Notes: * indicates stop codons;The letters in brackets are abbreviations of each amino acid;preferred codons in bold

2.4 鲂属鱼类线粒体基因组序列比对

对 4 种鲂属鱼类线粒体基因组全序列、13 个蛋白质编码基因序列、12S rRNA、16S rRNA 和 D-loop 区的序列分别进行种间两两 BLAST 比对。

线粒体基因组全序列比对结果显示,团头鲂、厚颌鲂和三角鲂的相似度极高(98.91%~99.66%)。其中,厚颌鲂和三角鲂的线粒体全序列相似度高达 99.66%,团头鲂与三角鲂的相似度为 99.06%,团头鲂与厚颌鲂的相似度也达 98.91%。广东鲂线粒体全序列与其他鲂属鱼类的相似度为 95.94%~96.05%。这表明三角鲂与厚颌鲂的亲缘关系最近,团头鲂与它们的亲缘关系相对较近,而

广东鲂与前述 3 种鲂属鱼类的亲缘关系均较远。

线粒体不同结构区的序列比对结果与全序列比对结果相同,团头鲂、厚颌鲂和三角鲂 3 种间各结构区的序列相似度极高。厚颌鲂与三角鲂的序列相似度高达 99.33%~100%,它们的 COX1 和 ATP8 序列完全相同,团头鲂与厚颌鲂和三角鲂的 ATP8 序列完全相同。在比对的 16 个线粒体结构区中,4 种鲂属鱼类 ND2 序列相似度最低,低于种间线粒体基因组全序列的相似度。另外,团头鲂、厚颌鲂和三角鲂 D-loop 区序列的种间相似度(98.30%~99.34%)略低于线粒体基因组全序列的相似度(98.91%~99.66%),但这 3 种鲂与广

东鲂 D-loop 区序列的相似度却略高于线粒体基因组全序列的相似度。

2.5 鲂属鱼类线粒体基因组变异位点分析

4种鲂属鱼类线粒体基因组全序列比对结果显示,共有758个变异位点,其中非简约性信息位点有691个,占总变异位点的91.16%,简约性信息位点有67个,仅占总变异位点的8.84%。其中由广东鲂单独变异引起的单现变异位点数就有571个,占4种鲂属鱼类的单现变异位点数的82.6%,可见广东鲂与其他3种鲂属鱼类(团头鲂、三角鲂和厚颌鲂)的亲缘关系比较远。简约性信息位点数最多的是ND2,其次为ND1、COX1和Cyt b,鲂属鱼类线粒体基因的种间分化主要体

现在这4个基因中(表5)。ND4L、ATP8和COX2及18个tRNA序列(除tRNA^{Phe}、tRNA^{Glu}和tRNA^{Gly}外)中均无简约性信息位点。

4种鲂属鱼类线粒体基因组13个蛋白质编码基因和2个核糖体RNA基因的变异位点的分析结果显示,12S rRNA和16S rRNA两个核糖体RNA基因的保守性最高,变异位点比例分别仅为2.29%和2.25%;其次是ATP8和COX1基因(分别为3.03%和3.35%)。ND2基因变异位点的比例最高达到7.46%;然后是Cyt b基因(7.36%)。变异位点数最多的基因为Cyt b基因,达84个;然后是ND4基因(83个)(表5粗体字表示)。

表5 鲂属鱼类线粒体基因组变异位点的比较

Tab. 5 Comparison of variation loci in mitochondrial genomes of the genus *Megalobrama*

| 基因 gene | 总位点数 total number of sites | 不变位点数 invariable sites | 变异位点数 variable sites | 单现变异位点数 singleton variable sites | 简约信息位点数 parsimony informative sites | 变异位点比例/% the percentage of the variable sites |
|--------------------------|----------------------------------|------------------------------|----------------------------|--|---|---|
| complete genome | 16 617 | 15 859 | 758 | 691 | 67 | 4.56 |
| 12S rRNA | 961 | 939 | 22 | 20 | 2 | 2.29 |
| 16S rRNA | 1689 | 1651 | 38 | 33 | 5 | 2.25 |
| ATP6 | 684 | 637 | 47 | 43 | 4 | 6.87 |
| ATP8 | 165 | 160 | 5 | 5 | 0 | 3.03 |
| Cytb | 1 141 | 1 057 | 84 | 78 | 6 | 7.36 |
| COX1 | 1 551 | 1 499 | 52 | 45 | 7 | 3.35 |
| COX2 | 691 | 663 | 28 | 28 | 0 | 4.05 |
| COX3 | 785 | 748 | 37 | 36 | 1 | 4.71 |
| ND1 | 975 | 907 | 68 | 60 | 8 | 6.97 |
| ND2 | 1 045 | 967 | 78 | 68 | 10 | 7.46 |
| ND3 | 349 | 326 | 23 | 22 | 1 | 6.59 |
| ND4 | 1 382 | 1 299 | 83 | 78 | 5 | 6.01 |
| ND4L | 297 | 280 | 17 | 17 | 0 | 5.72 |
| ND5 | 1 836 | 1 762 | 74 | 70 | 4 | 4.03 |
| ND6 | 522 | 492 | 30 | 25 | 5 | 5.75 |
| D-loop | 935 | 889 | 46 | 40 | 6 | 4.92 |
| OL | 32 | 31 | 1 | 1 | 0 | 3.13 |
| tRNA ^{Cys} | 68 | 67 | 1 | 1 | 0 | 1.47 |
| tRNA ^{Glu} | 69 | 68 | 1 | 0 | 1 | 1.45 |
| tRNA ^{Gly} | 72 | 70 | 2 | 1 | 1 | 2.78 |
| tRNA ^{His} | 69 | 66 | 3 | 3 | 0 | 4.35 |
| tRNA ^{Ile} | 72 | 69 | 3 | 3 | 0 | 4.17 |
| tRNA ^{Met} | 69 | 68 | 1 | 1 | 0 | 1.45 |
| tRNA ^{Phe} | 69 | 66 | 3 | 2 | 1 | 4.35 |
| tRNA ^{Pro} | 70 | 68 | 2 | 2 | 0 | 2.86 |
| tRNA ^{Ser(AGN)} | 69 | 68 | 1 | 1 | 0 | 1.45 |
| tRNA ^{Tyr} | 71 | 68 | 3 | 3 | 0 | 4.23 |
| tRNA ^{Val} | 72 | 69 | 3 | 3 | 0 | 4.17 |

注:总位点数不包含插入和缺失位点

Notes: Total number of sites excluding gaps and missing sites

2.6 不同基因中变异位点的分布

蛋白质编码基因中变异位点的分布 4 种鲂属鱼类线粒体基因组 13 个蛋白质编码基因密码子不同位点的变异分析显示,变异位点主要发生在密码子第三位点,占总变异位点的比例为 60.00%~92.86%;其次发生在密码子第一位点,占总变异位点的比例为 3.85%~40.00%;而密码

子第二位点最为保守,只有 *COX1*、*ND1*、*ND2*、*ND5* 和 *ND6* 基因发生了变异,仅占总变异位点比例的 0~7.69% (表 6)。4 种鲂属鱼类 *COX1*~3 的 3 个亚基在密码子第三位点发生变异比例都很高,最高的是 *COX2* (92.86%),其次是 *COX1* (92.31%),最低的是 *COX3* (91.89%)。

表 6 鲂属鱼类蛋白质编码基因密码子不同位点的变异比较

Tab.6 Comparison of variation of different loci in protein-coding genes of the genus *Megalobrama*

| 基因 gene | 总位 点数 total number of sites | 变异位 点数 variable sites | 密码子第 一位点变 异位点数 codon 1 st variable sites | 密码子第 二位点变 异位点数 codon 2 nd variable sites | 密码子第 三位点变 异位点数 codon 3 rd variable sites | 变异位点 比例/% the percentage of the variable sites | 密码子第一 位点变异 比例/% the percentage of the codon 1 st variable sites | 密码子第二 位点变异 比例/% the percentage of the codon 2 nd variable sites | 密码子第三 位点变异 比例/% the percentage of the codon 3 rd variable sites |
|------------|---|--------------------------------|--|--|--|---|---|---|---|
| ATP6 | 684 | 47 | 4 | 0 | 43 | 6.87 | 8.51 | 0.00 | 91.49 |
| ATP8 | 165 | 5 | 2 | 0 | 3 | 3.03 | 40.00 | 0.00 | 60.00 |
| Cytb | 1 141 | 84 | 14 | 0 | 70 | 7.36 | 16.67 | 0.00 | 83.33 |
| COX1 | 1 551 | 52 | 2 | 2 | 48 | 3.35 | 3.85 | 3.85 | 92.31 |
| COX2 | 691 | 28 | 2 | 0 | 26 | 4.05 | 7.14 | 0.00 | 92.86 |
| COX3 | 785 | 37 | 3 | 0 | 34 | 4.71 | 8.11 | 0.00 | 91.89 |
| ND1 | 975 | 68 | 8 | 2 | 58 | 6.97 | 11.76 | 2.94 | 85.29 |
| ND2 | 1 045 | 78 | 10 | 6 | 62 | 7.46 | 12.82 | 7.69 | 79.49 |
| ND3 | 349 | 23 | 5 | 0 | 18 | 6.59 | 21.74 | 0.00 | 78.26 |
| ND4 | 1 382 | 83 | 13 | 0 | 70 | 6.01 | 15.66 | 0.00 | 84.34 |
| ND4L | 297 | 17 | 3 | 0 | 14 | 5.72 | 17.65 | 0.00 | 82.35 |
| ND5 | 1 836 | 74 | 7 | 4 | 63 | 4.03 | 9.46 | 5.42 | 85.14 |
| ND6 | 522 | 30 | 8 | 1 | 21 | 5.75 | 26.67 | 3.33 | 70.00 |

注:总位点数不包含插入和缺失位点

Notes: Total number of sites excluding gaps and missing sites

4 种鲂属鱼类线粒体基因组的蛋白质编码基因中,由于碱基的转换或颠换,引起氨基酸序列变化的位点数共有 52 个,且 13 个蛋白质编码基因都存在着氨基酸的变异,其中 *ND2* 基因中的氨基酸变异比例最高,达到 4.02%,其次是 *ATP8* 和 *ND3* 基因,而 *COX1*、*COX2*、*COX3* 和 *ND4* 4 个基因的氨基酸变异比例都很低(表 7)。

转运 RNA (tRNA) 中变异位点的分布 4 种鲂属鱼类均编码 22 个 tRNA 基因,其中除了 4 个 tRNA [tRNA^{Leu} (UUN)、tRNA^{Leu} (CUN); tRNA^{Ser} (UCN)、tRNA^{Ser} (AGN)] 含有 2 个外,其余 18 个 tRNA 均只有 1 个,它们的排列顺序和长度完全一致,长度从 68~76 bp 不等(表 2),说明 4 个物种不存在基因重排,序列相似度高。用在线软件 tRNAscan-SE 1.2.1 和 RNA structure 5.2 软件预

测 4 种鲂属鱼类均编码的 22 个 tRNA 基因的二级结构,其二级结构也基本一致。除了 tRNA^{Ser} (AGN) 缺少二氢尿嘧啶臂 (DHU stem), 在相应的位置上只形成一个简单环外,其余的 tRNA 基因都形成典型的三叶草结构。22 个 tRNA 中只有 11 个 tRNA 相对比较保守,未发生变异(表 5),剩余的 11 个 tRNA (tRNA^{Cys}、tRNA^{Glu}、tRNA^{Gly}、tRNA^{His}、tRNA^{Ile}、tRNA^{Met}、tRNA^{Phe}、tRNA^{Pro}、tRNA^{Ser} (AGN)、tRNA^{Tyr} 和 tRNA^{Val}) 均发生了变异(表 8)。从表 8 可以看出,这 11 个 tRNA 总共有 23 个变异位点,主要分布在三叶草结构的 TΨC 和 DHU 臂环上,各分布着 8 个变异位点,反密码子臂环上只分布有 4 个变异位点,剩余 3 个变异位点分别分布在氨基酸接受臂上(2 个)和可变环上(1 个)。

表 7 鲂属鱼类蛋白质编码的氨基酸变异位点比较

Tab.7 Comparison of variation loci of protein-coding amino acids of the genus *Megalobrama*

| 基因 gene | 氨基酸数 | | |
|------------|---------------------------|--|---|
| | the number of amino acids | 氨基酸变异数 the number of variable amino acids | 氨基酸变异比例/% the percentage of variable amino acids |
| ATP6 | 227 | 4 | 1.76 |
| ATP8 | 54 | 2 | 3.70 |
| Cytb | 380 | 4 | 1.05 |
| COX1 | 516 | 2 | 0.39 |
| COX2 | 230 | 1 | 0.43 |
| COX3 | 261 | 1 | 0.38 |
| ND1 | 324 | 4 | 1.23 |
| ND2 | 348 | 14 | 4.02 |
| ND3 | 116 | 4 | 3.45 |
| ND4 | 460 | 2 | 0.43 |
| ND4L | 98 | 1 | 1.02 |
| ND5 | 611 | 9 | 1.47 |
| ND6 | 173 | 4 | 2.31 |
| Total | 3 798 | 52 | |

表 8 鲂属鱼类 11 个 tRNA 变异位点分布情况比较

Tab.8 Comparison of distribution of variation loci in the eleven tRNAs of the genus *Megalobrama*

| 基因 gene | 变异位点 variable sites | 变异位点的位置 the location of the variable site | | | | | | | |
|--------------------------|------------------------|---|----------------------|----------|--------|-----------|--------|----------------|--------|
| | | 接受臂 acceptor stem | 变异环 variable loop | 假尿嘧啶 TΨC | | 二氢尿嘧啶 DHU | | 反密码子 anticodon | |
| | | | | 臂 stem | 环 loop | 臂 stem | 环 loop | 臂 stem | 环 loop |
| tRNA ^{Cys} | 1 | 1 (UC) | | | | | | | |
| tRNA ^{Glu} | 1 | 1 (GA) | | | | | | | |
| tRNA ^{Gly} | 2 | 1 (UC) 1 (UC) | | | | | | | |
| tRNA ^{His} | 3 | 2 (UC) 1 (UC) | | | | | | | |
| tRNA ^{Ile} | 3 | 1 (UC) 1 (UC) 1 (GA) | | | | | | | |
| tRNA ^{Met} | 1 | 1 (GA) | | | | | | | |
| tRNA ^{Phe} | 3 | 2 (GA) 1 (GA) | | | | | | | |
| tRNA ^{Pro} | 2 | 1 (UC) 1 (UG) | | | | | | | |
| tRNA ^{Ser(AGN)} | 1 | 1 (GA) | | | | | | | |
| tRNA ^{Tyr} | 3 | 1 (GA) 2 (UA + UC) | | | | | | | |
| tRNA ^{Val} | 3 | 1 (UC) 1 (GA) 1 (UC) | | | | | | | |

OL 区和 D-loop 区中变异位点的分布 4 种鲂属鱼类的轻链复制起始区(OL 区)位于 WANCY 区域 (tRNA^{Trp}-tRNA^{Ala}-tRNA^{Asn}-tRNA^{Cys}-tRNA^{Tyr}) 的 tRNA^{Asn} 与 tRNA^{Cys} 之间,长度为 32 bp,这段序列有一段回文序列,可折叠成茎-环结构。OL 区的 1 个变异位点发生 loop 环上(G↔A)和茎区相对保守的特征与脊椎动物 OL 区序列 loop 环变异较大、茎的长度和组成均十分保守的特征相一致^[23]。

4 种鲂属鱼类线粒体基因组中,非编码区(D-loop 区)位于 tRNA^{Pro} 和 tRNA^{Phe} 基因之间,与其他硬骨鱼类的一致。4 种鲂属鱼类 D-Loop 区通过与其

他鲤科鱼类相比较,均识别出 1 个终止序列区(TAS)、3 个中央保守区(CSB-F、CSB-E 和 CSB-D)和 3 个保守序列区(CSB1、CSB2 和 CSB3)。TAS 有 4 个变异位点位于其中,是 D-loop 区中序列变异最大的区域;CSB-F 和 CSB-D 较 CSB-E 序列更为保守,没有变异,CSB-E 存在一个碱基 A 的插入;CSB1 和 CSB2 序列完全相同,显示出该序列的保守性,而 CSB3 序列存在一个碱基的转换(A↔G)。

2.7 系统发育树分析

以翘嘴鲌和蒙古鲌 2 种鲌属鱼类作为外群,分别采用最大似然法和贝叶斯法构建了 ML 树和

BI 树。系统发育树各分支的置信度以自举检验值(ML 树)和后验概率(BI 树)来表示。ML 树和 BI 树的拓扑结构是完全相同的,而且它们的置信度都极高(图 1)。系统发育树的结果显示:三角鲂和厚颌鲂首先聚在一起,然后与团头鲂聚在一起,最后与广东鲂聚在一起,形成一支(鲂属),表明三角鲂与厚颌鲂的亲缘关系最近,团头鲂与它们的亲缘关系相对较近,而广东鲂与前述 3 种鲂属鱼类的亲缘关系均较远。

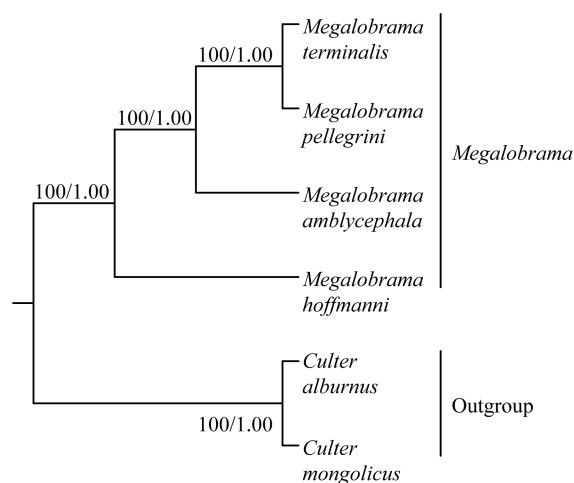


图 1 基于线粒体基因组全序列构建的 ML 树和 BI 树
各分支上的数字分别为自举检验值/后验概率

Fig. 1 The ML and BI trees based on the
complete mitochondrial genomes

Numbers on each node are bootstrap support values/Bayesian posterior probability values, respectively

3 讨论

本研究在 4 种鲂属鱼类线粒体基因组全序列的基础上,全面比较和分析了鲂属鱼类线粒体基因组的基本特征、碱基组成与偏好、蛋白质编码基因、变异位点分布情况及其系统进化关系等。结果发现,4 种鲂属鱼类线粒体基因组在基因组成和结构上高度保守,均编码 37 个基因,且基因排列顺序完全相同,这在其他脊椎动物中也存在这种情况^[24]。4 种鲂属鱼类线粒体基因组全序列长度只有 0~2 bp 的差异,与其他鱼类序列长度接近^[25-27]。基因排列顺序完全相同,没有发现基因的重排,大部分基因长度也基本相同,长度差异主要发生在 12S rRNA 和 16S rRNA 基因。在脊椎动物中,D-loop 区序列可能由于受到较小的选择压力,进化速度相对较快,其序列变异不仅有碱

基的替换,还有序列长度的变化,是线粒体基因组长度变异最大的区域^[16,28],然而 4 种鲂属鱼类间 D-loop 区变化相对较小,只存在 0~1 bp 的差异,整个线粒体基因组排列紧密,在相邻基因间只有少许的碱基重叠区存在,表明 RNA 转录和蛋白质翻译的效率高^[29]。在蛋白质编码基因序列中,ATP8/ATP6 基因和 ND4L/ND4 基因间各存在 7 bp 的重叠,这种特征在其他脊椎动物中很常见,只是重叠的碱基数不同。研究发现,鲤(*Cyprinus carpio*)^[25]和麦穗鱼(*Pseudorasbora parva*)^[30-31]的 ATP8/ATP6 基因以及 ND4L/ND4 基因间都存在 7 bp 的重叠,而爪蛙和鸡^[32]中这两组基因重叠 10 bp、哺乳动物^[33]之间达 40~46 bp。因此,通过这些信息的比较也可在一定程度上描述进化趋势和进行物种的鉴定。然而鲂属鱼类物种形成时间较晚,线粒体 DNA 变异的积累相对较少,进化速度较其他鲤科鱼类要慢^[34],本研究只对线粒体基因组进行了比对和系统进化分析,还不足以进行鲂属物种的鉴定。分子分类学常采用 DNA 分子标记技术进行物种的鉴定^[35],而选择合适、有效的分子标记是物种鉴别关键,本研究为鲂属物种的鉴定提供了良好的物质基础。鲂属鱼类线粒体基因组进化速度较快的 *Cyt b*、*ND2*、*ND4* 等基因区域可作为潜在分子标记来鉴别鲂属鱼类,这将为鲂属鱼类种质资源的保护提供有价值的信息,也将是我们今后研究的发展方向。

与其他脊椎动物一样,4 种鲂属线粒体基因组 13 个蛋白质编码基因中,除 *ND6* 基因外,均呈现出较强的 A+T 偏向性和 C 碱基偏好^[26]。*ND6* 基因由 L 链编码,可见 H 链与 L 链编码的基因存在较大的碱基组成差异。在 4 种鲂属鱼类蛋白质编码基因的起始和终止密码子方面,与大多数硬骨鱼类相同,起始密码子较为恒定。除 *COX1* 以 GTG 作为起始密码子外,其余 12 个蛋白质编码基因均以 ATG 作为起始密码子,这种现象在目前发现的硬骨鱼类中十分常见^[25,36]。相对于起始密码子稳定性,终止密码子在物种内和物种间则有较大变化。鲂属鱼类拥有完全终止密码子(TAA 或 TAG)和不完全终止密码子(T-或 TA-)两种终止密码子。*COX2*、*ND2*、*ND3*、*ND4* 和 *Cyt b* 5 个基因的终止密码子为不完全终止密码子,以 T-或 TA-终止,这些不完全的终止密码子在鱼类线粒体基因组中基本一致^[28],这种现

象也普遍存在于脊椎动物线粒体基因组中^[37]。虽然没有完全终止密码子,但其 DNA 序列转录产物末端为 T- 或 TA-, 加上线粒体 mRNA 3' 末端含有 PloyA 尾巴, 因此, 这些不完全的终止密码子可以通过转录后的 mRNA 加工过程中的多腺苷酸化形成 TAA 终止密码子^[29]。

4 种鲂属鱼类线粒体基因组 13 个蛋白编码基因的碱基平均含量为 A > C > T > G, 其 A + T 平均含量为 55.8%, 存在一定的 A + T 偏向性, 这与其他脊椎动物线粒体基因组 A + T 偏向性是一致的^[38-39]。它们还呈现出较强的 C 碱基偏好, 其中第三密码子的 A + T 平均含量高于第二密码子和第一密码子 A + T 平均含量, 而在 4 种碱基中 G 碱基含量最低, 特别是第三密码子, 仅含 12.9%, 因此在第三密码子呈现极强的 C 碱基偏好 (GC-skew = -0.39), 然而在第一密码子呈现较强的 G 碱基偏好 (GC-skew = 0.21)。在已测定的其它鲤科鱼类线粒体基因组蛋白质编码基因中也有相同情况, 这应该与密码子的稳定性有关系, 如麦穗鱼线粒体基因组蛋白质编码基因 A + T 含量为 59.5%, 存在一定的 A + T 偏向性, 也呈现出较强的 C 碱基偏好 (GC-skew = -0.213 8), 其中第三密码子的 A + T 含量最高 (69.1%), 而在 4 种碱基中 G 碱基含量最低, 特别是第三密码子, 仅含 8.1%, 呈现极强的 C 碱基偏好, 达到 -0.472 3, 然而在第一密码子呈现较强的 G 碱基偏好 (GC-skew = 0.048 0)^[30-31]。正是由于 13 个蛋白质编码基因具有较强的 C 碱基偏好导致了线粒体基因组全序列也存在着较强的 C 碱基偏好。

4 种鲂属鱼类线粒体基因组全序列共有 758 个变异位点, 而简约性信息位点只有 67 个, 仅占 8.84%。以往文献认为变异速度最快的 D-loop 区也只有 6 个简约性信息位点^[40]。整个线粒体基因组简约性信息位点的缺乏也预示线粒体基因组或基因在鲂属鱼类系统发育及种群遗传学研究中所能提供的信息将是十分有限。在蛋白质编码基因中, 变异位点主要发生在密码子第三位点 (60.00% ~ 92.86%), 鲂属鱼类密码子第三位点进化最快, 而密码子第二位点最为保守的这种特征在其他鱼类也是普遍存在的, 因而被广泛地用于系统进化研究^[41]。选择合适的分子标记在种群遗传学的研究中至关重要^[42], 基于鲂属鱼类线

粒体基因组主编码基因的变异位点分析, Cyt b 基因变异位点数最多 (达 84 个), 其次是 ND4 基因 (83 个)。因此, 在鲂属群体之间的遗传学研究中, Cyt b 和 ND4 基因可作为备选的分子标记, 从而为合理、高效利用其生物资源及其生物多样性的保护等方面提供参考。

4 种鲂属鱼类 22 个 tRNA 基因中只有 11 个存在种间变异, 共 23 个变异位点, 主要发生在 tRNA 三叶草结构的 TΨC 和 DHU 臂环上 (各 8 个), 这些变异主要是由于碱基 U ↔ C 和 A ↔ G 发生了转换, 只有 2 个变异分别是由于碱基 U ↔ G 和 U ↔ A 发生颠换而引起的, 这种碱基转换远远大于颠换的现象在鱼类线粒体基因组进化系统中很普遍^[2,38]。鱼类中通常被认为保守性较高的 tRNA^{His} 和 tRNA^{Met} 在鲂属鱼类间均存在变异, 变异位点分别为 3 个和 1 个, 而且 tRNA^{His} 的 3 个变异位点中有 1 个位于保守性较高的反密码子环, 而鱼类线粒体 22 个 tRNA 中变异最大的 tRNA^{Ser}^[43] 在鲂属鱼类中并没有表现出较大的变异, 体现出鲂属鱼类在线粒体 tRNA 变异方面的独特性。

4 种鲂属鱼类的 OL 区长 32 bp, 位于 WANCY 区域, 并含有一段回文序列, 可折叠成茎-环结构。在其他脊椎动物中也有类似的结构, 如鱼类^[44]、哺乳类^[45]、两栖类^[46]等, 但在鳄类^[47]、鸟类^[48]等尚未发现。蒲友光等^[49]认为该结构可能具有种属特异性, 因此推测其可能是脊椎动物线粒体基因组的一个进化特征。

4 种鲂属鱼类的 D-loop 区均识别出 1 个 TAS 序列 (TACATAT—ATGTATTATCACCAT—ATATTAACCAT), 符合刘焕章^[50]提出的鱼类 TAS 序列通式, 该序列共有 4 个变异位点, 是 D-loop 区中序列变异最大的区域, 与其他鱼类相符。中央保守区在 D-loop 区中是最保守的区域, 几乎在所有鱼类中都非常保守^[51]。4 种鲂属鱼类 CSB-F 和 CSB-D 序列完全一致, 分别为 ATGTAGTAAGAGACCACC 和 TATTACTTGC—ATCTGGCTT-A, 符合刘焕章^[50]以鳊鲂 (*Rhodeus ocellatus*) 为例描述的鱼类 CSB-F 序列。CSB-E 序列也仅存在 1 个碱基的变异, 团头鲂的 CSB-E 序列为 AGGG—GTGGGG, 而其他 3 种鲂属鱼类的 CSB-E 序列则变异为 AGGG—GTGAGGG (即 1 个 A 碱基插入)。4 种鲂属鱼类

的 CSB1 序列均为 ATTATTAAGACATA,与 鳊 鳊 CSB1 序 列 (TT-ATTATTGAA-GACATA)^[49]相比存在一定程度的变异,但与 鲂 属 鱼 类 的 翘 嘴 鲂 和 蒙 古 鲂 的 CBS1 序 列 完 全 相 同^[52]。同时,在 CBS1 序列后也存在一个含有多个 C 碱基的 TATGCCCCC 序列。4 种 鲂 属 鱼 类 的 CSB2 也 同 为 CAAACCCCTACCC,且 该 序 列 与 鳊 鳊、翘 嘴 鲂、蒙 古 鲂、银 鲌 (*Xenocypris argentea*) 和 黄 尾 鲌 (*Xenocypris davidi*) 等 鲤 科 鱼 类 的 CSB2 序 列 完 全 相 同,显 示 出 该 序 列 在 鲤 科 鱼 类 中 的 高 度 保 守 性^[51]。

基于线粒体基因组全序列构建的 ML 树和 BI 树的拓扑结构完全相同,置信度也极高,结果表明三角鲂与厚颌鲂的亲缘关系最近,团头鲂与它们的亲缘关系相对较近,而广东鲂与前述 3 种 鲂 属 鱼 类 的 亲 缘 关 系 均 较 远。李 思 发 等^[53]采 用 形 态 判 别、同 工 酶 分 析 和 RAPD 分 析 相 结 合 的 方 法,从 3 个 层 次 阐 明 了 团 头 鲂 与 三 角 鲂 亲 缘 关 系 较 近,它 们 与 广 东 鲂 亲 缘 关 系 较 远。谢 楠 等^[6]基 于 鲂 属 鱼 类 细 胞 色 素 b 片 段 序 列 构 建 的 系 统 发 育 树 显 示,三 角 鲂 和 厚 颌 鲂 首 先 聚 为 一 支,然 后 与 团 头 鲂 相 聚,最 后 与 广 东 鲂 聚 在 一 起。这 些 分 析 结 果 的 一 致 性 显 示 4 种 鲂 属 鱼 类 的 系 统 发 育 关 系 非 常 明 确,也 表 明 了 线 粒 体 基 因 组 序 列 适 用 于 鲂 属 鱼 类 的 系 统 发 育 分 析。

参考文献:

- [1] Yang M. Sequencing and analysis of the complete mitochondrial genome and phylogeny of Cyprinoidei of *Paracobiitis variegates* [D] Xi' an: Journal of Shaanxi Normal University, 2010. [杨敏. 红尾副鳊线粒体基因组序列测定和鲤亚目系统发育分析. 西安:陕西师范大学, 2010.]
- [2] Chen S H, Ou Y J, Li J E. Mitochondrial DNA and its progresses in fish [J]. Biotechnology Bulletin, 2011(3):13 - 20. [陈四海, 区又君, 李加儿. 鱼类线粒体 DNA 及其研究进展. 生物技术通报, 2011(3):13 - 20.]
- [3] Guo X H, Liu S J, Liu Q, et al. New progresses on mitochondrial DNA in fish [J]. Journal of Genetics and Genomics, 2004, 31(9):983 - 1000. [郭新红, 刘少军, 刘巧, 等. 鱼类线粒体 DNA 研究新进展. 遗传学报, 2004, 31(9):983 - 1000.]
- [4] Tian M, Shen X, Meng X P, et al. Comparison of mitochondrial genomes and genetic variation loci of seven starfish [J]. Journal of Oceanography in Taiwan Strait, 2012, 31(2):189 - 194. [田美, 申欣, 孟学平, 等. 7 个海星动物线粒体基因组比较及基因变异位点分析. 台湾海峡, 2012, 31(2):189 - 194.]
- [5] Tian M, Shen X, Meng X P, et al. Analysis of mitochondrial genome characteristics and genetic different loci in brachyuran [J]. Fisheries Science, 2011, 30(1):31 - 37. [田美, 申欣, 孟学平, 等. 短尾派线粒体基因组特征及基因差异位点分析. 水产科学, 2011, 30(1):31 - 37.]
- [6] Xie N, Liu X Y, Feng X Y, et al. Sequences analysis on mitochondrial cytochrome b gene fragment of *Megalobrama* spp. [J]. Modern Agricultural Science and Technology, 2012(1):290 - 292. [谢楠, 刘新轶, 冯晓宇, 等. 鲂属鱼类细胞色素 b 片段序列分析. 动物科学, 2012(1):290 - 292.]
- [7] Luo Y L. A revision of fishes of the cyprinid genus *Megalobrama* [J]. Acta Hydrobiologica Sinica, 1990, 14(2):160 - 165. [罗云林. 鲂属鱼类的分类整理. 水生生物学报, 1990, 14(2):160 - 165.]
- [8] Bai X H, Guo X W, Zhang X J, et al. Species identification and evolutionary inference of the genera *Megalobrama* and *Parabramis* (Cyprinidae: Cultrinae) in China [J]. Mitochondrial DNA, 2013 (Early online): 1 - 10. DOI: 10. 3109/19401736. 2013. 823166.
- [9] Hou X Y, Zhu F, Zhang L J. Phylogenetic relationships of 5 gobioid species based on sequences of mitochondrial DNA D-loop [J]. Marine Fisheries, 2013, 35(1):1 - 7. [侯新远, 祝斐, 张丽娟, 等. 基于线粒体 D-loop 基因序列研究我国 5 种虾虎鱼类的系统进化关系. 海洋渔业, 2013, 35(1):1 - 7.]
- [10] Tang Q Y, Huang Y, Wang J J, et al. The complete mitochondrial genome sequence of *Triplophysa bleekeri* (Teleostei, Balitoridae, Nemacheilinae) [J]. Mitochondrial DNA, 2013, 24(1):25 - 27.
- [11] Lussen A, Falk T M, Villwock W, et al. Phylogenetic patterns in populations of Chilean species of the genus *Orestias* (Teleostei: Cyprinodontidae): results of mitochondria DNA analysis [J]. Molecular Phylogenetics and Evolution, 2003, 1 (29): 151 - 160.
- [12] Li Y H, Wang W M, Liu X L, et al. DNA extraction from crayfish exoskeleton [J]. Indian Journal of Experimental Biology, 2011, 49(12):953 - 957.
- [13] Huang Z J, Xu X P, Tang J J, et al. Application and

- primer design of mitochondrial DNA D-loop of freshwater fishes [J]. *Acta Scientiarum Naturalium Universitatis Sunyatseni*, 2009, 48 (4) : 84 - 88. [黄志坚, 徐晓鹏, 唐晶晶, 等. 淡水鱼类线粒体 DNA D-loop 基因的引物设计 and 应用. 中山大学学报, 2009, 48 (4) : 84 - 88.]
- [14] Thompson J D, Gibson T J, Plewniak F, *et al.* The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools [J]. *Nucleic Acids Research*, 1997, 25 (24) : 4876 - 4882.
- [15] Hall T A. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT [J]. *Nucleic Acids Symposium Series*, 1998, 41 : 95 - 98.
- [16] Tamura K, Dudley J, Nei M, *et al.* MEGA4: Molecular evolutionary genetics software version 4.0 [J]. *Molecular Biology Evolution*, 2007, 24 (8) : 1596 - 1599.
- [17] Perna N T, Kocher T D. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes [J]. *Journal of Molecular Evolution*, 1995, 41 (3) : 353 - 358.
- [18] Geng R Q, Chang H, Li Y H, *et al.* Variation of complete coding sequence of GH gene in Chinese bovine species [J]. *Acta Veterinaria et Zootechnica Sinica*, 2008, 39 (12) : 1779 - 1784. [耿荣庆, 常洪, 李永红, 等. 中国牛亚科家畜 GH 基因编码区序列的遗传变异研究. 畜牧兽医学报, 2008, 39 (12) : 1779 - 1784.]
- [19] Guo Z C. Mitochondrial genome sequencing and phylogeny analyzation of Loach in *Cobitis granoei* [D]. Xi' an: Journal of Shaanxi Normal University, 2011. [郭忠超. 北方花鳅线粒体基因组测序及鳅类系统发育分析. 西安: 陕西师范大学, 2011.]
- [20] Rozas J, Sanchez-DelBarrio J C, Messeguer X, *et al.* DNA polymorphism analyses by the coalescent and other methods [J]. *Bioinformatics (Oxford, England)*, 2003, 19 (18) : 2496 - 2497.
- [21] Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood [J]. *Systems Biology*, 2003, 52 (5) : 696 - 704.
- [22] Ronquist F, Huelsenbeck J P. MrBayes 3: Bayesian phylogenetic inference under mixed models [J]. *Bioinformatics*, 2003, 19 (12) : 1572 - 1574.
- [23] Chen D X. Phylogenetic studies of the sinipercid fishes (Perciformes: *Sinipercidae*) based on three complete mitochondrial DNA sequences [D]. Guilin: Guangxi Normal University, 2009. [陈敦学. 三种鳊类线粒体基因组的克隆及其系统进化分析. 桂林: 广西师范大学, 2009.]
- [24] Noack K, Zardoya R, Meyer A. The complete mitochondrial DNA sequence of the *Bichir* (*Polypterus ornatipinnis*), a Basal Ray-Finned Fish: ancient establishment of the consensus vertebrate gene order [J]. *Genetics*, 1996, 144 (3) : 1165 - 1180.
- [25] Chang Y S, Huang F L, Lo T B. The complete nucleotide sequence organization of carp (*Cyprinus carpio*) mitochondrial genome [J]. *Journal of Molecular Evolution*, 1994, 38 (2) : 138 - 155.
- [26] Broughton R E, Milam J E, Roe B A. The complete sequence of the zebra fish (*Danio rerio*) mitochondrial genome evolutionary patterns in vertebrate mitochondrial DNA [J]. *Genome Research*, 2001, 11 (11) : 1958 - 1967.
- [27] Wang C H, Chen Q, Lu G Q, *et al.* Complete mitochondrial genome of the grass carp (*Ctenopharyngodon idella*, Teleostei): Insight into its phylogenic position within Cyprinidae [J]. *Gene*, 2008, 424 (1 - 2) : 96 - 101.
- [28] Liu H Y, Yu L N, Zhang F R. The progress of the molecular structure and its application in fish mitochondrial DNA control region [J]. *Reservoir Fisheries*, 2008, 28 (2) : 4 - 8. [刘红艳, 余来宁, 张繁荣. 鱼类线粒体 DNA 控制区的分子结构及应用进展. 水利渔业, 2008, 28 (2) : 4 - 8.]
- [29] Anderson S, Bankier B G, Barrell B G, *et al.* Sequence and organization of the human mitochondrial genome [J]. *Nature*, 1981, 290 (5806) : 457 - 465.
- [30] Chen T. Sequence and analysis of complete mitochondrial genome of *Pseudorasbora parva* [D]. Xi' an: Journal of Shaanxi Normal University, 2010. [陈涛. 麦穗鱼 (*Pseudorasbora parva*) 线粒体基因序列测定及分析. 西安: 陕西师范大学, 2010.]
- [31] Chen T, Shi Y R, You P. Sequence and analysis of complete mitochondrial genome of *Pseudorasbora parva* [J]. *Acta Zootaxonomica Sinica*, 2012, 37 (1) : 10 - 19. [陈涛, 史妍茹, 尤平. 麦穗鱼线粒体基因组序列测定及分析. 动物分类学报, 2012, 37 (1) : 10 - 19.]
- [32] Desjardins P, Morais R. Sequence and gene organization of the chicken mitochondrial genome: a novel gene order in higher vertebrates [J]. *Journal of*

- Molecular Evolution,1990,212(4):599-634.
- [33] Anderson S, de Bruijn M H L, Coulson A R, *et al.* Complete sequence of bovine mitochondrial DNA, conserved feature of the mammalian genome [J]. Journal of Molecular Evolution, 1982, 156(4):683-717.
- [34] Wang X Z. Molecular phylogeny of the East Asian cyprinids (Pisces: Cypriniformes) [D]. Wuhan: Institute of Hydrobiology, Chinese Academy of Sciences, 2005. [王绪祯. 东亚鲤科鱼类的分子系统发育研究. 武汉: 中国科学院水生生物研究所, 2005.]
- [35] Lin S H. Study on application of DNA markers to identify components and species of pufferfish [D]. Fuzhou: Fujian Agriculture and Forestry University, 2012. [林少华. DNA 标记在河豚鱼成分及其物种鉴定中的应用研究. 福州: 福建农林大学, 2012.]
- [36] Chang Y S, Huang F L, Lo T B. The complete nucleotide sequence organization of carp (*Cyprinus carpio*) mitochondrial genome [J]. Journal of Molecular Evolution, 1994, 38(2):138-155.
- [37] Liu Y, Cui Z. The complete mitochondrial genome sequence of the cutlassfish *Trichiurus japonicus* (Perciformes: Trichiuridae): genome characterization and phylogenetic considerations [J]. Marine Genomics, 2009, 2(2):133-42.
- [38] Lee W J, Kocher T D. Complete sequence of a sea lamprey (*Petromyzon marinus*) mitochondrial genome: early establishment of the vertebrate genome organization [J]. Genetics, 1995, 139(2):873-887.
- [39] Miya M, Takeshima H, Endo H, *et al.* Major patterns of higher teleostean phylogenies: sequences: a new perspective based on 100 complete mitochondrial DNA sequences [J]. Molecular Phylogenetics and Evolution, 2003, 26(1):121-138.
- [40] Nielsen J L, Gan C, Thomas W K. Difference in genetic diversity for mitochondrial DNA between hatchery and wild population of *Oncorhynchus* [J]. Canadian Journal of Fisheries Aquatic Sciences, 1994, 51(1):290-297.
- [41] Mojica C L, Meyei A, Barlow G W. Phylogenetic relationships of species of the genus *Brachyrhaphis* (Poeciliidae) inferred from partial mitochondrial DNA sequence [J]. Copeia, 1997(2):298-305.
- [42] Shen X, Wang H Q, Ren J F, *et al.* The mitochondrial genome of *Euphausia superba* (Prydz Bay) (Crustacea: Malacostraca: Euphausiacea) reveals a novel gene arrangement and potential molecular markers [J]. Molecular Biology Reports, 2010, 37(2):771-784.
- [43] Tseng C S, Hui C F, Shen S C, *et al.* The complete nucleotide sequence of the *Crossostoma iactstre* mitochondrial genome: conservation and variations among vertebrates [J]. Nucleic Acids Research, 1992, 20(18):4853-4858.
- [44] Rasmussen A S, Arnason U. Phylogenetic studies of complete mitochondrial DNA molecules place cartilaginous fishes within the tree of bony fishes [J]. Journal of Molecular Evolution, 1999, 48(1):118-123.
- [45] Horai S, Hayasaka K, Kondo R, *et al.* Recent african origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs [J]. Proceedings of the National Academy of Sciences, 1995, 92(2):532-536.
- [46] Zardoya R, Malaga-Trillo E, Veith M, *et al.* Complete nucleotide sequence of the mitochondrial genome of a salamander, *Mertensiella luschani* [J]. Gene, 2003, 317(1-2):17-27.
- [47] Wu X B, Wang Y Q, Zhou K Y, *et al.* The complete mitochondrial genome and crocodile phylogenetic in Chinese alligator [J]. Chinese Science Bulletin, 2003, 48(18):1954-1958. [吴孝兵, 王义权, 周开亚, 等. 扬子鳄的线粒体全基因组与鳄类系统发生. 科学通报, 2003, 48(18):1954-1958.]
- [48] Mindell D P, Sorenson M D, Dimcheff D E, *et al.* Interordinal relationships of birds and other reptiles based on whole mitochondrial genomes [J]. Systems Biology, 1999, 48(1):138-152.
- [49] Pu Y G, Peng Q L, Wang Z F, *et al.* Sequence and organization of the complete mitochondrial genome of the Chinese three-keeled pond turtle *Chinemys reevesi* [J]. Acta Zoologica Sinica, 2005, 51(4):691-696. [蒲友光, 彭巧玲, 王志方, 等. 乌龟线粒体全基因组序列和结构分析. 动物学报, 2005, 51(4):691-696.]
- [50] Liu H Z. Structure and evolution of the mitochondrial DNA control region in fish; in the case of bitterling fish [J]. Progress in Natural Science, 2002, 12(3):266-271. [刘焕章. 鱼类线粒体 DNA 控制区的结构与进化: 以鲢鳙鱼类为例. 自然科学进展, 2002, 12(3):266-271.]
- [51] Zou X J, Han X, Han H F. Research survey of mitochondrial DNA (mtDNA) and noncoding region of fish [J]. Guizhou Journal of Animal Husbandry & Veterinary Medicine, 2009, 6(4):23-25. [邹习俊,

- 韩雪,韩虎峰. 鱼类 mtDNA 及其非编码区的研究概况. 贵州畜牧兽医, 2009, 6(4): 23 - 25.]
- [52] You C P. Studies on the complete mitochondrial genomes and mitochondria-related genes in liver transcriptome of different polyploidy fish [D]. Changsha: Hunan Normal University, 2012. [尤翠平. 不同倍性鱼线粒体基因组即肝脏转录组中线粒体相关基因分析. 长沙: 湖南师范大学, 2012.]
- [53] Li S F, Zhu Z W, Zou S M, *et al.* Interspecific phylogenesis and intraspecific genetic differences of genus megalobrama; bluntnose black bream (*M. amblycephala*), Guangdong black bream (*M. hoffmanni*) AND BLACK BREAM (*M. terminalis*) [J]. Acta Zoologica Sinica, 2002, 48(3): 339 - 345. [李思发, 朱泽闻, 邹曙明, 等. 鲂属团头鲂、三角鲂及广东鲂种间遗传关系及种内遗传差异. 动物学报, 2002, 48(3): 339 - 345.]

Comparison of mitochondrial genomes of the genus *Megalobrama* and their phylogenetic analysis

LAI Ruifang¹, ZHANG Xiujie¹, LI Yanhe¹, WU Junjie¹, YANG Donghui², WANG Weimin^{1*}

(1. Key Lab of Agricultural Animal Genetics, Breeding and Reproduction of Ministry of Education, Key Lab of Freshwater Animal Breeding, Ministry of Agriculture, College of Fishery, Huazhong Agricultural University, Wuhan 430070, China;

2. Department of Fisheries Science, Xinyang College of Agriculture and Forestry, Xinyang 464000, China)

Abstract: This study cloned the complete mitochondrial genome sequences of *Megalobrama skolkovii*, *M. pellegrini* and *M. terminalis* based on the complete mitochondrial genome sequences of *M. amblycephala* and partial mitochondrial genome sequences of *M. hoffmanni*, *M. ellegrini* and *M. terminalis* in the GenBank database. The complete mitochondrial genome sequences of four species in the genus *Megalobrama* were analyzed. Results showed that all the mitochondrial genomes of the four *Megalobrama* fishes contained 13 protein-coding genes, 22 tRNA genes, 2 rRNA genes, a non-coding control region (D-loop) and a replication origin region of the light chain (OL), with identical arrangement and tightly packed. Eight tRNAs and ND6 genes were encoded on the light strand (L strand), and the remainders encoded on the heavy strand (H strand). All 13 protein-coding genes of the mitochondrial genomes from four species of the genus *Megalobrama* showed a strong bias toward A + T and a bias for C nucleotides. The comparison of the complete mitochondrial genome sequences showed that there were 758 variable sites, including 691 singleton variable sites (accounting for 91.19%), 67 parsimony informative sites (only accounting for 8.84%). Eleven of 22 tRNA genes had mutated in *Megalobrama*, and 23 variable sites were mainly distributed in the TΨC and DHU of the cloverleaf secondary structure. Additionally, 626 variable sites of the protein-coding genes were mainly distributed in the third codon position (accounting for 82.59%) and Cyt b gene had the largest number of different loci (84), followed by ND4 gene with 83 loci. Therefore, Cyt b and ND4 could be used as alternative molecular markers for genetics research in the genus *Megalobrama* groups, which provided a theoretical and technical basis for the rational and efficient use of biological resources and conservation of biological diversity. Both ML and BI phylogenetic trees based on the complete mitochondrial genome sequences from four species showed that *M. terminalis* and *M. pellegrini* had the closest genetic relationship, with *M. amblycephala* being more closely related to them, and *M. hoffmanni* was farther related to the other species.

Key words: *Megalobrama*; mitochondrial genomes; variation loci; molecular markers; phylogenetic

Corresponding author: WANG Weimin. E-mail: wangwm@mail.hzau.edu.cn