

# 鱼类线粒体 DNA 控制区的结构和进化： 以鳊鲃鱼类为例\*

刘焕章

中国科学院水生生物研究所, 武汉 430072

**摘要** 以鳊鲃鱼类为例, 研究了鱼类线粒体 DNA 控制区的结构和进化规律. 识别了终止序列区、中央保守区和保守序列区 3 个区域. 指出扩展终止相关序列(ETAS)的主体是 TACAT 和它的反向互补序列 ATGTA 形成的发夹结构. 给出了鱼类中若干重要保守序列的普遍形式. 研究表明, 一般情况下, 只有一个行使功能的 ETAS, 但可能会有多个复制的、不行使功能的 ETAS 存在. 鱼类的保守序列 CSB2 最为保守. 线粒体 DNA 控制区被认为是由各功能单位形成主体框架, 主体框架复制产生重复序列, 重复序列产生快速变异, 这样造成不同类群间线粒体 DNA 控制区巨大差异. 易突变点和二级结构的存在均可能与变异的发生相关.

**关键词** 鳊鲃鱼类 线粒体 DNA 控制区 进化

控制区是线粒体 DNA 中变化最复杂, 被了解最少, 也最吸引人的区域. 对控制区结构功能的研究将有助于了解 DNA 复制、转录的机制和进化规律. 目前对动物线粒体 DNA 控制区的研究主要是识别了众多的功能保守序列<sup>[1-5]</sup>, 例如终止相关序列(TAS), 保守序列 CSB1, CSB2, CSB3, CSB-A, -B, -C, -D, -E, -F 等. 另外是将控制区分为终止序列区(左区)、中央保守区和保守序列区(右区) 3 个区. 这些研究主要是针对哺乳动物进行的. Randi 等发现鸟类与哺乳动物线粒体 DNA 控制区结构相似但又有其特殊性<sup>[6]</sup>.

在鱼类方面, 对比哺乳动物的结构, 识别了一些类群的终止相关序列和部分保守序列<sup>[7-9]</sup>. 到目前为止, 还没有人给出各功能单位(包括终止相关序列 TAS 和各种保守序列 CSB)的普遍形式. 而且 Lee 等<sup>[10]</sup>对众多鱼类界定的中央保守区范围是错误的, 其后段比实际的区域少约 100 bp. 本文以鳊鲃鱼类为例, 参考已有研究结果, 研究鱼类线粒体 DNA 控制区的结构和进化规律, 给出保守序列的普遍形式和不同区域的划分标志, 为进一步对鱼类线粒体 DNA 控制区研究提供参考.

## 1 材料与方法

选取高体鳊鲃 (*Rhodeus ocellatus*)、彩副鲃 (*Paracheilognathus imberbis*) 和大鳍鲃 (*Acheilognathus macropterus*) 为代表. 标本用酒精固定, 取背部肌肉按 Kocher 等<sup>[11]</sup>的方法提取总 DNA, 消化温度改为 56℃, 时间为 12 h, 酚抽提去除蛋白. PCR 扩增控制区的完整片段. 引物序列为 DL1 (ACC CCT GGC TCC CAA AGC) 和 DH2 (ATC TTA GCA TCT TCA GTG), 分别位于脯氨酸和苯丙氨酸的 tRNA 上. 反应程序为: 94℃ 预变性 2 min; 94℃ 变性 30 s, 50℃ 退火 30 s, 72℃ 延伸 1.5 min, 共进行 40 个循环; 最后 72℃ 延伸 10 min. 采用 BioStar 的 glassmik 纯化试剂盒进行 PCR 产物的回收纯化, 并进行 DNA 测序(上海基康公司). 序列均已送 GenBank, 序列号为: AY017147, AY017148, AY017149. 用手工和计算机结合进行排序<sup>[12]</sup>, 先找出同源保守序列, 再进行调整完成排序. 以两个 tRNA 的结束分别作为控制区的起点和终点. 以 ECSB-F, CSB1 的起点分别作为终止序列区、中央保守区和保守序列区的分界线.

2001-05-28 收稿, 2001-06-16 收修改稿

\* 国家自然科学基金资助项目(批准号: 39670101, 49832010)

E-mail: hzliu@ihb.ac.cn

## 2 结果与讨论

### 2.1 序列长度与核苷酸组成

测得的3种鳊鱼类控制区长度分别为:高体鳊 1015 bp, 彩副鳊 915 bp, 大鳍鳊 1125 bp, 长度差异非常明显, 但是仍在 Lee 等研究的 856~1500 bp 的变异范围之内<sup>[10]</sup>. 核苷酸的平均组成为: T 32.2%, C 20.1%, A 33.6%, G 14.1%. 同其他鱼类 A, T 含量高, G 含量低的特点也很相似.

### 2.2 终止序列区

终止序列区, 包含与 DNA 复制终止相关的序列 TAS, 主要的变异一般出现在该区域. 由于序列

差异较大, 将这部分的序列分别分析. 彩副鳊的终止序列区长为 232 bp, 没有明显的重复、长片段的缺失和插入(图 1), 发现一个 TAS, 其中含有核心序列 ACAT. 高体鳊终止序列区长为 340 bp, 其中有两处重复, 一处位于 54~103 nt, 长为 50 bp, 另一处位于 105~129 nt, 长为 25 bp. 这两处均为不完全重复. 与源序列相比, 重复序列中的碱基有一定的变异, 都含有 TAS 的反向互补序列 ATGT. 高体鳊中有多个 TAS 被发现. 大鳍鳊终止序列区长为 442 bp, 其中有一段 59 bp 的不完全重复(46~103 nt 处), 重复序列也含有 TAS 的反向互补序列 ATGT. 紧接在重复序列后有一段 159 bp 的插入(104~262 nt 处), 其同源性尚不清楚.

彩副鳊(*Paracheilognathus imberbis*) 232bp  
 GATAGTAACCATATGGTGACAAACACGATATGTATAGTATATGCACATATATGTATGTATT  
 ATCACCATGCAATTATTTTAACTAAAAGCAAGTACTAAAATCCAAGAGTCCATAGACT  
 AAACCTCACAAGAACTTGGCAATGGGAGATCTTAATTTTATAAATTGATTATCCCACTTA  
 AAATGAATTACAAATAATTAATGATTGGATTACTCCACTAACGTTTAAATAA

大鳍鳊(*Acheilognathus macropterus*) 442bp  
 GGTAGAAACCATTTGATGTAGACAAATATGTATAATAAAGATATAACATGCTCTATGTAT  
 TATCACCATGCAATTATTTTAACTAAAAGCAAGTACTAATATTTTCAGAAGTACATAGAC  
 TAAATATTCACAATCAACCAAAATAGTTTTTAAATTATATTTAATATTTCTCCACCCCC  
 CCTCTGTAATAGTATTAAGTGCCAATATAACATTTAATTATAAAGAACATAAAATGTTT  
 AATAAAATATTAATGGTATATAATACATATCTATGTATTATCACCATGCAATTATTTTAA  
 CCTAAAAGCAAGTACTAATATATTAACGTTTCAGAACTTAAAATGAAAAATCAATTAATA  
 ATTATTTTCAAGTACTGTTTATTTCCACTAAGAATGGATCACAAATGTAAGATATCGAAT  
 TACTCCACTAATCTTTAAATAA

高体鳊(*Rhodeus ocellatus*) 340bp  
 GGTAGTAACCATATATGATTAGTACATCATATTCACATATAAACATTTTATGACATAT  
 ATTACCTATATGTATTAACACCATAGGCTTATTTTAACTAAATACATACATATGTATT  
 ATCCCATAAAATTATTTAGAGCTGCTCACATATTTATGTATTATCACCATTCAITTTATT  
 TTAACCTAAAAGCAAGTACTAAAATTTGTAATTCAAAAGCACTATCATTTAGATCCCCTA  
 TGGTTTATCTTTTCGTCTAAGTTTATGGAGGACTTAACTAACAAATCCACAGCCAATATTT  
 CCTCTGGTGGACCAACTAACATTTATATCGACAAGATTA

图 1 3种鳊鱼类线粒体 DNA 控制区终止序列区的序列

图中方框表示终止相关的序列, 阴影表示终止相关序列的反向互补序列, 双线表示重复序列和源序列所在位置, 单线表示 ETAS

Doda 等<sup>[1]</sup>最初给出的 TAS 为 ACATTAAYYAAT. 但是其他研究者发现这一序列在不同类群有较大的变异. Sbisà 等<sup>[5]</sup>对比 27 个哺乳动物的序列后, 将

TAS 变为扩展的 TAS (ETAS), 在哺乳动物中有两个, 其中比较稳定的是 ACAT 序列. Saccone 等<sup>[13]</sup>发现在哺乳类中, TACAT 和它的反向互补序列

ATGTA 常常一起出现, 形成发夹结构. 事实上, 这一现象在许多类群中存在. 所比较的鱼类序列中 都存在这一结构, 因此我们认为它可能是 ETAS 的 主体. 同时, 我们确定鱼类中的 ETAS 序列为: TACATAT-----ATGTATTATCACCAT-ATAT

TATATTAACCAT(图2)(“—”表示发生变异的碱 基, 即转换, 颠换或缺失, 下同). 由于很多类群 中 ETAS 只有一个, 我们认为一般情况下, 只有 一个 ETAS 行使功能, 其他的是复制的结果, 不行使 功能.

鲱 2	TACATCT-----ATGTATTATCACCATACGC-TGAAATTACCCTA-
鲱 3	TACATCT-----ATGTATTATCACCATACGC-TGAAATTACCCTA-
鲑 2	TACATTAT-----ATGTATTATCAACATACGC-TGATTTTAACCC-
鲈 2	TACATAT-----ATGTATTATCACCATATAT-AGAATTTAACCAT-
鲤	TACATAT-----ATGTATTATCACCACACTCA-CTATTTAACCAT-
台湾缨口鳅	TACATAT-----ATGTAATATCACCATTAAA-TTATTTAGACCAT-
黑线鲟	TACAT-----ATGTATAATCGCCATTAAT-TAAC-TTAACCATT
大西洋小鲑	TACAT-----ATGTATAATCACCATTAAAT-TAAC-TTAACCAAT
青鲑	TACAT-----ATGTATAATCACCATTAAAT-TAAC-TTAACCATA
大西洋鲑	TACAT-----ATGTATAATCACCATTAAAT-TAAG-TTAACCATA
丽鱼	TGCATAT-----ATGTATTATCACCATTATT-TTATGTTAAACAT-
鲈 1	TACATAT-----ATGTATTATCACCATTAAAT-TTATATTAACCAT-
丽鱼 b	TACATAT-----ATGTACTTACACCATGAAT-TTATATTAACCAT-
隆头鱼 1	TACATAT-----ATGTAATTACACCATAAAT-TTATATTAACCAT-
鲷	TACATGT-----ATGTAATAACACCATATAT-TTATAGTACCCAT-
黄盖鲈 2	TACATGT-----ATGTAATAACACCATATAT-TTATAGTAACCAT-
欧鲈	TACATCT-----ATGTAATAACACCATATAT-TTATAGTAACCAT-
西鲱	TACATACT-----ATGTATAATCACCATTATATATCAAGTA--
锯腹鲱	TACATACT-----ATGTATAATCCCATTCATATTATGTCAGGTA--
鲟	TACATAAGACATGCTATGTTAATCCCATTAAT-TTCTAGCCACCAT-
通式	TACATAT-----ATGTATTATCACCAT--ATATTATATTAACCAT-

图 2 19 种鱼类 ETAS 序列比较和普遍性序列

序列来源于 Lee 等<sup>[10]</sup>. 注意鲱有一个 ETAS 的完全重复. 许多种类有多个 ETAS 存在, 但图中只列出可能行使功能的 ETAS, 种类名字后面的数字表示该 ETAS 在出现多个 ETAS 时的编号

### 2.3 中央保守区

Southern 等<sup>[4]</sup>首次在哺乳动物线粒体 DNA 控制区的中央保守区识别了保守序列 B, C, D, E, F. Randi 等<sup>[6]</sup>发现保守序列 C, D, E, F 在鸟类中存在. 在对更多的哺乳动物进行比较时, Sbisà 等<sup>[5]</sup>发现保守序列并不明显. Lee 等<sup>[10]</sup>在对众多鱼类的序列进行比较时, 仅识别了 CSB-D 的存在. 对比哺乳动物的结构, 并参照其他鱼类序列, 我们研究了鳉鱼类中央保守区的结构.

首先有一段保守序列, 其位置相当于哺乳动物中的 CSB-F, 但序列不相似, 关键序列为 ATG

TAGTA---GAGACCACC, 我们将其命名为 ECSB-F, 它是分开终止序列区和中央保守区的标志(图 3). 紧接其后的是 CSB-E. CSB-E 有较大的变异, 识别的标志是关键序列 AGGG----GTGGGG 的存在, 其中 GTGGG-box 已被 Lee 等<sup>[10]</sup>所描述. 在 CSB-E 之后的是 CSB-D. 尽管 Lee 等<sup>[10]</sup>识别了 CSB-D, 但实际上 CSB-D 的变异也还是很大, 在鳉鱼类其普遍序列为: TATT-CTTG-ATCTG-T-A. 除此之外, 我们没有能够找到 CSB-B 和 CSB-C. 但是我们找到了一些保守片段, 特别是其中 GCAT 多次出现, 并且有的还形成发夹结构.

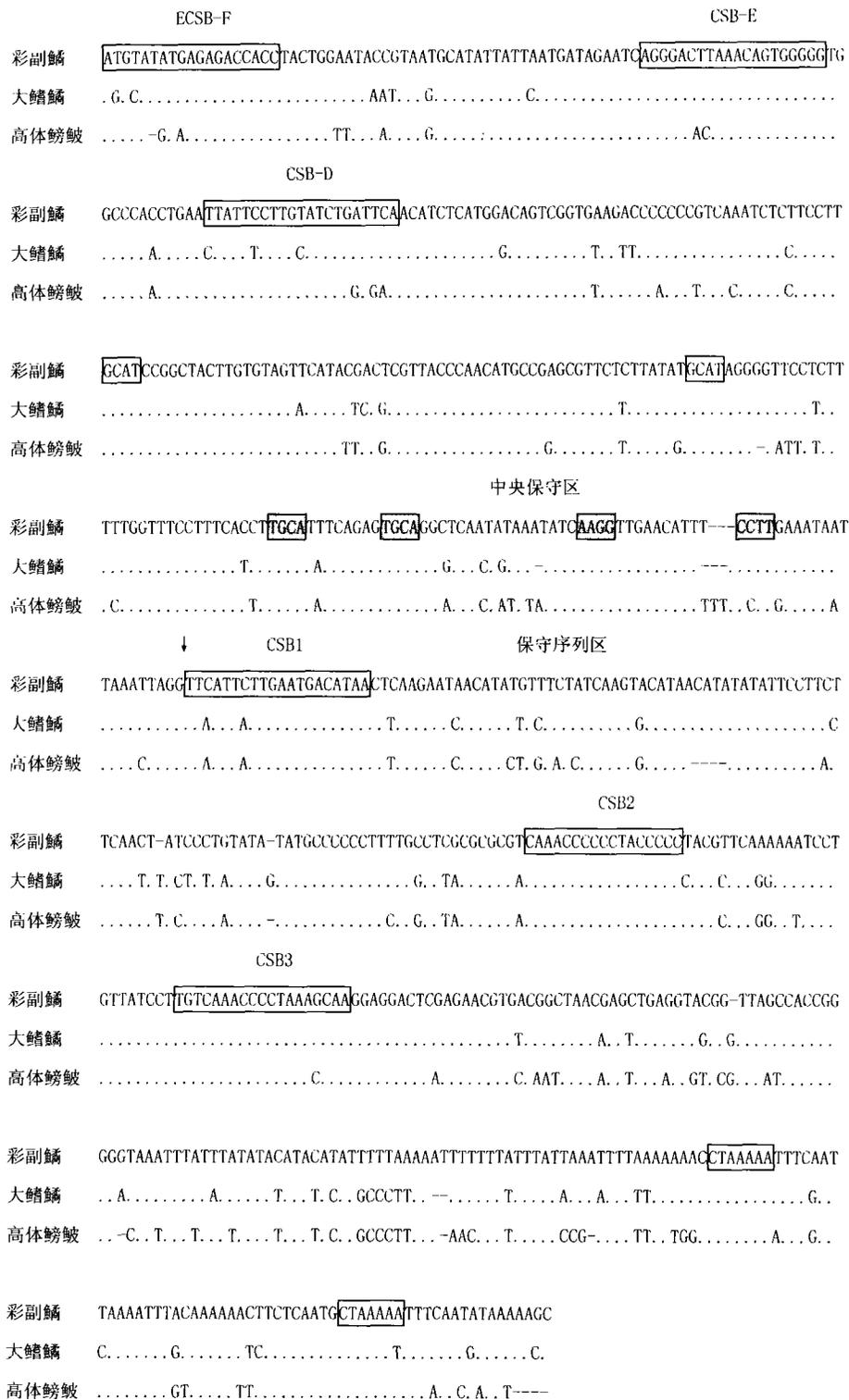


图3 3种鲟鱼类线粒体DNA控制区的中央保守区和保守序列区序列比较  
方框表示保守序列，阴影表示发夹结构。箭头为中央保守区和保守序列区的分界点。“-”表示缺失

## 2.4 保守序列区

哺乳动物的保守序列区一般存在保守序列 CSB1, CSB2, CSB3. Randi 等<sup>[13]</sup>发现鸟类中 CSB2 和 CSB3 合并到一起. 在鱼类仅 Broughton 等<sup>[8]</sup>成功地识别了雅罗鱼上的 CSB1, CSB2, CSB3; 其他人<sup>[9,10]</sup>则没有能识别 CSB1. 对比哺乳动物的序列, 参考 Broughton 等研究, 我们识别了鲮鱼鱼类上面的 CSB1, CSB2, CSB3.

研究表明 CSB1 与线粒体 DNA 的复制起点相关<sup>[14]</sup>, 在哺乳动物中特别保守, 其一般形式为: ATT-AATTAATG-T-GCAGGACATA. 对比哺乳动物和鱼类的序列, 我们发现其中的 GACATA 最为保守, 可以帮助识别 CSB1 的存在. 在鲮鱼鱼类中, 其一般形式为 TT-ATTATTGAA-GACATA. 在 CSB1 之后有一个含多个碱基 C 的序列, 其一般形式为 TATGCCCCC, 功能尚不清楚.

CSB2 在哺乳动物有一些变异, 在鱼类则最为保守, 最容易识别. 鲮鱼鱼类中其普遍形式为: CAAACCCCCTACCCCC, 其他鱼类中有一定的变异<sup>[9]</sup>. 鱼类的 CSB3 也较容易识别, 它比 CSB2 变异大一些, 一般形式为: TGTCAAACCCCTAAACAA. 在 CSB3 之后有两个保守序列 CTAAAAA 被发现, 其功能尚不清楚.

## 2.5 线粒体 DNA 控制区的进化规律

近年来, 大量的研究表明, 至少在脊椎动物中, 线粒体 DNA 控制区的结构是相似的. 一些序列非常保守, 在鱼类、鸟类、哺乳动物中都存在, 被认为是功能单位, 受自然选择影响较大. 但是线粒体 DNA 控制区又存在巨大的变异, 插入、缺失、串联重复等变异导致序列长度很不一样. 有研究表明, 串联重复序列常常与功能序列相似. 例如, Randi 等<sup>[6]</sup>发现鸟类中与 CSB1 相似的重复, Broughton 等<sup>[8]</sup>在雅罗鱼中发现与 ETAS 相似的重复. 在本研究中, 高体鲮鱼中出现了两次与 ETAS 相似的重复, 大鳍鲮鱼中出现了一次. 所以, 线粒体 DNA 控制区是由各功能单位形成主体框架, 这些主体框架会复制产生重复序列, 重复序列会产生快速变异, 这样造成不同类群间线粒体 DNA 控制区巨大差异.

对于变异的产生, Sbisà 等<sup>[5]</sup>认为存在容易发生缺失或扩展的变异点. Broughton 等<sup>[8]</sup>也认为将相

似的重复解释为独立复制更简约. 也有人认为复制可能与形成稳定的二级结构有关. 从鱼类情况看, 不同的区域机制可能不同, 易发变异点和二级结构的存在均可能与变异相关.

**致谢** 张唐林同志协助采样; 胡振新同学进行测序; 肖武汉博士帮助设计引物, 在此表示感谢.

## 参 考 文 献

- 1 Doda J N, et al. Elongation of displacement-loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. *Proc Natl Acad Sci USA*, 1981, 78: 6116
- 2 Walberg M W, et al. Sequence and properties of the human KB cell and mouse L cell d-loop regions of mitochondrial DNA. *Nucleic Acids Res*, 1981, 9: 5411
- 3 Saccone C, et al. Structural elements highly preserved during the evolution of the d-loop-containing region in vertebrate mitochondrial DNA. *J Mol Evol*, 1987, 26: 205
- 4 Southern S O, et al. Molecular characterization of a cloned dolphin mitochondrial genome. *J Mol Evol*, 1988, 28: 32
- 5 Sbisà E, et al. Mammalian mitochondrial D-loop region structural analysis: Identification of new conserved sequences and their functional and evolutionary implications. *Gene*, 1997, 205: 125
- 6 Randi E, et al. Organization and evolution of the mitochondrial DNA control region in the avian genus *Alectoris*. *J Mol Evol*, 1998, 47: 149
- 7 Buroker N E, et al. Length heteroplasmy of sturgeon mitochondrial DNA: An illegitimate elongation model. *Genetics*, 1990, 124: 157
- 8 Broughton R E, et al. Length variation in mitochondrial DNA of the minnow *Cyprinella spiloptera*. *Genetics*, 1994, 138: 179
- 9 Chen I, et al. Sequence length and variation in the mitochondrial control region of two freshwater fishes belong to *Rhinogobius* (Teleostei: Gobioidae). *J Fish Biol*, 1998, 53: 179
- 10 Lee W, et al. Structure and evolution of telost mitochondrial control regions. *J Mol Evol*, 1995, 41: 54
- 11 Kocher T D, et al. Dynamics of mitochondrial DNA evolution in animals: Amplification and sequencing with conserved primers. *Proc Natl Acad Sci USA*, 1989, 86: 6196
- 12 Thompson J D, et al. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, 1994, 22: 4673
- 13 Saccone C, et al. The main regulatory region of mammalian mitochondrial DNA: Structure-function model and evolutionary pattern. *J Mol Evol*, 1991, 33: 83
- 14 Chang D D, et al. Identification of primary transcriptional start sites of mouse mitochondrial DNA: Accurate *in vitro* initiation of both heavy and light-strand transcripts. *Mol Cell Biol*, 1986, 6(5): 1446