

## 硬骨鱼类线粒体基因系统发育信息效率分析

陈姝君<sup>1,2</sup>, 赫崇波<sup>2</sup>, 木云雷<sup>2</sup>, 刘卫东<sup>2</sup>, 周遵春<sup>2</sup>, 高祥刚<sup>2</sup>, 丛林林<sup>1,2</sup>

(1. 辽宁师范大学 生命科学学院, 辽宁 大连 116029; 2. 辽宁省海洋水产科学研究院, 辽宁省海洋水产分子生物学重点实验室, 辽宁省应用海洋生物技术开放实验室, 辽宁 大连 116023)

**摘要:** 采用 PCR 产物直接测序法获得圆斑星鲽 (*Verasper variegatus*) 和条斑星鲽 (*V. moseri*) 线粒体基因组的全部基因序列, 并从 GenBank 中下载已知分类地位相关鱼类的线粒体基因的核苷酸序列和蛋白质氨基酸序列, 用蛙、鸡、牛做外群, 采用 NJ 和 MP 法, 重建鱼类的系统发育树。通过计算各个基因在重建系统发育树时的准确率, 估算该基因所含有的系统发育信息。结果表明, 从氨基酸序列和核苷酸序列以及在目、科、属综合分析, 这些基因的系统发育信息大致分为好 (16S rRNA, ND2, ND4, 12S rRNA, ND6, ND5)、中 (Cytb, COII, COIII, ND1, COI) 和差 (ND3, ND4L, ATP6, ATP8) 3 个组。在目阶元, 当用核苷酸序列分析时, 12S rRNA, 16S rRNA, COII, ND4, ND5 和 ND6 最好, COI, COIII, Cytb, ND1 和 ND2 为中等, 而 ND3, ATPase6, ATPase8 和 ND4L 最差。当用氨基酸序列分析时, ND4 和 ND5 最好, ND1, ND2, Cytb, ND6, COI, COII 和 ND4L 为中等, ND3, ATPase6, COIII 和 ATPase8 最差。在科阶元, 当用核苷酸序列时, 12S rRNA, 16S rRNA, ND2 和 ND6 系统发育信息最好; 用氨基酸序列时, Cytb 和 ND2 最好。在属阶元, 所有基因的核苷酸序列都具有很好的系统发育信息, 而 COII, ND4L, ND1 和 ND3 的氨基酸序列系统发育信息最差。本研究结果有助于进一步利用线粒体基因研究分析鱼类系统进化关系。[中国水产科学, 2008, 15(1): 12-21]

**关键词:** 圆斑星鲽; 条斑星鲽; 硬骨鱼; 线粒体; 系统发育

**中图分类号:** S917.4      **文献标识码:** A      **文章编号:** 1005-8737-(2008)01-0012-10

由于线粒体中不同基因的进化速度不同, 所以对于脊椎动物系统发育研究而言, 其所具有的系统发育信息也不同<sup>[1-2]</sup>。系统发育信息<sup>[3]</sup> (Phylogenetic information) 也称作系统发生信息, 是用于估算生物有关类群或亲缘关系远近的基因序列信息, 一个基因所具有的系统发育信息的大小, 可以通过该基因在重建系统发育树时的准确率来估算<sup>[3]</sup>。自 PCR 技术和直接测序方法的广泛应用后, 线粒体 DNA 成为了系统发育研究的可靠标记<sup>[3]</sup>。Zardoya 等<sup>[3]</sup>对 19 种脊椎动物线粒体不同基因的进化速度进行研究后, 将 13 个蛋白编码基因分为好、中、差 3 个组, 好的一组包括 ND4, ND5, ND2, Cytb 和 COI, 中等的一组包括 COII, COIII, ND1 和 ND6, 差的一组包括 ATPase6, ND3, ATPase8 和 ND4L。鱼类在动物进化过程中处于承先启后的重要地位<sup>[4]</sup>。日本学者 Miya<sup>[5]</sup>对已知分类关系的 8 种硬骨鱼类的线粒体不同蛋白质编码基因所含有的系统发育信息进行

了研究后表明, 13 个蛋白编码基因所含有的系统发育信息明显不同, 大致分为非常好、好、中等、差和非常差 5 个类别, 非常好的一组包括 ND5, ND4, COIII 和 COI, 好的一组包括 COII 和 Cytb, 中等的一组包括 ND3 和 ND2, 差的一组包括 ND1 和 ATPase 6, 非常差的一组包括 ND4L, ND6 和 ATPase 8。这与 Zardoya 等的观点有所差异, Miya 认为观察的普遍性还不能确认, 需要进一步研究。关于鱼类线粒体各个基因的系统发育信息如何、鱼类线粒体不同基因的保守性及其在系统进化中所包含的信息如何、是否与其他脊椎动物相同等问题, 以及 Miya 和 Zardoya 等在脊椎动物的研究结论能否适应于鱼类这个庞大的类群、能否适应于鱼类目、科、属和种间的系统发育研究等尚需更深一步的研究。

本研究利用 PCR 产物直接测序法测定圆斑星鲽 (*Verasper variegatus*) 和条斑星鲽 (*V. moseri*) 的线粒体基因组全序列, 同时从 GenBank 中下载硬骨

收稿日期: 2006-11-20; 修订日期: 2007-04-20.

基金项目: 国家自然科学基金项目资助 (30571410); 海洋经济实施科技推进平台与运行项目 (2005-2007 年).

作者简介: 陈姝君 (1981-), 女, 硕士研究生, 研究方向: 动物分子遗传学.

通讯作者: 赫崇波 (1961-), 男, 博士, 研究员, 研究方向: 分子遗传学. E-mail: hechongbo@hotmail.com

鱼纲中已知分类关系的相关鱼类线粒体基因组序列,用蛙、鸡、牛做外群,采用生物信息学分析软件,对鱼类的线粒体各个蛋白质编码基因和核糖体基因的保守性以及它们在目、科、属不同分类阶元的系统发育信息进行了比较和研究,为进一步利用线粒体基因序列进行鱼类的分类、进化和遗传多样性等研究提供参考数据。

## 1 材料与方法

### 1.1 圆斑星鲈和条斑星鲈线粒体基因获得

实验鱼取自辽宁省海洋水产科学研究院旅顺海珍品繁育中试基地。取其背部肌肉组织约 500 mg,

按常规酚氯抽提方法<sup>[6]</sup>提取总 DNA。以 GenBank 中的圆斑星鲈 (GenBank Accession Nos: DQ242488、DQ242490、DQ242494、DQ242492、AY671919) 和条斑星鲈 (GenBank Accession Nos: DQ242489、DQ242491、DQ242495、DQ242493、AB207249)<sup>[7]</sup>, 以及石鲈 (*Platichthys bicoloratus*, AB028664) 的部分序列为基础,设计出 8 对扩增引物,对圆斑星鲈和条斑星鲈线粒体进行 PCR 扩增 (表 1)。利用扩增产物直接测序法和引物行走法 (Primer walking) 进行线粒体基因组测序 (大连宝生物)。通过 Blastn 同源保守序列搜索和 Clustal X 和 MEGA3.0<sup>[8]</sup> 和 FastPCR v3.6 等软件,确定各个基因序列。

表 1 圆斑星鲈和条斑星鲈线粒体扩增引物及 PCR 条件

Tab. 1 Oligo nucleotide primers for amplification and sequence *Verasper variegatus* and *V. moseri* mtDNA

引物 Primer	正向引物序列 Forward primer 5' -3'	反向引物序列 Reverse primer 5' -3'	产物大小 /bp Length	扩增条件 PCR profile (35cycles)
vm-CR-16S	gtccagtgtcatgcaatggat	aggatgtctgatccaacatc	4 300	96°C 1' -60°C 1' -72°C 3'
vv-ND5-Cytb	cgcattttctacatccagccag	gatgcgcrrttggcatgratgct	1 750	96°C 1' -60°C 1' -72°C 2'
v-16S-COI	cgcctgtttacaaaacatcgcctc	ggtttcgctgytyagtagyattg	4 100	96°C 1' -60°C 1' -72°C 3'
v-COI-ND5	agccggaatagtgggacagggcc	ctggctggatgtagaanaatgcg	7 300	96°C 1' -60°C 1' -72°C 5'
v-ND5-Cytb	cgcattttctacatccagccag	gatgcgcrrttggcatgratgct	1 750	96°C 1' -60°C 1' -72°C 2'
v-ND5	actctagcaccatagtcgttgc	agtatggctttgaagaaggcgtg	270	96°C 1' -60°C 1' -72°C 1'
v-Cytb-CR	ctccctgcccctctaatatct	actgatgagtgtcgtgttcggt	1 500	96°C 1' -60°C 1' -72°C 2'
vv-CR-16S	acgcagtgttcatacagatacgc	aggatgtctgatccaacatc	4 100	96°C 1' -60°C 1' -72°C 4'

注: vv 和 vm 分别代表圆斑星鲈和条斑星鲈的特异性引物; v- 为圆斑星鲈和条斑星鲈的共用引物。

Note: vv and vm- represent primers for *V. variegates* and *V. moseri*, respectively; v- primer for both *V. moseri* and *V. variegates*.

### 1.2 相关种类线粒体基因序列获取

为了研究不同线粒体基因的进化信息水平,我们从 GenBank 中下载鲈形目 (Pleuronectiformes)、鲈形目 (Perciformes)、鲤形目 (Cypriniformes)、鲑形目 (Salmoniformes)、鲇形目 (Siluriformes)、金眼鲷目 (Beryciformes)、鲅目 (Lophiiformes) 7 个目共已知分类关系的 27 种鱼以及黄牛 (*Bos taurus*)、红原鸡 (*Gallus gallus gallus*) 和马达加斯加彩蛙 (*Mantella madagascariensis*) 的线粒体蛋白质编码基因和 12S rRNA、16S rRNA 基因序列 (表 2)。

### 1.3 数据整理分析

将 30 种脊椎动物的线粒体 13 个编码基因的氨基酸序列和核苷酸序列以及 12S rRNA、16S rRNA 基因核苷酸序列,按基因种类归类。用 ClustalX 1.83<sup>[9]</sup>、MEGA3.0<sup>[10]</sup> 等软件,采用 NJ 法和 MP 法<sup>[11]</sup>,以牛、鸡和蛙作外群,利用线粒体 13 个蛋白编码基因和 12S rRNA、16S rRNA 的核苷酸序列和 13 个蛋白质

的氨基酸序列,分别重建 27 种鱼的系统发育树。在得到的 30 个核苷酸序列树 (NJ 法和 MP 法各 15 个) 和 26 个氨基酸序列树中 (NJ 法和 MP 法各 13 个),分别计算和分析出了各个基因在目、科、属 3 个阶元的分类准确率和系统发育信息。

### 1.4 基因系统发育信息的估算

**1.4.1 在目、科、属水平基因的信息计算** 利用公式: 分类准确率 = 正确分类个体数 / 总个体数 × 100%, 以百分率形式分别统计出每个线粒体基因的核苷酸、氨基酸序列在目、科和属各个水平上分类准确率,准确率越高,表示基因的系统发育信息越好。把 30 种脊椎动物构建所得的 56 个系统树 (MP 树和 NJ 树各 28 个) 与已知分类情况相比较,将分类符合已知情况,且自展值在 51 以上的个体视为正确分类个体,不符合已知分类或自展值低于 50 的视为不正确分类个体。

表 2 用于本研究的鱼类及外群的分类地位及其序列号  
Tab.2 Classification position and array number of samples used in this study

目 Order	科 Family	种 Species	序列号 Acc. No.
鲈形目 Pleuronectiformes	鲆科 Bothidae	牙鲆 <i>Paralichthys olivaceus</i>	AB028664
	鲽科 Pleuronectidae	圆斑星鲽 <i>Verasper variegatus</i>	DQ403797*
		条斑星鲽 <i>Verasper moseri</i>	EF025506*
鲈形目 Perciformes	鲹科 Carangidae	日本竹荚鱼 <i>Trachurus japonicus</i>	AP003092
		大西洋竹荚鱼 <i>Trachurus trachurus</i>	AB108498
		黑尻鲹 <i>Caranx melampygus</i>	AP004445
		刺鲹 <i>Carangoides armatus</i>	AP004444
		鲤鱼 <i>Cyprinus carpio</i>	X61010
鲤形目 Cypriniformes	鲤科 Cyprinidae	白鲫 <i>Carassius cuvieri</i>	AB045144
	鳅科 Cobitidae	台湾缨口鳅 <i>Crossostoma lacustre</i>	M91245
		日本八须鳅 <i>Lefua echigonia</i>	AB054126
鲇形目 Siluriformes	鲶科 Ictaluridae	斑点叉尾鲶 <i>Ictalurus punctatus</i>	AF482987
	鲿科 Bagridae	拟鲿 <i>Pseudobagrus tokiensis</i>	AB054127
	美鲇科 Callichthyidae	<i>Corydoras rabauti</i>	AB054128
	深海蛙科 Bathylagidae	深海蛙 <i>Bathylagus ochotensis</i>	AP004101
	银鱼科 Salangidae	小齿日本银鱼 <i>Salangichys microdon</i>	AP004109
鲑形目 Salmoniformes	鲑科 Salmonidae	虹鳟 <i>Oncorhynchus mykiss</i>	L29771
		大西洋鲑 <i>Salmo salar</i>	U12143
		白鲑 <i>Coregonus lavaretus</i>	AB034824
		美洲红点鲑 <i>Salvelinus fontinalis</i>	AF154850
		北极红点鲑 <i>Salvelinus alpinus</i>	AF154851
		红金眼鲷 <i>Beryx splendans</i>	AP002939
		十指金眼鲷 <i>Beryx decadactylus</i>	AP004430
金眼鲷目 Beryciformes	金眼鲷科 Berytidae	高体金眼鲷 <i>Anoplogaster cornuta</i>	AP004425
	高体金眼鲷科 Anoplogasteridae	阿部单棘躄鱼 <i>Chaunax abei</i>	AP004415
鲛鱈目 Lophiiformes	单棘躄鱼科 Chaunacidae	土佐单棘躄鱼 <i>Chaunax tosaensis</i>	AP004416
	黑角鲛鱈科 Melanocetidae	黑角鲛鱈 <i>Melanocetus murrayi</i>	AP004418
无尾目 Anura	树蛙科 Rhacophoridae	马达加斯加彩蛙 <i>Mantella madagascariensis</i>	AB212225
鸡形目 Galliformes	雉科 Phasianidae	红原鸡 <i>Gallus gallus gallus</i>	AP003322
反刍亚目 Ruminantia	牛科 Bovidae	黄牛 <i>Bos taurus</i>	AY526085

\* 为本研究测得序列。

\* Sequences determined in this study.

**1.4.2 总体基因系统发育信息的计算** 由公式<sup>[12]</sup>  
 $d_i = 2[\min(q_r, q_t) - r] + |q_r - q_t|$  计算出各个基因的核苷酸序列和氨基酸序列在目、科和属总体水平上的基因系统发育信息,以及在核苷酸序列和氨基酸序列总体水平上的基因系统发育信息。根据所计算出的值,确定线粒体各个基因的系统发育信息水平。公式中,  $d_i$  表示新建的系统树与真实系统树距离值,该值越小,表示越接近真实的系统树,基

因的系统发育信息越好。 $q_r$  表示重新构建系统树的分支数,  $q_t$  表示真实系统树的分支数,  $r$  表示重建系统树和真实系统树相同的分支数。

## 2 结果与分析

### 2.1 圆斑星鲽和条斑星鲽线粒体基因组序列

圆斑星鲽和条斑星鲽线粒体基因组全序列长度分别为 17 273bp 和 17 588bp, 序列已经提交到 GenBank 数据库中, 序列号分别为 DQ403797 和

EF025506。这两个基因组均包含 13 个蛋白质编码基因、1 个 12S rRNA、1 个 16S rRNA 基因、22 个 tRNA 基因和一个控制区。它们的核苷酸组成、基

因结构和排列顺序同其他硬骨鱼类线粒体基因组基本相似。圆斑星鲈和条斑星鲈线粒体 13 个蛋白质编码基因和 2 个 rRNA 基因的组成情况见表 3。

表 3 圆斑星鲈和条斑星鲈线粒体基因组的基因组成  
Tab. 3 Characteristics of *Verasper variegatus* and *V.moseri* mitochondrial genome

基因 Gene name	核苷酸长度 / bp 'Nucleotide Size	氨基酸序列长度 / aa Sequence size of amino acids	转录链 Transcribed from
12S rRNA	949 (948)		H
16S rRNA	1 716		H
ND1	975	324	H
ND2	1 046	348	H
COI	1 557	518	H
COII	691	230	H
COIII	785	261	H
ND3	349	116	H
ND4L	297	98	H
ND4	1 381	460	H
ND5	1 839	612	H
ND6	522	173	L
Cytb	1 141	380	H

注: 括号内为条斑星鲈数据, 没加括号表示两者相同。

Note: Data in the parentheses represent *V.moseri*.

## 2.2 系统树重建结果

首先把选择的 30 个物种, 根据已知的分类情况确定出期望的标准系统树 (图 1A)。在所得到的 56 个系统树中, 与期望的标准系统树最相近的系统树有 16S rRNA 和 ND2 (图 1B 和 C), 与期望的标准系统树差距最大的的系统树为 ATPase8 (图 1D)。

## 2.3 系统树准确率分析

**2.3.1 不同 mt 基因在目阶元的分类准确率** 在目一级分类阶元, 当用线粒体 13 蛋白编码基因和 12S rRNA、16S rRNA 基因的核苷酸序列构建系统关系树时, 12S rRNA、16S rRNA、COII、ND4、ND5 和 ND6 分类准确率最高, 都是 100%, COI、COIII、Cytb、ND1 和 ND2 次之, 准确率在 90% 以上, ND3、ATPase6、ATPase8、ND4L 最差, 准确率在 87%~90%。在用氨基酸序列建树时, ND4 和 ND5 最好, 准确率是 100%, ND1、ND2、Cytb、ND6、COI、COII 和 ND4L 次之, 准确率在 87% 以上, ND3、ATPase6、COIII 和 ATPase8, 最差, 准确率为 77%~83% (图 2A 和 B)。

**2.3.2 不同 mt 基因在科阶元的分类准确率** 根

据用 15 个基因的核苷酸序列和 13 个基因的氨基酸序列所构建的系统树中科之间准确率的计算结果, 同样可以将 15 个基因分为好、中、差 3 组 (图 2 中的 C 和 D): 用核苷酸序列时, 好的一组有 12S rRNA、16S rRNA、ND2 和 ND6, 准确率为 100%; 中等的一组有 Cytb、ATPase6、ATPase8、COI、COII、COIII、ND3、ND4 和 ND5; 差的一组有 ND1 和 ND4L。用氨基酸序列时, Cytb 和 ND2 最好, 准确率 100%, COI、ND4、ND5、ND6、ATPase8、COII 和 ND4L 次之, COIII、ND3、ND1、ATPase6 最差。

**2.3.3 不同 mt 基因在属阶元的分类准确率** 在属之间, 用核苷酸序列分类时 15 个基因都准确, 用氨基酸序列分类时 (图 2E), 只有 COII、ND4L、ND1、ND3 不完全准确, 其准确率分别为 97%、97%、93%、93%。

**2.3.4 核苷酸和氨基酸序列的  $d_i$  值** 在核苷酸水平上, 对目、科和属 3 个阶元的分类进行综合计算后, 在核苷酸水平上 (图 3 A) 线粒体 15 个基因的  $d_i$  值由低至高依次为 16S rRNA、ND6、Cytb、ND2、12S rRNA、ND4、ND5、COII、COIII、COI、ND3、ND1、ATPase6、ND4L、ATPase8。

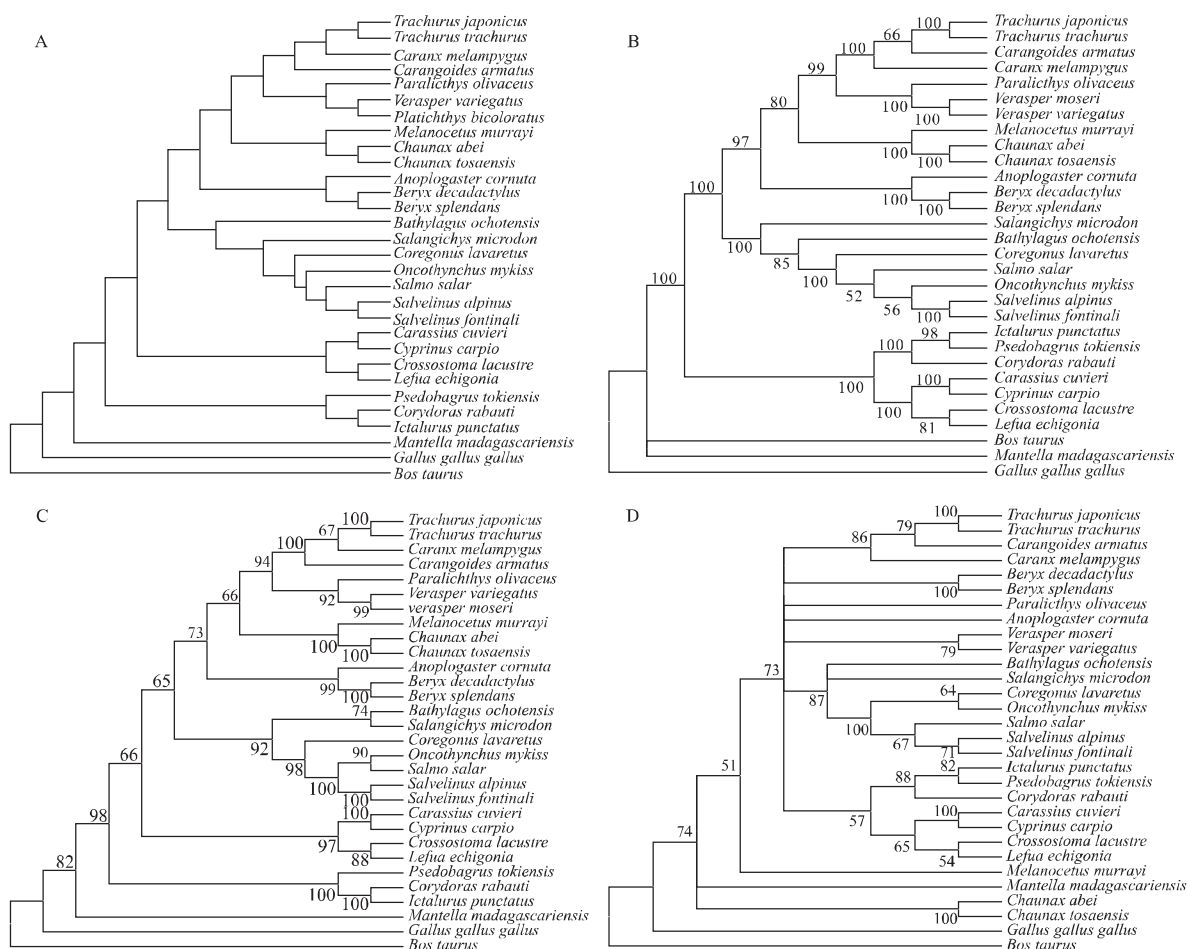


图1 系统发育树

A: 期望的系统发育树, B: 16S rRNA 核苷酸序列构建的系统发育树, C: ND2 核苷酸序列构建的系统发育树, D: ATPase8 核苷酸序列构建的系统发育树.

Fig.1 Phylogeny trees constructed in present study.

A: Expected phylogeny trees; B: Phylogeny trees of 16S rRNA sequence; C: Phylogeny trees of ND2 sequence; D: Phylogeny trees of ATPase8 sequence.

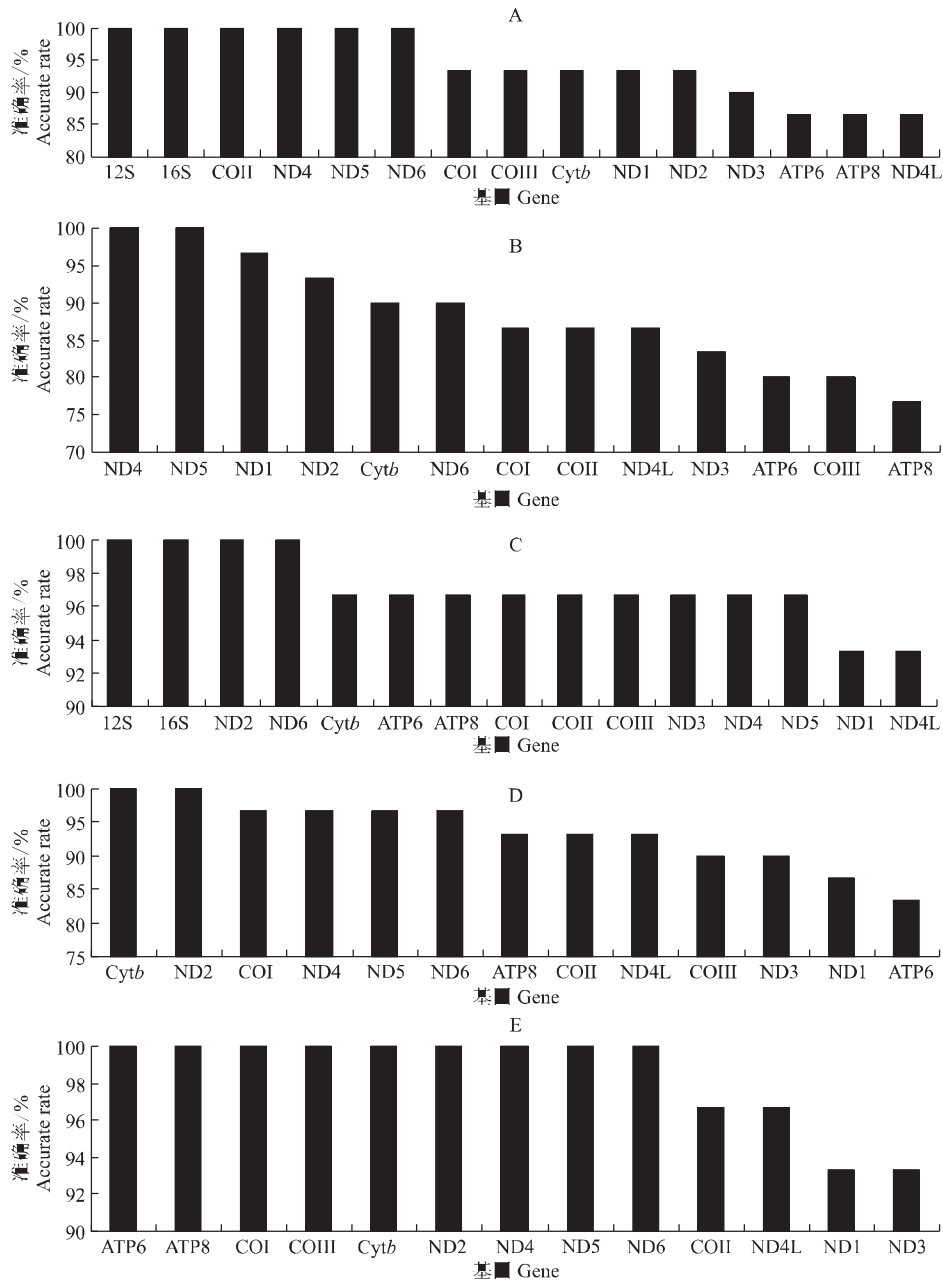
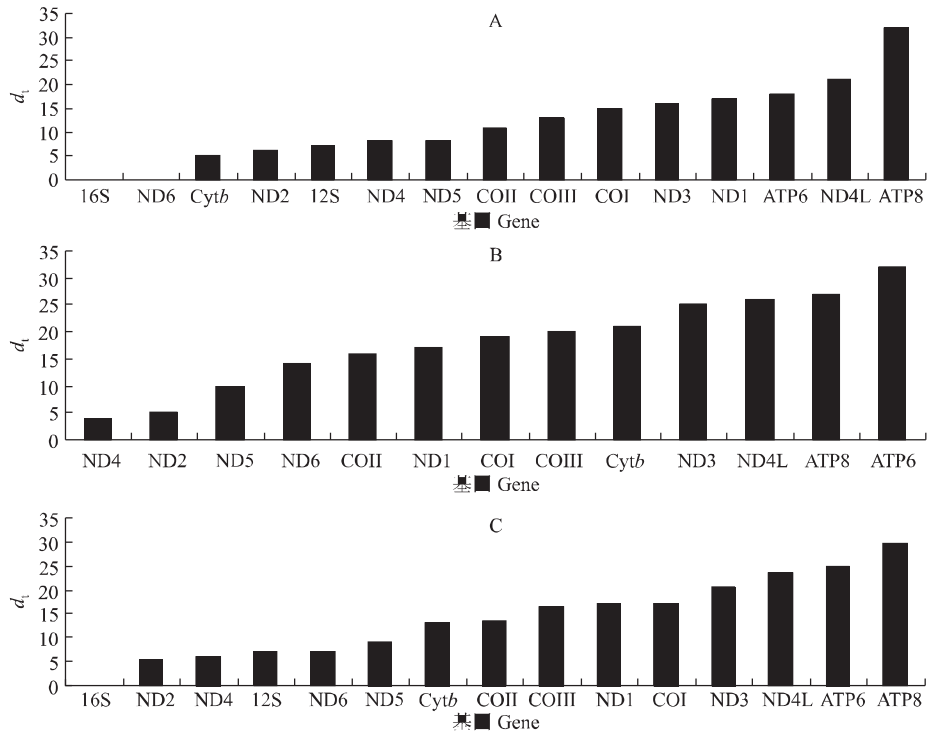


图2 不同线粒体基因的分类准确率

A: 核苷酸序列在目一阶元的准确率分析; B: 氨基酸序列在目一阶元的准确率分析; C: 核苷酸序列在科一阶元的准确率分析; D: 氨基酸序列在科一阶元的准确率分析; E: 氨基酸序列在属一阶元的准确率分析.

Fig.2 Accuracy of phylogenetic analysis using different mt genes

A: Accurate rate analysis of nucleotide sequence in order; B: Accurate rate analysis of amino acid sequence in order; C: Accurate rate analysis of nucleotide sequence in family; D: Accurate rate analysis of amino acid sequence in family; E: Accurate rate analysis of amino acid sequence in genera.

图3 不同线粒体基因的  $d_1$  值

A: 核苷酸序列的  $d_1$  值分析; B: 氨基酸序列的  $d_1$  值分析; C: 核苷酸和氨基酸序列的  $d_1$  值分析.

Fig.3 The  $d_1$  values of different mt genes.

A:  $d_1$  analysis of nucleotide sequence; B:  $d_1$  analysis of amino acid sequence; C:  $d_1$  analysis of nucleotide and amino acid sequence.

在氨基酸水平上(图3 B),线粒体13个蛋白编码基因的  $d_1$  值由低至高排列顺序为 ND4、ND2、ND5、ND6、COII、ND1、COI、COIII、Cytb、ND3、ND4L、ATPase8、ATPase6。

将核苷酸序列和氨基酸序列的  $d_1$  值放在一起统计计算综合  $d_1$  值,则15个基因的  $d_1$  值由低至高排列顺序为 16S rRNA、ND2、ND4、12S rRNA、ND6、ND5、Cytb、COII、COIII、ND1、COI、ND3、ND4L、ATPase6、ATPase8。由图3 C 可以将它们分为好、中、差3个组,好的一组包括 16S rRNA、ND2、ND4、12S rRNA、ND6、ND5,  $d_1$  都小于10,其中16S rRNA 最好,达100%,中间一组有 Cytb、COII、COIII、ND1、COI,  $d_1$  都小于20,剩下的 ND3、ND4L、ATPase6、ATPase8  $d_1$  都大于20,为较差的一组。

**2.3.4 基因系统发育信息的比较** 综上所述,可以看出,在目阶元,用核苷酸序列时,12S rRNA、16S rRNA、COII、ND4、ND5 和 ND6 最好,ND3、ATPase6、ATPase8、ND4L 最差(表3);在用氨基酸序列分类时,ND4 和 ND5 最好,ND3、ATP6、

COIII 和 ATP8 最差。在科阶元上,用核苷酸序列时,12S rRNA、16S rRNA、ND2 和 ND6 最好, Cytb、ATPase6、ATPase8、COI、COII、COIII、ND3、ND4 和 ND5 次之,ND1 和 ND4L 最差;用氨基酸序列分类时, Cytb 和 ND2 最好,COI、ND4、ND5、ND6、ATPase8、COII 和 ND4L 次之,COIII、ND3、ND1、ATPase6 最差。

### 3 讨论

#### 3.1 鱼类线粒体不同基因的进化及其具有的系统发育信息

鱼类线粒体 DNA 的不同区域的进化速率不同,线粒体各个基因的系统发育信息表现也有所不同<sup>[2]</sup>。一般认为在进行遗传分析时,应选择不同的区域进行不同时间尺度的进化分析。如 D-loop 区的进化速度最快,一般用于种内种群间的分化研究。2 个 rRNA 基因进化速度最慢,常用于种或种以上水平的检测,而 Cytb 和 ND5 基因进化速度适中,适合于从种间到种内遗传差异分析。

本实验中,综合  $d_i$  值的结果,与 Zardoya 等<sup>[3]</sup> 的结论基本一致,但是也有一定的差异(表4),其原因可能是,Zardoya 等<sup>[3]</sup> 研究的是脊椎动物门,范围大,并且得到的只是纲这个阶元的系统发育信息的比较。而本实验是在硬骨鱼纲中做研究,从目、科和属这3个阶元分别做研究,比前者更为细致。在本实验所研究的鱼类中,15个线粒体基因  $d_i$  值由低到高的排列顺序为 16S rRNA、ND2、ND4、12S rRNA、ND6、ND5、Cytb、COII、COIII、ND1、COI、

ND3、ND4L、ATPase6、ATPase8。好的一组包括 16S rRNA、ND2、ND4、12S rRNA、ND6、ND5,中等的有 Cytb、COII、COIII、ND1、COI,剩下的 ND3、ND4L、ATPase6、ATPase8 为较差的一组。Zardoya 等<sup>[3]</sup> 在研究四足类动物和哺乳动物时认为具有良好的系统发育信息的 Cytb 和 COI,在鱼类仅表现为具有中等系统发育信息,Zardoya 等认为 ND6 属于中等类型,而本结果则认为它们属于较好的类型。

表4 基因系统发育信息的比较  
Tab. 4 Comparison of phylogenetic information

基因 Gene	本实验中综合 $d_i$ 值 $d_i$ in present study	$d_i$ of Zardoya	本研究中目阶元的核苷酸的准确率 Accurate rate of nt in order in present study	目阶元的准确率 Accurate rate of nt in Miya
16 S rRNA	好 Good	—	好 Good	—
ND2	好 Good	好 Good	中等 Medium	中等 Medium
ND4	好 Good	好 Good	好 Good	非常好 Very good
12 S rRNA	好 Good	—	好 Good	—
ND6	好 Good	中等 Medium	好 Good	非常差 Very poor
ND5	好 Good	好 Good	好 Good	非常好 Very good
Cytb	中等 Medium	好 Good	中等 Medium	好 Good
COII	中等 Medium	中等 Medium	好 Good	好 Good
COIII	中等 Medium	中等 Medium	中等 Medium	非常好 Very good
ND1	中等 Medium	中等 Medium	中等 Medium	差 Poor
COI	中等 Medium	好 Good	差 Poor	非常好 Very good
ND3	差 Poor	差 Poor	中等 Medium	中等 Medium
ND4L	差 Poor	差 Poor	差 Poor	非常差 Very poor
ATPase6	差 Poor	差 Poor	差 Poor	差 Poor
ATPase8	差 Poor	差 Poor	差 Poor	非常差 Very poor

日本学者 Miya<sup>[5]</sup> 对8种硬骨鱼进行了研究,即目一阶元的研究,而本实验目一阶元的结果与 Miya 的研究结论出现不同点(表4)。本实验中,ND6 的系统发育信息属于好的类型,而在 Miya 的研究结果中,它属于非常差的类型;我们认为 Cytb、COIII 和 ND1 是中等类型,而 Miya 将它们依次归为好、非常好、差的类型;我们认为 COI 是差的类型,但 Miya 认为它是非常好的类型。

彭作刚等<sup>[13]</sup> 用 Cytb 基因序列研究中国鲈形目鱼类的系统发育时,发现细胞色素 b 不太适合作为鲈形目鱼类高级分类阶元系统发育重建的分子标记;Farias 等<sup>[14]</sup> 也指出 Cytb 在丽鱼科系统发育重建中存在局限性。本实验结论与其相符合。在本实验中,用氨基酸序列进行目的分类及用核苷酸序列、氨基酸序列在科一级进行分类时,Cytb 均未

被归到好的一组中。

Miya 和 Nishida<sup>[15-16]</sup> 认为,16S rRNA 和 12S rRNA 不能进行明确的比对,所以不能用于系统发育信息的分析。但 16S rRNA 和 12S rRNA 既有保守性,又有高变性,一般认为适用于种以上水平的变异分析<sup>[17]</sup>。在鱼类中,线粒体 16S rRNA 和 12S rRNA 基因序列被广泛用于研究不同阶元的系统发育关系<sup>[18-19]</sup>。在本实验中,16S rRNA 和 12S rRNA 在各个阶元都有良好的系统发育信息。

### 3.2 关于种内的鉴定标准

本研究在目间、科间、属间得出了线粒体各个基因的系统信息差异,但种内却没有得出明显的区别,这是因为本研究所使用的个体数量还十分有限,主要是选择几个不同的目,在每个目中选择几个科,在每个科中选择几个属,每个属中选择1个种,个别属



选择 2 个种。只能研究目、科和属间各个基因的情况。要想研究各个基因在种间和种内群体间系统分化的表现情况,必须获得同一个属的多个种、同一个种的不同亚种和不同地理种群的大量基因序列。

### 3.3 关于基因系统发育信息的标准

本实验从 4 个方面分析了硬骨鱼 15 个 mtDNA 基因所包含的系统发育信息情况,判别的标准是根据实际情况而定的。基本是根据分类准确率的高低和总体趋势,大概分为好、中、差 3 个水平,好、中、差是相对的。在研究目、科和属间的各个基因的系统信息差异时,计算方法是采用分类准确百分比。而在研究各个基因的核苷酸序列和氨基酸序列在目、科和属总体水平上以及在核苷酸序列和氨基酸序列总体水平上的分类情况时,因构建的系统树分支数与正确树的分支数目发生变化,可以采用计算  $d_i$  值的方法来估算。而在研究目、科、属各个阶元时,由于构建的树与真实树的分支数并不发生变化,所以只能用准确率来计算。

### 3.4 用线粒体全基因组序列和单个基因序列的准确率

在研究脊椎动物高级阶元的系统发育时,是利用 mtDNA 的单个基因序列还是采用全部基因序列连接起来分析,目前有两种观点。一种认为某些单个基因序列或其部分片段,可以代表整个 mtDNA 的进化水平和系统发育过程<sup>[3,20]</sup>;另一种观点认为单个基因或片段所包含的系统信息太少,不足以反映整个生物的分子进化状况,而应该采用全部编码区序列或基因序列连接起来分析更为可靠<sup>[5,21]</sup>。由于鱼类 mtDNA 不同基因的进化速度不同,即每个基因所包含的系统发育信息量不同,在进行不同分类阶元的研究时,应根据不同基因所包含的信息情况,选择最佳的基因,不宜盲目地选择全基因组序列,避免那些具有较差信息基因的干扰,同时避免分析工作和数据处理的繁琐。

#### 参考文献:

- [1] 郭新红,刘少军,刘巧,等. 鱼类线粒体 DNA 研究新进展[J]. 遗传学报,2004,31(9): 983-1000.
- [2] 赵凯. 鱼类线粒体 DNA (mtDNA) 及其在分子系统学中的应用[J]. 青海大学学报:自然科学版,2006,24(2): 49-53.
- [3] Zardoya R, Meyer A. Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates[J]. *Mol Biol Evol*, 1996, 13 (7): 933-942.
- [4] 余先觉,周敏,李康,等. 中国淡水鱼类染色体[M]. 北京: 科学出版社,1989.
- [5] Miya M, Nishida M. Use of mitogenomic information in teleostean molecular phylogenetics: a tree-based exploration under the maximum-parsimony optimality criterion[J]. *Mol Phylogenet Evol*, 2000, 17: 437-455.
- [6] 萨姆布鲁克 J, 拉塞尔 D W. 分子克隆实验指南: 第 3 版[M]. 北京: 科学出版社,2002.
- [7] 高天翔,陈四清,刘进贤,等. 圆斑星鲷与条斑星鲷线粒 DNA 部分片段的序列比较研究[J]. 高技术通讯,2004,14(增刊): 329-334.
- [8] 吕国庆,李思发. 鱼类线粒体 DNA 多态研究和应用进展[J]. 中国水产科学,1998,5(3): 94-103.
- [9] Thompson J D, Gibson T J, Plewniak F. The Clustal X windows interface: flexible strategies for multiple sequences alignment aided by quality analysis tools [J]. *Nucleic Acids Res*, 1997, 25(4): 4876-4882.
- [10] Kumar S, Tamura K, Nei M. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment [J]. *Briefings in Bioinformatics*, 2004, 5: 150-163.
- [11] Saitou N, Nei M. The neighbor-joining method: A new method of reconstruction phylogenetic trees [J]. *Mol Biol Evol*, 1987, 4: 406-425.
- [12] Russo C A M, Takezaki N, Nei M. Efficiencies of different genes and different tree-building methods in recovering a known vertebrate phylogeny[J]. *Mol Biol Evol*, 1996, 13(3): 525-536.
- [13] 彭作刚,张耀光,何舜平,等. 从细胞色素 b 基因序列变异分析中国鲷形目鱼类的系统发育[J]. 遗传学报,2005,32(2): 145-154.
- [14] Farias I P, Orti G, Sampaio I, et al. The cytochrome b gene as a phylogenetic marker: The limits of resolution for analyzing relationships among cichlid fishes [J]. *J Mol Evol*, 2001, 53: 89-103.
- [15] Miya M, Nishida M. Molecular phylogeny and evolution of the deep-sea fish genus *Sternoptyx* [J]. *Mol Phylogenetic Evol*, 1998, 10: 11-22.
- [16] Miya M, Nishida M. Molecular phylogeny of the deepsea fish genus *Gonostoma* (Teleostei: Stomiiformes): Two paraphyletic clades and resurrection of *Sigmops* [J]. *Copeia*, 2000: 378-389.
- [17] 丁少雄,王颖汇,王军,等. 基于 16S rDNA 部分序列探讨中国近海 30 种石斑鱼类的分子系统进化关系[J]. 动物学报,

- 2006, 52(3): 504-513.
- [18] Kuo C H, Huang S, Lee S C. Phylogeny of hagfish based on the mitochondrial 16S rRNA gene[J]. *Mol Phylogenet Evol*, 2003, 28(3): 448-457.
- [19] 李春枝, 张邦杰, 李本旺, 等. 尖塘鳢属鱼类线粒体 12SrRNA 基因序列分析 [J]. *生态科学*, 2006, 25(5): 433-436.
- [20] Meyer. Shortcomings of the cytochrome b gene as a molecular marker [J]. *Trends Ecol Evol*, 1994, 9: 278-280.
- [21] Miya M, Kawaguchi A, Nishida M. Mitogenomic exploration of higher teleostean phylogenies: a case study for moderate-scale evolutionary genomics with 38 newly determined complete mitochondrial dna sequences [J]. *Mol Biol Evo*, 2001, 18(11): 1993-2009.

## Informative efficiencies of mitochondrial genes in phylogenetic analysis of teleostean

CHEN Shu-jun<sup>1,2</sup>, HE Chong-bo<sup>2</sup>, MU Yun-lei<sup>2</sup>, LIU Wei-dong<sup>2</sup>, ZHOU Zunchun<sup>2</sup>, GAO Xiang-gang<sup>2</sup>, CONG Lin-lin<sup>1,2</sup>

(1. College of Life Science, Liaoning Normal University, Dalian 116029, China; 2. Liaoning Ocean and Fisheries Science Research Institute, Liaoning Key Laboratory of Marine Fishery Molecular Biology, Liaoning Open Laboratory of Applied Marine Biotechnology, Dalian 116023, China.)

**Abstract:** A large number of studies in evolutionary biology utilize phylogenetic information obtained from mitochondrial DNA. Researchers place trust in this molecule and expect it generally to be a reliable marker for addressing questions ranging from population genetics to phylogenies among teleostean lineages. Yet, regardless of the phylogenetic method and weighting treatment, individual mitochondrial genes might potentially produce misleading evolutionary inferences and hence might not constitute an adequate representation neither of the entire mitochondrial genome nor of the evolutionary history of the organisms from which they are derived. We investigated the performance of all mitochondrial protein-coding genes and ribosomal RNA genes to recover the expected phylogenies of teleostean. We sequenced the complete mitochondrial genome of *Verasper variegatus* and *Verasper moseri* using PCR direct sequencing method and reconstructed the phylogenetic trees by Neighbor-Joining (NJ) and Maximum Parsimony (MP) method with the nucleotide and amino acid sequences of all the protein-coding genes and ribosomal RNA genes of teleostean species download from GenBank, using *Bos Taurus*, *Gallus gallus gallus*, *Mantella madagascariensis* as outgroups and then calculated and analyzed the accuracy of the phylogenetic trees reconstructed. The results showed that the mitochondrial protein-coding genes and ribosomal RNA genes can be roughly classified into three groups of good (16S rRNA, ND2, ND4, 12S rRNA, ND6 and ND5), medium (*Cytb*, COII, COIII, ND1 and COI), and poor (ND3, ND4L, ATPase6 and ATPase8) phylogenetic performers in recovering these expected trees among the seven orders of teleostean fishes. At the level of Order, 12S rRNA, 16S rRNA, COII, ND4, ND5 and ND6 are of good groups, COI, COIII, *Cytb*, ND1 and ND2 are of medium, and ND3, ATPase6, ATPase8 and ND4L are of poor when using nucleotide sequence; ND4 and ND5 belong to good groups; ND1, ND2, *Cytb*, ND6, COI, COII and ND4L are medium, and ND3, ATPase6, COIII and ATPase8 are poor using amino acid sequence. When aligned by nucleotide sequence in family, 12S rRNA, 16S rRNA, ND2 and ND6 are in good groups; *Cyt b* and ND2 are in good group when aligned by amino acid sequence; in genera, all sequences are in good groups when aligned by nucleotide sequence; COII, ND4L, ND1 and ND3 are not in good groups when aligned by amino acid sequence. [Journal of Fishery Sciences of China, 2008, 15(1): 12-21]

**Key words:** *Verasper variegatus*; *Verasper moseri*; teleostean; mtDNA; phylogeny

**Corresponding author:** HE Chong-bo. E-mail: [hechongbo@hotmail.com](mailto:hechongbo@hotmail.com)