S7 核糖体蛋白基因序列变异及其在低等鲤科鱼类中的系统发育意义

王绪桢 何舜平* 陈宜瑜

(中国科学院水生生物研究所, 武汉 430072. * 联系人, E-mail: clad@ihb.ac.cn)

摘要 鲤科是世界鱼类中最大的科, 有 210 余属 2010 种. 为研究其系统发育关系需要筛选合适的 DNA 标记. 将 S7 核糖体蛋白基因用作遗传标记进行鲤科鱼类的系统发育分析. 通过 PCR 方法, 扩增了长度为 602 bp 的胭脂鱼以及长度为 655~859 bp 的 16 种鲤科鱼类的 S7 基因内含子 1 序列. 序列排列得到 925 个排列位点, 其中信息位点 499 个, 占全部位点的 54%. 结果表明, 鲤科鱼类 S7 基因内含子 1 序列具有丰富的信息位点, 并且在亲缘关系不同的物种间存在显著的序列差异. 基于 S7 基因内含子 1 序列的 NJ(neighbor-joining)和 MP(most-parsimony)分支树经 1000 次重复抽样检验后, 节点的自展支持率普遍高于以细胞色素 b 及线粒体控制区(d-loop)基因为遗传标记时所得到的分支树. 因此, S7 基因内含子 1 作为遗传标记在鲤科系统发育研究中的分辨率是比较高的, 它是一个适合鲤科内亚科水平系统发育研究的有用的分子遗传标记. 但 S7 基因内含子 1 是否适用于鲤形目科间或科以上水平的分子系统学分析, 有待进一步研究.

关键词 S7 核糖体蛋白基因 系统发育 鲤科 序列变异 内含子

鲤科是世界鱼类中最大的科, 由于其种类庞杂 (210 余属 2010 种), 对其进行系统发育研究一直颇具挑战性. 鲤科系统发育研究最初主要是基于形态学特征. 低等鲤科鱼类(鲴亚科和鲃亚科)是老第三纪类群的后裔^[1], 长期以来相似生活环境造成的平行演化使得它们在外部及骨骼形态上都表现出较高趋同性; 加上鲤科鱼类化石数据不足^[1], 用形态特征分析来处理这些外形相似物种间的亲缘关系时, 因可利用的同源性状缺乏而面临极大困难. 虽然目前也积累了一些形态特征分析资料^[2,3], 但低等鲤科鱼类的系统发育仍有许多不能解决的问题^[3,4]; 整个鲤科系统发育研究进展也因此受到较大制约. 近年来迅速发展的 DNA 序列分析为系统发育研究提供了一条较为客观的途径. 鲤科鱼类的分子系统学研究也有报道, 研究的基因集中在线粒体基因^[4~6]. 细胞色素 b 的研究表明低等鲤科鱼类中雅罗鱼亚科和鲴亚科不是单系群^[4], 但这两个亚科所包含类群间的相互关系仍不能明确界定. 因此目前的研究结果尚不足以解决低等鲤科鱼类的系统发育问题, 寻找新的 DNA 序列是解决问题的关键之一. 由于不同基因的进化历史存在差异, 采用多基因序列分析也是从基因树过渡到物种树的必然途径.

由一系列内含子和外显子串联而成, 其内含子中包含有核仁小分子 RNA(snoRNA)的编码序列. 研究表明, 除内含子序列长度外, S7 基因组结构在关系较远的脊椎动物类群中均表现出较高相似性; 而且鱼类 S7 基因明显较两栖类和人类结构紧凑而被用作宿主基因内含子编码功能研究的有效模型^[7]. 目前尚未见到关于 S7 基因的进化规律以及将其用于系统发育研究的报道. 本研究将 S7 基因用作系统发育研究的遗传标记, 探讨其用于鲤科鱼类系统演化研究的可行性.

由于内含子承受的进化压力小于相应的外显子, 内含子序列无论是在种间还是在种内都较外显子序列具有更明显的遗传多态现象. 基于内含子和外显子的系统发育分析的比较结果显示, 内含子序列可以用来对近缘物种进行进化分析^[8]. 同样基于内含子序列的分子进化分析可能会提供低等鲤科类群丰富的进化信息. 为低等鲤科鱼类的系统发育研究找到好的内含子遗传标记也是本研究的主要目的. 鱼类 S7 基因包括 7 个外显子和 6 个内含子, 其中内含子 1 的序列长度比较适中而容易得到全序列. 本研究特选择了 S7 基因内含子 1, 通过分析其序列为低等鲤科鱼类的系统发育研究寻找新的线索.

S7 核糖体蛋白基因是不连续的蛋白质编码基因, <http://www.scichina.com/doi/10.1360/scb2002-47-14-1089>

1 材料与方法

(i) 研究材料. 研究样本是鲤科部分亚科, 特别是鲴亚科、鲃亚科以及雅罗鱼亚科的一些代表种类, 具体种类的名称及其来源见表 1. 其中胭脂鱼在本研究中作为外类群. 所有标本都用 95%酒精保存.

(ii) DNA 的提取、PCR 扩增和序列测定. 用酚抽提法从肌肉组织中提取总 DNA. 依据 S7 基因内含子 1 两端外显子的保守区域设计了一对引物: 5'-TGGCCTCTTCCTTGGCCGTC-3' 和 5'-AACTCG-TCTGGCTTTTCGCC-3' [9]. 以总 DNA 为模板进行 PCR 扩增, 条件为 94℃预变性 3 min, 然后 94℃变性 30 s, 60℃退火 30 s, 72℃延伸 1 min 30 s, 35 个循环后, 72℃延伸 8 min, 回收长度为 900 bp 的目的条带. DNA 序列由上海联合基因公司进行测定.

(iii) 系统发育分析. 应用 Clustal X 软件^[10]进行 DNA 序列排定并人工校正. 使用 MEGA 2.1 软件^[11]中的 NJ(neighbor-joining)和 MP(most-parsimony)程序以及 PHYLIP^[12]软件包中的 DNAML 程序来进行系统发育分析.

在 NJ 分析中, 遗传距离模型选择 Kimura 二参数模型, 对于序列中的插入/缺失位点则采用 Pairwise Deletion. 而 MP 分支树的搜寻运用的是 Close-neighbor-interchange (CNI)程序, 同时采用了以下参数: 序列中的插入/缺失位点作为碱基取代位点对待并视为信息位点; 对所有位点一致性加权. 进行 ML (maximum-likelihood)分析时所采用的参数为转换/颠

换比 2.0; 采用经验的碱基频率以及采用一种类型的取代比率. NJ 和 MP 分支树可信度的测试使用 MEGA 软件中的自展检验, 经 1000 次重复抽样检验可得到分支树节点支持率.

2 研究结果

2.1 S7 基因内含子 1 的序列

通过 PCR 扩增, 在鲤科鱼类中获得了 655 ~ 859 bp 长度不等的内含子 1 序列. 其中鲴的内含子 1 序列最短, 而滇西低线鱮的最长. 胭脂鱼内含子 1 序列为 602 bp. 这些内含子序列经 Clustal X 软件排列后生成的 925 个核苷酸位点中, 变异位点数为 780 个, 信息位点 499 个, 分别占总位点数的 84%和 54%. 在得到的鲤科鱼类内含子 1 序列中, 各种碱基的百分比组成平均分别为: A 29.7%, T 35.1%, G 19.9%, C 15.3%. 当包括外类群胭脂鱼时, 各种碱基的百分比组成又平均为: A 29.8%, T 34.9%, G 20.0%, C 15.3%. 结果显示 G + C 含量明显小于 A + T 含量.

由全部排列位点的序列变异经 Pairwise Deletion 后计算出 Kimura 双因子遗传距离, 以确定鲤科鱼类间的遗传分化程度. 结果显示鲤科鱼类间的遗传距离值为 0.0232 ~ 0.6659(表 2).

2.2 序列中的保守区

真核生物蛋白质编码基因的内含子中通常包含 TACTAAC 盒序列, 因为功能上这一序列同编码基因内含子的剪切以及外显子的拼接密切相关, 因此它

表 1 研究标本名称及来源

分类阶元	物种	来源	DNA 长度/bp	GenBank 登记号
鲴亚科	胭脂鱼 <i>Myxocyprinus asiaticus</i> (Bleeker)	湖北	602	AY103173
	鲴 <i>Danio nigrofasciatus</i>	NRM33004 ^{a)}	655	AY103157
	波鱼 <i>Rasbora</i> sp.	NRM41684	771	AY103172
	宽鳍鱮 <i>Zacco platypus</i> (Temminck et Schlegel)	福建	760	AY103169
	台湾鱮 <i>Zacco taiwanensis</i> Chen	台湾	762	AY103167
	马口鱼 <i>Opsariichthys bidens</i> Gunther	陕西	749	AY103168
	滇西低线鱮 <i>Barilius barila</i> (Hamilton)	云南	859	AY103162
	细鲫 <i>Aphyocypris</i> sp.	日本	744	AY103171
鲃亚科	小鲃 <i>Puntius</i> sp.	NRM41713	806	AY103161
雅罗鱼亚科	<i>Sawbwa resplendens</i>	NRM51654	838	AY103159
	<i>Rhinichthys atratulus</i>	GN536	804	AY103163
	赤眼鳟 <i>Squaliobarbus curriculus</i> (Richardson)	湖南	810	AY103170
裂腹鱼亚科	裂腹鱼 <i>Schizothorax</i> sp.	云南	841	AY103158
野鲮亚科	野鲮 <i>Labeo senegalensis</i>	Tiassale1516196	821	AY103160
鲃亚科	贝氏鲃 <i>Hemiculter bleekeri</i> Warpachowsky	湖北	801	AY103164
鲃亚科	翘嘴鲃 <i>Culter alburnus</i> Basilewsky	湖北	822	AY103165
	鲃 <i>Xenocypris</i> sp.	湖北	797	AY103166

a) NRM 取自瑞典国家自然历史博物馆馆藏标本

表 2 16 种鲤科鱼类及外类群胭脂鱼种间的 Kimura 双因子遗传距离

编号与分类	遗传距离																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1 <i>Danio nigrofasciatus</i>																	
2 <i>Schizothorax</i> sp.	0.5800																
3 <i>Sawbwa resplendens</i>	0.6431	0.1705															
4 <i>Labeo senegalensis</i>	0.6152	0.2103	0.2508														
5 <i>Puntius</i> sp.	0.6659	0.3515	0.4251	0.4921													
6 <i>Barilius barila</i>	0.6554	0.4209	0.4845	0.4659	0.5955												
7 <i>Rhinichthys atratulus</i>	0.5706	0.4043	0.4495	0.4229	0.5203	0.4600											
8 <i>Hemiculter bleekeri</i>	0.5682	0.3880	0.4437	0.4136	0.4933	0.4092	0.2786										
9 <i>Culter alburnus</i>	0.5354	0.3651	0.4364	0.3809	0.4858	0.4049	0.2360	0.0913									
10 <i>Xenocypris</i> sp.	0.5790	0.3747	0.4406	0.3922	0.4651	0.4261	0.2260	0.1064	0.0858								
11 <i>Zacco taiwanensis</i>	0.5239	0.4129	0.4652	0.4343	0.5309	0.4344	0.3030	0.1778	0.1587	0.1769							
12 <i>Opsariichthys bidens</i>	0.5172	0.4050	0.4721	0.4384	0.5130	0.4262	0.2973	0.1615	0.1428	0.1691	0.0232						
13 <i>Zacco platypus</i>	0.5246	0.4191	0.4816	0.4389	0.5327	0.4313	0.3049	0.1766	0.1543	0.1759	0.0350	0.0261					
14 <i>Squaliobarbus curriculus</i>	0.5538	0.3709	0.4384	0.3827	0.4750	0.3905	0.2413	0.1055	0.0681	0.0859	0.1534	0.1406	0.1539				
15 <i>Aphyocypris</i> sp.	0.5731	0.3730	0.4482	0.4383	0.4873	0.4152	0.2810	0.2131	0.1710	0.1907	0.2437	0.2259	0.2450	0.1878			
16 <i>Rasbora</i> sp.	0.5013	0.5982	0.6622	0.5977	0.6354	0.6206	0.5459	0.5377	0.4787	0.5180	0.5274	0.5314	0.5498	0.5328	0.5584		
17 <i>Myxocyprinus asiaticus</i>	0.7405	0.5832	0.6811	0.6728	0.7308	0.6894	0.5840	0.5328	0.5085	0.5348	0.4838	0.4848	0.4868	0.5378	0.5789	0.8576	

<https://engine.scichina.com/doi/10.1360/csb2002-47-14-1089>

非常保守^[13]. 本研究中, 胭脂鱼和鲤科鱼类的内含子 1 中发现了序列为 GTTAGAT 的保守区域, 它靠近内含子 1 的 5' 端; 其下游约 20 bp 处存在另一保守序列——GGATGGC. 鲤科鱼类 S7 基因序列中上述区域在功能上可能同转录 mRNA 前体的剪接相关, 因此在进化上也相当保守.

2.3 分子系统树的构建

将获得的 925 个核苷酸排列位点用作系统发育的分析位点, 基于这些分析位点得到了表示鲤科鱼类相互关系的 NJ 树、MP 树以及 ML 树. 其中 MP 树共有两棵, 树长均为 1818, 图中给出的是其一致性树(图 1).

图 1 低等鲤科鱼类基于 S7 基因内含子 1 序列的 NJ 树(a)、MP 树的一致性树(b)和 ML 树(c)

节点显示 1000 次重复抽样检验的自展值 <https://engine.scichina.com/doi/pdf/10.1360/02-47-14-1089>

当以胭脂鱼(*M.asiaticus*)为外类群时, NJ 树中显示鲟(*D.nigrofasciatus*), 波鱼(*Rasbora* sp.), 滇西低线鳢(*B.barila*), 小鲃(*Puntius* sp.), *S.resplendens*, 野鲮(*L.senegalensis*), 裂腹鱼(*Schizothorax* sp.)等聚集成一个类群; 而马口鱼(*O.bidens*)、宽鳍鳢(*Z.platypus*)、台湾鳢(*Z.taiwanensis*)和东亚类群中的鲟(*Xenocypris* sp.)、翘嘴鲃(*C.alburnus*)、贝氏鲮(*H.bleekeri*)及赤眼鲮(*S.curriculus*)等聚在一起. MP 树和 ML 树的拓扑结构表现出高度的相似性, 其中鲟, 波鱼, 滇西低线鳢, 小鲃, *S.resplendens*, 野鲮以及裂腹鱼等分布于分支树的基部. 所有结果均显示目前所确认的鲟亚科不但没有聚成一个单系群; 而且它所包括的类群间的亲缘关系存在较大的差异, 表现为在分支树中它们聚成不同的类群.

本研究中得到的 NJ 树和 MP 树经 1000 次重复抽样检验后普遍得到较高的节点自展值(图 1).

3 讨论

3.1 S7 基因序列变异的系统发育意义

脊椎动物中, 鱼类的 S7 基因内含子序列长度比两栖类和人类显著要短^[7]. 因此低等脊椎动物的 S7 基因组结构表现得更为紧凑. 同样地, 在鱼类中低等类群的内含子 1 序列也比特化类群短, 本研究中内含子 1 序列长度从胭脂鱼的 602 bp 变化到鲤科鱼类的 856 bp. 可见 S7 基因内含子序列在 S7 基因进化过程中发生了大量的碱基插入和缺失现象, 而且其变化趋势可能是随物种的进化而不断插入核苷酸序列. 内含子 1 在鲤科鱼类中的序列插入和缺失主要集中在其靠近外显子 1 的一侧.

本研究将 S7 基因用作系统发育的遗传标记进行分子系统学研究. 通过比较 16 种鲤科鱼类的 S7 基因内含子 1 序列全部位点发现: 在近亲缘关系的物种中, 内含子 1 序列具有高度相似性, 序列的相似性可高达 96%; 而亲缘关系较远的物种间内含子 1 序列变异度超过 60%. 内含子 1 在亲缘关系不同的物种间序列同源性的差异表明, 它可以为研究物种之间的进化关系提供有用的信息. 而且 S7 基因用于系统发育研究也有其可靠性, 因为自展检验结果显示基于内含子 1 序列分支图中的分支节点普遍具有较高的支持率(图 1).

G+C 含量与序列的替代速率间存在一定的相关关系: 当序列的 G+C 含量低时, 它的替代速率就

高^[8]. 本研究中 S7 基因内含子 1 的 G+C 含量平均不超过 35.3%, 它表明内含子 1 的进化速率较高. 这是因为鱼类 S7 基因内含子 1 中包含的 U17RNA 编码序列属于异质性拷贝(pseudocopy)^[7], 它可能较其他规范性拷贝(canonical copy)的内含子具有更高的序列变异. 从 S7 基因内含子 1 在胭脂鱼和鲤科鱼类间的序列变异程度来看, 在 925 个核苷酸分析位点中, 信息位点为 499 个, 约占分析位点数的 54%. 讨论 DNA 序列用于分子系统学研究的可行性时, 有研究认为信息位点在序列中的比率接近 60%时不适宜于科及其以上水平的高阶元分子系统学研究^[14]. 此外, 内含子 1 序列在鲤形目科间的长度差异表明 S7 基因在进化过程中有序列的插入和缺失发生, 而且发生插入或丢失的情况可能各不相同. 因此 S7 基因内含子 1 是否适合鲤形目科间或科以上水平的分子系统学研究, 还有待进一步研究.

3.2 S7 基因在鲤科演化历史重建中的应用

基于 S7 基因内含子 1 的分支树表明鲃亚科小鲃属、*Sawbwa* 属和鲟亚科中的鲟属、波鱼属、低线鳢属等种类属于鲤科中起源较早的类群. 这些类群内含子 1 的核苷酸序列相似性普遍较低, 而且来自老第三纪的鲤科鱼类的化石记录表明鲃亚科和鲟亚科种类从第三纪早期就开始广布于欧亚大陆^[1], 因此它们由 DNA 序列变异产生的遗传分歧与其在系统演化上物种分化所经历的漫长历史是相一致的. 本研究结果支持马口鱼属和宽鳍鳢属形成一个单系群, 但它和东亚类群间序列的相似性更高. 在快速进化的内含子序列中, 物种间较高序列同源性通常表明物种分歧时间的相对短暂, 因此马口鱼类群与较晚起源的东亚类群在物种分化上应更为密切. S7 基因内含子 1 的序列差异揭示目前所确认的鲟亚科实际上包含着演化历史上原始的类群以及一些相对特化类群.

目前, 在鲤科系统发育研究中已经使用的分子标记主要是细胞色素 b, 12S RNA 以及 d-loop 等线粒体基因^[4-6]. 其中低等鲤科鱼类的系统发育分析中, 细胞色素 b 和 S7 基因内含子 1 所构建的分子系统树都显示目前确认的鲟亚科包括了鲟类和马口鱼类等类群. 但在这些所包括类群间的相互关系上, 基于细胞色素 b 序列的分支树所显示的相互关系是不确定的, 因为分支树的内部节点并不为自展测试所支持^[4];

相比之下, S7 基因内含子 1 序列却为类群间关系提供了较高的自展支持率(图 1). 由此可见, S7 基因内含子 1 在低等鲤科尤其是鲴亚科鱼类系统发育分析中有更高的分辨率; 因而它适合鲤科鱼类演化关系的推断. 此外, S7 基因内含子 1 序列在鲤科鱼类中有 462 个信息位点, 约占全部分析位点的 50%, 这样高度变异的序列能够提供充足的信息位点用于鲤科系统发育的重建. 因此 S7 基因可以作为鲤科系统发育研究中一个较合适的分子遗传标记, 基于 S7 基因序列的鲤科系统发育的进一步研究可能会为鲤科鱼类系统发育关系提供更明确更合理的推论.

致谢 感谢瑞典自然历史博物馆的方芳博士提供了部分珍贵的研究标本. 本工作为中国科学院 KSCX2-SW-101B 项目以及国家自然科学基金资助项目(批准号: 39830050).

参 考 文 献

1 陈宜瑜. 硬骨鱼纲, 鲤形目. 中国动物志. 中卷. 北京: 科学出版社, 1999. 1~16

2 陈湘彝, 乐佩琦, 林人端. 鲤科的科下分类及其宗系发生关系. 动物分类学报, 1984, 9(4): 424~440

3 Howes G J. Systematics and biogeography: an overview. In: Winfield I J, Nelson J S, eds. Cyprinid Fishes: Systematics, Biology and Exploitation. London: Chapman and Hall, 1991. 1~33

4 何舜平, 陈宜瑜, Nakajima T. 东亚低等鲤科鱼类细胞色素 b 基因序列测定及系统发育. 科学通报, 2000, 45(21): 2297~2302

5 Simons A M, Mayden R L. Phylogenetic relationships of North American cyprinids and assessment of homology of the

openposterior myodome. Copeia, 1999, (1): 13~21

6 Gilles A, Lecointre G, Faure E, et al. Mitochondrial phylogeny of European cyprinids: implications for their systematics, reticulate evolution, and colonization time. Mol Phyl Evo, 1998, 10(1): 132~143

7 Cecconi F, Crosio C, Mariottini P, et al. A functional role for some fugu introns larger than the typical short ones: the example of the gene coding for ribosomal protein S7 and snoRNA U17. Nucleic Acids Research, 1996, 24(16): 3167~3172

8 王宁, 陈润生. 基于内含子和外显子的系统发育分析的比较. 科学通报, 1999, 44: 2095~2102

9 Chow S, Hazama K. Universal PCR primers for S7 ribosomal protein gene introns in fish. Mole Ecol, 1998, 7: 1255~1256

10 Thompson J D, Gibson T J, Plewniak F. The Clustal_X windows interface: flexible strategies for multiple sequences alignment aided by quality analysis tools. Nucleic Acids Res, 1997, 25: 4876~4882

11 Kumar S, Tamura K, Jakobsen I B, et al. MEGA: Molecular Evolutionary Genetics Analysis Software. Version 2.1. Arizona: Arizona State University, 2001

12 Felsenstein J. PHYLIP: Phylogenetic Inference Package. Version 3.572c. Washington: University of Washington, 1997

13 Li W H, Graur D. Fundamentals of Molecular Evolution. Massachusetts: Sinauer Associates INC, 1991. 3~19

14 王金铃, 瞿礼嘉, 陈军, 等. CHS 基因外显子 2 的进化规律及其用于植物分子系统学研究的可行性. 科学通报, 2000, 45: 942~950

(2002-01-18 收稿, 2002-05-22 收修改稿)

· 动 态 ·

第 2 届中国生物信息学大会召开

2002 年 6 月 28~30 日, 由国家人类基因组北方研究中心、南方研究中心、中国科学院基因组信息学中心等 17 家单位主办, 北京生物技术和新医药产业促进中心和北京生物工程学会生物信息专业委员会共同承办的第 2 届中国生物信息学大会在北京大学隆重召开. 会议围绕功能基因组研究、序列和结构基因组研究、蛋白质结构模型模拟和预测、药物分子设计、生物信息学数据库、算法和软件的开发等几大主题展开讨论. 来自清华大学生物信息研究所的李衍达院士、英国 Imperial College Centre for Bioinformatics 的 Michael J. E. Sternberg 博士和美国华盛顿州立大学 School of Molecular Biosciences 的 Keith Dunker 教授等十余位国际著名生物信息学专家作了大会报告, 30 余位专家作了分组报告, 来自海内外的 500 余位专家、学者参加了此次会议, 100 余项成果以墙报的形式得到了交流.

生物信息学是生命科学、信息科学、数理科学等众多相关学科相互交融所形成的一门新兴边缘学科, 从其诞生之日起, 就展示出极其诱人的发展前景, 并迅速成为当今生物科学的重大前沿领域之一. 鉴于生物信息学技术的飞速发展和日新月异, 学术交流和互相借鉴就显得尤为重要, 此次中国生物信息学大会为海内外学者创造了一个良好的互动平台, 将对我国生物信息学的发展起到重要的推动作用.

<https://online.scichina.com/doi/10.1360/csb2002-47-14-1089>