

DOI: 10.3724/SP.J.1118.2017.16292

11 种鲈形目鱼类的核糖体基因 GC 含量及其与硬骨鱼类的特征比较

司李真^{1,2}, 武宝生^{1,2}, 孔晓瑜¹, 杨敏^{1,2}, 龚理^{1,2}, 时伟¹

1. 中国科学院 热带海洋生物资源与生态重点实验室, 广东 广州 510301;

2. 中国科学院大学, 北京 100049

摘要: 核糖体基因为串联重复多拷贝的基因, 包括 3 个编码基因(18S, 5.8S, 28S)和两个间隔区 ITS1(internal transcribed spacer 1)和 ITS2(internal transcribed spacer 2)。目前, 对核糖体基因的相关报道主要集中在个体内不同拷贝间的多态特征, 以及其作为分子标记在系统演化关系中的应用, GC 含量作为一项非常重要的核苷酸序列指标, 而鲜有报道。为了探讨鱼类的核糖体基因 GC 含量特征以及间隔区是否也存在 GC 平衡现象, 本研究选择了鲈形目(Perciformes) 5 科 11 种鱼类 5 个片段的核糖体基因进行研究, 包括尖吻鲈科(Latidae)、射水鱼科(Toxotidae)、军曹鱼科(Rachycentridae)、剑鱼科(Xiphiidae)、鲹科(Carangidae)。获得了 1651 个单克隆序列, 通过分析并比较已有的其他硬骨鱼序列片段的 GC 含量变化特征, 结果发现: 本研究鱼类的 18S 的 GC 含量为 52.6%~57.1% (平均 54.6%), 5.8S 为 55.6%~58.9% (平均 57.4%), 28S 为 64.2%~65.8% (平均 64.6%), ITS1 为 56.5%~73.0% (平均 65.0%), ITS2 为 62.3%~77.5% (平均 69.1%)。编码区的 GC 含量相对较保守, 变异范围较小, 18S 和 5.8S 变化范围明显小于间隔区, 28S 则位于间隔区的最低值和最高值之间。因此, 我们发现硬骨鱼核糖体 ITS 高于 60% 的 GC 含量是该类群的一个特征, 并且高 GC 含量的 ITS1 和 ITS2 序列中不存在明显的高 GC 富集区, 其含量高低的变化与序列长度也没有相关性。本研究 11 种鱼类的 ITS1 和 ITS2 的 GC 含量在种内的相似性既有大于也有小于种间相同片段的相似性, 因此 GC 平衡现象只存在部分种类中。本研究结果可为鱼类核糖体基因序列特征的进一步研究及利用提供科学依据。

关键词: 编码基因; 间隔区; 18S; 5.8S; ITS1; ITS2

中图分类号: S963

文献标志码: A

文章编号: 1005-8737-(2017)04-0657-12

核糖体基因在生物中有几十个甚至上万个串联重复单元, 每个单元包括编码区与非编码区, 编码区包括 18S、5.8S 和 28S 3 个基因, 非编码区则由位于 18S 与 5.8S 之间的 ITS1 及 5.8S 与 28S 之间的 ITS2 两个内转录间隔区组成。一般来说, 编码基因序列比较保守, 而间隔区序列变异相对较大^[1]。尽管核糖体基因为多拷贝基因, 但是其协同进化的方式使不同拷贝之间的序列倾向于保持一致, 因此, 一直以来对这些基因的研究及应用多以单条序列为主。然而, 越来越多的研究发现, 有时协同进化机制并不能使物种在基因组水平上的同质性速率与变异速率保持一致, 当变异速率

大于同质化速率时, 基因组内核糖体序列的不同拷贝之间则会出现不完全相同的现象, 从而导致其序列存在着极其丰富的多态性^[2-7]。因而, 要全面了解核糖体基因的特征, 就必须扩增多个单克隆子序列。目前, 对核糖体基因的相关报道主要集中在个体内不同拷贝间的多态特征^[8-15], 以及其作为分子标记在系统演化关系中的应用^[16-21], 而较少关注核糖体基因片段碱基的组成特征。

在分子生物学研究中, GC 含量是指 DNA 分子 4 种碱基中鸟嘌呤(G)和胞嘧啶(C)所占的比率。GC 含量是一项非常重要的核苷酸序列指标, 与许多指标特征都有着密切的关系。GC 含量与编码

收稿日期: 2016-09-30; 修订日期: 2017-03-17.

基金项目: 国家自然科学基金项目(31272273; 41276166).

作者简介: 司李真(1990-), 女, 硕士, 从事鱼类分类及系统进化研究. E-mail: lzsi314003@126.com

通信作者: 时伟, 助理研究员, 从事鱼类分类及系统演化研究. E-mail: shiwei@scsio.ac.cn

基因的第三位密码子成正相关性,并且与多种基因组特征相关,包括重复序列的分布、甲基化以及基因密度等^[22]。GC 含量还与进化地位有一定的关系,Rodriguez 等^[23]通过外推法分析指出,水果果蝇亚属(*Sophophora*)祖先基因的 GC 含量比其后裔物种更高。因此对 GC 含量的分析研究具有一定的理论意义。但是,目前对 GC 含量的研究多以编码基因为主,而对非编码基因的研究则较少^[23-25]。

为了研究编码基因和非编码基因的 GC 含量的特征及相互之间的异同,我们选择了同时具有编码基因和非编码基因的核糖体基因序列作为研究对象,对一些鱼类的核糖体基因序列的 GC 含量进行了初步分析。我们发现有些鱼类核糖体基因的序列不论是编码基因(18S, 5.8S, 28S)还是间隔区(ITS1, ITS2)的 GC 含量都大于 50%^[6, 8, 26-28],并且非编码区的 GC 含量高于编码基因,这与其他硬骨鱼类常见的基因组 GC 含量为 30%~50%^[29]有所不同,也与一般常见的非编码区的 GC 含量小于 50%现象有区别^[30]。

Torres 等^[31]测定了葫芦科 2 种植物绿皮西葫芦(*Cucurbita pepo*)和黄瓜(*Cucumis sativus*)的 ITS1 和 ITS2 全序列,并比较了已有的 18 个物种的 ITS 全序列,发现尽管不同物种之间 ITS 的 GC 含量变化很大(20% to 90%),但是同一个种的 ITS1 和 ITS2 的 GC 含量存在协同进化的现象,即同种的 ITS1 的 GC 含量接近于 ITS2 值,并将此现象称为 GC 平衡(GC balance)。

为了进一步探讨核糖体基因高 GC 含量的变化特征以及间隔区是否也存在 GC 平衡现象,我们选择了鲈形目中的 5 个科 11 种鱼类为研究对象并获得大量的核糖体基因不同片段的克隆序列,同时也汇总了目前已有的硬骨鱼类的核糖体基因序列,通过对这些片段的 GC 含量比较分析,发现鱼类中存在着高 GC 含量的序列变化特征,明确间隔区是否存在 GC 平衡现象,以期为今后硬骨鱼类的核糖体基因研究及利用提供科学依据。

1 材料与方法

1.1 实验材料

本研究选择了鲈形目的 5 科 11 种鱼类为研究

对象,包括鲹科(Carangidae)、尖吻鲈科(Latidae)、射水鱼科(Toxotidae)、剑鱼科(Xiphiidae)以及军曹鱼科(Rachycentridae),种类的详细信息见表 1。所获得的样品分别取肌肉放入 75%的酒精中保存待用。

1.2 DNA 提取、PCR 扩增和测序

取样品肌肉约 30 mg,切碎置于 1.5 mL 的离心管,加入 200 μ L 匀浆缓冲液和 20 μ L 蛋白酶 K,至肌肉组织完全裂解,然后用海洋动物组织基因组提取试剂盒提取 DNA(天根生化,北京);并用双蒸水溶解基因组 DNA,保存于-20℃冰箱中。根据已有近缘物种的核糖体基因片段序列,分别设计扩增引物,由于 ITS1 和 ITS2 相对不保守,因此引物设计在间隔区两侧的编码基因序列上(表 2)。扩增片段包括 18S 和 28S 部分序列、ITS1、5.8S 和 ITS2 全序列。为了最大程度的获得核糖体基因片段的各种类型的拷贝序列,本实验在 PCR 扩增时加入 8%的 DMSO。

PCR 反应总体积为 25 μ L,其中包括 2.5 μ L 10 $\times$ 缓冲液,2 μ L MgCl₂ (25 mmol/L),2 μ L dNTP (分别为 2.5 mmol/L),每个引物 1 μ L (10 μ mol/L),1 单位 rTaq 酶,1 μ L 模板 DNA (50 ng/ μ L),灭菌双蒸水补足至 25 μ L。反应程序为:94℃预变性 3 min,94℃变性 1 min,50℃退火 50 s,72℃延伸 50 s,进行 35 个循环后 72℃延伸 10 min。扩增产物使用 1%琼脂糖凝胶电泳检测,PCR 产物用 GenClean 柱式琼脂糖凝胶 DNA 回收试剂盒纯化回收,然后与 pMD 19-T 载体连接,转化至大肠杆菌 DH5 α 。每个物种的每个扩增片段分别挑选 12~36 个单克隆子进行测序,选择峰图准确清晰的序列用于分析。

1.3 数据分析

首先将测定的 11 种鱼类的核糖体基因序列片段利用 BLAST 网站(<http://www.ncbi.nlm.gov/BLAST/>)检索确定为目的片段。随后从 NCBI 数据库中下载所有其他硬骨鱼类的核糖体基因序列(表 3),并按目分类整理。然后,将以上两部分的数据分别利用 ClustalX 1.83^[32]并以种类或目为单位进行相同片段的序列比对。将获得的比对序列利用 MEGA 5.0^[33]统计不同分类阶元的核糖体基因片段的平均 GC 含量。由于 28S 序列长度约 3000 bp,

表 1 本研究 5 科 11 种鱼类的应用信息表
Tab. 1 The information of 11 species in 5 families used in present study

分类阶元 Taxonomic category	种名缩写 abbreviations of species	个体数 No.	采集地点 collecting location
尖吻鲈科 Latidae			
尖吻鲈属 <i>Lates</i>			
尖吻鲈 <i>Lates calcarifer</i>	JWL	2	广东湛江 Zhanjiang, Guangdong
军曹鱼科 Rachycentridae			
军曹鱼属 <i>Rachycentron</i>			
军曹鱼 <i>Rachycentron canadum</i>	JCY	3	浙江温岭 Wenling, Zhejiang
射水鱼科 Toxotidae			
射水鱼属 <i>Toxotes</i>			
射水鱼 <i>Toxotes chatareus</i>	SSY	2	广州花地湾市场 Huadiwan Market, Guangzhou
剑鱼科 Xiphiidae			
剑鱼属 <i>Xiphias</i>			
剑鱼 <i>Xiphias gladius</i>	JY	2	澳大利亚, 美国 Australia and America
鲷科 Carangidae			
副叶鲷属 <i>Alepes</i>			
吉打副叶鲷 <i>Alepes djedaba</i>	JDF	5	北海桥港鱼市 fish market in Beihai Port
圆鲷属 <i>Decapterus</i>			
蓝圆鲷 <i>Decapterus maruadsi</i>	LYS	2	北海桥港鱼市 fish market in Beihai Port
大甲鲷属 <i>Megalaspis</i>			
大甲鲷 <i>Megalaspis corayla</i>	DJS	3	北海桥港鱼市 fish market in Beihai Port
细鲷属 <i>Selaroides</i>			
金带细鲷 <i>Selaroides leptolepis</i>	JDX	3	北海桥港鱼市 fish market in Beihai Port
鲳鲷属 <i>Trachinotus</i>			
布氏鲳鲷 <i>Trachinotus blochii</i>	BSC	3	厦门第八市场 Xiamen Market 8
似鲷属 <i>Scomberoides</i>			
泰拉鲳鲷 <i>Scomberoides tala</i>	TLC	1	北海桥港鱼市 fish market in Beihai Port
竹荚鱼属 <i>Trachurus</i>			
日本竹荚鱼 <i>Trachurus japonicas</i>	ZJY	2	北海桥港鱼市 fish market in Beihai Port

表 2 核糖体基因不同片段 PCR 扩增的引物序列
Tab. 2 PCR primers for the amplification of different fragments in ribosomal genes

扩增序列 amplification primer	引物序列(5'-3') primer sequences (5'-3')
18S*	Z-18S-10/TCTGGTTGATTCTGCCAGTAG I8SF-1874/ATGATCCTTCCGCAGGTTCA
ITS1-5.8S	Z-18S-1720/TCGCTACTACCGATTGGATGGTTTA F-5.8S-150/AAGCGACCCTCAGACAGGCGTAG
5.8S-ITS2	Z-5.8S-10/DACAACCTCTTAGCGGTGGATCA F-28S-100/GCTCTTCCCTCTTCACTCG
28S*	Z-28S-10/ACGACCTCAGATCAGACGTGGC F-28S-1500/TCGGTTCATCCCGCAGCG

注: “*”表示不是全长序列。
Note: “*” represents the partial sequence.

目前报道的 NCBI 中有关硬骨鱼类的序列均为不完整的片段, 长度一致的数据比较少, 没有可比性, 因此, 在本研究中 28S 序列的比对分析只涉及本研究中的 11 种鲈形目鱼类, 不包括其他硬骨鱼类的该片段的序列。

2 结果与分析

2.1 11 种鱼类及其他硬骨鱼类核糖体基因的 GC 含量特征及比较

获得了 5 科 11 种鱼类核糖体基因的 1651 个单克隆序列, 包括长度为 1825~1845 bp 的 254 个 18S 单克隆序列, 443~662 bp 的 338 个 ITS1, 160 bp

表 3 本研究中其他硬骨鱼类核糖体基因片段应用信息
Tab. 3 Ribosomal gene statistics of each fragment of teleost in present study

目的种类 order	单克隆数 number of monoclon			
	18S	5.8S	ITS1	ITS2
鲟形目 Acipenseriformes	36	—	—	—
骨舌鱼目 Osteoglossiformes	28	—	—	—
鳗形目 Anguilliformes	9	8	14	—
鲱形目 Clupeiformes	6	64	100	—
鲤形目 Cypriniformes	71	—	199	—
鲮形目 Cyprinodontiformes	37	—	—	105
脂鲤目 Characiformes	16	—	36	37
鲇形目 Siluriformes	6	79	88	109
胡瓜鱼目 Osmeriformes	9	—	—	87
鲑形目 Salmoniformes	16	10	204	53
鳕形目 Gadiformes	—	12	37	17
银汉鱼目 Atheriniformes	14	—	—	—
颌针鱼目 Beloniformes	5	—	9	—
鲈形目 Perciformes	10	20	166	55
鲹形目 Pleuronectiformes	59	64	70	—
鲀形目 Tetraodontiformes	—	10	13	7

注：“—”表示无信息。
Note: “—”means no information.

的 339 个 5.8S, 310~515 bp 的 451 个 ITS2 和 1317~1437 bp 的 269 个 28S。

对这些单克隆序列的碱基组成分析, 结果显示 11 种鱼类的核糖体片段 GC 含量均超过 50%, 并且 ITS1、ITS2 以及 28S 的 GC 含量还大于 60%, 而 18S 和 5.8S 的 GC 含量则小于 60%。在编码基因中, 18S 的 GC 含量最低, 大多数的序列在 53.4% 到 54.2% 范围内(平均为 53.7%); 其次是 5.8S, 为 57.03%~58.2% (平均为 57.5%); 而 28S 的在 64.2%~65.8% 之间波动(平均为 64.6%)。而间隔区的 GC 含量变化较大, 其中大多数的 ITS1 从 61.5% 和 72.0% (平均为 66.9%), ITS2 则在 62.3% 和 77.0% 之间变化(平均为 69.6%; 图 1)。

由此可见, 编码基因的 GC 含量相对较保守, 变异范围较小, 间隔区的 GC 含量范围明显高于编码区的 18S 和 5.8S, 而 28S 的 GC 含量高于间隔区的最低值, 但低于间隔区最高值和平均值(图 1, 图 2)。

为了进一步了解 11 种鱼类的间隔区与编码基因 GC 含量之间的差别之处, 分别对每个种 4 种单碱基(A、T、C、G)的平均含量进行了统计分

析, 以期明确种类含量的变异来源以及是否存在碱基偏移现象(图 2)。结果显示, 18S 的 G 碱基和 C 碱基的含量分别为 28.4%~28.6%, 25.0%~25.6% (图 2a), 5.8S 为 28.8%~29.0%, 28.7%~29.0% (图 2b), 而 ITS1 的分别为 27.3%~31.7%, 33.4%~40.8% (图 2c), ITS2 的为 29.9%~35.1%, 35.2%~42.0% (图 2d)。由此可见, ITS 的高 GC 含量主要是由较高的 C 含量引起的, 但是 G、C 两个碱基之间不存在明显的偏移。

本研究的 11 种鱼类编码基因 GC 含量明显低于间隔区的, 这一现象引起了我们对 GC 含量的研究兴趣。为了探讨这种现象是否只存在于 11 种鱼类中, 还是也存在于其他硬骨鱼类的核糖体基因中, 我们下载了已有的 8~14 个目硬骨鱼类的 4 个片段(除 28S, 表 3), 并比较分析了 GC 含量(图 1, 图 2a~b)。结果表明, 其他硬骨鱼类不论编码基因还是间隔区的 GC 含量均超过 50%; 大多数 18S 的 GC 含量为 52.6%~57.1% (平均为 54.6%), 5.8S 为 55.6%~58.9% (平均为 57.4%), ITS1 为 56.5%~73.0% (平均为 65.0%), ITS2 的在 62.5%~77.5% (平均为 69.1%)。以上数据显示其他硬骨鱼类的 ITS 的 GC 含量平均值均高于 18S 和 5.8S 基因。同样为了更详细的了解每个目不同片段 G 或 C 含量是否存在碱基的偏移, 以及 ITS 高 GC 含量的来源, 我们又进一步按目对各片段的 G 和 C 含量分别进行了统计, 并按照目的进化顺序进行了比较分析。

结果显示, 18S 碱基的 G 和 C 含量分别为 27.4%~30.6%, 24.5%~27.0%, 5.8S 为 25.5%~31.6%, 25.2%~29.7%, 而 ITS1 的 G 和 C 含量分别在 27.5%~32.4%, 29.2%~37.6% 范围内, ITS2 为 31.0%~34.4%, 31.9%~39.3%。以上数据显示其他硬骨鱼类的 ITS 的 G 和 C 均较高, 特别是 ITS2 的最低 G 含量都高于 18S 和 5.8S 基因的最高含量; 这两个碱基之间不存在明显的偏移; 并且其 GC 含量的变化与物种进化地位无关, 而是随机变化的。这与获得的 11 种鱼类的 GC 含量结果是一致的。

2.2 GC 平衡

为了探讨 GC 平衡现象是否在本研究鱼类中存在, 即种内 ITS1 和 ITS2 两片段之间的相似性大于种间相同片段之间的相似性, 我们计算了 11

种鱼类种内两片段之间及种间相同片段之间的 GC 含量平均值的差异(表 3, 表 4), 结果表明种内两片段之间的差值分布在 1.20%~6.89%; ITS1 种间的 55 组数据中小于种内最小值有 7 组(0.35%~1.20%), 大多数是在最小值和最大值之间有 40 组(1.20%~6.58%), 大于最大值的有 8 组(>6.89%); ITS2 种间的 55 组中有 14 组小于种内最小值(0.01%~1.20%), 24 组数据分布在最小值和最大值之间

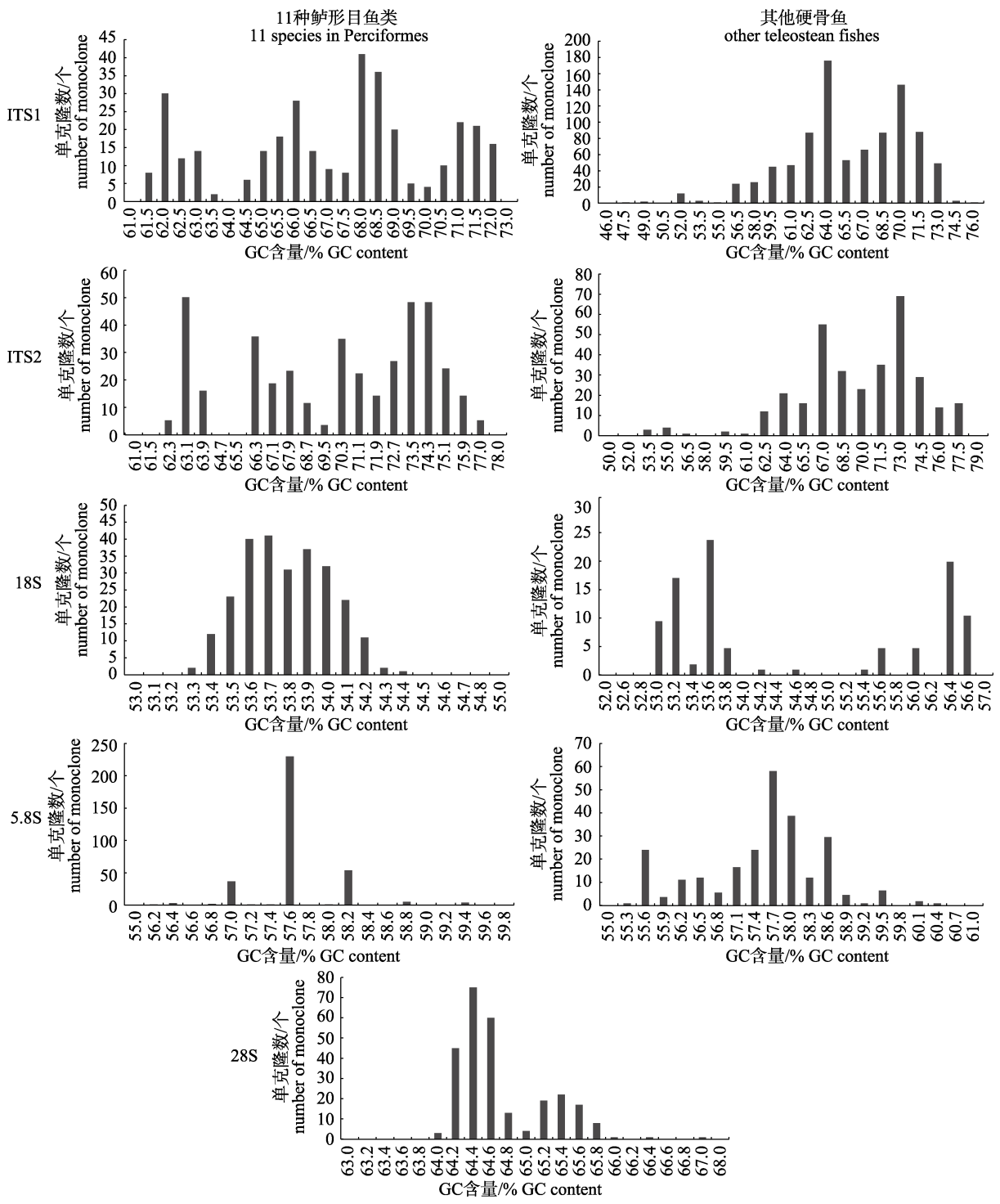


图 1 11 种鱼类及其他硬骨鱼类核糖体基因片段序列数与 GC 含量分布关系
Fig. 1 Distribution of the GC content of ribosomal genes from 11 species and other teleostean fishes

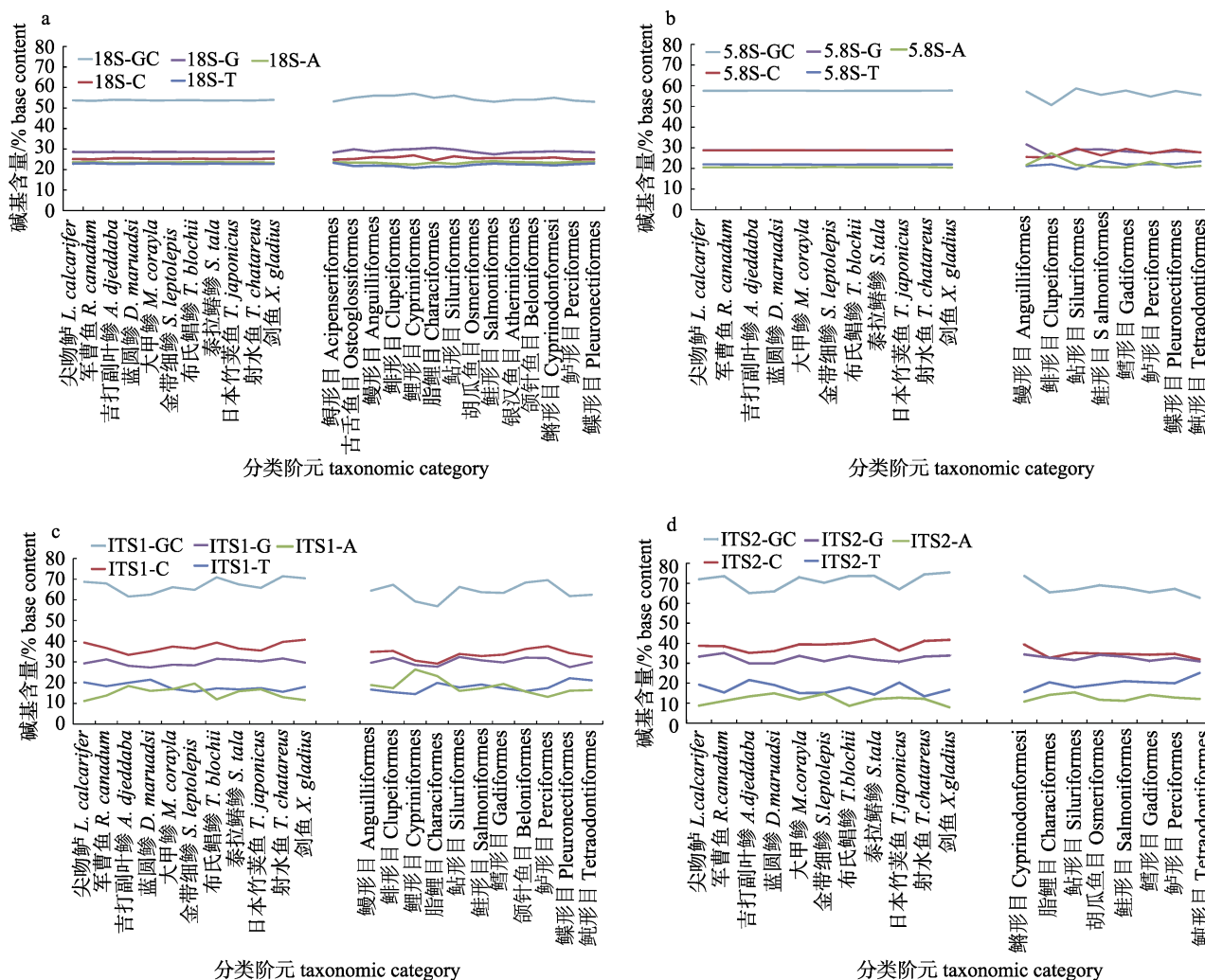


图 2 11 种鱼类及其他硬骨鱼类的核糖体基因片段各碱基平均含量图

Fig. 2 Mean GC content of ribosomal genes from 11 species and other teleostean fishes

(1.20%~6.75%), 17 组数据大于最大值(>6.89%)。

基于以上分析,我们认为在本研究的 11 种鲈形目的鱼类中,种内两片段之间的相似性与种间相同片段之间的相似性没有明显的规律性,例如,大甲鲈的种内两片段的差值全部大于种间相同片段的差值;吉打副叶鲈种内两片段的差值小于种间相同片段的差值的比例为 7 : 3;军曹鱼种内两片段的差值小于种间相同片段的差值的比例为 1 : 9。

因此,对这些种类是否存在 GC 平衡的现象,只能是单个种类和相应的比较种类间片段之间的差异而得出。例如尖吻鲈和军曹鱼比较时,种内的两片段相似性小于种间相同片段之间的相似性,不存在 GC 平衡现象;但是尖吻鲈和吉打副叶鲈

比较时,种内两片段的相似性则大于种间相同片段之间的相似性,因而存在 GC 平衡现象。

2.3 间隔区高 GC 含量分布特性

在本研究中发现了 ITS1 和 ITS2 存在高的 GC 含量现象,为了探讨这种现象是否与长度有关以及是否存在高 G、C 或 GC 含量富集簇,我们分别对已有其他硬骨鱼类的 917 条 ITS1(272~918 bp)和 333 条 ITS2 (128~694 bp)的每一条全长序列的 GC 含量进行了统计分析,比较不同长度的 GC 含量分布是否存在差异(图 3);同时我们对两片段序列位点的碱基组成进行了分析,以期发现不同位置序列与 GC 含量分布的关系,是否存在 GC 富集区域,进而揭示高 GC 含量的原因(图 3)。

对已有硬骨鱼类两片段全长序列的 GC 含量

表 4 11 种鱼类 ITS1 和 ITS2 GC 含量的种内及种间相同片段差异
Tab. 4 Differences of GC contents of ITS1 and ITS2 of 11 species in same species or between species

种名 species 种内差值 <i>D</i>			种间差值 difference within inter-species										
			JWL	JCY	SSY	JY	JDFY	LYS	DJS	JDXS	BSCS	TLCS	ZJY
JWL	3.32	ITS1	0	0.75	2.7	1.76	7.11	6.17	2.55	3.88	2.18	1.24	2.90
		ITS2	0	1.54	2.39	3.4	6.91	6.06	1.01	1.74	1.55	1.72	5.03
		PH					√	√					
JCY	5.62	ITS1	0.75	0	3.45	2.51	6.36	5.42	1.8	3.13	2.93	0.49	2.15
		ITS2	1.54	0	0.85	1.86	8.45	7.6	0.53	3.28	0.01	0.18	6.57
		PH					√						
SSY	3.02	ITS1	2.7	3.45	0	0.94	9.81	8.87	5.25	6.58	0.52	3.94	5.6
		ITS2	2.39	0.85	0	1.01	9.3	8.45	1.38	4.13	0.84	0.67	7.42
		PH					√	√		√			√
JY	5.97	ITS1	1.76	2.51	0.94	0	8.87	7.93	4.31	5.64	0.42	3	4.66
		ITS2	3.4	1.86	1.01	0	10.31	9.46	2.39	5.14	1.85	1.68	8.43
		PH					√	√					
JDFY	3.53	ITS1	7.11	6.36	9.81	8.87	0	0.94	4.56	3.23	9.29	5.87	4.21
		ITS2	6.91	8.45	9.3	10.31	0	0.85	7.92	5.17	8.46	8.63	1.88
		PH	√	√	√	√			√		√	√	
LYS	3.44	ITS1	6.17	5.42	8.87	7.93	0.94	0	3.62	2.29	8.35	4.93	3.27
		ITS2	6.06	7.6	8.45	9.46	0.85	0	7.07	4.32	7.61	7.78	1.03
		PH	√	√	√	√			√		√	√	
DJS	6.89	ITS1	2.55	1.8	5.25	4.31	4.56	3.62	0	1.33	4.73	1.31	0.35
		ITS2	1.01	0.53	1.38	2.39	7.92	7.07	0	2.75	0.54	0.71	6.04
		PH											
JDXS	5.47	ITS1	3.88	3.13	6.58	5.64	3.23	2.29	1.33	0	6.06	2.64	0.98
		ITS2	1.74	3.28	4.13	5.14	5.17	4.32	2.75	0	3.29	3.46	3.29
		PH											
BSCS	2.7	ITS1	2.18	2.93	0.52	0.42	9.29	8.35	4.73	6.06	0	3.42	5.08
		ITS2	1.55	0.01	0.84	1.85	8.46	7.61	0.54	3.29	0	0.17	6.58
		PH					√	√		√			√
TLCS	6.28	ITS1	1.24	0.49	3.94	3	5.87	4.93	1.31	2.64	3.42	0	1.66
		ITS2	1.72	0.18	0.67	1.68	8.63	7.78	0.71	3.46	0.17	0	6.75
		PH											
ZJY	1.2	ITS1	2.9	2.15	5.6	4.66	4.21	3.27	0.35	0.98	5.08	1.66	0
		ITS2	5.03	6.57	7.42	8.43	1.88	1.03	6.04	3.29	6.58	6.75	0
		PH	√	√	√	√	√				√	√	

注: 种名缩写请参考表 1; “√”代表符合 GC 平衡现象; “PH”为“GC 平衡”的缩写。
Note: “D” is difference within intra-species. Species abbreviations are given in Tab. 1; “√” represents which belong to GC balance; “PH” is short for “GC balance”.

分析结果显示, 不论在 ITS1 中(除个别种类的 GC 含量略低于 50%之外), 还是在 ITS2 序列中, 不存在长度变化而导致 GC 含量的变化关系。由此可见, GC 含量与两片段的序列长度变化没有相关性, GC 含量并不会随序列长度的变化而发生变化, 而是散状随机分布在 46.3%~75.7% (ITS1)和 52.7%~77.4% (ITS2)。

对两片段序列位点的碱基组成是通过分别计算每 20 个碱基中 G 和 C 的百分含量获得的。我们分别选择了 ITS1 和 ITS2 序列最长和最短的片

段为代表进行分析, ITS1 选择了最长的 918 bp 的结鱼(*Tor mosalmahanadicus*; GU568373)和最短的 272 bp 的西美洲鳊(*Notropis leuciodus*; JF523436), ITS2 选择 694 bp 的四川白甲鱼(*Onychostoma angustistomata*; DQ994151)和 128 bp 的圆花鲮(*Auxis rochei*; AB193740)。

首先利用 excel 分别计算单个序列每 20 个碱基中 G 和 C 的百分含量, 例如第 1~20 个碱基中

G 和 C 的含量, 第 2~21 个碱基中 G 和 C 的含量, 依次类推, 进而计算出该序列每 20 个碱基中 GC 所占的比例(图 4)。从图中可以看出, ITS1 序列的 GC 含量在 20%至 90%之间波动, ITS2 在 40%~90%。不论是 ITS1 还是 ITS2, GC 含量的变化呈现交替出现, 均没有出现高 GC 含量的区域。这表明在两片段的序列中, 都没有出现明显的 GC 富集簇, 碱基 G 或 C 是随机分布在位点上的。

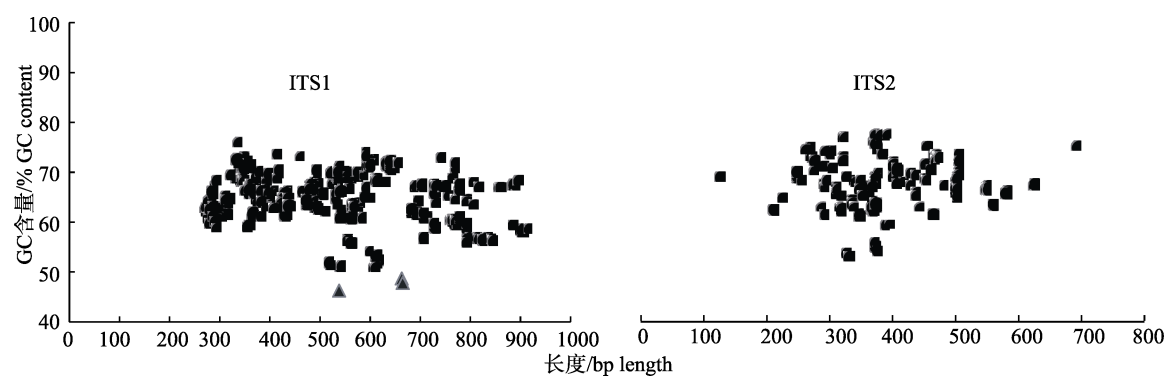


图 3 已有其他硬骨鱼类 ITS 片段序列长度与 GC 含量的关系图

●: 代表 GC 含量高于 50%的种类, ▲: 代表 GC 含量低于 50%的种类.

Fig. 3 Relationship between the length and GC content of ITS in other teleostean fishes

●: black circle represents GC content higher than 50%; ▲: triangle represents GC content below 50%.

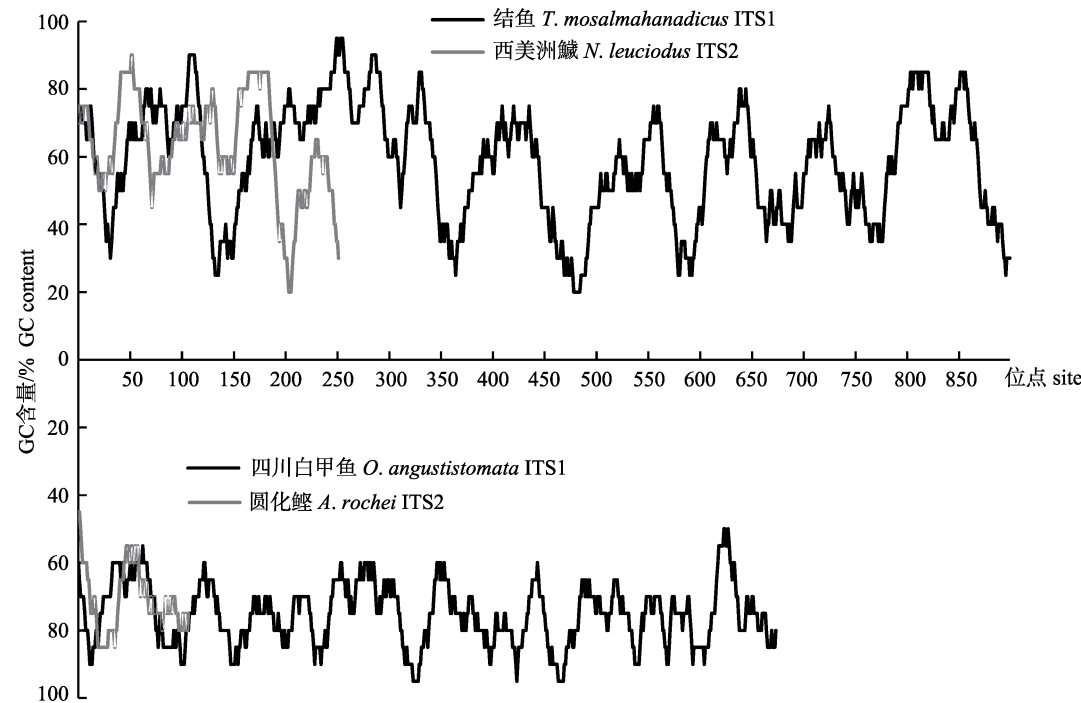


图 4 四种代表鱼类 ITS 序列位点与 GC 含量的关系图

Fig. 4 Relationship between the sites of nucleotides and GC content of ITS from four representative fishes

3 讨论

本研究对鲈形目5科的11种鱼类及部分硬骨鱼类的核糖体基因(18S、ITS1、5.8S、ITS2和28S)的GC含量特征进行了分析,结果表明这些片段的GC含量均大于50%,并且间隔区的大于60%大多数的间隔区片段的含量大于编码基因的,远超过鱼类基因组的GC含量(30%~50%);这些结果也在分析的部分硬骨鱼类核糖体基因GC含量的结果得到了证实。这与Torres等^[31]描述的一些种类如秀丽隐杆线虫(*Caenorhabditis elegans*)、黑腹果蝇(*Drosophila melanogaster*)、粟酒裂殖酵母(*Schizosaccharomyces pombe*)等小于50%还是有明显不同的。在分析ITS1和ITS2序列位点的碱基组成中,没有发现GC含量存在明显的高含量区域,并且含量大小也不会随序列长度的变化而发生变化,与序列长度没有相关性,但是片段不同位点之间GC含量的变异范围较大,在20%~90%。

GC平衡的概念是Torres等^[31]1990年提出的,作者根据葫芦科2种植物的ITS结果并比较了其他18个物种的ITS序列,发现不同物种之间ITS的GC含量虽然变化很大(20%~90%),但是种内的ITS1和ITS2的GC含量存在协同进化现象,即种内的ITS1的GC含量与ITS2相似,将此现象称为GC平衡(GC balance)。

自Torres等^[31]提出GC平衡概念后,也有学者^[34-36]相继在他们的研究中涉及到了GC平衡现象。不管是Torres等还是其他学者,涉及的研究对象包括了真菌、植物、动物,亲缘关系均较远,也没有比较种内两片段的GC含量值和种间相同片段之间的差异的大小,并依据这些差异的程度判断是否存在GC平衡。我们分析了Torres等^[31]已有的相关数据,发现有的种类和其研究结论中都存在GC平衡有不一致之处,例如研究中老鼠(mouse)的ITS1和ITS2的GC含量分别为70%和75%,绿海胆(*Lytechinus variegatus*)的分别为71%、75%,这两个种内两片段的差异分别为5%和4%,但两个种种间相同片段之间的差异ITS1为1%,ITS2的则为0,因此,种内的差值明

显大于种间相同片段之间的GC含量。进一步的分析发现之所以产生结果不一致的原因是作者在涉及GC平衡现象时,没有比较种内和种间相似性的差异大小关系,只是对所有的研究对象给出了统一的结论。

本研究中,我们比较了鲈形目5科11种鱼类ITS1和ITS2各种内及种间相同片段GC含量的差异(110组数据),所有种内两片段的差值小于其他10种鱼类种间相同片段差值的有34组,大于的有66组(表4),因此,不能就研究的所有种类给出一个结论,只能对具体种类的相应数据比对后才能判定。例如,尖吻鲈的种内两片段GC含量之间的差值是3.32,与其他10种类相同片段相比,只有吉打副叶鲈和蓝圆鲈的种间差值大于尖吻鲈两片段的种内差值,其余8种类至少有一个或者两片段种间的差异小于种内两片段之间的差异,因此,我们认为尖吻鲈与吉打副叶鲈和蓝圆鲈之间存在GC平衡现象,但与其他9种类之间不存在(表4)。

因而,我们认为GC平衡是相对的,需要有明确的相互比较的种类,并且种内两片段的GC含量之差小于种间相同片段的差异,存在GC平衡现象;相反,种内两片段的GC含量之差大于种间相同片段的差异,就不存在GC平衡现象。这种现象表现了种类的ITS片段在进化过程中协同进化所起的作用,但从我们的研究种类中可以发现,大多数的种类还是种间的相同片段更相似些。

在本研究中不论是本实验中的11种鱼类还是其他的硬骨鱼类,其ITS的GC含量都高于50%,远高于硬骨鱼基因组内的GC含量(30%~50%)^[29],我们认为核糖体ITS高GC含量是硬骨鱼系统中普遍存在的一个现象,这一研究结果将会对今后进一步的研究和利用鱼类核糖体间隔区提供了理论基础。

致谢: 感谢中国科学院海洋研究所刘静研究员所提供的样品以及在样品鉴定方面给予的重要支持。

参考文献:

- [1] Hillis D M, Dixon M T. Ribosomal DNA: Molecular evolu-

- tion and phylogenetic inference[J]. *Quart Rev Biol*, 1991, 66(4): 411–453.
- [2] Reed K M, Hackett J D, Phillips R B. Comparative analysis of intra-individual and inter-species DNA sequence variation in salmonid ribosomal DNA cistrons[J]. *Gene*, 2000, 249(2000): 115–125.
- [3] Aktas M, Bendele K G, Altay K, et al. Sequence polymorphism in the ribosomal DNA internal transcribed spacers differs among *Theileria* species[J]. *Veterin Parasitol*, 2007, 147(3–4): 221–230.
- [4] Xiao L Q, Zhu H. Intra-genomic polymorphism in the internal transcribed spacer (ITS) regions of *Cycas revoluta*: evidence of incomplete concerted evolution[J]. *Biodivers Sci*, 2009, 17(5): 476–481.
- [5] Harpke D, Peterson A. Non-concerted ITS evolution in *Mammillaria* (Cactaceae)[J]. *Molec Phylogenetics Evol*, 2006, 41(3): 579–593.
- [6] Xu H, Li J, Kong X Y, et al. Phylogenetic relationship and length variation in the first ribosomal internal transcribed spacer of *Cynoglossinae* species[J]. *Oceanologia et Limnologia Sinica*, 2008, 39(1): 35–41. [徐晖, 李军, 孔晓瑜, 等. 6种舌鳎亚科鱼类 ITS1 序列长度多态性及系统分析[J]. *海洋与湖泊*, 2008, 39(1): 35–41.]
- [7] Gong L, Shi W, Yang M, et al. Non-concerted evolution in ribosomal ITS2 sequence in *Cynoglossus zanzibarensis* (Pleuronectiformes: Cynoglossidae)[J]. *Biochem System Ecol*, 2016, (66): 181–187.
- [8] Gong L, Xu H, Li J, et al. Characterization of the first internal transcribed spacer of ribosomal DNA in *Paralichthys olivaceus* (♀) and *P. dentatus* (♂) hybrids[J]. *Journal of Fishery Sciences of China*, 2015, 22(1): 17–23. [龚理, 徐晖, 李军, 等. 褐牙鲆(♀)、夏鲆(♂)及其杂交子一代的 ITS1 序列特征分析[J]. *中国水产科学*, 2015, 22(1): 17–23.]
- [9] Worheide G, Nichols S A, Goldberg J. Intra-genomic variation of the rDNA internal transcribed spacers in sponges (Phylum Porifera): implications for phylogenetic studies[J]. *Molec Phylogen Evol*, 2004, 33(3): 816–830.
- [10] Fairley T L, Kilpatrick C W, Conn J E. Intra-genomic heterogeneity of internal transcribed spacer rDNA in neotropical malaria vector *Anopheles aquasalis* (Diptera : Culicidae)[J]. *J Med Entomol*, 2005, 42(5): 795–800.
- [11] Li C, Wilkerson R. Intra-genomic ITS2 variation and its impact on PCR identification of an anopheles malaria vector group[J]. *Am J Trop Med Hyg*, 2005, 73(6): 297–297.
- [12] Li C, Wilkerson R C. Intra-genomic rDNA ITS2 variation in the neotropical *Anopheles* (*Nyssorhynchus*) *albitarsis* complex (Diptera : Culicidae)[J]. *J Hered*, 2007, 98(1): 51–59.
- [13] Smith M E, Douhan G W, Rizzo D M. Intra-specific and intra-sporocarp ITS variation of ectomycorrhizal fungi as assessed by rDNA sequencing of sporocarps and pooled ectomycorrhizal roots from a *Quercus* woodland[J]. *Mycorrhiza*, 2007, 18(1): 15–22.
- [14] Bezzhonova O V, Goryacheva I I. Intra-genomic heterogeneity of rDNA internal transcribed spacer 2 in *Anopheles messeae* (Diptera : Culicidae)[J]. *J Med Entomol*, 2008, 45(3): 337–341.
- [15] Lindner D L, Banik M T. Intra-genomic variation in the ITS rDNA region obscures phylogenetic relationships and inflates estimates of operational taxonomic units in genus *Laetiporus*[J]. *Mycologia*, 2011, 103(4): 731–740.
- [16] Yao H, Song J Y, Liu C, et al. Use of ITS2 region as the universal DNA barcode for plants and animals[J]. *PLoS ONE*, 2010, 5(10): e13102.
- [17] Mu X D, Gu D D, Yang Y X, et al. Genetic diversity and phylogeny of the family Osteoglossidae by the nuclear 18S ribosomal RNA and implications for its conservation[J]. *Biochem System Ecol*, 2013, 51: 280–287.
- [18] Stock D, Moberg K, Maxson L, et al. A phylogenetic analysis of the 18S ribosomal RNA sequence of the coelacanth *Latimeria chalumnae*[J]. *Environ Biol Fishes*, 1991, 32(1–4): 99–117.
- [19] Pleyte K A, Duncan S D, Phillips R B. Evolutionary relationships of the salmonid fish genus *Salvelinus* inferred from DNA sequences of the first internal transcribed Spacer (ITS1) of ribosomal DNA[J]. *Molec Phylogen Evol*, 1992, 1(3): 223–230.
- [20] Huyse T, Van Houdt J, Volckaert F A M. Paleoclimatic history and vicariant speciation in the “sand goby” group (Gobiidae, Teleostei)[J]. *Molec Phylogen Evol*, 2004, 32(1): 324–336.
- [21] Chu K H, Li C P, Ho H Y. The first internal transcribed spacer (ITS1) of ribosomal DNA as a molecular marker for phylogenetic and population analyses in crustacea[J]. *Mar Biotechnol*, 2001, 3(4): 355–361.
- [22] Galtier N, Piganeau G, Mouchiroud D, et al. GC-content evolution in mammalian genomes: The biased gene conversion hypothesis[J]. *Genetics*, 2001, 159(2): 907–911.
- [23] Rodriguez-Trelles F, Tarrio R, Ayala F J. Evidence for a high ancestral GC content in *Drosophila*[J]. *Molec Biol Evol*, 2000, 17(11): 1710–1717.
- [24] Musto H, Naya H, Zavala A, et al. Genomic GC level, opti-

- mal growth temperature, and genome size in prokaryotes[J]. Biochem Biophys Res Communicat, 2006, 347(1): 1–3.
- [25] Mitchell D. GC content and genome length in chargaff compliant genomes[J]. Bioch Biophys Res Commun, 2007, 353(1): 207–210.
- [26] Chow S, Ueno Y, Toyokawa M, et al. Preliminary analysis of length and GC content variation in the ribosomal first internal transcribed spacer (ITS1) of marine animals[J]. Mar Biotechnol, 2009, 11(3): 301–306.
- [27] Freire R, Arias A, Mendez J, et al. Sequence variation of the internal transcribed spacer (ITS) region of ribosomal DNA in Cerastoderma species (Bivalvia: Cardiidae)[J]. J Mollus Stud, 2009, 76(1): 77–86.
- [28] Kumar R, Singh M, Kushwaha B, et al. Molecular characterization of major and minor rDNA repeats and genetic variability assessment in different species of mahseer found in North India[J]. Gene, 2013, 527(1): 248–258.
- [29] Tarallo A, Angelini C, Sanges R, et al. On the genome base composition of teleosts, the effect of environment and life-style[J]. BMC Genom, 2016, 17: 173.
- [30] Bohlin J, Snipen L, Hardy S P, et al. Analysis of intra-genomic GC content homogeneity within prokaryotes[J]. BMC Genom, 2010, 11: 464.
- [31] Torres R A, Ganai M, Hemleben V. GC balance in the internal transcribed spacers ITS1 and ITS2 of nuclear ribosomal RNAs[J]. J Molec Evol, 1990, 30(2): 170–181.
- [32] 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]. Nucl Acids Res, 1997, 25(24): 4876–4882.
- [33] Tamura K, Peterson D, Peterson N, et al. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods[J]. Molec Biol Evol, 2011, 28(10): 2731–2739.
- [34] Ghada B, Olfa S, Khaled C, et al. Sequence analysis of the internal transcribed spacers (ITSs) region of the nuclear ribosomal DNA (nrDNA) in fig cultivars (*Ficus carica* L.)[J]. Sci Horticult, 2009, 120(1): 34–40.
- [35] Yu D H, Chu K H. Species identity and phylogenetic relationship of the pearl oysters in *Pinctada* Röding, 1798 based on ITS sequence analysis[J]. Biochem System Ecol, 2006, 34(3): 240–250.
- [36] Mullineux T, Hausner G. Evolution of rDNA ITS1 and ITS2 sequences and RNA secondary structures within members of the fungal genera *Grosmannia* and *Leptographium*[J]. Fung Gen Biol, 2009, 46(11): 855–867.

Analysis of the GC content of ribosomal genes of 11 species of Perciformes and comparison with other teleostean fishes

SI Lizhen^{1,2}, WU Baosheng^{1,2}, KONG Xiaoyu¹, YANG Min^{1,2}, GONG Li^{1,2}, SHI Wei¹

1. CAS Key Laboratory of Tropical Marine Bio-resources and Ecology, South China Sea Institute of Oceanology, Guangzhou 510301, China;

2. University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: The ribosomal RNA gene (rDNA) cluster consists of multiple units of three coding genes (18S, 5.8S, and 28S) as well as two internal transcribed spacers (ITS1 and ITS2) separating the coding regions. Thus far, studies on these five fragments mainly focused on the polymorphism of different copies within each individual sample and identification of useful markers for phylogenetic relationship analysis. However, there are limited studies related to the features of GC content, which is a very important characteristic of ribosomal RNA gene. The characteristics of rDNA GC content and whether the GC balance phenomenon exists in the coding regions in fishes were investigated by selecting 11 species from five families of Perciformes, including Latidae, Toxotidae, Rachycentridae, Xiphiidae, and Carangidae. In all, 1651 monoclonal clones from the five fragments mentioned above were obtained. The GC content features were analyzed based on the sequences from the 11 species or the datasets from other teleostean fishes retrieved from GenBank. The results from the above two analyses were compared. The following results were obtained. First, the GC content of 18S, 5.8S, 28S, ITS1, and ITS2 ranged from 52.6% to 57.1% (average, 54.6%), 55.6% to 58.9% (average, 57.4%), 64.2% to 65.8% (average, 64.6%), 56.5% to 73.0% (average, 65.0%), and 62.3% to 77.5% (average, 69.1%), respectively. Second, compared with non-coding regions, coding regions were relatively conserved. The GC content of the coding genes varied in smaller ranges than those of the internal transcribed spacers. The GC contents of 18S and 5.8S were lower than those of ITS1 and ITS2, but that of 28S was between the lowest and highest values of ITS1 and ITS2. Therefore, we found that the GC content of non-coding regions was higher than 60%, which was a remarkable characteristic of these fishes, and no correlation was found between fragment length and higher GC content. Further, no obvious G, C, or GC rich block was found in the high-GC-content regions of ITS1 and ITS2 sequences. Third, the similarity of GC content between ITS1 and ITS2 within the same species could be higher or lower than that of the same fragment among different species within the 11 species. Therefore, the GC balance phenomenon is not universal and only exists in species whose intra-species GC content similarity is lower than the inter-species GC content similarity. The results of this study might provide a scientific basis for further studies and facilitate the utilization of the ribosomal gene characteristics of fish.

Key words: coding genes; internal transcribed spacers; 18S; 5.8S; ITS1; ITS2

Corresponding author: SHI Wei E-mail: Shiwei@scsio.ac.cn