

## 瘤背石磺线粒体基因组全序列分析

魏峦峦, 沈和定\*, 张雨, 方磊, 张坤霞, 陈诚

(上海海洋大学水产与生命学院, 上海 201306)

**摘要:** 采用 LA-PCR 技术对瘤背石磺线粒体基因组全序列进行了测定和分析。结果表明, 瘤背石磺线粒体基因组序列全长 13 957 bp, 由 22 个 tRNA、2 个 rRNA、13 个蛋白编码基因和 19 个长度为 2~138 bp 的非编码区组成。4 个蛋白质编码基因和 8 个 tRNA 基因从 L 链编码, 其余基因均从 H 链编码。蛋白质基因的起始密码子, 除 ND2 为 TTG 以外, 均为典型的起始密码子 ATN。CO III 和 Cytb 基因使用了不完全终止密码子 T, 其余基因均使用典型的 TAA 或 TAG。预测了 22 个 tRNA 基因的二级结构, 发现 tRNA<sup>Ser</sup> 缺少 DHU 臂, tRNA<sup>Ser</sup> 和 tRNA<sup>Thr</sup> 的反密码子环上有 9 个碱基, 而不是通常的 7 个碱基。最长的非编码区含有类似于 tRNA 的二级结构。基于线粒体基因组编码的 13 个蛋白质的氨基酸序列, 用 NJ、MP、ME 和 UPGMA 法构建系统进化树。分析 6 种软体动物之间的亲缘关系, 结果与传统的系统分类基本一致。研究初步确定瘤背石磺与平疣桑椹石磺的亲缘关系比与凯尔特石磺的亲缘关系近。

**关键词:** 瘤背石磺; 线粒体基因组; 序列分析; 系统进化树

**中图分类号:** Q 785; S 917

**文献标识码:** A

瘤背石磺 (*Onchidium struma*) 俗称土海参, 又名海癞子、土鸡等, 隶属软体动物门 (Mollusca), 腹足纲 (Gastropoda), 有肺类 (Pulmonata), 缩眼目 (Systellommatophora), 石磺科 (Onchidiidae)<sup>[1]</sup>。瘤背石磺栖息于海边的高潮区及潮上带, 是介于海洋与陆地过渡带的生物, 营穴居生活, 并摄食泥土表面的藻类、苔藓类和土壤中的有机营养物质<sup>[2]</sup>。瘤背石磺具有很高的营养价值和药用价值, 对滩涂湿地生物多样性和环境保护等诸学科研究具有重要意义<sup>[3]</sup>。石磺科贝类是海洋向陆地辐射的典型代表<sup>[4]</sup>, 不同属种其生活习性由深水向浅水、潮间带和潮上带区域的由深水至陆地生活的过程本身也是一个微型的由海洋向陆地辐射的过程。因此, 瘤背石磺线粒体基因组全序列的测定将为贝类由海洋向淡水及至陆地演化提供基础资料。

动物的线粒体基因组一般都较小, 通常在 15~20 kb, 除少数种类外, 主要包括 37 个基因, 即两个核糖体 RNA 基因, 13 个蛋白质编码基因

和 22 个转运 RNA 基因<sup>[5]</sup>。线粒体基因组属母系遗传, 是共价闭合的双链 DNA 分子, 核酸序列和组成比较保守, 基因的排列顺序比较稳定而且紧密, 无重组和单拷贝<sup>[6-7]</sup>。由于其结构和进化上的特点, mtDNA 已成为研究动物起源进化以及群体遗传分化的理想对象<sup>[8]</sup>。国内外有关石磺科贝类线粒体 DNA 全序列的报导较少, 迄今仅有平疣桑椹石磺 (*Platevindex mortoni*) 和凯尔特石磺 (*Onchidella celtica*)<sup>[4]</sup> 线粒体基因组全序列被测定出, 有关瘤背石磺线粒体基因组全序列仍未见报导。本研究采用 LA-PCR 技术, 测定瘤背石磺线粒体基因组 DNA 全序列, 分析其基因组组成及结构, 以期对石磺科贝类系统分类和分子进化研究提供重要参考。

### 1 材料与方法

#### 1.1 样本采集、mtDNA 提取

瘤背石磺样本于 2009 年 5 月采自上海崇明,

收稿日期: 2010-08-30 修回日期: 2010-11-19

资助项目: 国家自然科学基金项目 (30972259); 上海海洋大学博士科研启动基金; 上海市科委项目 (10DZ0503200); 上海市教委重点学科项目 (J50701)

通讯作者: 沈和定, Tel: 021-61900446, E-mail: hdshen@shou.edu.cn

取新鲜外套膜组织 50 ~ 100 mg, 使用高盐沉淀法<sup>[10]</sup>获得尽量完整的环状线粒体 DNA。线粒体 DNA 制备过程中尽量温和混匀, 避免环状线粒体 DNA 的断裂。制备好的线粒体 DNA 用 1% 的琼脂糖凝胶电泳检测, -20 °C 保存备用。

### 1.2 线粒体全序列的 PCR 扩增及测序

根据后生动物线粒体基因组通用引物扩增 *COI* 和 16S rRNA 基因的部分序列, 引物为 *COI*-L (5'-GGTCAACAAATCATAAAGTATTGG-3'), *COI*-H (5'-TAAACTTCAGGGTGACCAAAAATCA-3'), 16S-L (5'-CGCCTGTTTATCAAAAACAT-3'), 16S-H (5'-CCGGTCTGAACTCAGATCACGT-3')<sup>[11]</sup>。反应的模板 DNA 约为 100 ng, 反应体系总体积为 50  $\mu$ L, 其中 10  $\times$  Ex *Taq* Buffer 5  $\mu$ L, dNTPs 2  $\mu$ L (各 2.5 mmol/L), 引物各 1  $\mu$ L (20  $\mu$ mol/L), Ex *Taq* 酶 0.5  $\mu$ L (2 U)。PCR 反应条件: 94 °C 预变性 3 min; 94 °C 变性 30 s, 60 °C 退火 30 s, 72 °C 延伸 1 min, 35 个循环; 最后 72 °C 延伸 10 min。扩增产物经 1% 琼脂糖凝胶电泳检测。PCR 产物由上海生物工程技术有限公司纯化后双向测序。

瘤背石磺 mtDNA 扩增采用 LA-PCR。以前期得到的 *COI* 和 16S rRNA 两段序列为模板设计两对长 PCR 引物, 分别为 *COI*6-F (5'-G TCTATTTTGGTTACTGTTT-3'), *COI*6-R (5'-GTTATGTATTGTCTCCTTGC-3'); 16*CO*-F (5'-GGCAAGGAGACAATACATAA-3'), 16*CO*-R (5'-ACAGTAACCAAAATAGACCA-3')。反应体系为 50  $\mu$ L, 模板 mtDNA 约 100 ng, 引物 1  $\mu$ L (20  $\mu$ mol/L), dNTP 8  $\mu$ L (各 2.5 mmol/L), buffer mixture 5  $\mu$ L, LA-*Taq* 酶 (Takara 公司) 5 U, 灭菌水补足至 50  $\mu$ L。PCR 反应条件为 98 °C 预变性 10 s; 94 °C 变性 1 min, 51 °C/54.5 °C 退火 30 s, 68 °C 延伸 10 min, 35 个循环; 最后 72 °C 延伸 17 min。PCR 产物经 1% 的琼脂糖凝胶电泳检测后送大连宝生生物技术有限公司测序, 测序方法为鸟枪法。

### 1.3 mtDNA 片段拼接及序列分析

用 DNA STAR 和 MEGA 4.1 软件人工辅助进行序列拼接, 得到瘤背石磺线粒体基因组全序列。Open Reading Frame Finder (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>) 软件分析结合 BLAST (<http://ncbi.nlm.nih.gov/BLAST>)<sup>[12]</sup> 搜

索, 并与已知同源蛋白相比较, 确定并定位瘤背石磺 13 个蛋白质编码基因。用 Mega 4.1 软件分析蛋白质编码基因的密码子使用情况。通过近源物种 rRNA 序列比较及二级结构预测, 来定位瘤背石磺 rRNA 基因的起止点。大多数转运 RNA 基因通过在线 tRNA 注释工具 tRNAscan-SE1.21 软件 (<http://selab.janrelia.org/tRNAscan.SE/>)<sup>[13]</sup> 来注释得到, 剩余的转运 RNA 基因通过 RNA structure 4.6<sup>[14]</sup> 预测其茎环结构。全基因组序列经 Sequin 7.9 注释后提交 GenBank (登录号 HQ452830)。

## 2 结果

### 2.1 瘤背石磺 mtDNA 基因组结构

瘤背石磺线粒体基因组全长为 13957 bp, 基因组组成与典型的后生动物线粒体 DNA 相似。共编码 37 个基因, 包括 13 个蛋白质编码基因, 22 个 tRNA 基因, 2 个 rRNA 基因和一段 A + T 富集区。其中, tRNA<sup>Glu</sup>、tRNA<sup>Leu</sup> (UUR)、ATP8、tRNA<sup>Asn</sup>、ATP6、tRNA<sup>Arg</sup>、tRNA<sup>Glu</sup>、rrnS、tRNA<sup>Met</sup> ND<sub>3</sub>、tRNA<sup>Ser</sup> (UCN)、tRNA<sup>Thr</sup> 和 *COIII* 基因从 L 链编码, 其余基因均从 H 链编码 (图 1)。各基因的排列顺序与平疣桑椹石磺和凯尔特石磺<sup>[4]</sup> 相同。瘤背石磺含有高度紧凑的基因组, 大部分基因紧密相连或含有很少的基因间隔, 有的相邻基因还存在重叠区域。瘤背石磺线粒体共存在 7 处基因重叠, 其中 5 处为 tRNA 基因间的重叠。

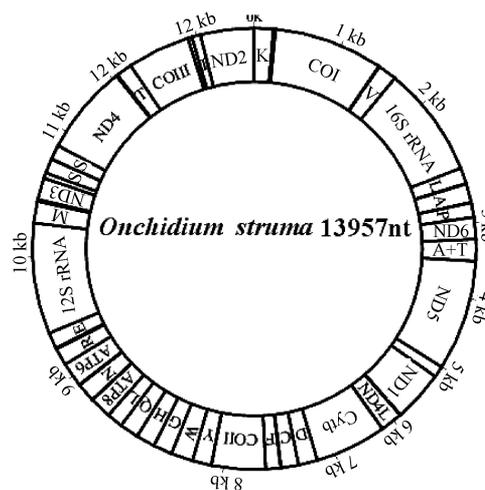


图 1 瘤背石磺线粒体基因组全序列简图

Fig. 1 Gene organization of *O. struma* mitochondrial genome

## 2.2 瘤背石磺 mtDNA 核苷酸组成

瘤背石磺线粒体 DNA 的碱基组成具有 A-T 偏好性<sup>[15]</sup>。其 A + T 含量为 64.67%, 4 种碱基的含量分别为 (A = 27.81%, T = 36.86%, C = 16.03%, G = 19.30%), 变化趋势与平疣桑椹石磺

和凯尔特石磺基本一致, T 含量最高而 C 含量最低。但瘤背石磺的 A + T 含量均高于凯尔特石磺的 59.32% 和平疣桑椹石磺的 62.99%。这种现象也普遍存在于蛋白质编码基因、tRNA 和 rRNA 基因, 尤其在 A + T 富集区表现更为明显(表 1)。

表 1 瘤背石磺、凯尔特石磺和平疣桑椹石磺 mtDNA 碱基组成比较

Tab. 1 Comparison of the base composition of *O. struma*, *O. celtica* and *P. mortoni* mitochondrial genome

碱基组成 base composition		物种 species		
		瘤背石磺 <i>O. struma</i>	凯尔特石磺 <i>O. celtica</i>	平疣桑椹石磺 <i>P. mortoni</i>
基因组全序列 complete genome	长度(bp) size AT(%)	13 957 64.67	14 150 59.32	13 991 62.99
13 个蛋白质基因 13 protein-coding	长度(bp) size AT(%)	10 366 65.01	10 692 59.08	10 557 63.11
22 个 tRNA 基因 22 tRNA gene	长度(bp) size AT(%)	1 693 64.70	1 438 58.31	1 579 60.79
2 个 rRNA 基因 2 rRNA gene	长度(bp) size AT(%)	1 760 63.85	1 764 61.79	1 737 63.86
非编码区 D-loop region	长度(bp) size AT(%)	138 70.59	173 64.29	118 73.91

## 2.3 瘤背石磺线粒体编码蛋白基因比较

在瘤背石磺线粒体基因组中, 参与蛋白质编码的核苷酸共 10 366 bp, 占全序列的 74.27%, 其中 26.95% A, 15.78% C, 19.21% G, 38.06% T, A + T = 65.01%, C + G = 34.99%。瘤背石磺蛋白质编码基因的起始密码子中, 除 *ND2* 使用 TTG 以外, 其余均为典型的起始密码子 ATN。所使用的终止密码子中, 13 个蛋白质基因中有 11 个是普通的 TAA 或 TAG, 而 *Cyrb* 和 *CO III* 则含有不完整的终止密码子, 这种现象在后生动物线粒体中比较常见<sup>[16]</sup>。其转录产物的 3' 末端为 U 或 UA, 可在转录后加工过程中加接多聚腺苷酸尾巴, 从而形成完整的 UAA 终止密码子<sup>[17]</sup>。

虽然瘤背石磺 mtDNA 蛋白质编码基因在序列上存在很大差异, 将其氨基酸序列进行排列比较时, 却发现密码子变化的一个很大部分是核苷酸的改变而编码的氨基酸种类并未发生变化, 即所谓“同义取代”。这种保守性的核苷酸变化主要发生在密码子的第三位上(表 2)。在密码子的第三位核苷酸的选择上, 线粒体基因组存在偏爱, A 或 T 的使用率远高于其他两种核苷酸, 在瘤背石磺线粒体基因组中, 此比率为 A = 32.88%, G = 16.12%, T = 33.32%, C = 12.68%, 这与线粒体细胞器中含有大量的 ATP 是一致的<sup>[18]</sup>。相比于核编码的一些单拷贝基因, 线粒体 DNA 进化

速度是比较快的, 因此可认为“同义取代”是线粒体在进化速度比较快的情况下维持其蛋白质结构与功能稳定性的一种方式<sup>[19]</sup>。

表 2 瘤背石磺、凯尔特石磺和平疣桑椹石磺 mtDNA 13 个蛋白质编码基因比较

Tab. 2 Comparison of the protein coding genes of *O. struma*, *O. celtica* and *P. mortoni* mitochondrial genome

蛋白 亚基 protein subunit	基因大小(起始密码子/终止密码子) genes size( initiation codons/termination codons)		
	瘤背石磺 <i>O. struma</i>	凯尔特石磺 <i>O. celtica</i>	平疣桑椹石磺 <i>P. mortoni</i>
<i>COI</i>	1494 (ATT/TAG)	1527 (TTG/TAG)	1494 (ATT/TAA)
<i>ND6</i>	468 (ATT/TAA)	465 (TTG/TAA)	468 (ATT/TAA)
<i>ND5</i>	1518 (ATT/TAG)	1643 (GTG/TAG)	1536 (ATG/TAG)
<i>ND1</i>	768 (ATT/TAA)	906 (TTG/TAA)	882 (ATT/TAA)
<i>ND4L</i>	285 (ATT/TAA)	268 (ATG/T)	288 (ATT/TAA)
<i>Cyrb</i>	1111 (ATT/T)	1122 (ATT/TAA)	1099 (ATT/T)
<i>COII</i>	666 (ATG/TAA)	681 (TTG/TAA)	666 (ATG/TAA)
<i>ATP8</i>	134 (ATT/TAA)	147 (ATG/TAA)	123 (ATT/TAA)
<i>ATP6</i>	632 (ATG/TAA)	645 (TTG/TAG)	645 (ATG/TAA)
<i>ND3</i>	279 (ATA/TAA)	352 (ATG/T)	279 (ATT/TAA)
<i>ND4</i>	1305 (ATG/TAA)	1308 (GTG/TAA)	1326 (ATG/TAA)
<i>COIII</i>	778 (ATG/T)	778 (ATG/T)	810 (ATA/TAA)
<i>ND2</i>	928 (TTG/TAA)	922 (ATG/T)	922 (GTG/T)

在所编码的 3 352 个氨基酸中,Leu(15.80%)、Ser(9.84%)、Phe(8.13%)和 Val(7.00%)为瘤背石磺常用的氨基酸,占有所有氨基酸的 42.45%;其它 2 种石磺的常用氨基酸与之相同。瘤背石磺最不用

的氨基酸为 Asp,而凯尔特石磺和平疣桑椹石磺最常用的氨基酸则分别为谷氨酰胺(Gln)和组氨酸(His)。瘤背石磺最常用密码子为 UUA,其次为 UUU,最不用密码子为 CGC(表 3)。

表 3 瘤背石磺、凯尔特石磺和平疣桑椹石磺 mtDNA 13 个蛋白质基因氨基酸和密码子偏好性比较

Tab.3 Comparison of the 13 protein amino acids and codon usage of *O. struma*, *O. celtica* and *P. mortoni* mitochondrial genome

氨基酸和密码子偏好性 amino acids and codon bias		物种 specis		
		瘤背石磺 <i>O. struma</i>	凯尔特石磺 <i>O. celtica</i>	平疣桑椹石磺 <i>P. mortoni</i>
氨基酸偏好性 amino acids bias	氨基酸总数(ind) total size	3 352	3 458	3 421
	最常用氨基酸(ind) the highest frequency amino acids	Leu/546	Leu/519	Leu /459
	百分率(%) percent	15.80	15.01	13.42
	第二常用氨基酸(ind) the second highest frequency amino acids	Ser/340	Ser/457	Ser /385
	百分率(%) percent	9.84	13.22	11.25
	最不用氨基酸(ind) the lowest frequency amino acids	Asp/57	Gln/61	His/59
	百分率(%) percent	1.65	1.76	1.72
密码子偏好性 codon usage bias	密码子总数(ind) total size	3 455	3564	3 519
	最常用密码子(ind) the highest frequency codon	UUA/234	UUU/192	UUU/223
	百分率(%) percent	6.77	5.39	6.52
	最不用密码子(ind) the lowest frequency codon	CGC/2	CGC /16	CGC /6
	百分率(%) percent	0.06	0.45	0.17

## 2.4 tRNA 基因

瘤背石磺 mtDNA 序列与其它物种一样,也包含 22 个 tRNA 基因,长度在 60 ~ 71 bp 变化。除亮氨酸(Leu)和丝氨酸(Ser)所对应的 tRNA 有 2 个以外,其他的氨基酸都只有一个 tRNA 与之对应。21 个 tRNA 均呈有规则的三叶草结构,仅 tRNA-Ser 无 D 臂,由 7 个核苷酸取代了该位置,这是后生动物线粒体基因组的一个普遍特征。根据摆动学说(wobble hypothesis) tRNA 反密码子 5'端摆动位上的 U 可与 mRNA 密码子 3'端的 A 和 G 两种碱基配对,在线粒体遗传系统中,甚至还可与 U、C 配对从而识别 4 种密码子。在预测的 tRNA 二级结构中总共出现了 31 对碱基 GU 的错配,这种错配现象在氨基酸接受臂、二氢尿嘧啶臂(DHU)及反密码子臂上均有出现。有人认为线粒体基因组 tRNA 基因的部分错配可以通过 RNA 编辑校正,不会引起氨基酸转运上的障碍<sup>[20]</sup>。

在臂和环的长度上,DHU 臂和 T $\psi$ C 臂在所有 tRNA 中变化较大,氨基酸接受臂和反密码子臂相对比较保守,分别为稳定的 7 对和 5 对碱基。但也存在例外情况,在所预测到的 tRNA<sup>Ala</sup>

(TGC)和 tRNA<sup>Asp</sup>(GTC)二级结构中,氨基酸接受臂上的碱基 U 不能配对从而形成凸环(bulge);另有一处异常情况是 tRNA<sup>Ser</sup>和 tRNA<sup>Thr</sup>反密码子环上为 9 个碱基,而不是通常的 7 个碱基。凯尔特石磺和平疣桑椹石磺均有类似情况出现。另外,除反密码子环为标准的 7 个碱基外,DHU 环和 T $\psi$ C 环的核苷酸数变化很大,长度为 2 ~ 13 bp 不等。瘤背石磺 mtDNA 的 tRNA 基因的反密码子序列与凯尔特石磺和平疣桑椹石磺相同,可能与它们的 mtDNA 利用相同的遗传密码子有关。

## 2.5 rRNA 基因和 A + T 富集区

瘤背石磺线粒体基因组上的 rRNA 基因具有较高的保守性。12S rRNA 基因位于 tRNA<sup>Glu</sup>和 tRNA<sup>Met</sup>之间,长度为 712 bp,而凯尔特石磺和平疣桑椹石磺长度分别为 708 和 695 bp;其 5'端和 3'端的确定,除了与近源物种进行同源序列比对之外,还因为它的 3'端存在一个茎环结构。CANNONE 等<sup>[21]</sup>认为 12S rRNA 和类似 12S rRNA 的二级结构都具有该茎环结构的存在。16S rRNA 基因位于 tRNA<sup>Val</sup>和 tRNA<sup>Leu</sup>之间,长度为 1 048 bp,介于凯尔特石磺(1 056 bp)和平疣桑

榅石磺(1 042 bp)之间;其5'端和3'端的确定,除了与近源物种进行同源序列比对之外,还因为它的3'端存在一个多聚腺苷酸位点。

A+T富集区是后生动物线粒体基因组中的一段非编码区,由于其富含A+T,称为A+T富集区<sup>[22]</sup>。瘤背石磺线粒体DNA的非编码区位于ND6和ND5基因之间,长度为138 bp。其最大非编码区的长度介于凯尔特石磺和平疣桑榅石磺之间。在近缘种类之间A+T富集区的长度差异则主要是由于串联重复拷贝的数目不同,虽然也有插入和缺失的存在,但影响相对较小<sup>[23]</sup>。另外,通过RNA structure 4.6预测其二级结构,发现这段序列存在一个类似于tRNA的二级结构(图2),它可能作为一种转录加工的信号<sup>[24]</sup>。

## 2.6 系统进化分析

以软体动物门多板纲(Polyplacophora)的半隐石鳖(*Katharina tunicata*)作为外群,使用NJ、ME、MP和UPGMA法构建系统进化树,结果都显示了相同的进化拓扑结构,图5是其中的MP进化树。从图中可以看出,6种腹足纲软体动物

分为3支,肺螺亚纲中的平疣桑榅石磺与瘤背石磺先聚为一支,然后与凯尔特石磺并为一支,3种石磺与菟齿菊花螺聚为一小支后再与后鳃亚纲的欧洲多角海牛聚为一大支;它们最后与前鳃亚纲澳大利亚黑鲍聚合在一起。

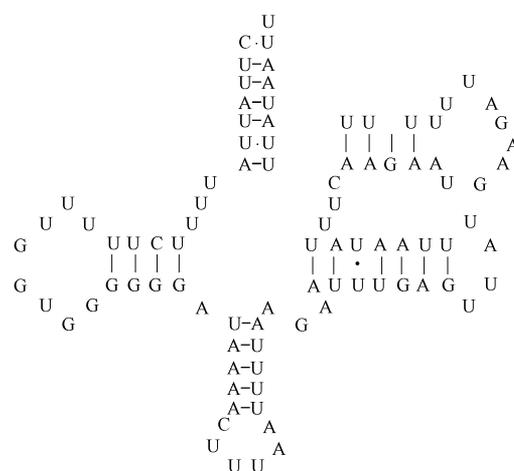


图2 瘤背石磺A+T富集区二级结构示意图  
Fig. 2 The secondary structure in the control region of the mt genome of *O. struma*



图3 基于13个蛋白质编码基因序列的腹足纲动物MP系统树  
Fig. 3 MP tree based on the 13 protein coding genes of gastropod

## 3 讨论

### 3.1 瘤背石磺线粒体基因组结构特点

本研究得到的瘤背石磺线粒体基因组与另外两种石磺科贝类(凯尔特石磺和平疣桑榅石磺)基因排列和基因序列水平上具有高度相似性,显示石磺科贝类线粒体基因组进化上的高度保守性。瘤背石磺线粒体基因组中一个显著的特征是蛋白质基因ND5、ND1、ND4L和Cytb紧密相连,没有tRNA基因的间隔,这也说明存在于蛋白质编码基因间的tRNA二级结构是多顺反子在转录过程中不进行蛋白质编码的间隔序列形成的<sup>[25]</sup>。

控制区(A+T富集区)是mtDNA中的非编码序列,在进化过程中选择压力相对较小,较线粒体其他区域具有更高的多态性,其碱基替换率是mtDNA其它区段的3~5倍,是线粒体DNA在长度和序列上最容易发生变异的区域,不同种类动物线粒体DNA大小的差异主要是由该区域的变化造成的<sup>[24]</sup>。三种石磺科贝类的控制区长度均在100~200 bp,其较短的控制区也是石磺科贝类线粒体基因组较小的主要原因。

线粒体tRNA基因同细胞质中的tRNA相比有许多不同之处,特别是其各个臂上均有高比例的碱基错配,除氨基酸接受臂和反密码子臂相对

比较保守外, DHU 臂和 T $\psi$ C 臂在所有 tRNA 中变化较大。瘤背石磺线粒体 tRNA 同样存在大量的碱基错配, 而且 DHU 环和 T $\psi$ C 环碱基数目变化较大<sup>[26]</sup>, 故推测这些线粒体 tRNA 三叶草结构的变化与物种的进化相关联。一般线粒体 DNA 只编码 22 个 tRNA 基因, 在这种情况下, 一种 tRNA 反密码子为了识别同一种氨基酸的一种以上的密码子, 氨基酸密码子和其 tRNA 反密码子配对应该具有更强的摆动性, 因为根据摇摆配对学说, 使用通用密码子的有机体至少具备 32 种 tRNA 分子<sup>[27]</sup>。在分析中发现, 氨基酸密码子具有偏好性, 但是这种偏好性与其 tRNA 携带的反密码子并不一致, 这可能是一种进化的结果, 因为在 RNA 中常见 G:U 配对, 再者瘤背石磺线粒体蛋白编码基因序列中 A + T 含量较高, 可能由此导致了密码子也大多由 A 或 U 组成<sup>[27]</sup>。

### 3.2 石磺科的分类与系统进化研究

石磺科贝类被视为海洋贝类向陆地辐射生活研究的代表<sup>[28]</sup>, 能为贝类由海洋向陆地进军的过程研究提供重要参考<sup>[29]</sup>。石磺科系统分类是其他研究工作的基础, 国外石磺分类工作是从 1800 年 BUCHANAN<sup>[30]</sup> 首次命名第一种石磺——香蒲石磺 (*Onchidium typhae*) 开始, 至今, 石磺科属种的分类命名工作一直比较混乱。2005 年 BOUCHET<sup>[31]</sup> 发表了腹足纲新的分类系统, 并把石磺科贝类划分为 6 个属分别是 *Onchidina*、*Onchidella*、*Peronina*、*Platevindex*、*Paraoncidium*、*Onchidium*, 目前已被广泛认可。长期以来我国石磺科贝类的研究工作并没有得到应有的重视, 导致国内石磺科贝类的分类、命名工作滞后, 同物异名现象普遍, 模式种资料不详。国内外有关石磺科系统发育和分类工作仍有待深入研究, 2009 年沈和定<sup>[9]</sup> 从石磺科的外部形态、内部结构和生态习性及其线粒体基因部分序列 (如 *CO I*、16S rRNA) 对我国大陆沿海石磺科贝类做了比较系统的研究。

近年来, 人们开始利用线粒体全基因组序列进行软体动物系统学分析, 国际上常采用将线粒体基因组中 13 个蛋白编码基因或轻链上 12 个蛋白编码基因的氨基酸序列按顺序连接起来构建系统进化树, 分析种间种内的系统进化关系<sup>[32]</sup>。本文基于 6 种贝类线粒体基因组 13 个蛋白编码基因构建的系统进化树中反映的系统进化关系与传

统的形态学分类基本一致。瘤背石磺与平疣桑椹石磺先聚为一支, 提示平疣桑椹石磺与瘤背石磺的亲缘关系比与凯尔特石磺更近, 并选取莴苣菊花螺、欧洲多角海牛和澳大利亚黑鲍分别作为肺螺亚纲、后鳃亚纲和前鳃亚纲的代表与上述三种石磺构建系统进化树, 聚类结果与传统的腹足纲系统分类相一致, 进一步验证了基于线粒体蛋白质编码基因进行系统进化的可靠性。相信随着越来越多石磺科贝类线粒体基因组全序列的测定, 将为石磺科贝类乃至软体动物分子进化和分子鉴定研究提供更多的标记基因。

### 参考文献:

- [1] 沈和定, 陈汉春, 陈贤龙, 等. 石磺繁殖生物学的实验研究[J]. 水产学报, 2006, 30(6): 753-760.
- [2] 黄金田, 沈伯平, 王资生. 瘤背石磺的生态习性观察[J]. 海洋渔业, 2004, 26(2): 103-109.
- [3] 邱立言. 苏沪沿海瘤背石磺的形态和习性[J]. 动物学杂志, 1991, 26(3): 33-36.
- [4] ANNETTE K, ANGELA D, KERSTIN K, et al. From sea to land and beyond new insights into the evolution of euthyneuran Gastropoda (Mollusca) [J]. BMC Evolutionary Biology, 2008, 8: 57.
- [5] WOLSTENHOLME D R. Animal mitochondrial DNA: structure and evolution [J]. Int Rev Cytol, 1992, 141: 173-216.
- [6] MINDELL D P, SORENSON M D, DIMCHEFF D E M, et al. Interordina relationships of birds and other reptiles based on whole mitochondrial genomes [J]. Syst Biol, 1999, 48(1): 138-152.
- [7] KUMAZAWA Y, OTA H, NISHIDA M, et al. The complete nucleotide sequence of a snake (*Dinodon semicarinatus*) mitochondrial genome with two identical control regions [J]. Genetics, 1998, 150(1): 313-329.
- [8] BEAGLEY C T, OKIMOTO R, WOLSTENHOLME D R. The mitochondrial genome of the sea anemone *Metridium senile* (Cnidaria): introns, a paucity of tRNA genes, and a near-standard genetic code [J]. Genetics, 1998, 148(3): 1091-1108.
- [9] 沈和定. 中国大陆沿海石磺生物学实验研究及系统分类[D]. 上海: 上海海洋大学, 2009.
- [10] 郑润玲, 李家乐, 牛东红. 缢蛏线粒体 DNA 提取方法的比较 [J]. 分子科学学报, 2008, 10(24): 312-315.
- [11] ZHENG X D, WANG R C, XIAO S, et al. Genetic diversity in populations of *Sepiella maindroni* using

- 16S rRNA gene [J]. Sequ Analys High Technol Lett,2003,9(1):1-4.
- [12] ALTSCHUL S F, MADDEN T L, SCHAFFER A A, *et al.* Gapped BLAST and PSI-BLAST: A new generation of protein database search programs [J]. Nucleic Acids Res,1997,25(17):3389-3402.
- [13] LOWE T M, EDDY S R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence [J]. Nucleic Acids Res, 1997, 25 (5):955-964.
- [14] MATHEWS D H. RNA secondary structure analysis using RNA structure [J]. Current Protocol Bioinform,2006,13:12.6.1-12.6.14.
- [15] BROWN W M. The mitochondrial genome of animals in molecular evolutionary genetics (RJ MacIntyre) [J]. Plenum(NY),1985,95-130.
- [16] BIBB M J, VAN ETTEN R A, WRIGHT C T, *et al.* Sequence and gene organization of mouse mitochondrial DNA [J]. Cell,1981,26:167-180.
- [17] CLARY D O, WOLSTENHOLME D R. The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization and genetic code [J]. J Mol Evol,1985,22:252-271.
- [18] MINDE U D P, SORENSON M D, DIMCHEFF D E. An extra nucleotide is not translated in mitochondrial of some birds and turtles [J]. J Mol Evol,1998,15:1568-1571
- [19] 张晓梅,单祥年,施燕峰,等. 小鹿线粒体基因组全序列的测定和分析 [J]. 遗传, 2004, 26 (6):849-853.
- [20] 陈念,赵树进,韩丽萍. 银环蛇线粒体基因组全序列分析 [J]. 中国生物化学与分子生物学报,2008, 24(12):1177-1182.
- [21] CANNONE J J, SUBRAMANIAN S, SCHNARE M N, *et al.* The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs [J]. BMC Bioinformatics,2002,3:2.
- [22] HARRISON R G. Animal mitochondrial DNA as a genetic marker in population and evolutionary biology [J]. Trends Ecol Evol,1989,4(1):6-11.
- [23] GRANDE C, TEMPLADO J, ZZARDOYA R. Evolution of gastropod mitochondrial genome arrangements [J]. BMC Evolutionary Biology,2008, 8:61.
- [24] ZHANG D X, HEWITT G M. Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies [J]. Biochem Syst Ecol,1997,25:99-120.
- [25] KURABAYASHI A, UESHIMA R. Complete sequence of the mitochondrial DNA of the primitive Opisthobranch gastropod *Pupa strigosa*: systematic implication of the genome organization [J]. Molecular Biology and Evolution,2000,17:2.
- [26] NARDI F, CARAPELLI A, DALLAI R. The mitochondrial genome of the olivefly *Bactrocera oleae*: two haplotypes from distant geographical locations [J]. Insect Mol Biol, 2003, 12 (6): 605-611.
- [27] 孙乃恩,孙东旭,朱德煦. 分子遗传学 [M]. 南京: 南京大学出版社,1990:248-257
- [28] TILLIER S. A new mountain *Platevindex* from Philippine islands (Pulmonata: Onchidiidae) [J]. J Moll Stud,1983,12A:198-202.
- [29] ANNETTE K, ANGELA D, KERSTIN K, *et al.* From sea to land and beyond new insights into the evolution of euthyneuran Gastropoda (Mollusca) [J]. BMC Evolutionary Biology,2008,8:57.
- [30] BUCHANAN F. An account of the *Onchidium*, a new genus of the class of vermes, found in Bengal [J]. Transactions of the Linnean Society of London, 2008,5(1):132-134.
- [31] BOUCHET P, ROCROI P, Fryda J, *et al.* Classification and nomenclator of gastropod families [J]. Malacologia,2005,47:1-397.
- [32] GRANDE C, TEMPLADO J, ZARDOYA R. Molecular phylogeny of Euthyneura (Mollusca: Gastropoda) [J]. Mol Biol Evol,2004,21(2):303-313.

**Analysis of complete mitochondrial genome of *Onchidium struma***  
**( Mollusca : Gastropoda : Pulmonata : Onchidiidae )**

WEI Luan-luan, SHEN He-ding\* , ZHANG Yu, FANG Lei, ZHANG Kun-xia, CHEN Cheng  
( College of Fisheries and Life Science, Shanghai Ocean University, Shanghai 201306, China )

**Abstract:** The complete mitochondrial genome of the *Onchidium struma* is important for general molecular and evolutionary studies in Onchidiidae. Via long and accurate polymerase chain reaction ( LA-PCR ), a complete mitochondrial genome sequence of the *O. struma* was determined and analyzed. The results indicated that the complete mitochondrial genome of *O. struma* is a circular molecule of 13 957 nucleotides. The gene content includes 22 tRNAs, 2 rRNAs and 19 non-coding regions from 2 bp to 138 bp in size. Two protein genes and 4 tRNA genes are transcribed from L strand, and the other 11 protein genes and 18 tRNA genes are transcribed from H strand. All protein coding genes of the *O. struma* mitochondrial genome start with a typical ATN codon, except for *ND2* with TTG. The usual TAA or TAG termination codons were found for eleven protein-coding genes. However, the *CO III* and *Cytb* genes have an incomplete termination codon (T). With respect to the secondary structures of 22 tRNA genes, the absence of DHU arm of tRNA<sup>Ser</sup> was observed as in *O. struma* tRNA<sup>Ser</sup> and tRNA<sup>Thr</sup> with 9 bases in anticodon loop was found, compared with typical 7 bases in others. The longest non-coding has secondary structures similar to the tRNA. Based on the deduced amino acid sequences of 13 protein-coding genes, the phylogenetic tree was constructed with Neighbor-joining (NJ) and Maximum parsimony (MP) methods and the phylogenetic relationships of six gastropoda were discussed. The results demonstrated that *O. struma* and *Platevindex mortoni* have closer phylogenetic relationship.

**Key words:** *Onchidium struma*; mitochondrial genome; sequence analysis; phylogenetic tree

**Corresponding author:** SHEN He-ding. E-mail: hdshen@shou.edu.cn