

# 鲸类线粒体基因组特征分析及分子标记探讨

田美<sup>1</sup>, 申欣<sup>1\*</sup>, 孟学平<sup>1</sup>, 程汉良<sup>1</sup>

(1. 淮海工学院 海洋学院, 江苏 连云港 222005)

**摘要:** 综合分析鲸类 32 个物种的线粒体基因组全序列, 全面揭示了鲸类线粒体基因组的基本特征、蛋白质编码基因、选择压力和差异位点等。鲸类线粒体基因组均编码后生动物标准的 37 个基因。绝大多数鲸类线粒体基因组的基因排列顺序完全相同, 而且与脊椎动物线粒体基因组的典型排列一致。4 个类群(真露脊鲸属、须鲸属、原海豚属和宽吻海豚属)线粒体基因组的 13 个蛋白质编码基因的  $Ka/Ks$  比值介于 0 和 0.587 7 之间, 均低于 1, 显示为纯化(负)选择。其中, *cox3* 基因的  $Ka/Ks$  均值最低, 其次为 *cox1*, *cox2* 和 *cob* 基因, 说明这些基因承受较强的选择压力和功能束缚; 而 *nad6* 基因的  $Ka/Ks$  均值最高, 其次为 *atp8*, *nad2* 和 *atp6* 基因, 说明这些基因的选择压力较弱。从 32 种鲸类物种间和属内(真露脊鲸属、须鲸属、原海豚属和宽吻海豚属)线粒体基因组主编码基因(13 个蛋白质编码基因和 2 个核糖体 RNA 基因)和 D-loop 区的变异位点分析显示, 在鲸类群体遗传的研究中, *nad5*, *nad4* 和 *nad2* 基因是理想的分子标记, 可以作为 *cox1* 基因辅助的分子标记, 用于分析鲸类不同群体之间的遗传多样性, 为其生物多样性的保护及其生物资源的合理利用提供参考。

**关键词:** 鲸目; 线粒体基因组; 差异位点; 分子标记

中图分类号: Q915.875

文献标志码: A

文章编号: 0253-4193(2011)05-0104-11

## 1 引言

线粒体基因组, 作为核外遗传物质, 具有结构简单、基因排列紧凑和严格的母系遗传等特点, 在过去的十几年中已经被广泛应用于后生动物分子系统发育和群体遗传的研究<sup>[1-4]</sup>。鲸类(Whales)隶属于脊索动物门 Chordata 有头动物亚门 Craniata 哺乳动物纲 Mammalia 鲸目 Cetacea, 现生鲸类可划分为 2 个亚目: 须鲸亚目 Mysticeti 和齿鲸亚目 Odontoceti<sup>[5]</sup>。长期以来由于非法捕猎、栖息地生境的持续性恶化以及渔业误捕等众多因素的影响, 数量大幅度下降, 对鲸类的保护已引起国际社会和公众的广泛关注<sup>[6]</sup>。在 GenBank 线粒体基因组的数据库中, 目前有 32 条鲸类的线粒体基因组。本文在 32 条鲸

类线粒体基因组研究的基础上, 全面分析鲸类线粒体基因组的基本特征、蛋白质编码基因、选择压力和差异位点等。本工作对于鲸类的生物多样性的保护及其生物资源合理利用等诸多方面提供遗传信息。鲸类线粒体基因组差异位点的分析, 为寻找合适的分子标记提供参考。

## 2 材料和方法

### 2.1 数据获取

从 GenBank 线粒体基因组数据库中检索、下载得到 32 条鲸类线粒体基因组全序列见表 1。分别隶属须鲸亚目下的露脊鲸科 Balaenidae、须鲸科 Balaenopteridae、灰鲸科 Eschrichtiidae 和小露脊鲸科 Neobalaenidae 以及齿鲸亚目下的海豚科 Delphini-

收稿日期: 2010-05-18; 修订日期: 2011-04-18。

基金项目: 国家自然科学基金(40906067); 江苏省海洋生物技术重点实验室基金(2009HS13); 淮海工学院自然科学基金(Z2009048); 江苏省“青蓝工程”人才基金。

作者简介: 田美(1979—), 女, 山东省成武县人, 实验师, 研究方向: 线粒体基因组学。E-mail: genomeresearch@163.com

\* 通信作者: 申欣, 男, 副教授, 博士, 研究方向: 线粒体基因组学和蛋白质组学。E-mail: shenthin@163.com

dae、亚河豚科 Iniidae、白鳍豚科 Lipotidae、一角鲸科 Monodontidae、鼠豚科 Phocoenidae、抹香鲸科

Physeteridae、恒河豚科 Platanistidae、普拉塔河豚科 Pontoporiidae 和剑吻鲸科 Ziphiidae<sup>[7-16]</sup>(图 1)。

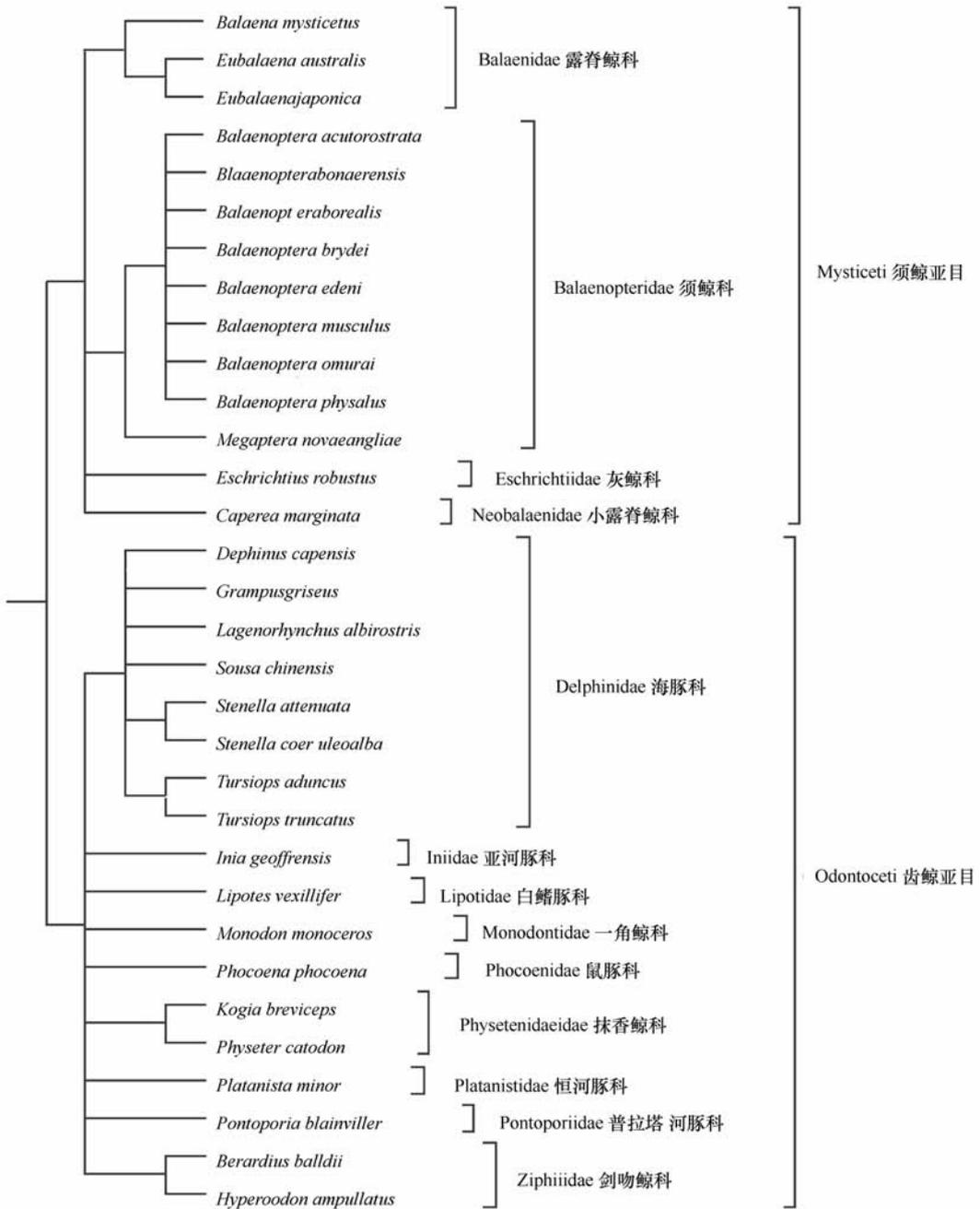


图 1 目前已测定线粒体基因组全序列的鲸类物种

## 2.2 同义替换率(Ks)和非同义替换率(Ka)分析

一般来讲,进化速率是受稳定性(负)选择、突变和定向(正)选择控制的。为了检验选择压力对于鲸类线粒体基因组的影响,以真露脊鲸属、须鲸属、原海豚属和宽吻海豚属为代表分析线粒体基因组中 13 个蛋白质编码基因的核酸序列通过 Clustal X<sup>[17]</sup>进行多

重序列比对。同义替换率(Ks)和非同义替换率(Ka)通过 PAML<sup>[18]</sup>和 DnaSP 4.10.7<sup>[19]</sup>进行了估算。

## 2.3 基因差异性分析

以 32 种鲸类线粒体基因组数据作为数据群进行了科间基因差异位点和基因变异分析,同时以真露脊鲸属、须鲸属、原海豚属和宽吻海豚属作为数据

群进行了属内基因差异位点和基因变异分析。13 个蛋白质编码基因、2 个核糖体 RNA 基因和 D-loop 区的核酸序列通过 Clustal X<sup>[17]</sup> 进行多重序列比对。然后通过 DnaSP 4.10.7<sup>[19]</sup> 分析了线粒体基因组主编码基因的差异位点和基因变异特征。

### 3 结果和讨论

#### 3.1 鲸类线粒体基因组的特征

鲸类 32 个物种线粒体基因组长度介于

16 324 bp(恒河豚)和 16 593 bp(普拉塔河豚)之间。32 个物种的线粒体基因组均编码后生动物标准的 37 个基因,包括 13 个蛋白质编码基因、22 个转运 RNA 基因和 2 个核糖体 RNA 基因(表 1)。绝大多数鲸类线粒体基因组的基因排列完全相同,而且与脊椎动物线粒体基因组的典型排列一致。线粒体基因组主编码链的 A+T 含量最高的为长吻真海豚 61.4%,最低的为小抹香鲸 56.8%。鲸类 32 个物种线粒体基因组的基本特征见表 1。

表 1 鲸类线粒体基因组的基本特征

物种名称	GenBank 接收号	线粒体基因组长度/bp	基因数目	A+T 含量 (%)
<i>Balaena mysticetus</i>	NC_005268	16 390	37	58.7
<i>Eubalaena australis</i>	NC_006930	16 385	37	58.8
<i>Eubalaena japonica</i>	NC_006931	16 386	37	58.9
<i>Balaenoptera acutorostrata</i>	NC_005271	16 417	37	59.6
<i>Balaenoptera bonaerensis</i>	NC_006926	16 421	37	59.3
<i>Balaenoptera borealis</i>	NC_006929	16 410	37	59.6
<i>Balaenoptera brydei</i>	NC_006928	16 408	37	59.7
<i>Balaenoptera edeni</i>	NC_007938	16 409	37	59.5
<i>Balaenoptera musculus</i>	NC_001601	16 402	37	59.4
<i>Balaenoptera omurai</i>	NC_007937	16 404	37	59.3
<i>Balaenoptera physalus</i>	NC_001321	16 398	37	59.4
<i>Megaptera novaeangliae</i>	NC_006927	16 398	37	59.2
<i>Eschrichtius robustus</i>	NC_005270	16 412	37	58.7
<i>Caperea marginata</i>	NC_005269	16 383	37	60.5
<i>Delphinus capensis</i>	NC_012061	16 385	37	61.4
<i>Grampus griseus</i>	NC_012062	16 386	37	60.9
<i>Lagenorhynchus albirostris</i>	NC_005278	16 393	37	61.3
<i>Sousa chinensis</i>	NC_012057	16 388	37	61.1
<i>Stenella attenuata</i>	NC_012051	16 386	37	61.3
<i>Stenella coeruleoalba</i>	NC_012053	16 384	37	61.2
<i>Tursiops aduncus</i>	NC_012058	16 385	37	61.2
<i>Tursiops truncatus</i>	NC_012059	16 388	37	61.1
<i>Inia geoffrensis</i>	NC_005276	16 588	37	58.8
<i>Lipotes vexillifer</i>	NC_007629	16 392	37	60.8
<i>Monodon monoceros</i>	NC_005279	16 383	37	59.8
<i>Phocoena phocoena</i>	NC_005280	16 382	37	59.4
<i>Kogia breviceps</i>	NC_005272	16 406	37	56.8
<i>Physeter catodon</i>	NC_002503	16 428	37	56.9
<i>Platanista minor</i>	NC_005275	16 324	37	59.8
<i>Pontoporia blainvillei</i>	NC_005277	16 593	37	59.6
<i>Berardius bairdii</i>	NC_005274	16 346	37	59.7
<i>Hyperoodon ampullatus</i>	NC_005273	16 346	37	60.4

### 3.2 鲸类线粒体基因组的蛋白质编码基因

在鲸类线粒体基因组的蛋白质编码基因中,对于 *atp6*, *atp8*, *cob*, *cox1*, *cox2*, *cox3*, *nad1*, *nad4* 和 *nad6* 基因,32 个鲸类线粒体基因组均以 ATG 作为起始密码子;对于 *nad2* 基因,除了白暨豚和普拉塔河豚以 ATC 起始外,余下的线粒体基因组以 ATA 或者 ATT 作为起始密码子;对于 *nad3* 基因,32 个鲸类线粒体基因组均以 ATA 作为起始密码子;对于 *nad4L* 基因,存在 GTG 和 ATG 两种起始密码子;对于 *nad5* 基因,除了白暨豚以 ATC 起始外,余下的线粒体基因组全部以 ATA 作为起始密码子(见表 2)。

而在终止密码子方面,对于 *atp6* 基因,除了抹香鲸以不完全终止密码子“T—”终止以外,余下的线粒体基因组均以 TAA 作为终止密码子;对于 *atp8* 基因,存在 TAG 和 TAA 两种终止密码子;对于 *cob* 基因,32 个鲸类线粒体基因组均以 AGA 作为终止密码子;对于 *cox1* 基因,除了鼠海豚和恒河豚分别以 AGG 和 TAA 终止外,余下的线粒体基因组均以 AGA 作为终止密码子;对于 *cox2*, *nad4L*, *nad5* 和 *nad6* 基因,32 个鲸类线粒体基因组均以 TAA 作为终止密码子;对于 *cox3* 基因,白暨豚以 TAA 作为终止密码子,同时,海豚科的 7 个物种(长吻真海豚、瑞氏灰海豚、印度太平洋驼背豚、点斑原海豚、条纹原海豚、印太瓶鼻海豚和瓶鼻海豚)以不完全终止密码子“TA—”终止,余下的 24 个线粒体基因组均以 TAG 作为终止密码子;对于 *nad1* 基因,鼠海豚和北瓶鼻鲸以 TAG 作为终止密码子,抹香鲸和普拉塔河豚分别以“T—”和 AGA 终止,余下的 28 个线粒体基因组均以 TAA 作为终止密码子;对于 *nad2* 基因,存在 TAG 和“T—”两种终止密码子;对于 *nad3* 基因,存在“TA—”和“T—”两种终止密码子;对于 *nad4* 基因,32 个鲸类线粒体基因组均以“T—”作为终止密码子(见表 2)。

鲸类线粒体基因组中的 8 个蛋白质编码基因(*atp6*, *cob*, *cox2*, *cox3*, *nad2*, *nad3*, *nad4* 和 *nad4L*) 编码氨基酸数目相同,分别为 226, 379, 227, 261, 347, 115, 459 和 98 个;其余 5 个蛋白质编码基因编码的氨基酸数目稍有差异(见表 2)。鲸类 32 个线粒体基因组蛋白质编码基因所编码氨基酸的数目及起始、终止密码子见表 2。

### 3.3 同义替代和非同义替代

为了分析鲸类线粒体基因组蛋白编码基因的选择压力,统计了真露脊鲸属、须鲸属、原海豚属和宽吻海

豚属的 13 个蛋白编码基因的同义替换率( $K_s$ )和非同义替换率( $K_a$ )。4 个类群线粒体基因组的 13 个蛋白质编码基因的  $K_a/K_s$  比值介于 0 和 0.587 7 之间,都远远低于 1,显示为纯化(负)选择(见图 2)。其中, *cox3* 基因的  $K_a/K_s$  均值最低,其次为 *cox1*, *cox2* 和 *cob* 基因,说明这些基因承受较强的选择压力和功能束缚;而 *nad6* 基因的  $K_a/K_s$  均值最高,其次为 *atp8*, *nad2* 和 *atp6* 基因,说明这些基因的选择压力较弱。

### 3.4 鲸类线粒体基因组的差异位点分析

在种群遗传的研究中,分子标记的选择是至关重要的,以 32 种鲸类线粒体基因组数据作为数据群进行了科间基因差异位点和基因变异分析,同时以真露脊鲸属、须鲸属、原海豚属和宽吻海豚属作为数据群进行了属内基因差异位点和基因变异分析。32 种鲸类线粒体基因组 13 个蛋白质编码基因、2 个核糖体 RNA 基因和 D-loop 区的差异位点分析详见表 3。从分析的结果可以看出,在 15 个主编码基因和 D-loop 区中, *lrRNA* 基因最为保守,差异位点的比例为 29.08%,其次为 *srRNA* 基因和 *cox1* 基因,差异位点比例分别为 34.43%和 39.49%。5 个基因(*cob*, *cox2*, *cox3*, *nad1* 和 *nad4*)差异位点的比例介于 40%~50%之间;7 个基因(*atp6*, *atp8*, *nad2*, *nad3*, *nad4L*, *nad5* 和 *nad6*)差异位点的比例介于 50%~60%之间;D-loop 区差异位点的比例最高,达到了 64.57%(见表 3)。差异位点数最多的基因为 *nad5* 基因,达到 969 个,其次为 *nad4* 和 *cox1* 基因,分别为 680 和 609 个。因此,在鲸类群体遗传的研究中, *nad5* 和 *nad4* 基因可以作为 *cox1* 基因辅助的分子标记。

从须鲸属线粒体基因组的差异位点分析,可以看出, *srRNA* 和 *lrRNA* 基因最为保守,差异位点的比例分别为 10.22%和 11.39%;差异位点数最多的基因是 *nad5* 基因,为 402 个,其次是 *nad4* 和 *cox1* 基因,分别为 309 和 270 个(见表 4)。因此,在须鲸群体遗传的研究中, *nad5* 和 *nad4* 基因可以作为 *cox1* 基因辅助的分子标记。

在真露脊鲸属线粒体基因组中, *cox3* 基因最为保守,差异位点的比例为 0.38%,其次为 *atp8* 基因和 *lrRNA* 基因,差异位点的比例分别为 0.53%和 0.64%。差异位点数最多的基因为 *nad5* 基因(26 个),其次为 *nad4* 基因、*nad2* 基因和 D-loop 区,差异位点数均为 20 个(见表 5)。因此,在真露脊鲸群体遗传的研究中, *nad5*, *nad4*, *nad2* 基因和 D-loop 区可以作为 *cox1* 基因辅助的分子标记。

表 2 鲸类线粒体基因组 13 个蛋白质编码基因的氨基酸长度及起始、终止密码子

物种名称	接收号	atp6	atp8	cob	cox1	cox2	cox3	nad1	nad2	nad3	nad4	nad4L	nad5	nad6
<i>Balaena mysticetus</i> 弓头鲸	No.	226	63	379	516	227	261	318	347	115	459	98	606	175
	I codon	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATA	ATA	ATG	GTG	ATA	ATG
	T codon	TAA	TAG	AGA	AGA	TAA	TAG	TAA	TAG	T-	T-	TAA	TAA	TAA
<i>Eubalaena australis</i> 南露脊鲸	No.	226	63	379	516	227	261	318	347	115	459	98	606	175
	I codon	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATA	ATA	ATG	GTG	ATA	ATG
	T codon	TAA	TAA	AGA	AGA	TAA	TAG	TAA	TAG	TA-	T-	TAA	TAA	TAA
<i>Eubalaena japonica</i> 北太平洋露脊鲸	No.	226	63	379	516	227	261	318	347	115	459	98	606	175
	I codon	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATA	ATA	ATG	GTG	ATA	ATG
	T codon	TAA	TAA	AGA	AGA	TAA	TAG	TAA	TAG	TA-	T-	TAA	TAA	TAA
<i>Balaenoptera acutorostrata</i> 小须鲸	No.	226	63	379	516	227	261	318	347	115	459	98	606	175
	I codon	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATA	ATA	ATG	GTG	ATA	ATG
	T codon	TAA	TAA	AGA	AGA	TAA	TAG	TAA	TAG	T-	T-	TAA	TAA	TAA
<i>Balaenoptera bonaerensis</i> 南极小须鲸	No.	226	63	379	516	227	261	318	347	115	459	98	606	175
	I codon	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATA	ATA	ATG	GTG	ATA	ATG
	T codon	TAA	TAA	AGA	AGA	TAA	TAG	TAA	TAG	TA-	T-	TAA	TAA	TAA
<i>Balaenoptera borealis</i> 塞鲸	No.	226	63	379	516	227	261	318	347	115	459	98	606	175
	I codon	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATA	ATA	ATG	GTG	ATA	ATG
	T codon	TAA	TAA	AGA	AGA	TAA	TAG	TAA	TAG	TA-	T-	TAA	TAA	TAA
<i>Balaenoptera brydei</i> 布氏鲸	No.	226	63	379	516	227	261	318	347	115	459	98	606	175
	I codon	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATA	ATA	ATG	GTG	ATA	ATG
	T codon	TAA	TAA	AGA	AGA	TAA	TAG	TAA	TAG	TA-	T-	TAA	TAA	TAA
<i>Balaenoptera edeni</i> 髯鲸	No.	226	63	379	516	227	261	318	347	115	459	98	606	175
	I codon	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATA	ATA	ATG	GTG	ATA	ATG
	T codon	TAA	TAA	AGA	AGA	TAA	TAG	TAA	TAG	TA-	T-	TAA	TAA	TAA

续表 2

物种名称	接收号	atp6	atp8	cob	cox1	cox2	cox3	nad1	nad2	nad3	nad4	nad4L	nad5	nad6
<i>Balaenoptera musculus</i> 蓝鲸	No.	226	63	379	516	227	261	318	347	115	459	98	606	175
	I codon	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATA	ATA	ATG	GTG	ATA	ATG
	T codon	TAA	TAA	AGA	AGA	TAA	TAG	TAA	TAG	T-	T-	TAA	TAA	TAA
<i>Balaenoptera omurai</i> 角岛鲸	No.	226	63	379	516	227	261	318	347	115	459	98	606	175
	I codon	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATA	ATA	ATG	GTG	ATA	ATG
	T codon	TAA	TAA	AGA	AGA	TAA	TAG	TAA	TAG	TA-	T-	TAA	TAA	TAA
<i>Balaenoptera physalus</i> 长须鲸	No.	226	63	379	516	227	261	318	347	115	459	98	606	175
	I codon	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATA	ATA	ATG	ATG	ATA	ATG
	T codon	TAA	TAA	AGA	AGA	TAA	TAG	TAA	TAG	T-	T-	TAA	TAA	TAA
<i>Megaptera novaeangliae</i> 座头鲸	No.	226	63	379	516	227	261	318	347	115	459	98	606	175
	I codon	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATA	ATA	ATG	ATG	ATA	ATG
	T codon	TAA	TAA	AGA	AGA	TAA	TAG	TAA	TAG	TA-	T-	TAA	TAA	TAA
<i>Eschrichtius robustus</i> 灰鲸	No.	226	63	379	516	227	261	318	347	115	459	98	606	175
	I codon	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATT	ATA	ATG	GTG	ATA	ATG
	T codon	TAA	TAA	AGA	AGA	TAA	TAG	TAA	TAG	T-	T-	TAA	TAA	TAA
<i>Caperea marginata</i> 小露脊鲸	No.	226	63	379	516	227	261	318	347	115	459	98	606	175
	I codon	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATA	ATA	ATG	GTG	ATA	ATG
	T codon	TAA	TAA	AGA	AGA	TAA	TAG	TAA	TAG	T-	T-	TAA	TAA	TAA
<i>Delphinus capensis</i> 长吻真海豚	No.	226	63	379	516	227	261	318	347	115	459	98	606	175
	I codon	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATT	ATA	ATG	ATG	ATA	ATG
	T codon	TAA	TAA	AGA	AGA	TAA	TA-	TAA	T-	T-	T-	TAA	TAA	TAA
<i>Grampus griseus</i> 瑞氏灰海豚	No.	226	63	379	516	227	261	318	347	115	459	98	606	175
	I codon	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATT	ATA	ATG	ATG	ATA	ATG
	T codon	TAA	TAA	AGA	AGA	TAA	TA-	TAA	T-	T-	T-	TAA	TAA	TAA
<i>Lagenorhynchus albirostris</i> 白喙斑纹海豚	No.	226	63	379	516	227	261	318	347	115	459	98	606	175
	I codon	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATT	ATA	ATG	ATG	ATA	ATG
	T codon	TAA	TAA	AGA	AGA	TAA	TAG	TAA	TAG	T-	T-	TAA	TAA	TAA
<i>Sousa chinensis</i> 印度太平洋驼背豚	No.	226	63	379	516	227	261	318	347	115	459	98	606	175
	I codon	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATT	ATA	ATG	ATG	ATA	ATG
	T codon	TAA	TAA	AGA	AGA	TAA	TA-	TAA	T-	TA-	T-	TAA	TAA	TAA

续表 2

物种名称	接收号	atp6	atp8	cob	cox1	cox2	cox3	nad1	nad2	nad3	nad4	nad4L	nad5	nad6
<i>Stenella attenuata</i> 点斑原海豚	No.	226	63	379	516	227	261	318	347	115	459	98	606	175
	I codon	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATT	ATA	ATG	ATG	ATA	ATG
	T codon	TAA	TAA	AGA	AGA	TAA	TA-	TAA	T-	T-	T-	TAA	TAA	TAA
<i>Stenella coeruleoalba</i> 条纹原海豚	No.	226	63	379	516	227	261	318	347	115	459	98	606	175
	I codon	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATT	ATA	ATG	ATG	ATA	ATG
	T codon	TAA	TAA	AGA	AGA	TAA	TA-	TAA	T-	T-	T-	TAA	TAA	TAA
<i>Tursiops aduncus</i> 印太瓶鼻海豚	No.	226	63	379	516	227	261	318	347	115	459	98	606	175
	I codon	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATT	ATA	ATG	ATG	ATA	ATG
	T codon	TAA	TAA	AGA	AGA	TAA	TA-	TAA	T-	T-	T-	TAA	TAA	TAA
<i>Tursiops truncatus</i> 瓶鼻海豚	No.	226	63	379	516	227	261	318	347	115	459	98	606	175
	I codon	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATT	ATA	ATG	ATG	ATA	ATG
	T codon	TAA	TAA	AGA	AGA	TAA	TA-	TAA	T-	T-	T-	TAA	TAA	TAA
<i>Inia geoffrensis</i> 亚马孙海豚	No.	226	66	379	516	227	261	318	347	115	459	98	606	175
	I codon	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATT	ATA	ATG	ATG	ATA	ATG
	T codon	TAA	TAA	AGA	AGA	TAA	TAG	TAA	TAG	T-	T-	TAA	TAA	TAA
<i>Lipotes vexillifer</i> 白暨豚	No.	226	63	379	516	227	261	318	347	115	459	98	606	175
	I codon	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATC	ATA	ATG	ATG	ATC	ATG
	T codon	TAA	TAA	AGA	AGA	TAA	TAA	TAA	TAG	TA-	T-	TAA	TAA	TAA
<i>Monodon monoceros</i> 一角鲸	No.	226	66	379	516	227	261	318	347	115	459	98	606	175
	I codon	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATT	ATA	ATG	ATG	ATA	ATG
	T codon	TAA	TAA	AGA	AGA	TAA	TAG	TAA	TAG	T-	T-	TAA	TAA	TAA
<i>Phocoena phocoena</i> 鼠海豚	No.	226	63	379	517	227	261	318	347	115	459	98	606	175
	I codon	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATT	ATA	ATG	ATG	ATA	ATG
	T codon	TAA	TAG	AGA	AGG	TAA	TAG	TAG	TAG	T-	T-	TAA	TAA	TAA
<i>Kogia breviceps</i> 小林香鲸	No.	226	66	379	516	227	261	318	347	115	459	98	606	175
	I codon	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATA	ATA	ATG	GTG	ATA	ATG
	T codon	TAA	TAG	AGA	AGA	TAA	TAG	TAA	TAG	T-	T-	TAA	TAA	TAA
<i>Physeter catodon</i> 抹香鲸	No.	226	63	379	516	227	261	318	347	115	459	98	606	175
	I codon	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATA	ATA	ATG	ATG	ATA	ATG
	T codon	T-	TAA	AGA	AGA	TAA	TAG	T-	T-	T-	T-	TAA	TAA	TAA

续表 2

物种名称	接收号	atp6	atp8	cob	cox1	cox2	cox3	nad1	nad2	nad3	nad4	nad4L	nad5	nad6
<i>Platanista minor</i> 恒河豚	No.	226	66	379	514	227	261	318	347	115	459	98	609	175
	I codon	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATA	ATA	ATG	GTG	ATA	ATG
	T codon	TAA	TAG	AGA	TAA	TAA	TAG	TAA	TAG	T-	T-	TAA	TAA	TAA
<i>Pontoporia blainvilliei</i> 普拉塔河豚	No.	226	63	379	516	227	261	319	347	115	459	98	606	173
	I codon	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATC	ATA	ATG	ATG	ATA	ATG
	T codon	TAA	TAA	AGA	AGA	TAA	TAG	AGA	TAG	T-	T-	TAA	TAA	TAA
<i>Berardius bairdii</i> 贝氏喙鲸	No.	226	66	379	516	227	261	318	347	115	459	98	606	175
	I codon	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATA	ATA	ATG	GTG	ATA	ATG
	T codon	TAA	TAG	AGA	AGA	TAA	TAG	TAA	TAG	T-	T-	TAA	TAA	TAA
<i>Hyperoodon ampul-</i> <i>latus</i> 北瓶鼻鲸	No.	226	65	379	516	227	261	318	347	115	459	98	606	175
	I codon	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATA	ATA	ATG	GTG	ATA	ATG
	T codon	TAA	TAA	AGA	AGA	TAA	TAG	TAG	TAG	TA-	T-	TAA	TAA	TAA

表 3 鲸类线粒体基因组的差异位点分析

基因名称	atp6	atp8	cob	cox1	cox2	cox3	nad1	nad2	nad3	nad4	nad4L	nad5	nad6	srRNA	trRNA	D-loop
总位点数*	678	186	1 137	1 542	681	783	954	1 041	345	1 377	294	1 815	519	941	1 527	748
不变位点数	319	78	603	933	391	446	524	489	165	697	139	846	222	617	1 083	265
差异位点数	359	108	534	609	290	337	430	552	180	680	155	969	297	324	444	483
单现突变数	63	30	109	103	48	71	96	141	33	152	26	221	88	101	136	147
简约位点数	296	78	425	506	242	266	334	411	147	528	129	748	209	223	308	336
差异位点比例(%)	52.95	58.06	46.97	39.49	42.58	43.04	45.07	53.03	52.17	49.38	52.72	53.39	57.23	34.43	29.08	64.57

注: \* 总位点数不包含插入和缺失位点。

表 4 须鲸属线粒体基因组的差异位点分析

基因名称	atp6	atp8	cob	cox1	cox2	cox3	nad1	nad2	nad3	nad4	nad4L	nad5	nad6	srRNA	trRNA	D-loop
总位点数*	678	189	1 137	1 548	681	783	954	1 041	345	1 377	294	1 818	525	969	1 571	923
不变位点数	512	144	897	1 278	548	648	775	814	261	1 068	216	1 416	412	870	1 392	743

续表 4

基因名称	<i>atp6</i>	<i>atp8</i>	<i>cob</i>	<i>cox1</i>	<i>cox2</i>	<i>cox3</i>	<i>nad1</i>	<i>nad2</i>	<i>nad3</i>	<i>nad4</i>	<i>nad4L</i>	<i>nad5</i>	<i>nad6</i>	<i>srRNA</i>	<i>trRNA</i>	D-loop
差异位点数	166	45	240	270	133	135	179	227	84	309	78	402	113	99	179	180
单现突变数	81	28	124	125	52	66	98	114	46	171	37	201	55	45	101	91
简约位点数	85	17	116	145	81	69	81	113	38	138	41	201	58	54	78	89
差异位点比例(%)	24.48	23.81	21.11	17.44	19.53	17.24	18.76	21.81	24.35	22.44	26.53	22.11	21.52	10.22	11.39	19.50

注：\* 总位点数不包含插入和缺失位点。

表 5 真露脊鲸属线粒体基因组的差异位点分析

基因名称	<i>atp6</i>	<i>atp8</i>	<i>cob</i>	<i>cox1</i>	<i>cox2</i>	<i>cox3</i>	<i>nad1</i>	<i>nad2</i>	<i>nad3</i>	<i>nad4</i>	<i>nad4L</i>	<i>nad5</i>	<i>nad6</i>	<i>srRNA</i>	<i>trRNA</i>	D-loop
总位点数*	678	189	1 137	1 548	681	783	954	1 041	345	1 377	294	1 818	525	973	1 574	918
不变位点数	670	188	1 119	1 537	664	780	944	1 021	338	1 357	291	1 792	514	964	1 564	898
差异位点数	8	1	18	11	17	3	10	20	7	20	3	26	11	9	10	20
差异位点比例(%)	1.18	0.53	1.58	0.71	2.50	0.38	1.05	1.92	2.03	1.45	1.02	1.43	2.10	0.92	0.64	2.18

注：\* 总位点数不包含插入和缺失位点。

表 6 原海豚属线粒体基因组的差异位点分析

基因名称	<i>atp6</i>	<i>atp8</i>	<i>cob</i>	<i>cox1</i>	<i>cox2</i>	<i>cox3</i>	<i>nad1</i>	<i>nad2</i>	<i>nad3</i>	<i>nad4</i>	<i>nad4L</i>	<i>nad5</i>	<i>nad6</i>	<i>srRNA</i>	<i>trRNA</i>	D-loop
总位点数*	678	189	1 137	1 548	681	783	954	1 041	345	1 377	294	1 818	525	974	1 577	912
不变位点数	650	177	1 097	1 501	661	755	922	994	329	1 318	280	1 748	503	966	1 561	878
差异位点数	28	12	40	47	20	28	32	47	16	59	14	70	22	8	16	34
差异位点比例(%)	4.13	6.35	3.52	3.04	2.94	3.58	3.35	4.51	4.64	4.28	4.76	3.85	4.19	0.82	1.01	3.73

注：\* 总位点数不包含插入和缺失位点。

表 7 宽吻海豚属线粒体基因组的差异位点分析

基因名称	<i>atp6</i>	<i>atp8</i>	<i>cob</i>	<i>cox1</i>	<i>cox2</i>	<i>cox3</i>	<i>nad1</i>	<i>nad2</i>	<i>nad3</i>	<i>nad4</i>	<i>nad4L</i>	<i>nad5</i>	<i>nad6</i>	<i>srRNA</i>	<i>trRNA</i>	D-loop
总位点数*	678	189	1 137	1 548	681	783	954	1 041	345	1 377	294	1 818	525	974	1 577	914
不变位点数	652	185	1 096	1 509	662	765	927	979	329	1 340	285	1 748	504	961	1 562	887
差异位点数	26	4	41	39	19	18	27	62	16	37	9	70	21	13	15	27
差异位点比例(%)	3.83	2.12	3.61	2.52	2.79	2.30	2.83	5.96	4.64	2.69	3.06	3.85	4.00	1.33	0.95	2.95

注：\* 总位点数不包含插入和缺失位点。

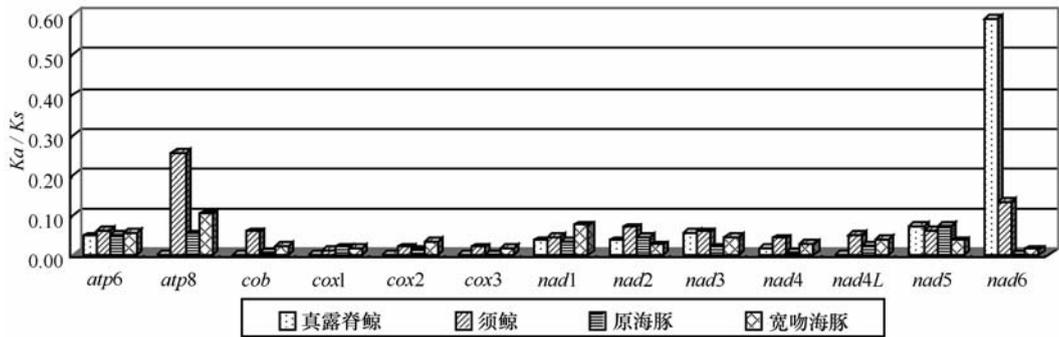


图2 真露脊鲸属、须鲸属、原海豚属和宽吻海豚属线粒体基因组13个蛋白质编码基因的  $Ka/Ks$  分析

从原海豚属线粒体基因组主编码基因的差异位点分析, *srRNA* 和 *lrRNA* 基因最为保守, 差异位点的比例分别为 0.82% 和 1.01%; 差异位点数最多的基因为 *nad5* 基因(70 个), 其次为 *nad4* 基因, 差异位点数达到 59 个(表 6)。因此, 在原海豚群体遗传的研究中, *nad5* 和 *nad4* 基因可以作为 *cox1* 基因辅助的分子标记。

在宽吻海豚属线粒体基因组中, *lrRNA* 和 *srRNA* 基因最为保守, 差异位点的比例分别为 0.95% 和 1.33%; 差异位点数最多的基因为 *nad5* 基因(70 个), 其次为 *nad2* 基因, 差异位点数达到 62 个(见表 7)。因此, 在宽吻海豚群体遗传的研究中, *nad5* 和 *nad2* 基因可以作为 *cox1* 基因辅助的分子标记。

综上所述, 在鲸类群体遗传的研究中, *nad5*, *nad4* 和 *nad2* 基因是理想的分子标记, 可以作为 *cox1* 基因辅助的分子标记, 用于分析鲸类不同群体之间的遗传多样性, 为其生物多样性的保护及合理利用其生物资源提供更多保障。

## 4 结论

本论文在 32 个鲸类线粒体基因组全序列的基

础上, 全面分析鲸类线粒体基因组的基本特征、蛋白质编码基因、选择压力和差异位点等。鲸类 32 个物种线粒体基因组长度介于 16 324 bp 和 16 593 bp 之间。32 个物种的线粒体基因组均编码后生动物标准的 37 个基因。鲸类线粒体基因组的基因排列完全相同, 而且与脊椎动物线粒体基因组的典型排列一致。4 个类群(真露脊鲸属、须鲸属、原海豚属和宽吻海豚属)线粒体基因组的 13 个蛋白质编码基因的  $Ka/Ks$  比值介于 0 和 0.587 7 之间, 都远远低于 1, 显示为纯化(负)选择。其中, *cox3* 基因的  $Ka/Ks$  均值最低, 其次为 *cox1*, *cox2* 和 *cob* 基因, 说明这些基因承受较强的选择压力和功能束缚; 而 *nad6* 基因的  $Ka/Ks$  均值最高, 其次为 *atp8*, *nad2* 和 *atp6* 基因, 说明这些基因的选择压力较弱。从科间(32 种鲸类)和属内(真露脊鲸属、须鲸属、原海豚属和宽吻海豚属)线粒体基因组数据作为数据群的基因差异位点和基因变异分析显示, 在鲸类群体遗传的研究中, *nad5*, *nad4* 和 *nad2* 基因是理想的分子标记, 可以作为 *cox1* 基因辅助的分子标记, 用于分析鲸类不同群体之间的遗传多样性, 为其生物多样性的保护及合理利用其生物资源提供更多保障。

## 参考文献:

- [1] DELLAPORTA S L, XU A, SAGASSER S, et al. Mitochondrial genome of *Trichoplax adhaerens* supports placozoa as the basal lower metazoan phylum [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2006; 103(23):8751—8756.
- [2] HELFENBEIN K G, FOURCADE H M, VANJANI R G, et al. The mitochondrial genome of *Paraspadella gotoi* is highly reduced and reveals that chaetognaths are a sister group to protostomes [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2004, 101(29):10639—10643.
- [3] BOORE J L. Animal mitochondrial genomes [J]. *Nucleic Acids Res*, 1999, 27(8):1767—1780.
- [4] BOORE J L, BROWN W M. Big trees from little genomes: mitochondrial gene order as a phylogenetic tool [J]. *Current Opinion in Genetics & Development*, 1998, 8(6):668—674.
- [5] 杨光, 季国庆, 魏辅文. 鲸类分子系统学研究进展 [J]. *动物分类学报*, 2003, 28(3):367—373.
- [6] 刘珊, 杨光. 鲸类保护遗传学研究进展 [J]. *动物学杂志*, 2002, 37(5):83—86.

- [7] ARNASON U, GULLBERG A. Comparison between the complete mtDNA sequences of the blue and the fin whale, two species that can hybridize in nature[J]. *Journal of Molecular Evolution*, 1993, 37(4):312—322.
- [8] ARNASON U, GULLBERG A, GREYARSDOTTIR S, et al. The mitochondrial genome of the sperm whale and a new molecular reference for estimating eutherian divergence dates[J]. *Journal of Molecular Evolution*, 2000, 50(6):569—578.
- [9] ARNASON U, GULLBERG A, JANKE. Mitogenomic analyses provide new insights into cetacean origin and evolution[J]. *Gene*, 2004, 333:27—34.
- [10] ARNASON U, GULLBERG A, WIDEGREN B. The complete nucleotide sequence of the mitochondrial DNA of the fin whale, *Balaenoptera physalus*[J]. *Journal of Molecular Evolution*, 1991, 33(6):556—568.
- [11] ARNASON U, GULLBERG A, WIDEGREN B. Cetacean mitochondrial DNA control region: sequences of all extant baleen whales and two sperm whale species[J]. *Molecular Biology and Evolution*, 1993, 10(5):960—970.
- [12] SASAKI T, NIKAIDO M, HAMILTON H, et al. Mitochondrial phylogenetics and evolution of mysticete whales[J]. *Systematic Biology*, 2005, 54(1):77—90.
- [13] SASAKI T, NIKAIDO M, WADA S, et al. *Balaenoptera omurai* is a newly discovered baleen whale that represents an ancient evolutionary lineage[J]. *Molecular Phylogenetics and Evolution*, 2006, 41(1):40—52.
- [14] VALVERDE J R, MARCO R, GARESSE R. A conserved heptamer motif for ribosomal RNA transcription termination in animal mitochondria[J]. *Proc Natl Acad Sci U S A*, 1994, 91(12):5368—5371.
- [15] XIONG Y, BRANDLEY MC, XU S, et al. Seven new dolphin mitochondrial genomes and a time-calibrated phylogeny of whales[J]. *BMC Evolutionary Biology*, 2009, 9:20.
- [16] YAN J, ZHOU K, YANG G. Molecular phylogenetics of 'river dolphins' and the baiji mitochondrial genome[J]. *Molecular Phylogenetics and Evolution* 2005; 37(3):743—750.
- [17] THOMPSON J D, GIBSON T J, PLEWNIAC F, et al. The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools[J]. *Nucleic Acids Res*, 1997, 25(24):4876—4882.
- [18] YANG Z. PAML 4: phylogenetic analysis by maximum likelihood[J]. *Mol Biol Evol*, 2007, 24(8):1586—1591.
- [19] ROZAS J, SANCHEZ-DELBARRIO J C, MESSEGUER X, et al. DnaSP, DNA polymorphism analyses by the coalescent and other methods[J]. *Bioinformatics (Oxford, England)*, 2003, 19(18):2496—2497.
- [20] SCHINDEL D E, MILLER S E. DNA barcoding a useful tool for taxonomists[J]. *Nature*, 2005, 435(7038):17.

## Analysis of the characteristics of whales mitochondrial genomes and exploration of molecular markers

TIAN Mei<sup>1</sup>, SHEN Xin<sup>1</sup>, MENG Xue-ping<sup>1</sup>, CHENG Han-liang<sup>1</sup>

(1. College of Marine Science, Huaihai Institute of Technology, Lianyungang 222005, China)

**Abstract:** The basic characteristics (including protein coding genes, selection pressure and different loci etc.) of whales mitochondrial genomes were fully revealed by comprehensive analysis of 32 whales mitochondrial genomes. Whales mitochondrial genomes contain 37 standard metazoan genes and their gene order is identical. The  $Ka/Ks$  ratio of the whales (*Eubalaena*, *Balaenoptera*, *Stenella* and *Tursiops*) 13 mitochondrial protein coding genes is lower than 1 (range between 0 and 0.5877), indicating for the purifying selection (negative selection). The average  $Ka/Ks$  of *cox3* gene is the lowest, followed by *cox1*, *cox2* and *cob* genes, indicating that these genes bear a strong selection pressure and functional constraints. Meanwhile, the average  $Ka/Ks$  of *nad6* gene is the highest, followed by *atp8*, *nad2* and *atp6* gene, indicating these genes bear weaker selection pressure. The genetic variation analysis of main genes (13 protein coding genes and 2 ribosomal RNA genes) and D-loop region within 32 species and intra-genus (*Eubalaena*, *Balaenoptera*, *Stenella* and *Tursiops*) showed that *nad5*, *nad4* and *nad2* gene are ideal molecular markers and can be used as supplementary molecular markers to *cox1* gene, which providing a reference for the conservation of whales biological diversity and use of the biological resources rationally.

**Key words:** Cetacea; mitochondrial genome; genetic different loci; molecular marker