

尾索动物线粒体基因组特征比较及分子系统发育

申欣, 田美, 孟学平, 程汉良, 李士虎

(淮海工学院 海洋学院, 江苏 连云港 222005)

摘要: 为全面揭示尾索动物线粒体基因组的基本特征, 作者系统分析了 12 种尾索动物的线粒体基因组全序列, 尾索动物线粒体基因组所有的基因均在同一链上编码, 这与脊椎动物、头索动物线粒体基因组的基因分布特征显著不同。与后生动物线粒体基因组标准的基因组成相比, 尾索动物线粒体基因组存在转运 RNA 基因的增加以及部分物种 *atp8* 基因的缺失。在尾索动物线粒体基因组中发生了大规模的基因重排, 即使在同属物种的线粒体基因组中, 也存在主编码基因的基因重排及基因缺失。尾索动物线粒体基因组蛋白质编码基因的起始密码子、终止密码子及氨基酸长度存在明显差异。2 个玻璃海鞘 (*Ciona intestinalis* 和 *C. savignyi*) 线粒体基因组 12 个蛋白质编码基因的 *Ka/Ks* 比值都低于 1 (0.0927~0.6752), 显示出一定的负选择。在所有的主编码基因中, *nad5* 基因和 *nad4* 基因的变异位点最多, 可以作为备选的分子标记, 用于分析尾索动物不同物种之间的生物多样性。基于线粒体基因组的系统发育分析, 支持尾索动物在科级的亲缘关系为: (((柄海鞘科 Pyuridae+芋海鞘科 Styelidae)+((三段海鞘科 Polyclinidae+星骨海鞘科 Didemnidae)+簇海鞘科 Clavelinidae)+玻璃海鞘科 Cionidae))+长纹海鞘科 Ascidiidae)+海樽科 Doliolidae。

关键词: 尾索动物; 线粒体基因组; 基因排列; 分子标记; 分子系统发育

中图分类号: Q959.284 文献标识码: A 文章编号: 1000-3096(2012)11-0030-08

与单个基因或者基因的片段相比, 线粒体基因组是一个完整的体系, 具有信息量丰富和系统发育树结构稳定等优点, 在过去的十几年间已被广泛应用于后生动物分子系统演化和种群遗传的研究^[1-2]。尾索动物隶属于脊索动物门(Chordata)、尾索动物亚门(Tunicata 或者 Urochordata), 为海洋生境中所特有的生物类群, 广泛栖息于世界各大海洋中, 从潮间带至深海均有分布。由于尾索动物在后生动物进化中处于特殊的关键位置, 其在免疫学、胚胎发育学、细胞学等各个学科领域对研究脊椎动物的起源与系统发育都有着不可取代的作用^[3-4]。在 GenBank 线粒体基因组的数据库中, 目前有 12 个尾索动物的线粒体基因组全序列。作者在 12 个尾索动物线粒体基因组研究的基础上, 全面分析此类群线粒体基因组的基本特征、基因排列、蛋白质编码基因、选择压力、变异位点和分子系统发育等。

1 材料和方法

1.1 数据获取

从 GenBank 线粒体基因组数据库中检索、下载

得到 12 个尾索动物线粒体基因组全序列。分别是隶属于海鞘纲(Ascidiacea)、内性目(Enterogona)的灯泡簇海鞘(*Clavelina lepadiformis*)、群体海鞘(*Diplosoma listerianum*)、胃海鞘(*Aplidium conicum*)、黑色次口海鞘(*Phallusia fumigata*)、具疣次口海鞘(*Phallusia mammilata*)、玻璃海鞘(*Ciona intestinalis*)和萨氏海鞘(*Ciona savignyi*)、褶鳃目(Stolidobranchia)的真海鞘(*Halocynthia roretzi*)、莫马斯赫海鞘(*Herdmania momus*)、小海鞘(*Microcosmus sulcatus*)和皱瘤海鞘(*Styela plicata*)以及樽海鞘纲(Thaliacea)的樽海鞘(*Doliolum nationalis*)^[5-11]。

1.2 选择压力分析

为了检验选择压力对于尾索动物(以玻璃海鞘属

收稿日期: 2011-10-08; 修回日期: 2011-11-04

基金项目: 国家自然科学基金资助项目(40906067); 中国博士后科学基金项目(2012M510054); 中央财政支持地方高校发展专项资金项目(CXTD04、CXTD01); 江苏省海洋生物技术重点实验室资助项目(2009HS13、2011HS009); 江苏省“青蓝工程”人才基金资助项目(苏教师[2010]27号)

作者简介: 申欣(1981-), 男, 山东成武人, 副教授, 博士, 主要从事海洋动物基因组学和蛋白质组学研究, 电话: 15062988145, E-mail: shenthin@163.com

为代表)线粒体基因组的影响, 通过 Clustal X^[12] 对线粒体基因组蛋白质编码基因的核苷酸序列进行多重序列比对。同义替换率(*Ks*)和非同义替换率(*Ka*)通过 PAML^[13]和 DnaSP 4.10.7^[14]进行了估算。

1.3 基因变异特征分析

分别以 12 个尾索动物、2 个玻璃海鞘属物种(玻璃海鞘和萨氏海鞘)和 2 个次口海鞘属物种(黑色次口海鞘和具疣次口海鞘)线粒体基因组作为数据群进行了基因变异位点分析, 蛋白质编码基因和核糖体 RNA 基因的核苷酸序列通过 Clustal X^[12]进行多重序列比对。然后通过 DnaSP 4.10.7^[14]分析了线粒体基因组主编码基因变异位点特征。

1.4 系统发育分析

线粒体基因组 12 个蛋白质编码基因(由于部分物种 *atp8* 缺失, 所以将其排除在外)的氨基酸序列通过 Clustal X^[12]进行多重序列比对。然后, 通过两种方法(贝叶斯法和最大似然法)进行系统发育重建, 使用的软件分别为 MrBayes^[15] 和 PHYML^[16]。在贝叶斯分析中, 马尔可夫链需要运行 1 000 000 个世代(构树频率=1000 世代), 从而保证达到收敛, 并估算

贝叶斯的先验概率(以 BPP 表示)。在最大似然法中, 通过自展法(Bootstrap = 1000)评估节点的可靠性(以 BPM 表示)。

2 结果和讨论

2.1 线粒体基因组的基本特征

在脊椎动物线粒体基因组中, *nad6* 基因和一系列转运 RNA 基因通常在负链上编码。在头索动物线粒体基因组中, 也存在与脊椎动物线粒体基因组类似的基因分布特征^[17-18]。然而, 12 个尾索动物线粒体基因组的所有基因均在同一链上编码, 这与脊椎动物、头索动物线粒体基因组的基因分布特征显著不同。

尾索动物线粒体基因组长度介于 13 648bp(群体海鞘)和 16 351bp(樽海鞘)之间。尾索动物线粒体基因组的基因组成与后生动物线粒体基因组标准的基因组成并不完全一致, 存在转运 RNA 基因数目的增加, 同时在灯泡簇海鞘、萨氏海鞘和真海鞘的线粒体基因组中缺乏 *atp8* 基因(表 1)。线粒体基因组主编码链的 A + T 含量差别很大, 最高的达到 80.8%(群体海鞘), 最低的仅为 52.9%(具疣次口海鞘)。尾索动物线粒体基因组的基本特征见表 1。

表 1 尾索动物线粒体基因组的基本特征

Tab. 1 Basic features of tunicates mitochondrial genomes

物种名称	全长(bp)	PCGs 数量	tRNA 数量	rRNA 数量	A+T 含量(%)
灯泡簇海鞘	14 484	12	24	2	76.2
群体海鞘	13 648	13	24	2	80.8
胃海鞘	15 183	13	24	2	76.9
黑色次口海鞘	15 535	13	26	2	53.2
具疣次口海鞘	14 579	13	23	2	52.9
玻璃海鞘	14 790	13	24	2	78.6
萨氏海鞘	14 737	12	24	2	77.3
真海鞘	14 771	12	23	2	68.3
莫马斯赫海鞘	15 816	13	24	2	63.4
小海鞘	15 028	13	25	2	70.2
皱瘤海鞘	14 414	13	24	2	70.1
樽海鞘	16 351	13	24	2	61.0

2.2 基因排列

在 12 个尾索动物线粒体基因组中, 发生了大规模的基因重排。即使排除容易发生易位的转运 RNA 基因, 它们的基因排列顺序也显著不同(图 1)。在同属物种的线粒体基因组中, 也存在主编码基因的基因重排和基因缺失。在玻璃海鞘属, 与玻璃海鞘线粒

体基因组相比, 萨氏海鞘的 *nad1* 基因发生易位, 同时 *atp8* 基因发生了缺失(图 1)。在次口海鞘属, 仅存在 2 个保守的基因区块: (1)*cox2-cob* 和 (2)*nad2-nad4L-cox3*, 而其他基因的排列顺序均发生了改变(图 1)。

与尾索动物线粒体基因组相比, 在头索动物线

粒体基因组中, 基因排列顺序较为保守^[17-18]。尽管 2 个类群(尾索动物和头索动物)均为脊椎动物的近亲, 但在线粒体基因组基因排列的特征上却有如此显著

的差别, 说明在 2 个类群之间, 线粒体基因组的进化机制和选择压力可能会存在明显差异, 在今后应该得到进一步的关注和探究。

灯泡簇海鞘														
<i>cox1</i>	<i>nad5</i>	<i>srRNA</i>	<i>nad2</i>	<i>cob</i>	<i>lrRNA</i>	<i>nad4</i>	<i>atp6</i>	<i>nad4L</i>	<i>nad6</i>	<i>nad1</i>	<i>cox2</i>	<i>nad3</i>	<i>cox3</i>	
群体海鞘														
<i>cox1</i>	<i>nad6</i>	<i>cox3</i>	<i>atp8</i>	<i>nad4</i>	<i>atp6</i>	<i>lrRNA</i>	<i>srRNA</i>	<i>cox2</i>	<i>nad3</i>	<i>nad5</i>	<i>nad4L</i>	<i>nad1</i>	<i>cob</i>	<i>nad2</i>
胃海鞘														
<i>cox1</i>	<i>nad2</i>	<i>atp8</i>	<i>lrRNA</i>	<i>atp6</i>	<i>nad4</i>	<i>cox2</i>	<i>nad5</i>	<i>nad6</i>	<i>nad4L</i>	<i>cox3</i>	<i>nad1</i>	<i>cob</i>	<i>nad3</i>	<i>srRNA</i>
黑色次口海鞘														
<i>cox1</i>	<i>nad1</i>	<i>lrRNA</i>	<i>srRNA</i>	<i>nad4</i>	<i>nad5</i>	<i>cox2</i>	<i>cob</i>	<i>atp6</i>	<i>nad6</i>	<i>nad3</i>	<i>nad2</i>	<i>nad4L</i>	<i>cox3</i>	<i>atp8</i>
具疣次口海鞘														
<i>cox1</i>	<i>atp6</i>	<i>cox2</i>	<i>cob</i>	<i>lrRNA</i>	<i>nad1</i>	<i>nad2</i>	<i>nad4L</i>	<i>cox3</i>	<i>nad3</i>	<i>nad6</i>	<i>nad5</i>	<i>atp8</i>	<i>nad4</i>	<i>srRNA</i>
玻璃海鞘														
<i>cox1</i>	<i>nad5</i>	<i>srRNA</i>	<i>nad6</i>	<i>atp6</i>	<i>nad2</i>	<i>cox2</i>	<i>cob</i>	<i>nad4L</i>	<i>cox3</i>	<i>nad1</i>	<i>lrRNA</i>	<i>nad3</i>	<i>nad4</i>	<i>atp8</i>
萨氏海鞘														
<i>cox1</i>	<i>nad5</i>	<i>srRNA</i>	<i>nad6</i>	<i>atp6</i>	<i>nad2</i>	<i>cox2</i>	<i>cob</i>	<i>nad1</i>	<i>nad4L</i>	<i>cox3</i>	<i>lrRNA</i>	<i>nad3</i>	<i>nad4</i>	
真海鞘														
<i>cox1</i>	<i>nad3</i>	<i>nad4</i>	<i>nad6</i>	<i>cox3</i>	<i>nad4L</i>	<i>srRNA</i>	<i>cox2</i>	<i>cob</i>	<i>nad2</i>	<i>nad5</i>	<i>lrRNA</i>	<i>nad1</i>	<i>atp6</i>	
莫马斯赫海鞘														
<i>cox1</i>	<i>nad4L</i>	<i>atp8</i>	<i>nad1</i>	<i>nad6</i>	<i>lrRNA</i>	<i>cob</i>	<i>nad2</i>	<i>nad5</i>	<i>srRNA</i>	<i>nad4</i>	<i>atp6</i>	<i>cox2</i>	<i>cox3</i>	<i>nad3</i>
小海鞘														
<i>cox1</i>	<i>cox2</i>	<i>cob</i>	<i>nad5</i>	<i>nad2</i>	<i>nad4</i>	<i>nad3</i>	<i>atp6</i>	<i>nad1</i>	<i>srRNA</i>	<i>atp8</i>	<i>nad6</i>	<i>lrRNA</i>	<i>cox3</i>	<i>nad4L</i>
皱瘤海鞘														
<i>cox1</i>	<i>nad6</i>	<i>nad2</i>	<i>nad5</i>	<i>nad3</i>	<i>cox3</i>	<i>atp6</i>	<i>atp8</i>	<i>nad4</i>	<i>lrRNA</i>	<i>srRNA</i>	<i>cox2</i>	<i>nad1</i>	<i>nad4L</i>	<i>cob</i>
樽海鞘														
<i>cox1</i>	<i>atp8</i>	<i>nad1</i>	<i>nad2</i>	<i>nad4</i>	<i>lrRNA</i>	<i>nad6</i>	<i>srRNA</i>	<i>cox3</i>	<i>cox2</i>	<i>cob</i>	<i>nad4L</i>	<i>nad3</i>	<i>nad5</i>	<i>atp6</i>

图 1 尾索动物线粒体主编码基因的基因排列(不包含转运 RNA 基因)

Fig. 1 Gene order of tunicates mitochondrial major coding genes (transfer RNA genes are excluded)

2.3 蛋白质编码基因

在 3 个物种(灯泡簇海鞘、萨氏海鞘和真海鞘)的线粒体基因组中存在 *atp8* 基因的缺失, 其余的 9 个物种均包含后生动物线粒体基因组标准的 13 个蛋白质编码基因。在樽海鞘线粒体基因组的蛋白质编码基因中, 最为常用的起始密码子为“TTG”; 而在 11 个海鞘纲物种的线粒体蛋白质编码基因中, “ATG”出现的比例最高。除了这两种起始密码子外, 同时还存在“ATA”、“ATT”和“GTG”等 3 种起始密码子。真海鞘线粒体基因组所有的蛋白质编码基因均以完全终止密码子“TAA”或者“TAG”终止, 余下的 11 个物种部分蛋白质编码基因中, 存在不完全的终止密码子(“TA-”或者“T-”)(表 2)。尾索动物线粒体基因组蛋白质编码基因所编码的氨基酸数目并不保守。群体海鞘的 *cox2* 基因编码 313

氨基酸, 比萨氏海鞘的 *cox2* 基因超出 92 个氨基酸; 而同时, 玻璃海鞘的 *nad2* 基因编码 369 个氨基酸, 与群体海鞘的 *nad2* 基因相比, 多了 83 个氨基酸(表 2)。

2.4 选择压力

为了分析尾索动物线粒体基因组蛋白编码基因的选择压力, 统计了玻璃海鞘和萨氏海鞘蛋白质编码基因的同义替换率(*K_s*)和非同义替换率(*K_a*)。2 个玻璃海鞘线粒体基因组 12 个蛋白质编码基因的 *K_a/K_s* 比值都低于 1(介于 0.0927 和 0.6752 之间), 显示出一定的负(纯化)选择(图 2)。其中, *nad6* 基因的 *K_a/K_s* 比值最高(0.6752), 其次是 *nad3* 和 *nad2* 基因(分别为 0.4321 和 0.3916), 因此表明, 在 13 个蛋白质编码基因中, *nad6*、*nad3* 和 *nad2* 基因承受较小的选择压力和功能束缚。同时, *cox1* 基因的 *K_a/K_s* 比值

最低(0.0927), 其次是 *nad1* 和 *cob* 基因(分别为 0.1353 和 0.1484), 从而说明, 在尾索动物线粒体基因组中, *cox1*、*nad1* 和 *cob* 基因承受较强的选择压力和功能束缚。

表 2 尾索动物线粒体基因组蛋白质编码基因的氨基酸数量及起始、终止密码子

Tab. 2 Amino acids number, initiation and termination codons in protein-coding genes of tunicates mitochondrial genomes

物种名称	蛋白质编码基因													
		<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>
灯泡簇海鞘	N	207	-	362	513	228	261	298	318	114	440	86	568	148
	I	ATG	-	ATG	ATG	TTG	TTG	TTG	ATG	ATT	ATT	ATG	ATG	ATG
	T	TAA	-	TAA	TAA	T-	TAA	TAA	TAA	T-	T-	TAG	TA-	TAA
群体海鞘	N	193	30	356	512	313	259	284	286	113	420	88	545	146
	I	ATG	ATA	TTG	ATG	ATA	ATG	ATG	ATG	ATA	ATG	ATA	ATT	ATG
	T	TA-	TAA	TAA	TAA	TAA	T-	T-	T-	TAA	TA-	TAA	TAG	T-
胃海鞘	N	205	40	362	518	225	261	303	314	128	446	87	564	152
	I	ATG	ATG	ATG	TTG	TTG	ATG	ATG	TTG	TTG	ATG	ATT	ATT	ATG
	T	TAA	TAA	TAG	TAA	TAA	TAA	T-	T-	TAG	T-	TAA	T-	TAG
黑色次口海鞘	N	212	38	363	514	228	263	305	340	126	459	89	550	159
	I	GTG	ATT	ATG	GTG	GTG	TTG	GTG	ATG	GTG	GTG	GTG	ATG	GTG
	T	TAG	TAG	T-	TAG	TAA	TAG	TAG	TAA	TAG	TAG	TAG	TAG	TAG
具疣次口海鞘	N	212	38	363	509	228	262	302	337	124	463	85	547	157
	I	ATG	ATT	ATG	GTG	GTG	GTG	ATA	ATA	ATG	GTG	ATA	GTG	TTG
	T	TAG	TAA	TAG	TAG	TAA	TAA	TAG	TAA	TAA	TAG	T-	TAA	TAG
玻璃海鞘	N	210	43	361	520	229	262	302	369	113	442	92	578	159
	I	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATT	ATG	ATA	ATT	ATT	ATG
	T	TAA	TAG	TA-	TAA	TAA	TAA	TAA	TAA	TAA	TAA	TAG	T-	TAA
萨氏海鞘	N	209	-	361	516	221	262	344	330	115	441	88	578	165
	I	ATG	-	ATG	ATG	ATG	ATG	ATG	ATG	GTG	ATA	ATG	ATA	GTG
	T	TAG	-	TAA	TAG	T-	T-	TAG	TAG	T-	TAA	TAA	TAA	TAA
真海鞘	N	211	-	362	517	225	264	298	333	111	435	106	551	147
	I	GTG	-	ATT	ATG	ATA	GTG	ATA	ATG	ATG	GTG	ATG	TTG	GTG
	T	TAA	-	TAG	TAG	TAA	TAA	TAA	TAA	TAG	TAA	TAG	TAA	TAA
莫马斯赫海鞘	N	209	38	365	515	224	262	304	343	118	443	84	582	145
	I	ATT	GTG	TTG	ATG	ATG	TTG	ATA	GTG	GTG	GTG	GTG	GTG	ATG
	T	TAG	TAG	TAG	TAG	TAG	TAA	T-	TAG	TAA	T-	T-	TAG	TAG
小海鞘	N	213	40	367	518	227	261	295	337	134	447	91	575	148
	I	ATG	ATG	ATG	ATA	ATG	ATG	ATG	ATG	ATA	ATG	ATT	ATG	ATG
	T	TAA	TAA	TAA	TAA	TAA	TA-	TAA	TAA	TAA	TAG	TAA	TA-	TAA
皱瘤海鞘	N	208	37	367	516	226	260	303	328	126	443	80	576	149
	I	ATG	ATG	ATG	GTG	ATG	ATG	TTG	TTG	TTG	TTG	TTG	TTG	GTG
	T	TAG	TAA	TAG	TA-	TAG	TAG	TAA	TAG	TAG	TAG	TAG	TAA	TAG
樽海鞘	N	202	48	359	510	239	264	312	319	105	434	101	561	157
	I	TTG	ATG	GTG	TTG	TTG	TTG	TTG	TTG	ATT	TTG	GTG	GTG	TTG
	T	T-	TA-	TAG	TAG	TAA	TAG	TAG	T-	TAG	T-	TAG	TAG	TAA

注: N. 氨基酸数量; I. 起始密码子; T. 终止密码子

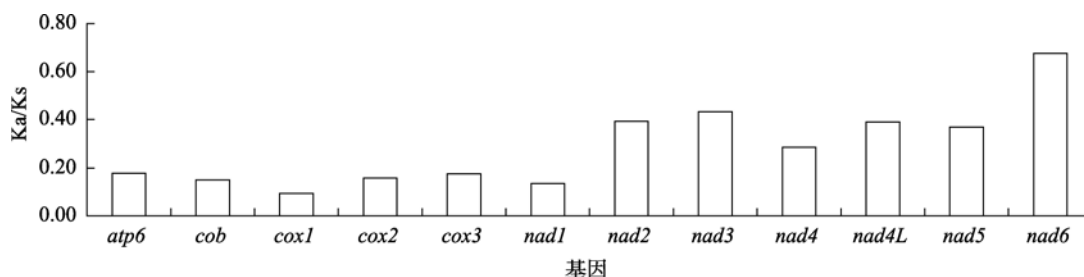


图 2 玻璃海鞘线粒体基因组蛋白质编码基因的 Ka/Ks 分析

Fig. 2 The ratio of Nonsynonymous and synonymous substitutions rates (Ka/Ks) was estimated in protein-coding genes of *Ciona* mitochondrial genomes

2.5 基因变异位点特征

在种群遗传学的研究中,选择合适的分子标记至关重要。以 12 个尾索动物、2 个玻璃海鞘属物种(玻璃海鞘和萨氏海鞘)和 2 个次口海鞘属物种(黑色次口海鞘和具疣次口海鞘)线粒体基因组作为数据群,分别进行了基因变异位点分析。在尾索动物线粒体基因组 14 个主编码基因中, *cox1* 基因最为保守,变

异位点的比例为 61.57%。*Cob*、*cox2-3*、*nad1* 和 *lrRNA* 等 5 个基因变异位点的比例介于 70%~80%; *atp6*、*nad3-5*、*srRNA* 和 *nad4L* 等 6 个基因变异位点的比例介于 80%~90%; *nad2* 和 *nad6* 变异位点的比例超过 90%(表 3)。变异位点数最多的基因为 *nad5* 基因(1120 个),其次为 *nad4* 和 *cox1* 基因,变异位点数分别达到 1005 和 931 个。因此,在尾索动物种群遗传的研究中, *nad5* 和 *nad4* 基因可以作为备选的分子标记。

表 3 尾索动物线粒体基因组蛋白质编码基因和核糖体 RNA 基因的变异位点分析

Tab. 3 Analysis of different loci of protein-coding genes and ribosomal RNA gene in tunicates mitochondrial genomes

项目	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>lrRNA</i>
总位点数*	552	63	1038	1512	636	768	810	750	249	1173	207	1356	387	473	793
不变位点数	100	11	301	581	142	218	192	55	38	168	23	236	26	82	118
变异位点数	452	52	737	931	494	550	618	695	211	1005	184	1120	361	391	621
单现突变数	58	13	113	138	74	83	86	63	26	100	18	114	30	54	66
简约位点数	394	39	624	793	420	467	532	632	185	905	166	1006	331	337	555
变异位点比例 (%)	81.88	82.54	71.00	61.57	77.67	71.61	76.30	92.67	84.74	85.68	88.89	82.60	93.28	82.66	78.31

注: *. 总位点数 (Total number of sites) 不包含插入和缺失位点; **. 仅包括 9 个线粒体基因组

在玻璃海鞘属线粒体基因组的主编码基因中, *cox1* 和 *cox2* 基因最为保守, 变异位点的比例分别为 16.21% 和 18.85%; 变异位点数最多的基因为 *nad5* 基因(462 个), 其次为 *nad4* 基因(331 个)(表 4)。与玻璃海鞘属线粒体基因组类似, 在次口海鞘属线粒体基因组主

编码基因中, *cox2* 和 *cox1* 基因最为保守, 变异位点的比例分别为 23.68% 和 23.97%; 变异位点数最多的基因为 *nad5* 基因(618 个), 其次为 *nad4* 基因(527 个)(表 5)。因此, 在玻璃海鞘和次口海鞘种群遗传的研究中, *nad5* 和 *nad4* 仍然可以作为备选的分子标记。

表 4 玻璃海鞘属线粒体基因组蛋白质编码基因和核糖体 RNA 基因的变异位点分析

Tab. 4 Analysis of different loci of protein-coding genes and ribosomal RNA gene in *Ciona* mitochondrial genomes

项目	基因名称													
	<i>atp6</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>lrRNA</i>
总位点数*	627	1083	1542	663	786	906	990	339	1323	264	1719	474	672	1075
不变位点数	470	856	1292	538	622	701	689	235	992	178	1257	321	473	837
变异位点数	157	227	250	125	164	205	301	104	331	86	462	153	199	238
变异位点比例 (%)	25.04	20.96	16.21	18.85	20.87	22.63	30.40	30.68	25.02	32.58	26.88	32.28	29.61	22.14

注: *. 总位点数不包含插入和缺失位点(表 5 同)

表 5 次口海鞘属线粒体基因组蛋白质编码基因和核糖体 RNA 基因的变异位点分析

Tab. 5 Analysis of different loci of protein-coding genes and ribosomal RNA gene in *Phallusia* mitochondrial genomes

项目	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>lrRNA</i>
总位点数*	633	96	1089	1527	684	786	903	987	360	1365	255	1629	459	709	1117
不变位点数	407	56	790	1161	522	555	618	556	231	838	164	1011	250	488	750
变异位点数	226	40	299	366	162	231	285	431	129	527	91	618	209	221	367
变异位点比例 (%)	35.70	41.67	27.46	23.97	23.68	29.39	31.56	43.67	35.83	38.61	35.69	37.94	45.53	31.17	32.86

综上所述,在尾索动物种群遗传的研究中, *nad5* 和 *nad4* 可以作为备选的分子标记,用于分析不同物种之间的遗传多样性,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。

2.6 系统发育关系分析

目前在国际上,基于线粒体基因组数据分析尾索动物内部的系统发育关系才刚刚兴起^[5, 8],对 12 个尾索动物线粒体基因组进行全面的系统发育分析尚未见报道。作者基于 12 个线粒体基因组蛋白质编码基因的氨基酸序列,通过两种方法(贝叶斯法和最

大似然法)所构建系统发育树的拓扑结构完全一致(图 3)。从系统发育树的结果可以看出,尾索动物分为两个分支:海鞘纲和樽海鞘纲(BPP=100, BPM=100)。在海鞘纲内部,小海鞘、皱瘤海鞘、真海鞘和莫马斯赫海鞘聚为一支,从而支持褶鳃目为单系群(BPP=100, BPM=100)。具疣次口海鞘和黑色次口海鞘单独聚为一支(BPP=100, BPM=100),从而导致内性目并非单系群。线粒体基因组的数据支持尾索动物在科级的亲缘关系为:(((柄海鞘科+芋海鞘科)+(((三段海鞘科+星骨海鞘科)+簇海鞘科)+玻璃海鞘科))+长纹海鞘科)+海樽科(图 3)。

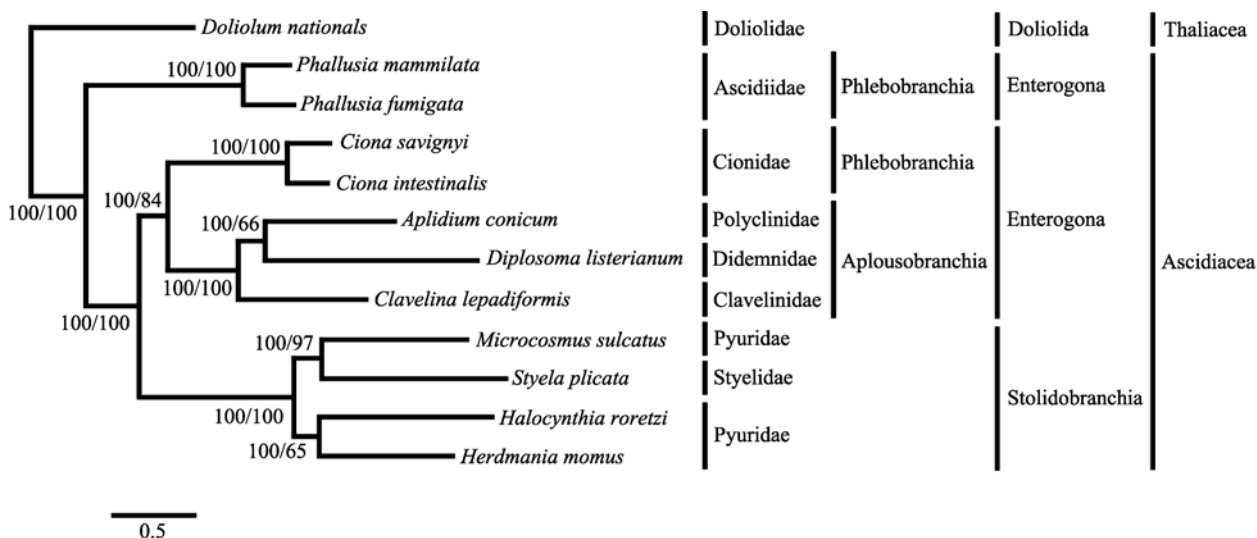


图 3 基于线粒体基因组蛋白质编码基因的氨基酸序列构建的系统发育树

Fig. 3 Phylogenetic tree based on aligned amino acids sequences from mitochondrial protein coding genes. Two system reconstruction methods were used, and the topologies were identical. The first and second numbers represent the Bayesian inference (BPP) and maximum likelihood analysis (BPM) values, respectively.

Tree topologies produced by the two methods were identical. The first numbers are from Bayesian inferences (BPP) and the second numbers are from maximum likelihood analyses (BPM)

3 结论

作者在 12 个尾索动物线粒体基因组全序列的基础上,全面分析尾索动物线粒体基因组的基本特征、

基因排列、蛋白质编码基因、选择压力、变异位点和分子系统发育等。12 个尾索动物线粒体基因组的所有基因均在同一链上编码,这与脊椎动物、头索动物线粒体基因组的基因分布特征显著不同。尾索动

物线粒体基因组的基因组成与后生动物线粒体基因组标准的基因组成并不完全一致,存在转运 RNA 基因数目的增加,同时在灯泡簇海鞘、萨氏海鞘和真海鞘的线粒体基因组中缺乏 *atp8* 基因。在 12 个尾索动物线粒体基因组中,发生了大规模的基因重排,即使在同属物种的线粒体基因组中,也存在主编码基因的基因重排和基因缺失。尾索动物线粒体基因组蛋白质编码基因使用的起始密码子、终止密码子及氨基酸长度存在差异。2 个玻璃海鞘线粒体基因组 12 个蛋白质编码基因的 *Ka/Ks* 比值都低于 1(介于 0.0927 和 0.6752 之间),显示出一定的负选择。在尾索动物群体遗传的研究中, *nad5* 和 *nad4* 可以作为备选分子标记,用于分析不同物种和种群之间的生物多样性。基于线粒体基因组蛋白质编码基因的氨基酸序列,支持将尾索动物分为两个分支:海鞘纲和樽海鞘纲。尾索动物在科级的亲缘关系为:(((柄海鞘科+芋海鞘科)+((三段海鞘科+星骨海鞘科)+簇海鞘科)+玻璃海鞘科))+长纹海鞘科)+海樽科。

参考文献:

- [1] Shen X, Ma X, Ren J, et al. A close phylogenetic relationship between Sipuncula and Annelida evidenced from the complete mitochondrial genome sequence of *Phascolosoma esculenta*[J]. BMC Genomics, 2009, 10: 136.
- [2] Shen X, Ren J F, Cui Z X, et al. The complete mitochondrial genomes of two common shrimps (*Litopenaeus vannamei* and *Fenneropenaeus chinensis*) and their phylogenomic considerations[J]. Gene, 2007, 403(1-2): 98-109.
- [3] Dehal P, Satou Y, Campbell R K, et al. The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins[J]. Science, 2002, 298(5601): 2157-2167.
- [4] Delsuc F, Brinkmann H, Chourrout D, et al. Tunicates and not cephalochordates are the closest living relatives of vertebrates[J]. Nature, 2006, 439(7079): 965-968.
- [5] Singh T R, Tsagkogeorga G, Delsuc F, et al. Tunicate mitogenomics and phylogenetics: peculiarities of the *Herdmania momus* mitochondrial genome and support for the new chordate phylogeny[J]. BMC Genomics, 2009, 10: 534.
- [6] Gissi C, Pesole G, Mastrotoaro F, et al. Hypervariability of ascidian mitochondrial gene order: exposing the myth of deuterostome organelle genome stability[J]. Molecular Biology and Evolution, 2010, 27(2): 211-215.
- [7] Iannelli F, Griggio F, Pesole G, et al. The mitochondrial genome of *Phallusia mammillata* and *Phallusia fumi-gata* (Tunicata, Ascidiacea): high genome plasticity at intra-genus level[J]. BMC Evolutionary Biology, 2007, 7:155.
- [8] Yokobori S, Oshima T, Wada H. Complete nucleotide sequence of the mitochondrial genome of *Doliolum nationalis* with implications for evolution of urochordates[J]. Molecular Phylogenetics and Evolution, 2005, 34(2): 273-283.
- [9] Gissi C, Iannelli F, Pesole G. Complete mtDNA of *Ciona intestinalis* reveals extensive gene rearrangement and the presence of an *atp8* and an extra *trnM* gene in ascidians[J]. Journal of Molecular Evolution, 2004, 58(4): 376-389.
- [10] Yokobori S, Watanabe Y, Oshima T. Mitochondrial genome of *Ciona savignyi* (Urochordata, Ascidiacea, Enterogona): comparison of gene arrangement and tRNA genes with *Halocynthia roretzi* mitochondrial genome[J]. Journal of Molecular Evolution, 2003, 57(5): 574-587.
- [11] Yokobori S, Ueda T, Feldmaier-Fuchs G, et al. Complete DNA sequence of the mitochondrial genome of the ascidian *Halocynthia roretzi* (Chordata, Urochordata) [J]. Genetics 1999, 153(4): 1851-1862.
- [12] Thompson J D, Gibson T J, Plewniak F. 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.
- [13] Yang Z. PAML 4: phylogenetic analysis by maximum likelihood[J]. Mol Biol Evol, 2007, 24(8): 1586-1591.
- [14] Rozas J, Sanchez-DelBarrio J C, Messeguer X, et al. DnaSP, DNA polymorphism analyses by the coalescent and other methods[J]. Bioinformatics, 2003, 19(18): 2496-2497.
- [15] Ronquist F, Huelsenbeck J P. MrBayes 3: Bayesian phylogenetic inference under mixed models[J]. Bioinformatics, 2003, 19(12): 1572-1574.
- [16] Guindon S, Gascuel O. A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood[J]. Syst Biol, 2003, 52(5): 696-704.
- [17] Nohara M, Nishida M, Miya M. Evolution of the mito-

chondrial genome in cephalochordata as inferred from complete nucleotide sequences from two epigonichthys species[J]. Journal of Molecular Evolution, 2005, 60(4): 526-537.

[18] Kon T, Nohara M, Yamanoue Y, et al. Phylogenetic position of a whale-fall lancelet (Cephalochordata) inferred from whole mitochondrial genome sequences[J]. BMC Evolutionary Biology, 2007, 7: 127.

Comparison on characteristics of tunicates mitochondrial genomes and molecular phylogeny

SHEN Xin, TIAN Mei, MENG Xue-ping, CHENG Han-liang, LI Shi-hu

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

Received: Oct., 8, 2011

Key words: tunicates; mitochondrial genome; gene order; molecular marker; molecular phylogeny

Abstract: In this paper, a comprehensive analysis based on 12 tunicates mitochondrial genomes has been done and the basic characteristics of tunicates mitochondrial genomes are fully revealed. In tunicates mitochondrial genomes, all genes are encoded on the same DNA strand, which is significantly different with the genes distribution feature of vertebrates and amphioxus mitochondrial genomes. Compared with typical gene composition of metazoan mitochondrial genomes, there are increase of transfer RNA genes and missing of *atp8* gene in some tunicates species. Gene order is extremely rearranged among tunicates mitochondrial genomes, there are gene rearrangement and gene loss of major genes even in same genus. There are differences in the start codon, stop codon and amino-acid number in protein-coding genes of tunicates mitochondrial genomes. The ratio of *Ka/Ks* of 12 protein-coding genes from 2 *Ciona* mitochondrial genomes is less than 1 (between 0.0927 and 0.6752), indicating a negative selection. In major coding genes, *nad5* and *nad4* genes have the largest number of different loci, which can be used as alternative molecular markers to analyze biodiversity among different tunicates species. The phylogenetic trees based on mitochondrial genomes support the relationship at family level: (((Pyuridae+Styelidae)+(((Polyclinidae+Didemnidae)+Clavelinidae)+Cionidae))+ Ascidiidae)+Doliolidae.

(本文编辑: 谭雪静)