

鸮形目两种鸟类线粒体基因组全序列测定 与比较研究*

孙 毅 马 飞 肖 冰 郑俊杰 袁晓东 汤敏谦
王 黎 于业飞 李庆伟 **

(辽宁师范大学生命科学学院, 大连 116029; 宝生物工程(大连)有限公司, 大连 116600;
蛇岛老铁山国家级自然保护区管理处, 大连 116041)

摘要 利用 Long-PCR 和 Primer Walking 结合克隆测序法对短耳鸮和长耳鸮线粒体基因组进行了全序列测定. 结果表明: 短耳鸮 mtDNA 序列全长为 18858 bp, 长耳鸮 mtDNA 全长为 18493 bp, 其中短耳鸮 mtDNA 是目前已知最长的鸟类线粒体基因组. 两种鸮类的基因组结构和基因排列顺序与家鸡相同, 无假控制区, 在 ND3 基因 174 位点都存在一个额外插入的胞苷酸(C). 控制区序列异常增大是造成这两种鸟类 mtDNA 增大的主要原因, 短耳鸮控制区长度为 3288 bp, 长耳鸮为 2926 bp, 这是目前已知脊椎动物线粒体基因组中仅次于盲鳗的最大的控制区. 在其控制区 3' 端存在大量的串联重复序列, 分析发现这两种鸮类的重复序列和 Mt5 调控元件有较高的序列相似性, 且能形成多重的茎环二级结构, 这表明该重复序列可能具有一定的生理功能, 影响线粒体基因组的复制或转录表达, 从而使相应物种具有更大的选择优势, 以适应环境和生存竞争.

关键词 短耳鸮 长耳鸮 线粒体基因组 控制区 重复序列

作为从分子水平研究进化和系统发育问题的理想材料, 线粒体DNA(mitochondrial DNA, mtDNA)近年来已成为研究分子系统学和进化的重要手段^[1]. 随着PCR技术的日趋成熟和DNA序列测定技术的广泛使用, 利用线粒体DNA全序列研究进化关系的方法受到越来越多研究者的青睐^[2~4]. 自 1981 年Anderson

等人^[5]首次测定人的线粒体基因组以来, 已经有 342 种脊椎动物的mtDNA全序列被测定. 到目前为止, 在GenBank上登陆mtDNA全序列的鸟类共有 29 种, 包括: 鸵鸟目(Struthioniformes)、美洲鸵目(Rheiformes)、鹤鸵目(Casuariiformes)、恐鸟目(Dinornithiformes)、鸮形目(Tinamiformes)、几维目(Apterygiformes)、企

2004-06-28 收稿, 2004-09-21 收修改稿

* 国家自然科学基金资助项目(批准号: 30170505)

** 联系人, E-mail: Liqw@263.net

鵝目(Sphenisciformes)、鸛形目(Ciconiiformes)、鴿形目(Charadriiformes)、雁形目(Anseriformes)、鸡形目(Galliformes)、鸚形目(Psittaciformes)、隼形目(Falconiformes)、鴞形目(Strigiformes)和雀形目(Passeriformes)共 15 个目.

脊椎动物线粒体基因组的大小约为 15~19 kb, 家鸡(*Gallus gallus*)mtDNA 基因组测序结果表明其大小为 16775 bp, 美洲鸵(*Rhea americana*)为 16714 bp, 在已测定 mtDNA 全序列的鸟类中, 最大的普通鵟(*Buteo buteo*)为 18674 bp. 1996 年李庆伟等人^[6]对鸞形目鸞科的短耳鸞(*Asio flammeus*)进行了 14 种限制性内切酶的单酶解和双酶解消化后的电泳图谱研究, 结果显示, 短耳鸞 mtDNA 长度有异常增大的现象.

在隼形目已测定线粒体 DNA 全基因组的普通鵟^[7]和游隼(*Falco peregrinus*)^[8]中, 发现以下特点: 两种鸟类 mtDNA 基因组较大, 均为 18 kb 以上, 两种鸟类中均存在假控制区, 且基因排列顺序和传统鸟类(家鸡)相比较为独特, 即控制区转座到 tRNA^{Thr}和 tRNA^{Pro}基因之间, 并在 tRNA^{Glu}和 tRNA^{Phe}基因间产生了一个长度多变的假控制区. 而作为猛禽类另一代表的鸞形目, 目前仅有鸞科的斑布克鵟(*Ninox novaeseelandiae*)已测定了 mtDNA 基因组序列^[9], 但其仍有部分 tRNA 基因和控制区序列缺失, 完整的线粒体全基因组则尚未有报道. 鸞形目鸟类的 mtDNA 基因组大小、结构特点、控制区的组成等都有待进一步研究. 本文测定了鸞形目鸞科耳鸞属的短耳鸞(*Asio flammeus*)及其近缘物种长耳鸞(*Asio otus*)的线粒体基因组全序列, 以期揭示鸞形目鸟类 mtDNA 的结构特点, 并从基因水平研究短耳鸞 mtDNA 异常增大的原因和机制.

1 材料和方法

1.1 材料

试验用短耳鸞(*Asio flammeus*)和长耳鸞(*Asio otus*)为从大连老铁山国家级自然保护区采集的两只雄性个体, 取其新鲜的肝脏组织, 保存于-80℃ 冰箱备用.

1.2 方法

(1) mtDNA 的提取及制备: mtDNA 的提取按照 Arnason 等人^[10]的方法进行, 提取的 mtDNA 在-20

冰箱中保存备用.

(2) mtDNA 全序列测定: 以纯化后的 mtDNA 为模板, 用鸟类的通用引物^[11]分别扩增 12S rRNA, COI, ND4 和 Cytb 基因部分片段, 并测序. 根据测得的上述序列分别设计 4 对引物, Long-PCR 扩增 12SrRNA-COI(5 kb), COI-ND4(4.6 kb), ND4-Cytb(3 kb)和 Cytb-12SrRNA(5.8 kb)4 个长片段, 引物序列见表 1. 然后采取 Primer Walking 和克隆测序的方法分别得到 4 个长片段的序列, 进而拼接成 mtDNA 全序列. 测序时每一个片段与下一个片段重叠约 80~120 bp. 测序策略如图 1 所示.

测序试剂采用宝生物工程(大连)有限公司 TaKaRa TaqTM Cycle Sequencing Kit, 序列测定采用 ABI377 型全自动遗传分析仪(PE Applied Biosystems). 每条序列均经过正反链双向测定, 以确保序列的准确性. DNA 测序工作主要在宝生物工程(大连)有限公司完成.

(3) 序列分析: 将所测得的两种鸞类 mtDNA 全序列与 GenBank 中斑布克鵟(*Ninox novaeseelandiae*; AY309457)、家鸡(*Gallus gallus*; X52392)、靛蓝维达雀(*Vidua chalybeata*; AF090341)和普通鵟(*Buteo buteo*; AF380305)的 mtDNA 序列进行比较, 采用 Clustal X1.8 软件^[12]对序列进行排列, 再进行人工校对. 使用 DNASIS 和 DNASTAR 软件包的 Editseq 程序对序列进行开放阅读框的查找、蛋白编码序列的翻译、氨基酸编码情况的调查、序列互补链的获得、序列核苷酸组成的统计等. 使用 RNAstructure3.5 软件构建核苷酸序列的二级结构.

2 结果

2.1 基因组结构

短耳鸞 mtDNA 全长为 18858 bp, 长耳鸞为 18493 bp, 均为环形双链 DNA, 基因组结构见表 2. 与大部分脊椎动物相似, 两种鸞类的 mtDNA 由 22 个 tRNA, 2 个 rRNA, 13 个编码蛋白基因和 1 个非编码的控制区所组成. 其基因排列顺序与斑布克鵟及家鸡^[13]相同, 而有别于隼形目的普通鵟和游隼的基因排列顺序(tRNA^{Thr}/control region/tRNA^{Pro}/ND6/tRNA^{Glu}/noncoding)^[14].

表 1 短耳鸚和长耳鸚线粒体基因组扩增引物^{a)}

引物		序列(5'~3')	来源
1	12S-F	CAAAGCTGGGATTAGATACCCCACTAT	Sorenson, et al. (1999)
2	12S-R	TTGCGAGGACGCTGACGGGCGGTATGT	Sorenson, et al. (1999)
3	CoI-F	AACAACATAAGCTTCTGACT	Sorenson, et al. (1999)
4	CoI-R	CTACGTGCGAGATAATTCCAAAACC	Sorenson, et al. (1999)
5	ND4-F	TTCGAAGCAACCCTAATCCCAAC	Sorenson, et al. (1999)
6	ND4-R	GGTGCTTCTACGTGGGCTTTAGG	Sorenson, et al. (1999)
7	Cytb-F	CCACCAAACATCTCCGCATGATGAAA	Sorenson, et al. (1999)
8	Cytb-R	GCCCTCAGAATGATATTTGTCTCTCA	Sorenson, et al. (1999)
9	S-12S-738F	ATCAAGCTTGAGGTGGAAGAGATGGGCTACATTCT	本研究
10	S-CoI-103R	CTCGAATTCATTGCTAGGTCTACTGAAGCTCCTGC	本研究
11	S-CoI-364F	GGTGACCCTATTCTCTACCA	本研究
12	S-ND4-53R	GGGTGTAGAATAGTAGGTA	本研究
13	S-ND4-192F	ATCAAGCTTACTTAACTCCTCATGAACCGACCTCC	本研究
14	S-Cytb-182R	ATCGAATTCATGTGAGACGGATGTGAAGGCTAGGG	本研究
15	S-Cytb-13354F	GCTGACTACTCCGCAACCTACACGCAAATG	本研究
16	S-12S-256R	GAATGTAGCCCATCTCTCCACCTCATAGG	本研究
17	L-12S-738F	ATCAAGCTTGAGGTGGAAGAGATGGGCTACATTCT	本研究
18	L-CoI-103R	CTCGAATTCATTGCTAGGTCTACTGAAGCTCCTGC	本研究
19	L-CoI-258F	GCAATCAACTTCATCACCACCGCCATCAAC	本研究
20	L-ND4-366R	GGCTATTGTGAGGAATGGGTAGTGTAGGTG	本研究
21	L-ND4-155F	ACTCAACACCTCATGGGCTGACCTC	本研究
22	L-Cytb-141R	GAAAAGCGAGGCTCCATTGCGTGG	本研究
23	L-Cytb-185F	ACATCGGACGAGGACTATACTACGG	本研究
24	L-12S-112R	TGGTTGGGTGTATCGTGGGTATCG	本研究

a) 1~8 为扩增部分基因片段的通用引物, 9~24 为扩增长片段的引物; S 示短耳鸚引物, L 示长耳鸚引物; 短耳鸚的引物 9 和 10 与长耳鸚的引物 17 和 18 为同一对引物

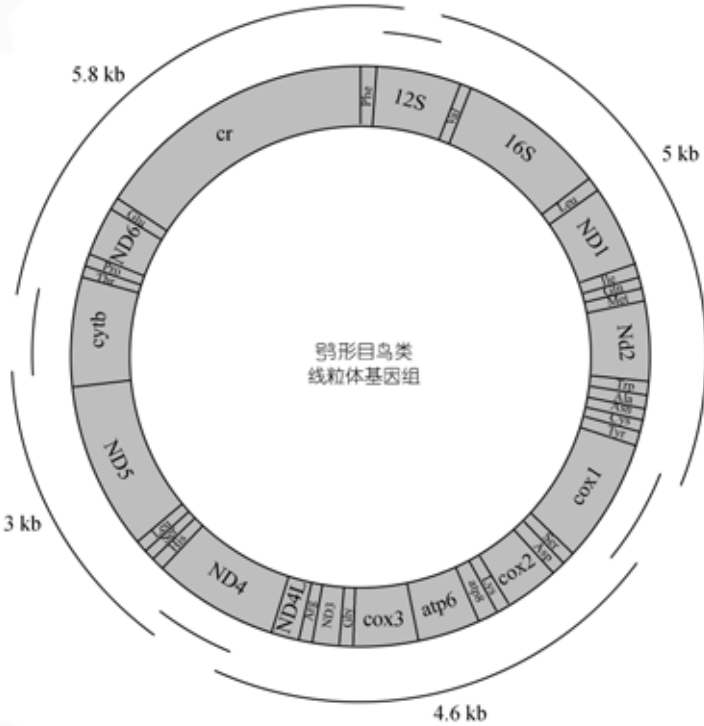


图 1 鸚形目两种鸟类线粒体基因组测序策略

表 2 短耳鸮(*Asio flammeus*)和长耳鸮(*Asio otus*)线粒体基因组结构

基因	<i>Asio flammeus</i>				<i>Asio otus</i>			
	位置	长度/bp	起始子	终止子	位置	长度/bp	起始子	终止子
控制区	1~3288	3288			1~2926	2926		
tRNA-Phe	3289~3359	71			2927~2998	72		
12S rRNA	3360~4349	990			2999~3972	974		
tRNA-Val	4350~4421	72			3973~4044	72		
16S rRNA	4422~6004	1583			4045~5651	1607		
tRNA-Leu(UUR)	6005~6078	74			5652~5725	74		
ND1	6089~7048	960	ATG	AGG	5736~6695	960	ATG	AGG
tRNA-Ile	7047~7118	72			6694~6765	72		
tRNA-Gln*	7130~7200	71			6777~6847	71		
tRNA-Met	7200~7268	69			6847~6915	69		
ND2	7269~8309	1041	ATG	TAG	6916~7956	1041	ATG	TAG
tRNA-Trp	8308~8384	77			7955~8029	75		
tRNA-Ala*	8386~8454	69			8031~8099	69		
tRNA-Asn*	8456~8530	75			8101~8175	75		
tRNA-Cys*	8533~8601	69			8178~8246	69		
tRNA-Tyr*	8601~8672	72			8246~8317	72		
CO	8683~10233	1551	ATG	AGG	8319~9869	1551	GTG	AGG
tRNA-Ser(UCN)*	10225~10298	74			9861~9934	74		
tRNA-Asp	10302~10370	69			9938~10006	69		
CO	10373~11056	684	ATG	TAA	10009~10692	684	ATG	TAA
tRNA-Lys	11062~11131	70			10700~10770	71		
ATP8	11133~11300	168	ATG	TAA	10772~10939	168	ATG	TAA
ATP6	11291~11974	684	ATG	TAA	10930~11613	684	ATG	TAA
CO	11974~12757	784	ATG	T - -	11613~12396	784	ATG	T - -
tRNA-Gly	12758~12825	68			12397~12464	68		
ND3	12826~13177	352	ATA	TAA	12465~12816	352	ATA	TAA
tRNA-Arg	13180~13248	69			12819~12887	69		
ND4L	13250~13546	297	ATG	TAA	12889~13185	297	ATG	TAA
ND4	13540~14917	1378	ATG	T - -	13179~14556	1378	ATG	T - -
tRNA-His	14918~14988	71			14558~14628	71		
tRNA-Ser(AGY)	14993~15057	65			14633~14697	65		
tRNA-Leu(CUN)	15062~15132	71			14705~14775	71		
ND5	15133~16953	1821	ATG	TAA	14776~16596	1821	ATG	TAA
Cytb	16958~18100	1143	ATG	TAA	16601~17743	1143	ATG	TAA
tRNA-Thr	18100~18169	70			17743~17813	71		
tRNA-Pro*	18185~18254	70			17820~17889	70		
ND6*	18261~18782	522	ATG	TAG	17896~18417	522	ATG	TAG
tRNA-Glu*	18786~18858	73			18421~18493	73		

* 示由轻链编码基因

两种鸮类mtDNA基因间的间隔和重叠见表 3(部分数据来自Slack, et al., 2003^[15]). 与其他鸟类相比差异较明显的有: 短耳鸮tRNA^{Tyr}和CO 基因间间隔为

10 bp, 其余鸟类除鸮面鸱(*Strigops habroptilus*)为 9 bp外, 均为 1 bp间隔; 短耳鸮CO 和tRNA^{Lys}基因间间隔为 8 bp, 长耳鸮为 10 bp, 其他古鸮鸟均

表3 鸟类线粒体基因组基因间隔和基因重叠

	Emu	GSKiwi	Rhea	Ostrich	GiantMoa	Tinamou	Penguin	ShortOwl	LongOwl	Morepork	Kakapo	Buzzard	Duck	Turnstone	Stork	Indigobird	Rifleman	Range
tRNA-Phe/ <i>12S rRNA</i>	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0
<i>12S rRNA</i> /tRNA-Val	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0
tRNA-Val/ <i>16S rRNA</i>	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0
<i>16S rRNA</i> /tRNA-Leu (UUR)	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0
tRNA-Leu (UUR)/ <i>ND1</i>	8	9	11	9	11	9	5	10	10	13	7	9	4	16	17	8	19	2~19
<i>ND1</i> /tRNA-Ile	3	1	11	1	4	19	1	1	1	1	1	1	1	1	1	9	1	1~24
tRNA-Ile/tRNA-Gln (L)	6	9	14	11	8	7	9	11	11	9	8	13	8	8	8	5	8	5~14
tRNA-Gln (L)/tRNA-Met	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap; 12
tRNA-Met/ <i>ND2</i>	—	—	—	—	—	—	8	—	—	—	—	—	—	—	—	—	—	0; 1, 8
<i>NADH2</i> /tRNA-Trp	1	1	1	1	1	1	1	1	1	1	2	1	1	1	2	2	1	1~5
tRNA-Trp/tRNA-Ala (L)	1	1	1	—	1	1	1	1	1	1	1	1	3	1	1	1	1	0~10
tRNA-Ala (L)/tRNA-Asn (L)	3	1	2	4	1	1	2	1	1	1	1	2	2	3	2	9	4	0~10
tRNA-Asn (L)/tRNA-Cys (L)	3	2	3	1	2	2	2	2	2	2	2	2	—	2	2	—	13	0~13
tRNA-Cys (L)/tRNA-Tyr (L)	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap
tRNA-Tyr (L)/ <i>CO</i>	1	1	1	1	1	1	1	10	1	1	9	1	1	1	1	1	1	1; 9,10
<i>COI</i> /tRNA-Ser (UCN) (L)	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap	6 overlap
tRNA-Ser (UCN) (L)/tRNA-Asp	2	2	2	3	2	2	4	3	3	4	3	4	2	2	2	5	3	2~5
tRNA-Asp/ <i>CO</i>	1	1	1	1	1	1	2	2	2	2	4	2	1	1	1	10	7	1~10
<i>CO</i> /tRNA-Lys	1	1	1	1	1	1	4	8	10	4	4	4	4	4	4	14	4	1~14
tRNA-Lys/ <i>ATPase8</i>	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
<i>ATPase8</i> / <i>ATPase6</i>	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap	7 overlap
<i>ATPase6</i> / <i>CO</i>	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	8	2	2; 8~12
<i>CO</i> /tRNA-Gly	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
tRNA-Gly/ <i>ND3</i>	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0
<i>ND3</i> /tRNA-Arg	4	4	5	4	4	3	7	5	5	4	2	7	4	7	7	4	5	2~7
tRNA-Arg/ <i>ND4L</i>	—	—	—	—	—	—	1	1	1	1	1	1	—	1	1	1	1	0~1
<i>ND4L</i> / <i>ND4</i>	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap	4 overlap
<i>ND4</i> /tRNA-His	4	10	4	4	4	5	2	1	2	1	1	1	1	1	1	1	13	1~13
tRNA-His/tRNA-Ser (AGY)	—	—	—	—	—	—	—	4	4	2	—	—	—	—	—	—	—	0; 2,4
tRNA-Ser (AGY)/tRNA-Leu (CUN)	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	1 overlap	4	7	4	1 overlap	—	1 overlap	1 overlap	1 overlap	1 overlap	1	1 overlap-7
tRNA-Leu (CUN)/ <i>ND5</i>	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	0; 2
<i>Cytb</i> /tRNA-Thr	7	4	8	5	4	4	6	2	2	4	3	5	5	4	4	6	8	3~8
tRNA-Thr/tRNA-Pro (L)	82	83	9	55	27	85	9	15	6	11	6	N/A	10	7	11	12	6	0~85; N/A
tRNA-Thr/CR	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	—	N/A	N/A	N/A	N/A	N/A	N/A; 0
CR/tRNA-Pro (L)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	—	N/A	N/A	N/A	N/A	N/A	N/A; 0
tRNA-Pro (L)/ <i>ND6</i> (L)	23	30	18	13	14	12	16	9	9	15	12	9	13	11	13	29	23	8~30; N/A
<i>ND6</i> (L)/tRNA-Glu (L)	3	3	3	3	—	—	3	3	3	3	2	3	—	3	3	1	2	0~3
tRNA-Glu (L)/CR #	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0
CR/tRNA-Phe #	—	—	—	—	—	—	—	—	—	unknown	unknown	—	—	—	—	—	unknown	0
Total Length of Genome	16711	16816	16714	16591	17070	16703	17611	18858	18493	>16223	>15599	18674	16616	16725	17622	16895	>16069	16589~8858

基因间间隔和重叠包括了前一个编码蛋白基因的终止密码子碱基的长度; tRNA-Glu/CR#和 CR/tRNA-Phe#: 在普通鸚(Buzzard)中为 tRNA-Glu/Additional Noncoding Region 和 Additional Noncoding Region/tRNA-Phe; Emu = *Dromaius novaehollandiae*, GSKiwi = *Apteryx haastii*, Rhea = *Rhea americana*, Ostrich = *Struthio camelus*, GiantMoa = *Dinornis giganteus*, Tinamou = *Tinamus major*, Penguin = *Eudyptula minor*, ShortOwl = *Asio flammeus*, LongOwl = *Asio otus*, Morepork = *Ninox novaeseelandiae*, Kakapo = *Strigops habroptilus*, Buzzard = *Buteo buteo*, Duck = *Aythya americana*, Turnstone = *Arenaria interpres*, Stork = *Ciconia boyciana*, Indigobird = *Vidua chalybeata*, Rifleman = *Acanthisitta chloris*.

为 1 bp, 新腭鸟中除靛蓝维达雀(*Vidua chalybeata*)为 14 bp 外, 其余均为 4 bp 间隔; 两种鸚类 tRNA^{His}和 tRNA^{Ser(AGY)}基因间间隔为 4 bp, 斑布克鸚鵒(*Ninox novaeseelandiae*)为 2 bp, 其他无间隔序列; 短耳鸚 tRNA^{Ser(AGY)}和 tRNA^{Leu(CUN)}基因间间隔为 4 bp, 长耳鸚为 7 bp, 斑布克鸚鵒为 4 bp, 刺鸚(*Acanthisitta chloris*)为 1 bp 间隔, 其他均无间隔序列或有 1 bp 的基因重叠; 两种鸚类 *Cytb* 和 tRNA^{Thr} 基因间间隔仅为 2 bp, 其余鸟类中间隔最小的为鸚面鸚鵒的 3 bp.

短耳鸚和长耳鸚 tRNA^{Asn}和 tRNA^{Cys} 基因间间隔为 2 bp, 未发现在哺乳动物及两栖动物该位置中存在的 L 链复制起始点结构^[13].

2.2 碱基组成

短耳鸚和长耳鸚的碱基组成与其他鸟类相似: 4 种碱基中 G 含量最少, 且均表现出较明显的 AT 偏向, 特别是在控制区序列中表现的尤为明显. 在脊椎动物 mtDNA 序列中普遍存在的这种现象表明了它们较多利用 AT 编码的偏好性. 全基因组序列、13 种编码蛋白基因、2 种 rRNA 基因、22 种 tRNA 基因和控制区的碱基组成见表 4.

2.3 编码蛋白基因和密码子

与其他鸟类相比, 两种鸚类的 13 个编码蛋白基因的长度比较保守, 只有 *ND1* 基因有明显差别, 在两种鸚类中均为 319 个氨基酸, 而同为鸚形目的斑布克鸚鵒为 324 个氨基酸, 其他鸟类为 324~326 个氨基酸. 此外, 在短耳鸚和长耳鸚 *ND3* 基因的 174 位点均存在一个额外插入的胞苷酸(C), 这在斑布克鸚鵒及其他多种鸟类^[16]中也都有发现.

该两种鸚类与斑布克鸚鵒 13 个编码蛋白基因最常用的起始密码子为 ATG, 但 *ND3* 基因均为 ATA. 此

外, 长耳鸚和斑布克鸚鵒 *CO* 基因的起始密码为 GTG. 最常用的终止密码子为 TAA, 其他的终止密码有: 3 种鸚类的 *ND2* 和 *ND6* 基因终止密码子均为 TAG, *ND1* 和 *CO* 基因为 AGG, 而在 *CO* 和 *ND4* 中为不完全的终止密码子 T. 此外, 斑布克鸚鵒的 *ND5* 基因终止密码子为 AGA.

2.4 控制区序列

短耳鸚控制区序列的长度为 3288 bp, 长耳鸚为 2926 bp, 这大大超过了其他鸟类 mtDNA 控制区的长度. 两种鸚类控制区的碱基组成和基本结构与其他鸟类相似, AT 含量丰富. 按照碱基变化速率的不同, 可以将其分为 3 个区: domain ~ ^[17,18]. 与其他鸟类相应区域相比, domain 和 domain 区碱基变异程度较大, 而 domain 区则相对保守. 三个区内也均存在一些保守序列, 如在其他鸟类 domain 区中发现的 ETAS1, ETAS2 序列和 CSB1-like 框, domain 区的 B-box, C-box, D-box, E-box, F-box 和 Bird similarity box, domain 区的 CSB-1 序列等^[19], 在短耳鸚和长耳鸚中也同样发现了相应结构. 但在 domain 区未能找到在部分鸟类相应区域存在的 H 链复制起始序列(O_H)和双向复制起始点(LSP/HSP)结构.

两种鸚类控制区的 3 端均出现了大片段的重复序列拷贝, 其中在短耳鸚中为 126 bp 单元重复 7 次和 78 bp 单元重复 14 次, 两段重复区之间有 99 bp 间隔, 为两种重复单元的不完全拷贝; 长耳鸚中为 127 bp 单元重复 8 次和 78 bp 单元重复 7 次, 重复区之间有 55 bp 两种重复单元的不完全拷贝, 重复单元的碱基序列见表 5. 短耳鸚和长耳鸚的重复区与其他鸟类相应重复片段序列的相似性不高, 在每种鸚类控制区内, 126 bp(127 bp)单元重复区与 78 bp 单元重复区之间也没有太大的相似性. 但在两种鸚类之间相应单

表 4 短耳鸚(*Asio flammeus*)和长耳鸚(*Asio otus*)线粒体基因组不同部分的核苷酸组成

mtDNA	<i>Asio flammeus</i>					<i>Asio otus</i>				
	%A	%C	%G	%T	%A+T	%A	%C	%G	%T	%A+T
全基因组	30.79	32.08	13.06	24.07	54.86	30.35	32.75	13.32	23.58	53.93
13 个蛋白基因	29.76	34.47	11.92	23.86	53.61	29.27	36.00	12.34	22.39	51.66
2 个 rRNA 基因	32.69	28.57	19.12	19.63	52.31	32.39	29.52	19.33	18.75	51.14
22 个 tRNA 基因	32.54	26.78	17.36	23.32	55.86	31.88	27.27	17.86	22.98	54.87
D-Loop 区	32.42	28.44	10.52	28.62	61.04	32.30	25.09	9.81	32.81	65.11

表 5 短耳鸱(*Asio flammeus*)和长耳鸱(*Asio otus*)线粒体控制区重复单元序列

种类	重复单元	碱基序列
<i>Asio flammeus</i>	126 bp 重复单元	5 'TAACATCGGCGCCCAAATTCGCTAACAATTTTGATCCCATCTATCC ATTTTCTATTCAACTTTCATC AAAACTTCTAGGCAAACCTAACTGTT TGCCATTCGATCTTATCACTCACCTTACCC-3 '
	78 bp 重复单元	5 'TAGGCAAACCTATAACAATCAATTATCGTCCACCATCAAAAATTTTAC CGATCAATCGAACGAACGATCA AAAACCTTC-3 '
<i>Asio otus</i>	127 bp 重复单元	5 'TAGGCAAACCTATAACTGTTTGCCAATCAACTTTATCGTTAATTCCTTAT CATAATATCGGCACCCAAATCCCGCTAATAAATCTATCTTTTATCCATTT ATTTTTATCCAAATTTACTCGACACTTC-3 '
	78 bp 重复单元	5 'CAAACGAACGATAAGAACTTTCTAGGCAAACCTATAACAATCAATT ATCATCCATCATCAAAAATTTATTGATCAAC-3 '

元重复区的序列相似性却较高，其中短耳鸱和长耳鸱 78 bp 单元重复区之间序列相似性高达 90%左右，而 126 bp (127 bp)单元区序列相似性也在 75%左右。

3 讨论

3.1 基因组结构

脊椎动物中已知mtDNA序列最大的是圆口纲的盲鳗^[20](*Myxine glutinosa*, 18909 bp)和鱼纲的宽咽鱼(*Eurypharynx pelecanoides*, 18978 bp); 鸟类中已知最大的是普通鵟^[7]，为18674 bp。与它们相比，短耳鸱mtDNA成为目前鸟类中已知最大的线粒体基因组。

该两种鸱类及斑布克鹰鸱的线粒体基因组与大部分鸟类在基因排列顺序上完全相同，这显示了鸟类线粒体基因排列的保守性。然而，与同为猛禽类的普通鵟和游隼相比，基因顺序则有所差异，但两种鸱类与普通鵟及游隼各编码基因间的序列相似性却较高。这一方面表明鸱形目与隼形目间的亲缘关系较近；另一方面也表明在鸟类各目间或目内存在的基因重排现象，有着多重的独立起源^[14,21,22]。这为进一步探讨线粒体基因组的演化关系提供了实验依据。

从基因间的间隔或重叠情况来看，短耳鸱和长耳鸱与其他鸟类差别不大，但比较突出的是在tRNA^{His}与tRNA^{Ser(AGY)}基因间(均为4 bp 间隔)以及tRNA^{Ser(AGY)}与tRNA^{Leu(CUN)}基因间(4 bp 或 7 bp 间隔)存在独特的间隔序列，在鸟类中只有同为鸱形目的斑布克鹰鸱与之相似，因此，这两组 tRNA 基因间特定长度的间隔序列可能是鸱形目鸟类的一个共同特征。

在多种鸟类ND3 基因的特定位点都发现存在一个插入的胞苷酸(C)，一般认为，这个额外的核苷酸在翻译过程中是不表达的，而是通过RNA自我剪切来恢复基因的功能，从而避免了移码突变导致的ND3 基因转译提前终止^[16]。两种鸱类ND3 基因排除 174 位点胞苷酸(C)后所得的氨基酸序列与其他鸟类该基因氨基酸序列相似程度较高，也进一步证实了该观点。此前在家鸡ND3 基因中未发现有此插入的核苷酸^[13]，但在随后不同作者^[16,23]的工作中，证实家鸡中亦存在此插入碱基，然而在鸱形目的近缘物种普通鵟中却并未发现该插入核苷酸的存在。关于ND3 基因中的插入碱基是否是鸟类的一个共同特征，还是具有一定的种属特异性，尚有待更多的数据加以证明。

部分鸟类编码蛋白基因的起始、终止密码子见表 6，比较发现：起始密码子可分为 3 类：编码甲硫氨酸(Met)的ATR(ATG和ATA)、编码异亮氨酸(Ile)的ATY(ATT和ATC)、以及编码缬氨酸(Val)的GTG，共 5 种，其中ATG的使用频率最高，而ATT的频率最低；终止密码子亦可以分为 3 类：AGR(AGG和AGA)，TAR(TAG和TAA)和不完全的T-或TA-，共 6 种，以TAA最为常用，而TA-的使用频率最低，目前只在鸱面鹦鹉的ND3 基因和小企鹅(*Eudyptula minor*)的ND4 基因中有过发现。不完全的终止密码子(T-或TA-)被认为是由转录后mRNA 3 端的PolyA所补齐^[24]。从鸟类线粒体基因组密码子的使用情况来看，基本符合“摆动学说”^[25]，即密码子第三位点变异较大。线粒体基因组的起始(5 个)和终止密码子(6 个)的数目均大于核基因组的相应数目(1 个，3 个)，这可能是由于线粒体起源于原始的原核生物所造成的^[26]。

表 6 部分鸟类编码蛋白基因起始和终止密码子的使用统计^{a)}

	Asf	Aso	Nin	But	Fal	Gal	Ayt	Hae	Smi	Vid	Rhe	Str	Eud
起始密码子													
ATG	12	11	11	10	11	12	10	11	11	12	10	9	10
GTG	—	1	1	2	1	1	3	1	—	—	2	3	2
ATC	—	—	—	1	—	—	—	—	1	—	—	—	1
ATA	1	1	1	—	1	—	—	—	1	—	1	1	—
ATT	—	—	—	—	—	—	—	1	—	1	—	—	—
终止密码子													
TAA	7	7	6	7	8	9	7	8	8	7	6	6	7
TAG	2	2	2	2	1	1	2	1	1	1	3	2	2
T—	2	2	2	2	2	2	2	2	2	2	2	2	1
AGG	2	2	2	2	1	1	2	2	2	1	2	2	2
AGA	—	—	1	—	1	—	—	—	—	2	—	1	—
TA—	—	—	—	—	—	—	—	—	—	—	—	—	1

a) Asf = *Asio flammeus*; Aso = *Asio otus*; Nin = *Ninox novaeseelandiae*; But = *Buteo buteo*; Fal = *Falco peregrinus*; Gal = *Gallus gallus*; Ayt = *Aythya americana*; Hae = *Haematopus ater*; Smi = *Smithornis sharpei*; Vid = *Vidua chalybeata*; Rhe = *Rhea americana*; Str = *Struthio camelus*; Eud = *Eudiptula minor*

3.2 控制区序列

鸟类线粒体控制区的大小一般为 1~1.5 kb，例如家鸡为 1227 bp，靛蓝维达雀为 1295 bp。在脊椎动物中，目前已知最大的控制区序列是圆口纲的盲鳗，为 3628 bp，其余超过 3 kb 的物种极为少见，因此短耳鸮和长耳鸮的线粒体基因组控制区就成为目前脊椎动物中仅次于盲鳗，而在鸟类中为最大的线粒体控制区。此外，在非编码区所占全基因组的比例上，短耳鸮为 17.4%，长耳鸮为 15.8%，其中短耳鸮也大于了目前鸟类中已知最大的普通鸮(16.7%)，而一般鸟类此比例仅为 6.3%~11.7%左右；其他如爬行动物中的扬子鳄(*Alligator sinensis*)、哺乳动物中的绵羊(*Ovis aries*)等亦仅为 6.2%和 7.1%。这说明，无论是在绝对长度还是相对长度上，两种鸮类的控制区大小，在鸟类、甚至是整个脊椎动物中都是非常罕见的。

控制区 3 端存在的大片段重复序列是造成这两种鸮类线粒体基因组异常增大的主要原因。由于 mtDNA 是一个高度经济有效的基因组，绝大多数区域为编码区，仅有控制区为非编码区，传统认为，该

区域所受进化压力较小，遗传变异较大^[27]。近年来在脊椎动物多种类群的mtDNA中都发现有重复序列的存在，说明在控制区中出现重复序列是一种比较常见的现象，但类似两种鸮类中如此大规模的重复片段，却较为罕见。两种鸮类重复区之间较高的相似性表明它们的重复序列可能有着共同的起源，相应序列应该在短耳鸮和长耳鸮从共同的祖先分化出来之前就已经出现，其后由于各自进化，导致现在两种鸮类中相应重复单元的数目有较大差异。关于重复序列的产生机制，目前的观点倾向于由链的滑动错配所导致^[28,29]。由于鸮形目mtDNA基因组的数据目前还比较缺乏(斑布克鹰鸮mtDNA只有部分控制区序列^[9])，尚无法判断这种大片段重复的现象是否是鸮形目鸟类的共同特征。

对重复区序列进行研究还发现：两种鸮类的重复序列均与控制区元件——Mt5^[30]有着较高的相似性(图 2)，而Mt5 元件则被认为是一种可能起着调节控制区转录终止作用的核编码蛋白的特异性结合区域^[31,32]。这种重复序列的大量出现，有可能是为Mt5 元件提供备份或增强Mt5 元件被蛋白识别的能力。

Asio flammeus (126 bp 单元): A A A C T T C T A G G C A A A
Asio flammeus (78 bp 单元): A A C C T T C T A G G C A A A
Asio otus (127 bp 单元): A C A C T T C T A G G C A A A
Asio otus (78 bp 单元): A A C T T T C C A G G C A A A
Human: A T G C T T A C A A G C A A G
* * * * *

图 2 短耳鸮(*Asio flammeus*)、长耳鸮(*Asio otus*)与人类线粒体控制区 Mt5 元件的核酸序列比较
* 示与参考序列(Human)相同的碱基

利用RNAstructure软件构建核苷酸序列二级结构发现：两种鸢类的重复序列均可形成多重的茎环二级结构，且短耳鸢的126 bp和78 bp重复单元与长耳鸢相应单元间二级结构的形状也非常相似，而这种结构被认为和DNA合成的起始或终止是有关的^[33]。

目前在多个动物类群的线粒体控制区中都发现了重复序列，且出现的位置比较固定，为控制区的5'或3'端(即domain和domain区)^[34~37]。传统观点认为，重复序列是线粒体基因组中的冗余序列，缺乏生物学功能^[38]。但也有人^[20,33]认为，重复序列可能在mtDNA的复制和转录过程中执行一定的调控作用。在一些脊椎动物重复序列中发现了包含特定保守框的相似序列^[32,39,40]，也进一步证实了该观点。本研究中，两种鸢类的重复序列中包含特定调控元件的相似序列，这可能有两重含义：第一，由于控制区序列碱基突变较多，且缺乏相应的修复机制，当突变发生在调控元件的序列中，并影响其功能时，重复片段可以为其提供多个拷贝，从而避免影响基因组的正常功能；第二，调控元件多个序列拷贝的存在，可能会增强相应元件的调控、表达效率，从而增加线粒体基因组的选择优势。此外，在控制区序列中多存在回文序列或发夹结构，这对于维持或增强控制区核苷酸序列的二级结构是必需的，而在不少动物的重复序列中也发现了大量回文序列或发夹结构的存在，且能形成保守而稳定的茎环结构。因此，大片的重复序列很可能参与并促进控制区序列复杂而保守、稳定的空间结构的形成，这在另一个层面上对于控制区的功能——线粒体基因组的复制和转录，可能会起到促进和增强的作用。

根据线粒体起源的“内共生假说”^[41]，线粒体基因组在进化的早期，大量的遗传信息以相当快的速度丢失或转移，特别是与氨基酸核苷酸生物合成、无氧糖酵解和调控有关的基因更容易从原有的基因组中删除，因为它们逐渐成为非必需的或可由宿主的功能补充^[42]，而且，从节省能源、提高增殖效率从而有利于竞争的角度来看，较小的基因组也是有利的^[43]。然而，在脊椎动物多个物种线粒体控制区中发现的大片段重复序列，似乎违背了这一原则。且存在重复序列的物种范围分布较大，在鱼类、两栖、爬行、鸟

类和哺乳动物中均有发现，说明这种现象也并非只是线粒体基因组原始性状的残迹。通过比较还发现：在同为猛禽类的普通鵟、游隼和两种鸢类的线粒体控制区中，均存在着比较大的重复片段，这可能与猛禽类独特的生活方式有关，其肉食性的取食方式、在食物链中次级消费者的地位，决定了它们的线粒体必须提供更多的能量和有更强的复制、表达能力，从而使其具有更大的选择优势，以适应环境和生存竞争。

参 考 文 献

- 1 Stoneking M, Soodyall H. Human evolution and the mitochondrial genome. *Curr Opin Genet Dev*, 1996, 6(6): 731~736[DOI]
- 2 Zardoya R, Meyer A. The complete nucleotide sequence of the mitochondrial genome of the lungfish (*Protopterus dolloi*) supports its phylogenetic position as a close relative of land vertebrates. *Genetics*, 1996, 142: 1249~1263
- 3 Naylor G J, Brown W M. Amphioxus mitochondrial DNA, chordate phylogeny, and the limits of inference based on comparisons of sequences. *Syst Biol*, 1998, 47: 61~76[DOI]
- 4 Rasmussen A S, Arnason U. Phylogenetic studies of complete mitochondrial DNA molecules place cartilaginous fishes within the tree of bony fishes. *J Mol Evol*, 1999, 48: 118~123
- 5 Anderson S, Bankier A T, Barrell B G, et al. Sequence and organization of the human mitochondrial genome. *Nature*, 1981, 290: 457~465[DOI]
- 6 李庆伟, 陈宜峰. 一个增大的鸟类线粒体基因组. *动物学研究*, 1996, 17(4): 376, 384, 392
- 7 Haring E, Kruckenhauser L, Gamauf A, et al. The complete sequence of the mitochondrial genome of *Buteo buteo* (Aves, Accipitridae) indicates an early split in the phylogeny of raptors. *Mol Biol Evol*, 2001, 18(10): 1892~1904
- 8 Mindell D P, Sorenson M D, Dimcheff D E, et al. Interordinal relationships of birds and other reptiles based on whole mitochondrial genomes. *Syst Biol*, 1999, 48(1): 138~152[DOI]
- 9 Harrison G L, Mclenachan P A, Phillips M J, et al. Four new avian mitochondrial genomes help get to basic evolutionary questions in the late Cretaceous. *Mol Biol Evol*, 2004, 21(6): 974~983[DOI]
- 10 Arnason U, Gullberg A, Widegren B. The complete nucleotide sequence of the mitochondrial DNA of the fin whale, *Balaenoptera physalus*. *J Mol Evol*, 1991, 33: 556~568
- 11 Sorenson M D, Ast J C, Dimcheff D E, et al. Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. *Mol Phylogenet Evol*, 1999, 12(2): 105~114[DOI]
- 12 Thompson J D, Gibson T J, Plewniak F, et al. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*, 1997, 24: 4876~4882[DOI]

- 13 Desjardins P, Morais R. Sequence and gene organization of the chicken mitochondrial genome. *J Mol Biol*, 1990, 212: 599~634[DOI]
- 14 Mindell D P, Sorenson M D, Dimcheff D E. Multiple independent origins of mitochondrial gene order in birds. *Proc Natl Acad Sci USA*, 1998, 95: 10693~10697[DOI]
- 15 Slack K E, Janke A, Penny D, et al. Two new avian mitochondrial genomes (penguin and goose) and a summary of birds and reptile mitogenomic features. *Gene*, 2003, 302: 43~52[DOI]
- 16 Mindell D P, Sorenson M D, Dimcheff D E. An extra nucleotide is not translated in mitochondrial ND3 of some birds and turtles. *Mol Biol Evol*, 1998, 15(11): 1568~1571
- 17 Brown G G, Gadaleta G, Pepe G, et al. Structural conservation and variation in the D-loop containing region of vertebrate mitochondrial DNA. *J Mol Biol*, 1986, 192: 503~511[DOI]
- 18 Saccone C, Pesole G, Sbisà E. The main regulatory region of mammalian mitochondrial DNA: structure-function model and evolutionary pattern. *J Mol Evol*, 1991, 33: 83~91
- 19 Ruokonen M, Kvist L. Structure and evolution of the avian mitochondrial control region. *Mol Phylogenet Evol*, 2002, 23: 422~432[DOI]
- 20 Delarbre C, Rasmussen A S, Arnason U, et al. The complete mitochondrial genome of the hagfish *Myxine glutinosa*: unique features of the control region. *J Mol Evol*, 2001, 53: 634~641[DOI]
- 21 Bensch S, Harlid A. Mitochondrial genomic rearrangements in songbirds. *Mol Biol Evol*, 2000, 17(1): 107~113
- 22 Haddrath O, Baker A J. Complete mitochondrial DNA genome sequences of extinct birds: ratite phylogenetics and the vicariance biogeography hypothesis. *Proc R Soc Lond B*, 2001, 268: 939~945[DOI]
- 23 Nishibori M, Hayashi T, Tsudzuki M, et al. Complete sequence of the Japanese quail (*Coturnix japonica*) mitochondrial genome and its genetic relationship with related species. *Anim Genet*, 2001, 32(6): 380~385[DOI]
- 24 Ojala D, Montoya J, Attardi G. tRNA punctuation model of RNA processing in human mitochondria. *Nature*, 1981, 290: 470~474[DOI]
- 25 Crick F H. Codon-anticodon pairing: the Wobble Hypothesis. *J Mol Biol*, 1966, 19: 548~555
- 26 陈惟昌, 陈志华, 王自强, 等. 线粒体遗传密码及基因组遗传密码的对称分析. *生物物理学报*, 2002, 18: 87~94
- 27 Moritz C, Dowling T E, Brown W M. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Ann Rev Ecol Syst*, 1987, 18: 269~292[DOI]
- 28 Buroker N E, Brown J R, Gilbert T A, et al. Length heteroplasmy of sturgeon mitochondrial DNA: an illegitimate elongation model. *Genetics*, 1990, 124: 157~163
- 29 Faber J E, Stepien C A. Tandemly repeated sequences in the mitochondrial DNA control region and phylogeography of the pike-perches *Stizostedion*. *Mol Phylogenet Evol*, 1998, 10: 310~322[DOI]
- 30 Ohno K, Tanaka M, Suzuki H, et al. Identification of a possible control element, Mt5, in the major noncoding region of mitochondrial DNA by intraspecific nucleotide conservation. *Biochem Int*, 1991, 24(2): 263~271
- 31 Kumar S, Suzuki H, Onoue S, et al. Rat mitochondrial mtDNA binding proteins to inter-specifically conserved sequences in the displacement loop region of vertebrate mtDNA. *Biochem Mol Biol Int*, 1995, 36(5): 973~981
- 32 Nesbo C L, Arab M O, Jakobsen K S. Heteroplasmy, length and sequence variation in the mtDNA control region of three percid fish species (*Perca fluviatilis*, *Acerina cernua*, *Stizostedion lucio-perca*). *Genetics*, 1998, 148: 1907~1919
- 33 Delport W, Ferguson J W, Bloomer P. Characterization and evolution of the mitochondrial DNA control region in Hornbills (*Bucerotiformes*). *J Mol Evol*, 2002, 54: 794~806[DOI]
- 34 Randi E, Lucchini V. Organization and evolution of the mitochondrial DNA control region in the Avian Genus *Alectoris*. *J Mol Evol*, 1998, 47: 449~462
- 35 Fumagalli L, Taberlet P, Favre L, et al. Origin and evolution of homologous repeated sequences in the mitochondrial DNA control region of shrews. *Mol Biol Evol*, 1996, 13(1): 31~46
- 36 Crochet P A, Desmarais E. Slow rate of evolution in the mitochondrial control region of gulls (Aves: Laridae). *Mol Biol Evol*, 2000, 17(12): 1797~1806
- 37 Sbisà E, Tanzariello F, Reyes A, et al. Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. *Gene*, 1997, 205: 125~140[DOI]
- 38 Casane D, Denneboug N, Rochambeau H, et al. Nonneutral evolution of tandem repeats in the mitochondrial DNA control region of Lagomorphs. *Mol Biol Evol*, 1997, 14(8): 779~789
- 39 Ravago R G, Monje V D, Juinio-Menez M A. Length and sequence variability in mitochondrial control region of the Milkfish, *Chanos chanos*. *Mar Biotechnol*, 2002, 4: 40~50[DOI]
- 40 Douzery E, Randi E. The mitochondrial control region of Cervidae: evolutionary patterns and phylogenetic content. *Mol Biol Evol*, 1997, 14(11): 1154~1166
- 41 Margulis L. *Origin of Eukaryotic Cells*. New Haven: Yale University Press, 1970
- 42 Andersson S G, Zomorodipour A, Andersson J O, et al. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature*, 1998, 396(6707): 133~140[DOI]
- 43 Cavalier-Smith T. The simultaneous symbiotic origin of mitochondria, chloroplasts, and microbodies. *Ann N Y Acad Sci*, 1987, 503: 55~71