

地山雀线粒体基因组全序列的测定和分析

杨超¹, 雷富民², 黄原^{1,*}

(1. 陕西师范大学 生命科学学院, 陕西 西安 710062; 2. 中国科学院动物研究所 动物进化与系统学重点实验室, 北京 100101)

摘要: 基于长距 PCR 扩增及保守引物步移法测定并注释了地山雀(*Pseudopodoces humilis*)的线粒体基因组全序列。结果表明, 地山雀线粒体基因组全长 1.6 809 万 bp, A+T 含量为 52.9%, 37 个基因排列顺序与红原鸡一致。蛋白质基因的起始密码子中, 除 *COI* 基因为 GTG 外, 其余均为 ATG。 *ND1* 和 *ND5* 基因终止密码子为 AGA; *COII* 基因为 AGG; *COIII* 和 *ND4* 基因为不完全终止密码子 T; 其余基因均为典型的 TAA 或 TAG。预测了 22 个 tRNA 基因的二级结构, 发现 tRNA^{Ser} (AGN) 缺少 DHU 臂, tRNA^{Phe} 的 T ψ C 臂存在一单核苷酸插入。预测的地山雀 12S 和 16S rRNA 二级结构分别包括 3 个结构域 47 个茎环和 6 个结构域 60 个茎环。控制区位于 tRNA^{Glu} 和 tRNA^{Phe} 之间, 长度 1240 bp。控制区结构为高变 I 区、中央保守 II 区和保守序列 III 区 3 个结构域。

关键词: 地山雀; 雀形目; 山雀科; 地山雀属; 线粒体全基因组; RNA 二级结构; 中央保守区
中图分类号: Q344.13; Q786; Q959.739 **文献标志码:** A **文章编号:** 0254-5853-(2010)04-0333-12

Sequencing and Analysis of the Complete Mitochondrial Genome of *Pseudopodoces humilis* (Aves, Paridae)

YANG Chao¹, LEI Fu-Min², HUANG Yuan^{1,*}

(1. School of Life Sciences, Shaanxi Normal University, Xi'an 710062, China;

2. Key Laboratory of Zoological Evolution and Systematics, Institute of Zoology, the Chinese Academy of Sciences, Beijing 100101, China)

Abstract: The complete sequence of *Pseudopodoces humilis* mitochondrial genome was determined by using long PCR and conserved primers walking approaches. The results showed that the entire mitochondrial genome of *Pseudopodoces humilis* was 16809 bp in length with 52.9% A+T content, and the 37 genes had the same gene order with that of *Gallus gallus*. All protein coding genes of the *Pseudopodoces humilis* mitochondrial genome started with ATG codon, except for *COI* with GTG. For terminate codon usage, most of genes terminate with codons TAA or TAG, *ND1* and *ND5* were AGA, *COII* was AGG, and the *COIII* and *ND4* genes had an incomplete termination codon (T). The secondary structures of 22 tRNA were predicted, all tRNA can form typical secondary structure, except tRNA^{Ser} (AGN) with an absence of DHU arm, and tRNA^{Phe} with an extra nucleotide inserted in the T ψ C arm. The predicted secondary structures of 12S rRNA and 16S rRNA exhibit 47 helices in 3 structural domains and 60 helices in 6 domains, respectively. The control region of *Pseudopodoces humilis* located between tRNA^{Glu} and tRNA^{Phe} with 1240 bp in length, and has structure domains found in that of the other birds.

Key words: *Pseudopodoces humilis*; Passeriformes; Paridae; *Pseudopodoces*; Complete mitochondrial genome; RNA secondary structure; Central conserved domain

鸟类线粒体基因组 (mitochondrial genome, mtDNA) 为闭合环状分子, 长度为 16.3~23.5 kb (Li et al, 2002), 与其他脊椎动物线粒体 DNA (mtDNA) 一样, 鸟类 mtDNA 也是一种双链 DNA 分子, 含有 13 个左右的蛋白编码基因、2 个核糖体 RNA 和 22 个左右的 tRNA 基因 (Boore, 1999)。基因排列高

度简洁, 无内含子和转座子, 除 1~2 kb 的非编码区外, 整个基因组都有编码功能 (Quinn, 1997)。在鸟类线粒体基因组所编码的 37 个基因中, ND6 和 8 个 tRNA 基因 (*Glu*、*Gln*、*Ala*、*Asn*、*Cys*、*Tyr*、*Pro* 和 *Ser-UCN*) 由轻链编码, 其余皆为重链编码。由于具有分子结构简单、母性方式遗传、碱基突变

收稿日期: 2010-01-11; 接受日期: 2010-04-07

基金项目: 国家杰出青年科学基金和自然科学基金资助项目 (30925008; 30870270); 中国科学院动物进化与系统学重点实验室开放课题资助项目 (O529YX5105)

*通讯作者 (Corresponding author), E-mail: yuanh@snnu.edu.cn

率高和进化速度快等特点,线粒体基因组已成为从分子水平研究起源、进化及亲缘关系的重要遗传标记(Stoneking & Soodyall, 1996; Ramirez et al, 1993; Avise et al, 1990)。截至2009年12月,GenBank公布的已测出全线粒体基因组的雀形目鸟类仅为9种。随着测序技术的飞速发展及基因组研究的不断深入,近些年来,线粒体ND2、12/16S rRNA、Cyt b、COI和D-loop区等材料相继被用于雀形目鸟类的系统学研究(Dimcheff et al, 2002; Spicer & Dunipace, 2004; Ma et al, 2005; Liang et al, 2007; Singh et al, 2008),然而,这些研究仅仅是基于某个基因或部分序列,大规模、系统性的采用线粒体全基因组进行研究则并不多见。此外,有关该类群的起源以及各科演化关系,也一直是鸟类学界研究的热点(Ericson et al, 2000)。

由于雀形目鸟类具有极大的多样性,测序的物种较少,加之趋同进化的干扰和建树时的长枝吸引等问题,导致在推断该类群分类和进化关系上存在困难。因而,补充该类群线粒体全序列有利于正确地解决其存在的系统发育问题。此外,目前在鸟类中还没有完整的线粒体16S和12S rRNA的二级结构报道,而二级结构的信息是指导rRNA基因序列比对、数据划分和加权分析的前提。

分布于青藏高原的地山雀(*Pseudopodoces humilis*)为地栖性鸟类,隶属于雀形目(Passeriformes)山雀科(Paridae)地山雀属(*Pseudopodoces*),为我国特有种。关于地山雀分类地位问题一度引起颇多争议,曾被作为地鸦属(*Podoces*)的亚属(Zarudny & Loudon, 1902)或被独立列为拟地鸦属(*Pseudopodoces*) (Riley, 1930; Vaurie, 1959)。近些年,有学者结合骨骼比较特征与线粒体DNA以及核DNA的综合研究,建议将该物种定为“地山雀”(James et al, 2003),从鸦科(Corvidae)移到山雀科,得到了国际上广泛的承认。本文测定了地山雀的线粒体全基因组,在对序列进行拼接、注释的基础上,对线粒体各基因的起始位置、组成含量、结构特征等进行了详细的阐述,预测和分析了22种tRNA和两种rRNA的二级结构。地山雀全线粒体基因组的测定和完整的16S和12S rRNA的二级结构预测,将为雀形目鸟类线粒体谱系基因组学的研究提供新的数据资料。

1 材料与方法

1.1 材料

地山雀(*Pseudopodoces humilis*)标本于2003年8月6日采集于青海省青海湖鸟岛,浸泡于无水乙醇中,并保存于-20℃的冰柜;而凭证标本(4783)现保存于中国科学院动物研究所动物进化与系统学重点实验室鸟类标本馆。

1.2 方法

1.2.1 总DNA提取 采用传统的酚-氯仿提取法,将地山雀肌肉置于600 μL匀浆缓冲液(0.01 mol/L Tris, pH 7.8; 5 mmol/L EDTA, 5% SDS, 50 ng proteinase K/μL)中研磨,65℃水浴消化3~5 h。用平衡酚(pH 7.6~7.8)提纯两次,再用CI(氯仿:异戊醇=24:1)提纯一次。最后用-20℃预冷的100%乙醇沉淀和70%乙醇洗涤总DNA(Dang et al, 2008)。

1.2.2 引物设计和PCR扩增 根据已公布的红原鸡(*Gallus gallus*)mtDNA序列(NC_001323),参考相关文献(Sorenson et al, 1999; Sorenson, 2003)提供的引物位置和比对保守区确定引物对,并对每条引物进行评价和修改。最后确定29对Sub-PCR引物和6对候补引物,并从中选取5对作为长PCR引物,它们在整个基因组中的分布、扩增片段大小等详见表1,表2。

以所提取的总DNA为模板,用5对长PCR引物扩增覆盖整个线粒体基因组并相互重叠的片段D1~D5。长PCR反应程序为:93℃预变性2 min; (92℃ 10 s, 58~53℃ 30 s, 68℃ 10 min)×20; (92℃ 10 s, 53℃ 30 s, 68℃ 10 min, 且每一循环增加20 s)×20; 68℃ 7 min; 4℃保温。反应体系为15 μL: 10×LA PCR Buffer 1.5 μL, 25 mmol/L MgCl₂ 1.5 μL, 2.5 mmol/L dNTP 2.4 μL, 10 μmol/L引物各2.1 μL, 5 U/μL LA Taq DNA聚合酶0.18 μL,总DNA(模板)1~3 μL,然后加ddH₂O补足。

以长PCR的纯化产物为模板进行Sub-PCR,扩增出长度为500~1300 bp的短片段。Sub-PCR反应程序:95℃预变性,4 min; (95℃ 45 s, 53~58℃ 60 s, 72℃ 60 s)×30/35; 72℃ 7/4 min, 4℃保温。Sub-PCR反应体系(25 μL)为:10×PCR Buffer 2.5 μL, 25 mmol/L MgCl₂ 2.5 μL, 2.5 mmol/L dNTP 2.5 μL, 10 μmol/L引物各2.0 μL, 5 U/μL Taq DNA聚合

表 1 长 PCR 扩增引物序列
Tab. 1 Primer sequences for Long PCR amplification

片段 Fragments	引物名称 Name	引物序列(5'- 3') Primer sequences
D1	L2260	CAA GGT AAG TGT ACC GGA AGG TG
D1	H6681	GGT ATA GGG TDC CRA TGT CTT TRT G
D2	L6615	CCT CTG TAA AAA GGA CTA CAG CC
D2	H10884	GGG TCR AAW CCR CAT TCG TAT GG
D3	L10635	CAC CAC TTY GGC TTY GAR GCA GC
D3	H13563	TGR AGG GCD GCR GTG TTR GC
D4	L13040	ATC CAA TGG TCT TAG GAR CCA
D4	H16064	CTT CAA TCT TTG GYT TAC AAG ACC
D5	L15725	AAA CCH GAA TGA TAC TTC CTM TTY GC
D5	H1530	GGT GGC TGG CAC ARG ATT TAC C

兼并位点说明(degenerate sites): G, A, T=D; A, g=R; A, C=M; A, T=W; C, T=Y; A, T, C=H.

表 2 扩增线粒体全基因组片段引物对列表
Tab. 2 The list of partial pair primers for amplifying the mitochondrial genome

片段 Fragments	引物名称 Name	预计扩增长度 Expected length (bp)	片段 Fragments	引物名称 Name	预计扩增长度 Expected length (bp)
B01	L1263 (tPhe)	596	A06	L10635 (COX3)	1202
	H1859 (12S)			H11837 (ND4)	
B02	L1754 (12S)	540	A07	L11458 (ND4L)	886
	H2294 (tVal)			H12344 (ND4)	
B03	L2260 (12S)	631	A08	L12156 (ND4)	891
	H2891 (16S)			H13047 (tLeu)	
B04	L2725 (16S)	567	A09	L13040 (tLeu)	523
	H3292 (16S)			H13563 (ND5)	
B05	L3218 (16S)	503	A10	L13525 (ND5)	602
	H3784 (16S)			H14127 (ND5)	
B06	L3722 (16S)	448	A11	L14080 (ND5)	969
	H4170 (ND1)			H15049 (CYTB)	
B07	L3803 (16S)	841	A12	L14770 (ND5)	525
	H4644 (ND1)			H15295 (CYTB)	
B08	L4500 (ND1)	701	B14	L14996 (CYTB)	650
	H5201 (tMet)			H15646 (CYTB)	
B09	L5143 (tGln)	623	B15	L15413 (CYTB)	651
	H5766 (ND2)			H16064 (tTHR)	
B10	L5758 (ND2)	923	B16	L15725 (CYTB)	2000
	H6681 (CO1)			H1530 (12S)	
B11	L6615 (tTyr)	507	B17	L16087 (tTHR)	50
	H7122 (CO1)			H16137 (tPRO)	
B12	L7036 (CO1)	512	B18	L15710	补充引物
	H7548 (CO1)			H16728	
B13	L7525 (CO1)	596	B20	L16225	补充引物
	H8121 (CO1)			H614	
A01	L7987 (CO1)	641	B21	L537	补充引物
	H8628 (CO2)			H1860	
A02	L8386 (CO2)	849	B26	L16206	补充引物
	H9235 (ATP6/ATP8)			H739	
A03	L8929 (CO2)	797	B29	L16525	补充引物
	H9726 (ATP6)			H1530	
A04	L9700 (ATP6)	643	B34	L436	补充引物
	H10343 (CO3)			H1530	
A05	L10236 (CO3)	648			
	H10884 (ND3)				

酶 0.25 μ L, 长片段产物(模板)0.5~1 μ L, 然后加 ddH₂O 补足。

所有长 PCR 产物和 Sub-PCR 产物均用 1.0%的琼脂糖凝胶电泳检测, 使用 DNA 凝胶回收试剂盒进行纯化。

1.2.3 测序 Sub-PCR 产物进行直接测序或者克隆测序, 克隆使用的载体为 pMD18-T。测序均由上海生工生物工程技术服务有限公司完成。

1.2.4 序列拼接、注释和分析 使用 Staden Package 1.7 (Bonfield et al, 1995)进行序列拼接和注释, 通过与斑胸草雀 *Taeniopygia guttata*(NC_007897)线粒体 DNA 序列进行比对确定蛋白质基因、tRNA 基因和 rRNA 基因的位置。使用 MEGA 3.0 统计线粒体基因组碱基组成, 蛋白质基因密码子使用频率等。使用 tRNAScan-SE ver.1.21(Desjardins & Morais, 1990)进行 tRNA 二级结构预测。参照从 RNA 数据库(<http://www.rna.cccb.utexas.edu/>)下载的扁嘴鹅 (*Coscoroba coscoroba*)、绿头鸭(*Anas platyrhynchos*)、红原鸡(*Gallus gallus*)的 srRNA 二级结构, 并结合所发表的五彩绿咬鹃(*Pharomachrus pavoninus*) srRNA 二级结构(Espinosa de los Monteros, 2003), 对 srRNA 基因二级结构进行了预测。参照从 RNA 数据库(<http://www.rna.cccb.utexas.edu/>)下载的非洲爪蟾(*Xenopus laevis*)的 lrRNA 二级结构模板, 并参考黄牛(*Bos taurus*)的 lrRNA 二级结构(Burk et al, 2002), 对 lrRNA 基因二级结构进行了预测。

2 结果

2.1 线粒体基因组组成与基因排列

地山雀线粒体基因组全序列已提交至 GenBank, 登录号为: HM535648。拼接好的地山雀线粒体基因组全长 1.6 809 万 bp, 由 13 个蛋白质编码基因(protein-coding genes, PCGs)、22 个 tRNA 基因、2 个 rRNA 基因(lrRNA 和 srRNA)组成, 基因排列顺序(表 3)与斑胸草雀和红原鸡相同。基因排列非常紧密, 存在 6 处总长度为 29 bp 的基因重叠, 重叠区长度从 1~10 bp 不等; 基因间隔序列总长度为 81 bp, 共计 19 处, 大小在 1~12 bp 之间, 基因间隔最大的在 ND1-tRNA^{Ile}之间和 tRNA^{Pro}-ND6 之间, 达到了 12 bp; 既没有重叠, 也没有间隔的紧密排列基因对共计 12 处。

2.2 核苷酸组成

从表 4 可以看出, 地山雀 mtDNA 碱基百分含

量为 C>A>T>G, A+T%略大于 G+C%, 并无明显差异。4 种碱基的含量都非常接近, 这与已报道的绝大多数鸟类线粒体基因组碱基含量分布一致。蛋白编码基因 4 种碱基百分含量与整个线粒体基因组最为相似。在不同密码子碱基含量的问题上, 地山雀第三位密码子的碱基含量波动较大, 该结论与 Gao et al(2009)对 74 种鸟类全线粒体基因组不同密码子第三位碱基含量波动最大的描述一致, 表明线粒体基因组中碱基变化主要由密码子第三位碱基变化造成。

2.3 蛋白编码基因及密码子使用

地山雀 13 个蛋白编码基因长度与已公布的雀形目鸟类一致。13 个蛋白编码基因所用起始密码子除 *COI* 基因为 GTG 外, 其余均为 ATG; 最常用终止密码子为普通的 TAA, 其它情况为: *NDI* 和 *ND5* 基因终止密码子为 AGA; *COII* 基因终止密码子为 AGG; *COIII* 和 *ND4* 基因为不完全终止密码子 T; *ATP6* 基因和 *ND6* 基因终止密码子均为 TAG; 在所编码的 3 787 个氨基酸中, Ala、Ile、Leu、Ser 和 Thr5 种氨基酸的使用频率最高, 占到了所有氨基酸的 49.44%。相对同义密码子使用也和其它鸟类相似。

2.4 tRNA 二级结构

地山雀线粒体基因组有 22 种 tRNA, 其中 14 种位于重链上, 8 种位于轻链上。所用反密码子与红原鸡相同。用 tRNAScan-SE 1.21 预测出地山雀除 tRNA^{Ser}(AGN)以外的所有 tRNA 二级结构, 其均能形成典型的三叶草结构。在所预测的 22 种 tRNA 二级结构中总共出现了 37 处错配, 其中有 28 对为 GU 错配; 其余为 5 对 AC 配对, 发生在 tRNA^{Asp}、tRNA^{His}、tRNA^{Phe} 和 tRNA^{Ser}(AGN)的氨基酸接受臂和 tRNA^{Ser}(AGN)的反密码子臂; 2 对 UU 错配, 发生在 tRNA^{Gly}的反密码子臂和 tRNA^{Met}的 T ψ C 臂; 2 对 CC 错配, 发生在 tRNA^{Gly}的反密码子臂和 tRNA^{Leu}(UUR)的氨基酸接受臂。

2.5 rRNA 二级结构

地山雀 srRNA 基因位于 tRNA^{Phe}和 tRNA^{Val}之间, 长度为 976 bp。预测的 srRNA 基因二级结构(图 1)与所发表的其他鸟类 srRNA 二级结构大体一样, 包含 3 个结构域 47 个茎环。

lrRNA 基因位于 tRNA^{Val}和 tRNA^{Leu}(UUR)之间, 长度为 1597 bp。预测的地山雀 lrRNA 基因二级结构包含 6 个结构域 60 个茎环(图 2)。

表 3 地山雀线粒体基因组组成
 Tab. 3 The gene contents of *Pseudopodoces humilis* mitochondrial genome

基因 Gene	编码链 Coding strand	起始位置 Start position	终止位置 End position	碱基 长度 Length	间隔 Intergenic length	起始密码子 Initiation condons	终止密码子 Termination condons
<i>tRNA^{Phe}</i>	H	1	69	69			
<i>SrRNA</i>	H	70	1045	976	0		
<i>tRNA^{Val}</i>	H	1046	1115	70	0		
<i>LrRNA</i>	H	1116	2712	1597	0		
<i>tRNA^{Leu}(UUR)</i>	H	2713	2787	75	0		
<i>ND1</i>	H	2793	3770	978	5	ATG	AGA
<i>tRNA^{Ile}</i>	H	3783	3854	72	12		
<i>tRNA^{Gln}</i>	L	3861	3931	71	6		
<i>tRNA^{Met}</i>	H	3935	4003	69	3		
<i>ND2</i>	H	4004	5044	1041	0	ATG	TAA
<i>tRNA^{Trp}</i>	H	5044	5114	71	-1		
<i>tRNA^{Ala}</i>	L	5116	5185	70	1		
<i>tRNA^{Asn}</i>	L	5191	5263	73	5		
<i>tRNA^{Cys}</i>	L	5264	5330	67	0		
<i>tRNA^{Tyr}</i>	L	5330	5400	71	-1		
<i>CO I</i>	H	5402	6952	1551	1	GTG	AGG
<i>tRNA^{Ser}(UCN)</i>	L	6944	7018	75	-9		
<i>tRNA^{Asp}</i>	H	7022	7090	69	3		
<i>CO II</i>	H	7098	7781	684	7	ATG	TAA
<i>tRNA^{Lys}</i>	H	7783	7850	68	1		
<i>ATP8</i>	H	7852	8019	168	1	ATG	TAA
<i>ATP6</i>	H	8010	8693	684	-10	ATG	TAG
<i>CO III</i>	H	8699	9482	784	5	ATG	T
<i>tRNA^{Gly}</i>	H	9483	9551	69	0		
<i>ND3</i>	H	9552	9902	351	0	ATG	TAA
<i>tRNA^{Arg}</i>	H	9904	9973	70	1		
<i>ND4L</i>	H	9975	10271	297	1	ATG	TAA
<i>ND4</i>	H	10265	11642	1378	-7	ATG	T
<i>tRNA^{His}</i>	H	11643	11713	71	0		
<i>tRNA^{Ser}(AGN)</i>	H	11714	11779	66	0		
<i>tRNA^{Leu}(CUN)</i>	H	11779	11849	71	-1		
<i>ND5</i>	H	11850	13667	1818	0	ATG	AGA
<i>Cyt b</i>	H	13676	14818	1143	8	ATG	TAA
<i>tRNA^{Thr}</i>	H	14822	14890	69	3		
<i>tRNA^{Pro}</i>	L	14896	14965	70	5		
<i>ND6</i>	L	14978	15496	519	12	ATG	TAG
<i>tRNA^{Glu}</i>	L	15498	15569	72	1		
D-loop		15570	16809	1240	0		

表 4 地山雀 mtDNA 核苷酸组成
 Tab. 4 The nucleotide composition of *Pseudopodoces humilis* mitochondrial genome

Genes or partitions	A(%)	T(%)	C(%)	G(%)	A+T(%)	n
Whole genome	29.7	23.2	32.6	14.6	52.9	16809
All protein genes	27.6	24.4	33.7	14.3	52.0	11361
1 st codon position	27.5	20.3	28.6	23.6	47.8	3787
2 nd codon position	18.2	40.0	28.8	12.9	58.2	3787
3 rd codon position	37.2	12.9	43.5	6.4	50.1	3787
tRNA genes	28.7	27.8	21.3	22.2	56.5	1548
rRNA	33.0	20.5	26.2	20.3	53.5	2573
A+T rich region	26.4	28.2	31.5	13.9	54.6	1240

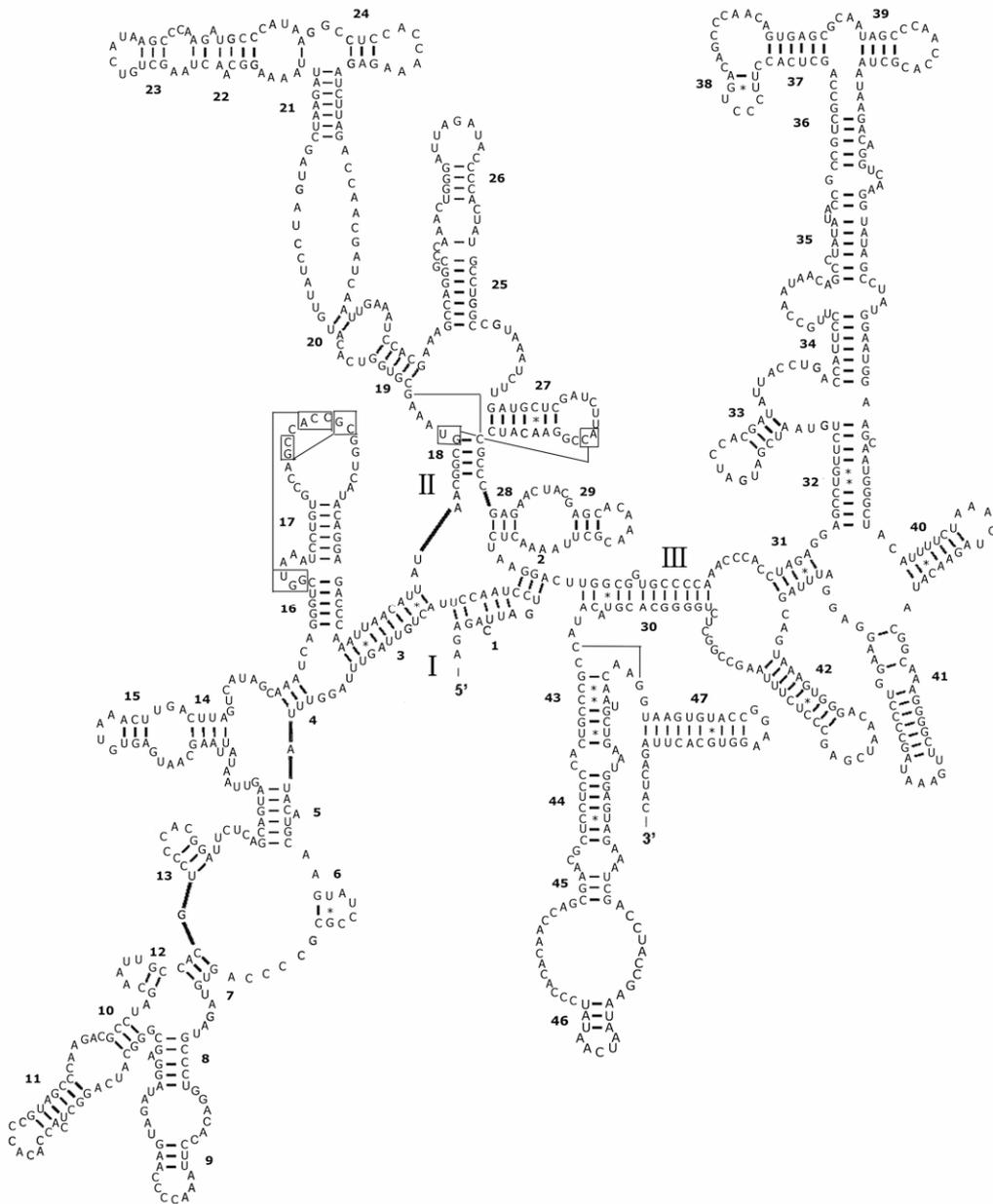


图 1 地山雀 srRNA 二级结构预测结果

Fig. 1 Predicted secondary structures for srRNA of *Pseudopodoces humilis* mispairing are indicted by *

2.6 控制区

地山雀的控制区位于 tRNA^{Glu} 和 tRNA^{Phe} 之间, 长度 1240 bp。本文以红原鸡(Eberhard et al, 2001) 和蓝灰蚱鹌(*Poliopitila caerulea*)(Ruokonen & Kvist, 2002)为基准序列预测了地山雀的线粒体控制区中央保守区 F、E、D、C、B 结构域(图 3), 以及推断了保守序列区(conserved sequence blocks, CSB)中 CSB-1 结构域存在 20 bp 长度碱基序列插入现象(图 4)。

3 讨论

3.1 ND3 基因的胞嘧啶插入现象

据 Gao et al(2009)统计, 古颌超目中几乎所有的物种都发现 ND3 基因存在胞嘧啶插入现象, 而在今颌超目中大多数鸟类都存在不携带胞嘧啶插入的物种, 并且在当时已测定线粒体基因组全序列的 9 个雀形目物种中都不存在胞嘧啶的插入现象。

地山雀 ND3 基因序列 174 位点处不存在胞嘧

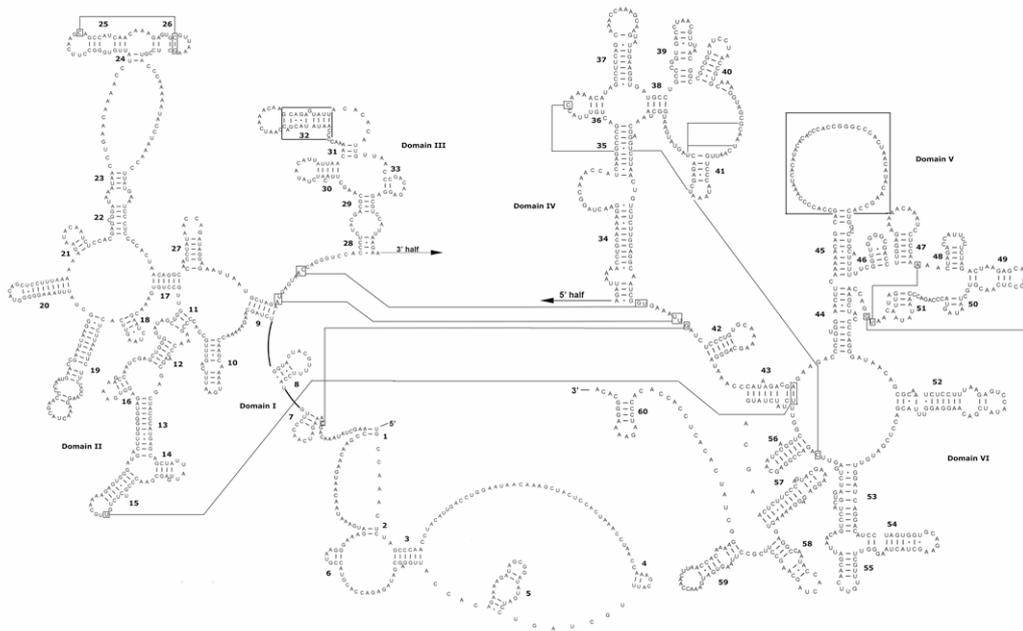


图 2 地山雀 lrRNA 二级结构预测结果

Fig. 2 Predicted secondary structures for lrRNA of *Pseudopodoces humilis* mispairing are indicted by *

啉的插入现象(图 5)。原鸡(图 5 的 3_ *G. gallus*, NC_001323)为 12 个已公布的鸡形目鸟类中惟一的一个不存在胞嘧啶插入的物种;但后续的研究工作发现在原鸡(图 5 的 4_ *G. gallus*, NC_007235)中也存在该现象(Mindell et al, 1998)。Tu et al(2009)对北京鸭(2_ *Anas platyrhynchos*)及 Sun et al(2004)对两种鸚类(5_ *Asio otus*)的线粒体基因组全序列描述中都涉及到了 *ND3* 基因序列存在碱基插入现象。对这一现象的解释,通常认为这个额外的核苷酸在翻译过程中是不参与表达的,而是被 RNA 的自我编辑后剪切掉,使得基因功能得以恢复,从而避免了移码突变导致的 *ND3* 基因转录提前终止 (Mindell et al, 1998)。

3.2 tRNA 二级结构

tRNA 序列变异主要发生在环区,茎区相对保守。其中一些变异如双链的互补性碱基突变、G-U 配对等在所预测的地山雀 tRNA 二级结构中都有出现,大部分碱基对符合 Waston-Click 配对原则,另有一些 GU 变偶碱基对符合 G-U 摆动配对原则,这对于维系 tRNA 二级结构的稳定性起到非常重要的作用。地山雀的所有 tRNA 基因中,有 37 处错配,其中 28 对为 GU 错配。一些学者认为线粒体基因组 tRNA 基因的部分错配可以通过 RNA 自我剪切来恢复基因的功能,不会引起氨基酸转运上的障碍 (Yokobori & Pääbo, 1995)。

地山雀没有发现像位于脊椎动物线粒体基因组的 $tRNA^{Asn}$ 和 $tRNA^{Cys}$ 之间相当于轻链复制起始的发夹结构, Ramirez et al(1993)和 Tu et al(2009)对北京鸭线粒体基因研究也有类似报道;地山雀线粒体 $tRNA^{Phe}$ 的 T Ψ C 臂有一处嘧啶的插入,为不配对的“C”,其组成和配对形式符合 Harrison et al(2004)描述的鸟类线粒体 $tRNA^{Phe}$ 的 T Ψ C 臂通常存在的三种配对形式(图 6a);地山雀 $tRNA^{Ser}(UCN)$ 、 $tRNA^{Lys}$ 、 $tRNA^{Asn}$ 和 $tRNA^{Leu}(CUN)$ 四种 tRNA 基因临近氨基酸接受臂的第八位上均为腺嘌呤“A”,这些有别于其它脊椎动物,可能是鸟类普遍存在的一个特征(Hälid et al, 1997);关于地山雀 $tRNA^{Ser}(AGN)$ 基因缺少 DHU 臂,这种情况在包括鸟类在内的脊椎动物线粒体中比较常见。 $tRNA^{Ser}(AGN)$ 基因均缺少 DHU 臂的长度由低等无脊椎动物到高等脊椎动物逐渐变短,对原鸡的 $tRNA^{Ser}(AGN)$ 二级结构预测发现其 DHU 臂仅有一对碱基对(Lowe & Eddy, 1997),对哺乳动物这类 tRNA 的研究表明,它们可以形成潜在的 L 型结构以维持反密码子与 CCA 接受臂之间的距离(Hanada et al, 2000),tRNA 转运特定氨基酸的功能及其 L 形的三级结构主要由氨基酸臂和反密码子环决定,而 T Ψ C 环和 DHU 环,及其相应的臂,似乎不会抑制 tRNA 的功能(Dirheimer et al, 1995; Wolstenholme et al, 1994)。

3.3 rRNA 二级结构

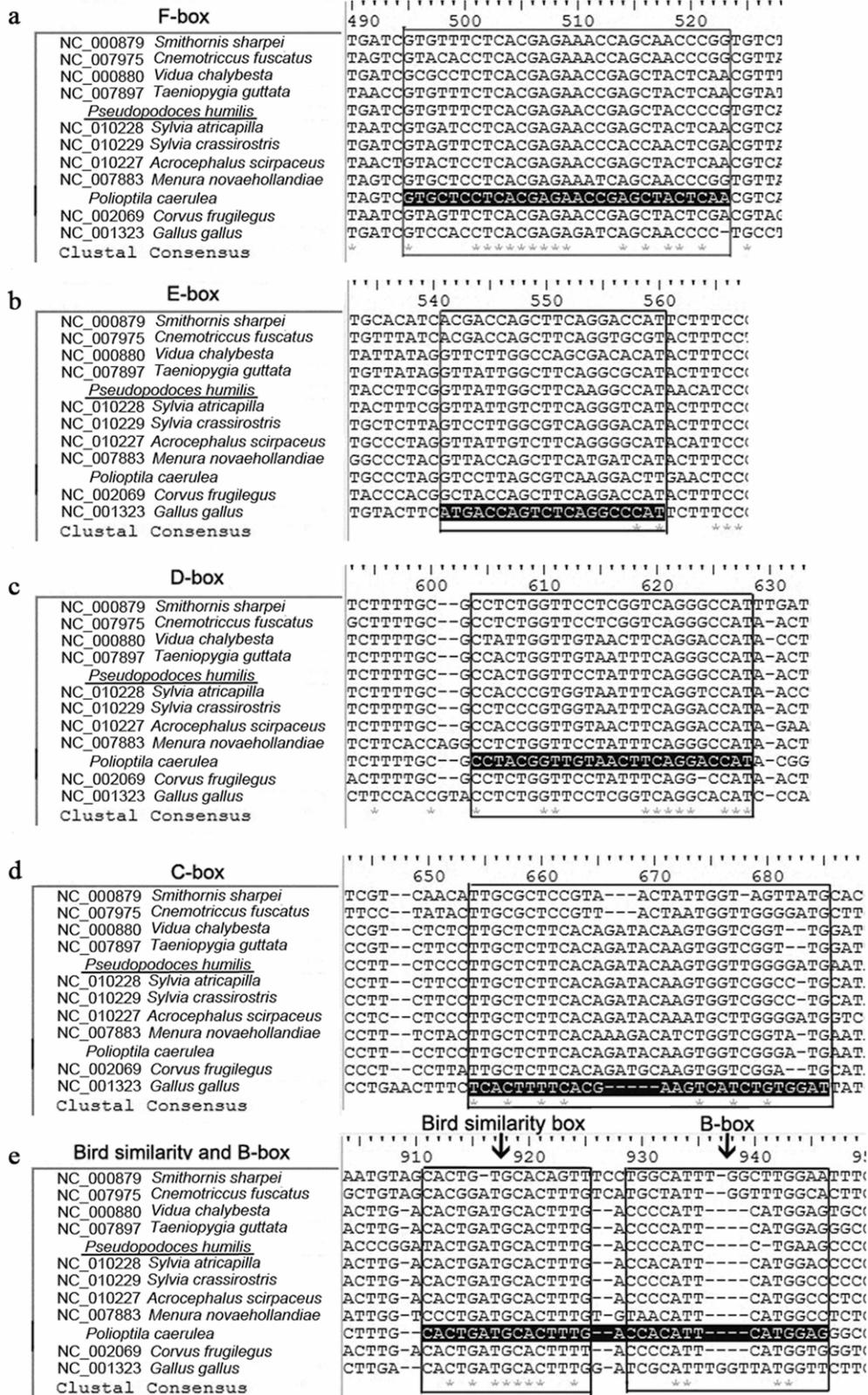


图 3 11 种雀形目鸟类及红原鸡线粒体基因组控制区结构域 (中央保守区)

Fig. 3 The structure domains of control region in 11 Passeriformes birds and *Gallus gallus* (central conserved domain)

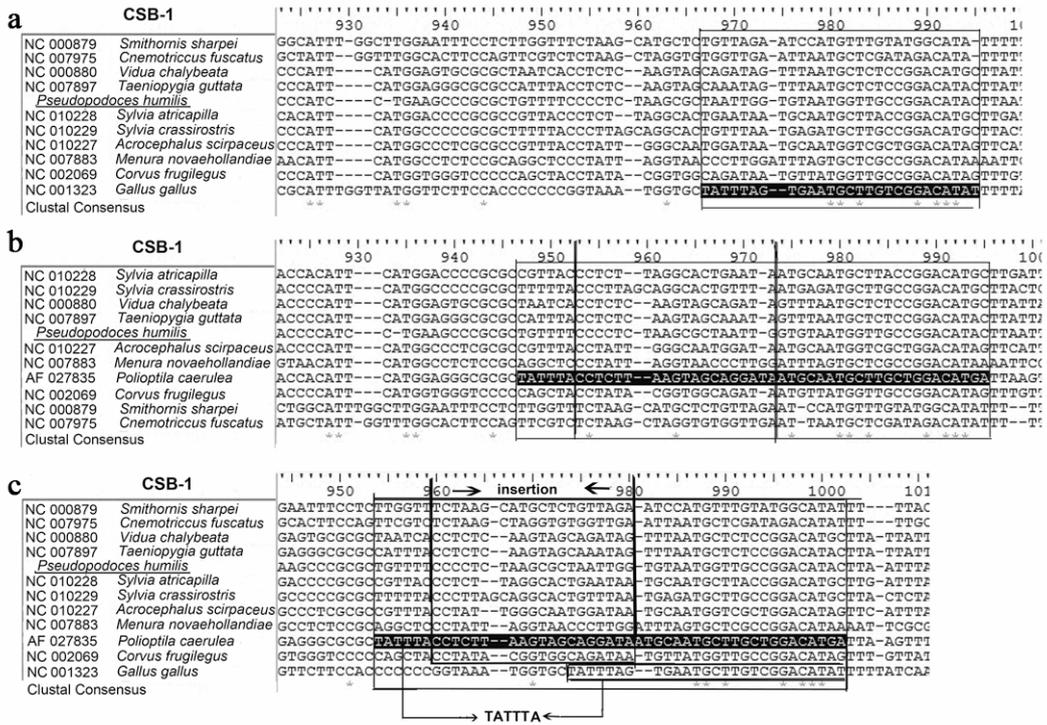


图 4 11 种雀形目鸟类及红原鸡线粒体基因组控制区结构域比较(保守序列区-1)

Fig. 4 The structure domains of control region in 11 Passeriformes birds and *Gallus gallus* (CSB-1)



图 5 5 种鸟类线粒体 ND3 基因序列 174 位点比对情况

Fig. 5 Compared results on the 174 site for ND3 gene of five avian mitochondrial genome

rRNA 中的单链环区和双链茎区交替排列, 受到不同的选择压。一般地, 环区的进化非常快, 在物种间表现出高度的变异性, 而茎区则相对保守 (Noller, 1984; Woese et al, 1980)。但也有一些茎区含有同环区一样的可变长度, 某些位点却有较高的可变性, 而一些环区相反则具有十分保守的序列 (Vawter & Brown, 1993; Simon et al, 1994)。通过预测 rRNA 的二级结构, 有助于了解地山雀 rRNA 基因不同区段行使的功能和所蕴含的进化信息, 从而可以有力的应用于解决其系统发育问题。

srRNA: 地山雀 srRNA 基因二级结构整体上与红原鸡、扁嘴鹅等相似。不同之处在于其 srRNA 二级结构中茎 1 区由 6 个碱基对组成, 比红原鸡、扁嘴鹅等多一对, 并在第三位有一胞嘧啶插入, 形成

单核苷酸突环, 这与五彩绿咬鹃相一致(图 6b); 地山雀结构域 I 中的茎 4 仅形成 3 bp 的配对, 红原鸡、绿头鸭、扁嘴鹅中, 该区域则为 7—8 bp 碱基的配对。通过对上述四种鸟 srRNA 基因序列进行比对, 发现地山雀该区域与其它鸟类相比缺失了“UUUGC”一小段序列, 致使连接茎 3 和茎 4 的内部环异常增大, 且与 U/GUUGC 互补的上游序列 AAACG 依然存在(图 6c)根据双链的互补性碱基突变, 可见地山雀茎 4 并没有发生共变 (covariation); 关于茎 24, 所预测的地山雀该区域形成 3 bp 的碱基配对, 已发表的绿头鸭和扁嘴鹅 srRNA 基因二级结构中该处同样为 3 bp 的碱基配对, 但在左端茎第一、二位碱基对之间存在一单核苷酸胞嘧啶插入形成的凸出, 而红原鸡该处则只有

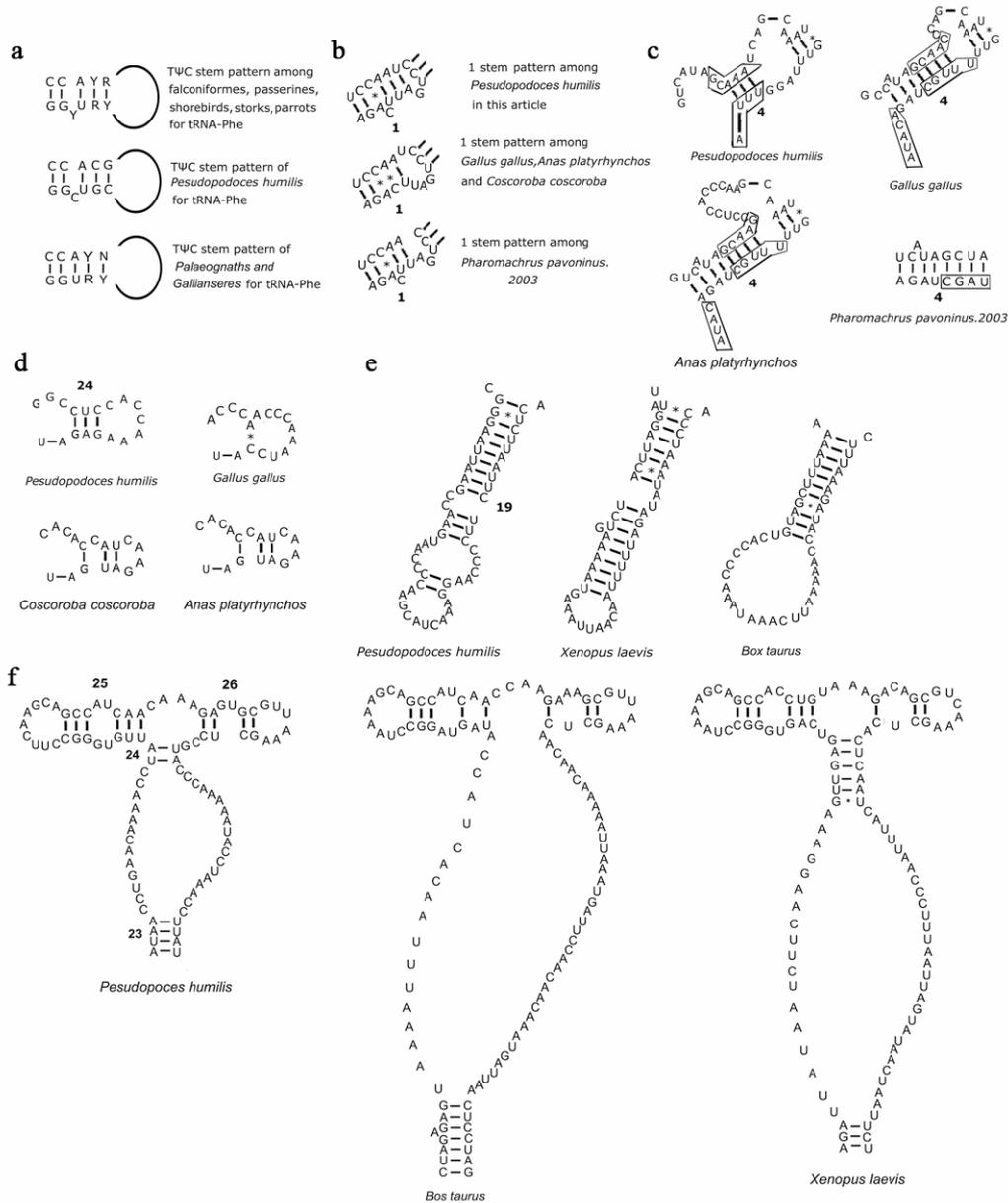


图 6 rRNA 二级结构变化区比较图

Fig. 6 The variable regions of rRNA secondary structures among the species compared

1 bp 的碱基配对，且为弱的 AC 错配(图 6d)；地山雀茎 35 和 36 被认为是一个具有多凸出的连续结构，至少存在 8 bp 的碱基配对才能满足这种排列所需的最小自由能(Hickson et al, 1996)，在五彩绿咬鹃和红原鸡中也有同样的发现。

lrRNA: 由于缺乏鸟类 lrRNA 二级结构资料，本研究参考非洲爪蟾和黄牛的 lrRNA 二级结构进行预测。地山雀 lrRNA 二级结构骨架与非洲爪蟾和黄牛相似，但局部存在许多差异，包括：茎 19，非洲爪蟾形成一个 17 bp 碱基配对的茎，黄牛该处则为 11 bp 碱基配对，同时环区比非洲爪蟾大，地山

雀该处则形成了具有发夹结构的颈环，碱基配对数介于非洲爪蟾和黄牛之间(图 6e)；茎 24，地山雀仅形成 2 bp 碱基配对，黄牛该处没有形成配对的茎，而非洲爪蟾对应的茎，则形成了 6 bp 的配对，且对应的下端内部环黄牛和非洲爪蟾要比本实验的地山雀要大得多(图 6f)。结构域 III 中的茎 32 在所比较的三种物种中变异性很大，呈现出不同的排列形式。变异性较大的区域还有茎 45 的顶端环。结构域 IV 和 V 包含核糖体肽基转移酶中心(peptidyltransferase center)的一部分(Zimmerman et al, 1990)，并且证实紧密连接这两个结构域的是一

种三级结构相互作用(tertiary interaction)(Stiege et al, 1983)。Gutell & Woese(1990)同样发现了该位置上 UU 代替 CC 的共变。Larsen(1992)发现了对应于地山雀线粒体 1rRNA 基因二级结构的茎 49 的发夹环和茎 51 的 3' 端未配对区之间的三级结构相互作用, 我们的研究支持雀形目鸟类中这种三级结构的相互作用(相关发生三级结构之处已在图 1 和图 2 中标出)。

3.4 控制区

地山雀和绝大多数鸟类一样有一个控制区, 位于 tRNA^{Glu} 和 tRNA^{Phe} 之间。但是在近来相继出现了关于具有两个控制区的报道, Singh et al(2008)对芦莺 (*Acrocephalus scirpaceus*) 和黑头莺 (*Sylvia atricapilla*) 的线粒体基因组的描述中存在两个完全相同的控制区, Abbott et al(2005)对 5 种信天翁 (*Thalassarche* spp.) 的全线粒体描述中也有类似双控制区的报道。

控制区结构方面, 先前报道的包括鸟类在内的脊椎动物线粒体控制区结构可分为高变 I 区、中央保守 II 区和保守序列 III 区 3 个结构域。绝大多数鸟类线粒体控制区中央保守区(central conserved

domain) 均可发现保守域 (conserved sequence blocks, CSB) B、D、F, 部分鸟类发现 C、E 区。本文根据 Randi & Lucchini(1998)对石鸡属 (*Alectoris*) 鸟类、Eberhard et al (2001)对 3 种亚马逊鸚鵡 (*Amazona* spp.)、Quinn & Wilson(1993)对雪雁 (*Anser caerulescens*) 以及 Ruokonen & Kvist(2002)对 68 个鸟类物种控制区的分析, 预测了地山雀的线粒体控制区中央保守区结构域的情况(图 3 所示)。

以红原鸡和蓝灰蚋鹌为基准序列所预测的地山雀控制区 CSB-1 结构域(图 4a,b), 结果显示: 图 4b 所预测的结构域序列长度要比图 4a 所预测的情况长得多, 为 46~48 bp 之多。红原鸡和蓝灰蚋鹌基准序列的 5'端均为 TATTTA, 比对的 12 种鸟类 Block-1 区 3'端序列大体一致, 并且核苷酸保守位点几乎全部位于此端。综合比较发现, 与红原鸡相比, 其他 11 种雀形目鸟类均多出一段长度为 19~21 bp 的序列, 如图 4c 所示。目前, 已有学者分析得出雀形目鸟类鹑科、山雀科、莺科、旋木雀科和雀科中, 该区存在被 14~21 bp 序列插入所阻断的现象, 并推测真正行使 CSB-1 功能的序列长度约 20 bp (Ruokonen & Kvist, 2002)。

参考文献:

- Abbott CL, Double MC, Trueman JW, Robinson A, Cockburn A. 2005. An unusual source of apparent mitochondrial heteroplasmy: Duplicate mitochondrial control regions in *Thalassarche albatrosses* [J]. *Mol Ecol*, **14**(11): 3605-3613.
- Avise JC, Ankney CD, Nelson WS. 1990. Mitochondrial genotypes and the evolutionary relationship of mallard and black ducks [J]. *Evolution*, **44**: 1109-1119.
- Bonfield JK, Smith KF, Staden R. 1995. A new DNA sequence assembly program [J]. *Nucleic Acids Res*, **23**(24): 4992-4999.
- Boore JL. 1999. Animal mitochondrial genomes [J]. *Nucleic Acids Res*, **27**(8): 1767-1780.
- Burk A, Douzery EJP, Springer MS, Springer. 2002. The secondary structure of mammalian mitochondrial 16S rRNA molecules: refinements based on a comparative phylogenetic approach [J]. *J Mammal Evol*, **9**(3): 225-252.
- Dang JP, Liu N, Ye W, Huang Y. 2008. Complete mitochondrial genome sequence of *Gastrimargus marmoratus* (Thunberg) (Orthoptera : Acridoidea) [J]. *Acta Entomol Sin*, **51**(7): 671-680. [党江鹏, 刘念, 叶伟, 黄原. 2008. 云斑车蝗线粒体基因组全序列测定与分析. 昆虫学报, **51**(7): 671-680.]
- Desjardins P, Morais R. 1990. Sequence and gene organization of the chicken mitochondrial genome. A novel gene order in higher vertebrates [J]. *J Mol Biol*, **212**(4): 599-634.
- Dimcheff DE, Drovetski SV, Mindell DP. 2002. Phylogeny of Tetraoninae and other galliform birds using mitochondrial 12S and ND2 genes [J]. *Mol Phylogenet Evol*, **24**(2): 203-215.
- Dirheimer G, Keith G, Dumas F, Westhof E. 1995. Primary, Secondary and tertiary Structures of tRNAs [M]// Hill WE, Dahlbert A, Garrett RA, Moore PB, Schlessinger D, Warner JR. tRNA: Structure, Biosynthesis and Function. Washington DC: American Society for Microbiology Press, 93-126.
- Eberhard JR, Wright TF, Bermingham E. 2001. Duplication and concerted evolution of the mitochondrial control region in the parrot genus *Amazona* [J]. *Mol Biol Evol*, **18**(7): 1330-1342.
- Ericson PGP, Johansson US, Parsons TJ. 2000. Major divisions in oscines revealed by insertions in the nuclear gene c-myc: A novel gene in avian phylogenetics [J]. *Auk*, **117**(4): 1069-1078.
- Espinosa de los Monteros A. 2003. Models of the primary and secondary structure for the 12S rRNA of birds: A guideline for sequence alignment [J]. *DNA Seq*, **14** (4): 241-256.
- Gao YK, Miao YW, Su XQ, Chi ZF, Yu B, Jiang F. 2009. A comprehensive analysis on 74 avian mitochondrial genome base compositions [J]. *J Yunnan Agric Univ*, **24**(1): 51-58. [高英凯, 苗永旺, 苏小茜, 池振奋, 俞贇, 姜枫. 2009. 74 种鸟类线粒体基因组碱基组成及特征分析. 云南农业大学学报, **24**(1): 51-58.]
- Gutell RR, Woese CR. 1990. Higher order structural elements in ribosomal RNAs: pseudo-knots and the use of noncanonical pairs [J]. *Proc Natl Acad Sci USA*, **87**(2): 663-667.
- Hanada T, Suzuki T, Watanabe K. 2000. Translation activity of mitochondrial tRNA with unusual secondary structure [J]. *Nucleic Acids Symp Ser*, **44**: 249-250.
- Hälid A, Janke A, Arnason U. 1997. The mtDNA sequence of the ostrich and the divergence between paleognathous and neognathous birds [J]. *Mol Biol Evol*, **14**(7): 754-761.

- Harrison GL, McLenachan PA, Phillips MJ, Slack KE, Cooper A, Penny D. 2004. Four new avian mitochondrial genomes help get to basic evolutionary questions in the late cretaceous [J]. *Mol Biol Evol*, **21**(6): 974-983.
- Hickson RE, Simon C, Cooper A, Spicer GS, Sullivan J, Penny D. 1996. Conserved sequence motifs, alignment and secondary structure for the third domain of animal 12S rRNA [J]. *Mol Biol Evol*, **13**: 150-169.
- James HF, Ericson PGP, Slikas B, Lei FM, Gill FB and Olson SL. 2003. *Pseudopodoces humilis*, a misclassified terrestrial tit (Aves: Paridae) of the Tibetan Plateau: Evolutionary consequences of shifting adaptive zones [J]. *Ibis*, **145**: 185-202.
- Li QW, Li S, Tian CY, Wang YJ, Guo YM. 2002. Molecular evolution and variability in mitochondrial DNA in 10 species of Passeriforme [J]. *Acta Zool Sin*, **48** (5): 625-632. [李庆伟, 李爽, 田春宇, 王勇军, 郭玉梅. 2002. 雀形目 10 种鸟类线粒体的 DNA 变异及分子进化. *动物学报*, **48** (5): 625-632.]
- Liang G, Zhang W, Lei FM, Yin ZH, Huang Y, Li TX. 2007. Comparison of Cyt *b* and *COI* gene sequences from 15 species in Passeriformes [J]. *Acta Zootax Sin*, **32**(3): 613-620. [梁刚, 张卫, 雷富民, 尹祚华, 黄原, 李天宪. 2007. 雀形目 15 种鸟类 *COI* 与 Cyt *b* 基因序列的比较. *动物分类学报*, **32**(3): 613-620.]
- Larsen N. 1992. Higher order interactions in 23S rRNA [J]. *Proc Natl Acad Sci USA*, **89**(11): 5044-5048.
- Lowe TM, Eddy SR. 1997. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence [J]. *Nucleic Acids Res*, **25**(5): 955-964.
- Ma YK, Guo HY, Niu LM. 2005. Primers design and PCR optimization for amplifying mitochondrial Cytochrome *b* gene of 38 Passeriformes [J]. *Forest Sci Tech*, **30**(6): 29-32. [马玉堃, 国会艳, 牛黎明. 2005. 雀形目 38 种鸟类线粒体 Cyt *b* 的引物设计及 PCR 优化. *林业科技*, **30**(6): 29-32.]
- Mindell DP, Sorenson MD, Dimcheff DE. 1998. An extra nucleotide is not translated in mitochondrial ND3 of some birds and turtles [J]. *Mol Biol Evol*, **15**(11): 1568-1571.
- Noller HF. 1984. Structure of ribosomal RNA [J]. *Ann Rev Biochem*, **53**(253): 119-162.
- Quinn MTW. 1997. Molecular Evolution of the Mitochondrial Genome [C]// Mindell DP. Avian Molecular Evolution and Systematics. SanDiego: Academic Press, 5: 3-28.
- Quinn TW, Wilson AC. 1993. Sequence evolution in and around the mitochondrial control region in birds [J]. *J Mol Evol*, **37**(4): 417-425.
- Ramirez V, Savoie P, Morais R. 1993. Molecular characterization and evolution of a duck mitochondrial genome [J]. *J Mol Evol*, **37**(3): 296-310.
- Randi E, Lucchini V. 1998. Organization and evolution of the mitochondrial DNA control region in the avian genus *Alectoris* [J]. *J Mol Evol*, **47**(4): 449-462.
- Riley JH. 1930. Birds collected in Inner Mongolia, Kansu and Chili by the National Geographic Society's Central China Expedition under the direction of FR Wulsin [J]. *Proc US Nat Mus*, **77**: 20-21.
- Ruokonen M, Kvist L. 2002. Structure and evolution of the avian mitochondrial control region [J]. *Mol Phylogenet Evol*, **23**: 422-432.
- Simon C, Frati F, Bekenbach A, Crespi B, Liu H, Flook P. 1994. Evolution, weighting and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers [J]. *Ann Entomol Soc Am*, **87**: 651-701.
- Singh TR, Shneor O, Huchon D. 2008. Bird mitochondrial gene order: Insight from 3 warbler mitochondrial genomes [J]. *Mol Biol Evol*, **25**(3): 475-477.
- Sorenson MD, Ast JC, Dimcheff DE, Yuri T, Mindell DP. 1999. Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates [J]. *Mol Phylogenet Evol*, **12**(2): 105-114.
- Sorenson MD. 2003-02-14 [2010-1-8]. Avian mtDNA Primers [M/OL]. Boston: Boston University, <http://people.bu.edu/msoren/Bird.mt.Primers.pdf>.
- Spicer GS, Dunipace L. 2004. Molecular phylogeny of songbirds (Passeriformes) inferred from mitochondrial 16S ribosomal RNA gene sequences [J]. *Mol Phylogenet Evol*, **30**(2): 325-335.
- Stiege W, Glotz C, Brimacombe R. 1983. Localisation of a series of intra-RNA cross-links in the secondary and tertiary structure of 23S RNA induced by ultraviolet irradiation of *Escherichia coli* 50S ribosomal subunits [J]. *Nucleic Acids Res*, **11**(6): 1687-1706.
- Stoneking M, Soodyall H. 1996. Human evolution and the mitochondrial genome [J]. *Curr Opin Genet Dev*, **6**(6): 731-736.
- Sun Y, Ma F, Xiao B, Zheng JJ, Yuan XD, Tang MQ, Wang L, Yu YF, Li QW. 2004. The complete mitochondrial genomes sequences of *Asio flammeus* and *Asio otus* and comparative analysis [J]. *Sci Chin C: Life Sci*, **47**(6): 510-520.
- Tu JF, Huang YH, Liu SF, Li N. 2009. Complete sequence determination and analysis of Beijing Duck mitochondrial genome [J]. *Chn J Zool*, **44**(2): 28-33. [涂剑锋, 黄银花, 刘三凤, 李宁. 2009. 北京鸭线粒体基因组全序列测定和分析. *动物学杂志*, **44**(2): 28-33.]
- Vawter L, Brown WM. 1993. Rates and patterns of base change in the small subunit ribosomal RNA gene [J]. *Genetics*, **134**(2): 597-608.
- Vaurie C. 1959. The Birds of the Palaearctic Fauna: A Systematic Reference, Order Passeriformes[M]. London: HF & G Witherby Limited.
- Woese CR, Magrum LJ, Gupta R, Siegel RB, Stahl DA, Kop J, Crawford N, Brosius J, Gutell R, Hogan JJ, Noller HF. 1980. Secondary structure model for bacterial 16S ribosomal RNA phylogenetic, enzymatic and chemical evidence [J]. *Nucleic Acids Res*, **8**(10): 2275-2293.
- Wolstenholme DR, Okimoto R, Macfarlane JL. 1994. Nucleotide correlations that suggest tertiary interactions in the TV-replacement loop-containing mitochondrial tRNAs of the nematodes, *Caenorhabditis elegans* and *Ascaris suum* [J]. *Nucleic Acids Res*, **22**(20): 4300-4306.
- Yokobori S, Pääbo S. 1995. Transfer RNA editing in land snail mitochondria [J]. *Proc Natl Acad Sci USA*, **92**(22): 10432-10435.
- Zarudny N, Loudon HB. 1902. Uber Einteilung des genus Podoces in subgenera [J]. *Orn Monatsb*, **10**: 185.
- Zimmerman RA, Thomas CL, Wower J. 1990. Structure and Function of rRNA in the Decoding Domain and in the Peptidyltransferase Center [M]// Hill WE, Dahlbert A, Garrett RA, Moore PB, Schlessinger D, Warner JR. The Ribosome: Structure, Function and Evolution. Washington DC: American Society for Microbiology Press, 331-347.