

黄脸油葫芦线粒体基因组：一种新的基因排列方式

叶伟¹, 党江鹏¹, 谢令德², 黄原^{1,*}

(1. 陕西师范大学 生命科学学院, 西安 710062; 2. 武汉工业学院食品学院, 武汉 430023)

摘要: 采用长距 PCR 扩增及保守引物步移法测定并注释了黄脸油葫芦 (*Teleogryllus emma*) 线粒体基因组全序列。结果表明, 黄脸油葫芦的线粒体基因组全长 15660bp, A+T 含量为 73.1%。谷氨酸、色氨酸及天冬酰胺的转运 RNA 基因由 N 链编码, 形成了直翅目中的第三种基因排列顺序, 其余结构与其它螞蟓亚目昆虫的线粒体结构一致。基因间隔序列共计 73bp, 间隔长度从 1—24bp 不等; 有 14 对基因间存在共 54bp 重叠, 重叠碱基数在 1—11 bp 之间。13 个蛋白质编码基因中 12 个基因 (除 CO I 基因外) 的起始密码子为标准的 ATN 组成, COI 基因的起始密码子为 TTA。有 10 个基因在基因 3' 端能找到完全的 TAA 或 TAG 终止密码子, 而有三个基因 (COII, ND5 和 ND4) 终止密码子为不完整的 T。除 *tRNA^{Ser}* (AGN) 外, 其余 21 个 tRNA 基因的二级结构均属典型的三叶草结构。黄脸油葫芦 940bp 的 A+T 富集区中存在一个被认为与复制起始有关的保守的二级结构, 该结构不仅存在于直翅目昆虫中, 而且也存在于双翅目、鳞翅目和膜翅目中, 但是未见于昆虫纲的早期分化类群——弹尾目中。

关键词: 黄脸油葫芦; 线粒体基因组; 基因顺序; 蟋蟀科; 直翅目

中图分类号: Q969.26+34; Q754 文献标识码: A 文章编号: 0254-5853-(2008)03-0236-09

Complete Mitochondrial Genome of *Teleogryllus emma* (Orthoptera: Gryllidae) with a New Gene Order in Orthoptera

YE Wei¹, DANG Jiang-peng¹, XIE Ling-de², HUANG Yuan^{1,*}

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

2. College of Food Science, Wuhan Polytechnic University, Wuhan 430023, China)

Abstract: The complete sequence of *Teleogryllus emma* mitochondrial genome was determined using long PCR and conserved primer walking approaches. The results showed that the entire mitochondrial genome of *T. emma* was 15660 bp in length, with an A+T content of 74.3%. The most distinct feature of this mtDNA is that the tRNA genes of Glutamic acid, Serine and Asparagine are located in the minor strand, and thus constitutes the third type of gene order of mitochondria in Orthoptera. The intergenic spacers range from 1 to 24bp in size, totaling 73bp. A total of 54 overlapping nucleotides exist among 14 genes, ranging from 1 to 11bp in size. The TAA and TAG codons are used as stop codons in all the protein coding genes (PCGs) except *COII*, *ND5* and *ND4*. The *COI* gene has an unusual start codon, TTA, instead of ATN or NTG. Most of the tRNA genes, except *tRNA^{Ser}* (AGN), form typical secondary structures. An A+T rich region of *T. emma* mitochondrial genome with 940bp in length exists as an ubiquitous secondary structure found in the Orthoptera and other orders of Insecta, such as Diptera, Lepidoptera, Hymenoptera, but not in Collembola, which is considered the basal group of Insecta.

Key words: *Teleogryllus emma*; Mitochondrial genome; Gene order; Orthoptera; Gryllidae

昆虫的线粒体基因组为双链闭合环状分子, 长度约为 13—19kb (*Onychiurus orientalis* 12984bp; *Trialeurodes vaporariorum*, 18414bp), 一般包括 37 个基因, 其中 13 个蛋白质编码基因 (protein-coding gene, PCG), 22 个转运 RNA (tRNAs) 基因和两

个编码核糖体 RNA (ribosomal RNAs) 的基因 (Wolstenholme, 1992; Boore, 1999); 同时, 基因组中具有一个被认为是与基因组的复制与转录有关的非编码序列, 即所谓的 A+T 富集区 (Wolstenholme, 1992)。

收稿日期: 2008-01-14; 接受日期: 2008-04-08

基金项目: 国家自然科学基金 (30470238; 30670279)

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

第一作者简介: 叶伟, 男, 硕士研究生, 研究方向为昆虫分子进化。

在脊椎动物中, A+T 富集区被称为控制区, 包含线粒体 DNA 轻链的复制起点 (O_L), 并且在复制起点的两侧具有一些保守结构, 这些结构被认为是参与了线粒体 DNA 轻链的复制或转录; 而重链的复制起始则开始于轻链复制完成三分之二的位点 (Tapper & Clayton, 1981; Hixson et al, 1986)。昆虫线粒体基因组的 A+T 富集区的研究相对较少, 且主要集中于模式生物果蝇中。果蝇 N 链的复制起点 (O_N) 位于 A+T 富集区内, 且有 20bp 左右的 Poly(T) 结构 (T-stretch) 作为识别 O_N 的标志, 同时在 O_N 的两侧翼具有一些保守结构, 被认为参与了 N 链的复制或者起调控作用 (Brehm et al, 2001; Tsujino et al, 2002; Saito et al, 2005)。但由于果蝇的 A+T 富集区的结构与半变态昆虫的结构不完全一样, 所以对昆虫线粒体基因组内的 A+T 富集区的结构与功能还有待进一步的研究。

黄脸油葫芦 (*Teleogryllus emma*) 隶属于直翅目 (Orthoptera) 蟋蟀科 (Gryllidae), 是我国常见的一种蟋蟀。蟋蟀科目前还没有一种线粒体基因组全序列被测出, 而在 GenBank 中, 整个直翅目也只有 4 种线粒体基因组全序列的报道, 即非洲飞蝗 (*Locusta migratoria*) (Flook et al, 1995) 东方蝼蛄 (*Gryllotalpa orientalis*) (Kim et al, 2005) 疑钩额蚤 (*Ruspolia dubia*) (Zhou et al, 2007), 以及摩门螽斯 (*Anabrus simplex*) (Fenn et al, 2007)。这些被测出的线粒体基因组的基因排列次序仅在二个亚目之间存在转运 RNA 基因 K 和 D 的排列顺序差别。

本文测定了黄脸油葫芦线粒体基因组全序列 (GenBank accession No EU557269), 在对序列进行拼接、注释和详细分析的基础上, 发现了直翅目的第三种线粒体基因排列方式, 并详细比较了昆虫 A+T 富集区的序列组成特点和能够形成的保守的二级结构。希望通过对黄脸油葫芦线粒体基因组的测序与分析, 为研究蟋蟀科以及整个直翅类群的系统发育关系提供新的信息。

1 材料与方法

1.1 标本来源

黄脸油葫芦标本由谢令德教授于 2005 年 8 月采自河北省保定市, 浸泡于 100%乙醇中, 在冰箱中于 4℃ 保存。

1.2 DNA 提取方法

本实验选用蟋蟀后足股节肌肉为实验材料, 采用常规酚-氯仿-异戊醇抽提法提取 DNA。

1.3 PCR 扩增

从 GenBank 下载了 20 余种半变态类昆虫的线粒体基因组全序列并采用 ClustalX (Thompson, 1997) 进行比对, 以寻找最保守的区域, 设计长 PCR 引物。同时参照 Simon et al (1994) 发表的线粒体通用引物序列设计二次 PCR 引物 (Rychlik, 2007)。

首先利用两对引物 (LP03 与 LP04, LP05 与 LP06) 将线粒体 DNA 扩增为两条相互重叠的片段 A 和 B。L-PCR 引物、反应体系和条件见 Liu et al (2006)。经凝胶纯化的 L-PCR 产物作为后续 PCR (Sub-PCR) 的模板。Sub-PCR 扩增了相互重叠且覆盖整个线粒体基因组的 28 个片段。所有 PCR 产物采用 1.0% 的琼脂糖凝胶电泳检测及纯化, 使用安徽优品 (U-Gene) 公司的凝胶回收试剂盒进行 DNA 回收。

1.4 DNA 测序

对 Sub-PCR 的纯化产物, 采用直接测序或者克隆测序的方法来获得序列信息。克隆载体为 Takara PMD18-T。测序试剂盒为 ABI PRISM BigDye Terminator v3.1, 直接测序反应的引物与 PCR 引物相同。大部分测序工作在本实验室进行操作, 测序仪采用 ABI 公司 3100-Avant 型遗传分析仪 (3100-Avant Genetic Analyzer), 个别较长的片段送往北京奥科生物技术公司进行测序。

1.5 序列拼接

序列拼接由 Staden Package (Ver.1.5.) (Bonfield et al, 1995) 中的 Pregap 4 完成, 通过设置自动删除了序列中的载体序列。组装后的序列以 *.seq 格式输出用于线粒体基因组序列的注释。

1.6 线粒体基因组序列的注释与基因定位

线粒体基因组注释参考非洲飞蝗和东方蝼蛄等线粒体基因组全序列的注释结果, 用 Staden package 中的 Spin 进行注释, 同时参考 ClustalX1.83 和 NCBI 的 Blast 比对来确定出蛋白质基因、tRNA 基因、rRNA 基因、srRNA 基因和 A+T 富集区在线粒体基因组中的位置。其中 tRNA 注释及二级结构的预测参考了 tRNA-scan-SE (ver.1.21) (<http://lowelab.ucsc.edu/tRNAscan-SE/>) (Lowe & Eddy, 1997) 的搜索结果, 部分 tRNA 通过比对, 手工画出了它们的二级结构。A+T 富集区保守二级结构的预测, 使用 RNAdraw1.10 (Hofacker et al,

1995), 选择默认参数进行计算。此外, 序列碱基组成等数据由 MEGA3.0 (Kumar et al, 2004) 完成。

2 结果

2.1 基因组组成与基因排列

黄脸油葫芦线粒体基因组全长 15 660 bp, A+T 含量为 73.1%。这一线粒体基因组大小在直翅目已发表的序列中处于一个中间水平[摩门螽斯 15 766 bp (Fenn et al, 2007); 疑钩额螽 14 971 bp (Zhou et al, 2007)]。整个基因组包括 13 个蛋白质基因、22 个 tRNA 基因、两个核糖体 RNA 基因, 还有一个介于 *tRNA^{Ile}* 基因和 12S *rRNA* 基因之间的非编码区, 即 A+T 富集区 (图 1 和表 1)。基因位置与非洲飞

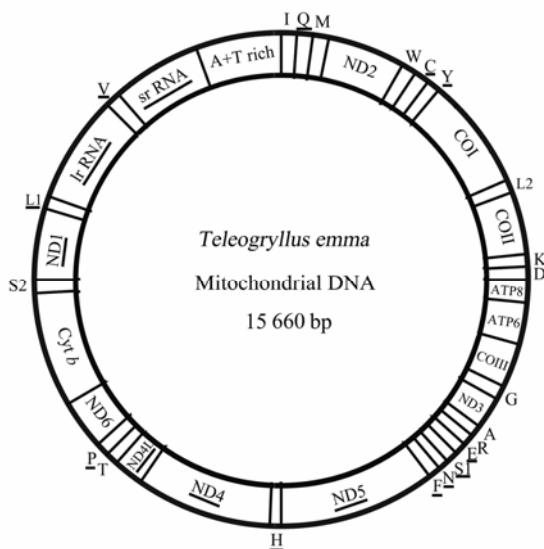


图 1 黄脸油葫芦线粒体基因组的结构

Fig. 1 Gene map of the mitochondrial genome of *Teleogryllus emma*

由 N 链编码的基因用加下划线表示, tRNA 基因用所编码氨基酸的单字母表示。L1, L2, S1 和 S2 分别表示 *tRNA^{Leu(CUN)}*, *tRNA^{Leu(UUR)}*, *tRNA^{Ser(AGN)}* 和 *tRNA^{Ser(UCN)}*。

Those genes encoded by the N strand are underlined. The tRNA genes are designated by single letter amino acid codes. L1, L2, S1 and S2 denote the *tRNA^{Leu(CUN)}*, *tRNA^{Leu(UUR)}*, *tRNA^{Ser(AGN)}*, *tRNA^{Ser(UCN)}*, respectively.

蝗和其他直翅目昆虫不同处是 *tRNA^{Glu}*、*tRNA^{Ser(AGN)}* 和 *tRNA^{Asn}* 由 N 链编码, 即不是通常的 ARNSEF 结构, 而是 ARESNF 结构; 同时, *tRNA^{Lys}* 与 *tRNA^{Asp}* 之间的结构, 也与非洲飞蝗 DK 结构相反, 而与直翅目螽亚目的 KD 结构一致 (图 2)。除 A+T 富集区外, 整个基因组的基因间隔区总长 77 bp, 间隔长度从 1—24 bp 不等; 有 15 对基因间存在共 56 bp 重叠, 重叠碱基数在 1—11 bp 之间 (表 1)。

黄脸油葫芦线粒体基因组与其他昆虫种类一样, 具有很高的碱基组成偏向性, A+T 含量为 73.1%, 介于非洲飞蝗 (75.3%) 与东方蝼蛄 (70.5%) 之间。此外, 蛋白质编码基因的密码子第三位 A+T 的含量 (81.5%) 明显高于第一位及第二位 (分别为 68.2% 和 68.1%) (表 2), 这与密码子第三位的选择压较小有关, 同时也符合线粒体基因组中的 GC→AT 突变的进化趋势。A+T 富集区的 A+T 含量 (73.9%) 没有非洲飞蝗 (86%) 那样高, 而是与整个基因组 A+T 含量基本相当。

2.2 蛋白质编码基因

黄脸油葫芦线粒体基因组中 13 个蛋白质编码基因中的 12 个 (除 CO I 基因外) 的起始密码都为标准的 ATN 组成 (表 1), 其中 6 个为 ATG (*CO II*, *ATP6*, *CO III*, *ND4*, *ND4L*, *Cyt b*); 5 个为 ATT (*ND2*, *ATP8*, *ND3*, *ND5*, *ND6*); 1 个为 ATA (*ND1*)。而 *CO I* 的起始密码较为特殊, 为 TTA。

黄脸油葫芦线粒体基因组中有 10 个基因使用了完全终止密码 (8 个为 TAA, 2 个为 TAG, 见表 1), 有 3 个基因 (*CO II*, *ND5* 和 *ND4*) 与其下游的 tRNA 基因重叠, 使用了不完全终止密码子 T。目前普遍认为, 不完全终止密码子 T 可能通过 mRNA 加工过程中的多聚腺苷酸化形成完整的终止密码子 TAA (Anderson et al, 1981; Ojala et al, 1981; Lu et al, 2002)。

2.3 tRNA 基因

黄脸油葫芦线粒体 tRNA 的长度范围为 61—70 nt, 全部的 22 个 tRNA (Leu 和 Ser 各有两种 tRNA, 其余的 18 种氨基酸各一种 tRNA) 中, 除 *tRNA^{Ser(AGN)}* 外都能形成标准的三叶草结构 (图 3), 氨基酸接受臂 (AA stem) 为 7 个碱基对, 反义密码子臂 (AC stem) 有 5 个碱基对。唯一一个无法形成三叶草结构的 *tRNA^{Ser(AGN)}*, 其 DHU 臂上也形成了典型的 Type-9 结构 (Zhou et al, 2007)。同时, 除 *tRNA^{Gly}* 之外, 其余 21 种 tRNA 的反义密码子环 (AC loop) 都是标准的 7 nt 结构。DHU 臂的长度为 3—5 bp, 而 T 臂为 3—6 bp; 长度变化较大的是 DHU 环、T 环以及可变环 (V loop), 长度分别为 3—8 nt、3—11 nt 和 3—5 nt。

与细胞质中的 tRNA 相比, 线粒体 tRNA 在各个臂上存在较多的错配现象。在黄脸油葫芦的 22 个 tRNA 中, 共有 40 对错配, 其中 26 个以 G-U 配对方式存在, 部分的碱基错配可以通过 RNA 编辑

进行校正 (Yokobori, 1995)。tRNA 结构中存在的这些变异被认为是与物种的进化相关联的 (Takashi, 1991; Watanabe, 1994)。

2.4 rRNA 基因

黄脸油葫芦的两个 rRNA 基因同其他节肢动物一样, 位于两个保守的区段之间。即 16S rRNA 位于 *tRNA^{Leu(CUN)}* 与 *tRNA^{Val}* 之间, 12S rRNA 位于 *tRNA^{Val}* 和 A+T 富集区之间, 长度分别是 1293 bp 和 812 bp, A+T 含量分别为 75.25% 和 71.68%。

2.5 A+T 富集区

黄脸油葫芦的 A+T 富集区全长 940 bp, 结构与非洲飞蝗的一样, 也在 A+T 富集区和相邻的 12S rRNA 基因区域内有两段重复序列 Rpt1(187 bp)和 Rpt2(188 bp)。两段重复序列之间只有个别碱基的差异, 但存在 17 bp 间隔, 且 Rpt1 与 12S rRNA 基因有 57 bp 的重叠。而在重复区以外的部分, 存在一个长度为 6 bp 的 5' 端为“TA”的茎-环结构 (图 4A)。

表 1 黄脸油葫芦线粒体基因组
Tab. 1 Mitochondrial gene profile of *Teleogryllus emma*

Feature	Region ^a		Size (bp)	Codon		Intergenic nucleotides ^b
	From	Stop		Start	To	
<i>tRNA^{Ile}</i>	1	68	68			
<i>tRNA^{Gln}</i>	(66	134)	69			-3
<i>tRNA^{Met}</i>	136	204	69			1
<i>ND2</i>	205	1203	999	ATT	TAA	0
<i>tRNA^{Trp}</i>	1202	1269	68			-2
<i>tRNA^{Cys}</i>	(1262	1322)	61			-8
<i>tRNA^{Tyr}</i>	(1328	1394)	67			5
<i>COX I</i>	1390	2931	1542	TTA	TAA	-5
<i>tRNA^{Leu(UUR)}</i>	2927	2990	64			-5
<i>COX II</i>	2992	3670	679	ATG	T	1
<i>tRNA^{Lys}</i>	3671	3740	70			0
<i>tRNA^{Asp}</i>	3741	3805	65			0
<i>ATP8</i>	3806	3961	156	ATT	TAA	0
<i>ATP6</i>	3955	4638	684	ATG	TAA	-11
<i>COX III</i>	4638	5420	783	ATG	TAG	-1
<i>tRNA^{Gly}</i>	5423	5489	67			0
<i>ND3</i>	5490	5840	351	ATT	TAG	0
<i>tRNA^{Ala}</i>	5840	5905	66			-1
<i>tRNA^{Arg}</i>	5905	5966	62			-1
<i>tRNA^{Glu}</i>	(5965	6031)	67			-2
<i>tRNA^{Ser(AGN)}</i>	(6033	6099)	67			1
<i>tRNA^{Asn}</i>	(6100	6164)	65			0
<i>tRNA^{Phe}</i>	(6174	6237)	64			9
<i>ND5</i>	(6238	7951)	1715	ATT	T	0
<i>tRNA^{His}</i>	(7969	8031)	63			17
<i>ND4</i>	(8035	9376)	1342	ATG	T	3
<i>ND4L</i>	(9370	9666)	297	ATG	TAA	-7
<i>tRNA^{Thr}</i>	9669	9732	64			2
<i>tRNA^{Pro}</i>	(9733	9799)	67			0
<i>ND6</i>	9810	10319	510	ATT	TAA	10
<i>CYTB</i>	10319	11455	1137	ATG	TAA	-1
<i>tRNA^{Ser(UCN)}</i>	11455	11517	63			-1
<i>ND1</i>	(11542	12486)	945	ATA	TAA	24
<i>tRNA^{Leu(CUN)}</i>	(12481	12546)	66			-6
<i>lrRNA(16s)</i>	(12547	13839)	1293			0
<i>tRNA^{Val}</i>	(13840	13908)	69			0
<i>srRNA(12s)</i>	(13909	14720)	812			0
D-Loop(CR)	14721	15660	940			0

a: 基因起始位置由括号括起来表示该基因由 N 链编码 (Brackets denote that the genes are encoded in the N strand); b: IGN: Intergenic nucleotides, 基因间的核酸序列, 负号表示基因间发生重叠 (Numbers correspond to the nucleotides separated by different genes. Negative numbers indicate overlapping nucleotides between adjacent genes)。

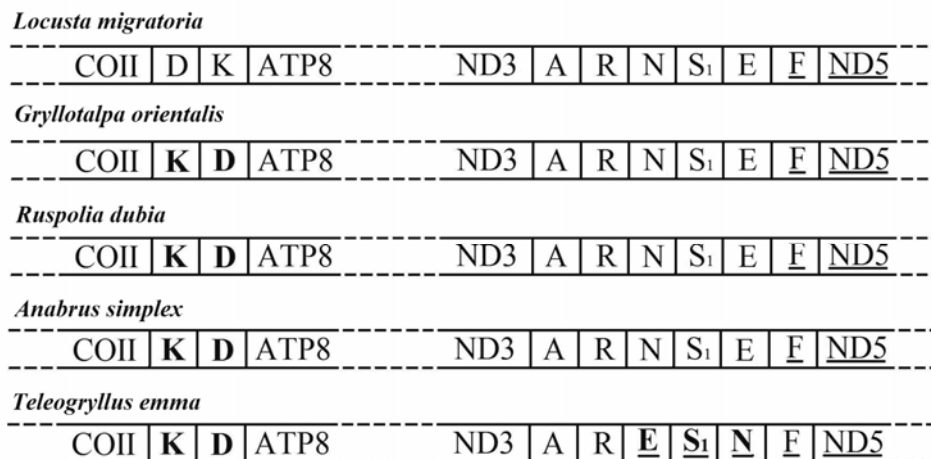


图 2 5 种直翅目昆虫线粒体基因组的基因排列顺序

Fig. 2 Mitochondrial genes arrangements of five species in Orthoptera

由 N 链编码的基因用下划线标出。tRNA 基因以编码的氨基酸缩写表示，虚线表示具有相同排列而未显示的基因。All genes are encoded by the J strands except genes indicated by underbars. The tRNA genes are designated by single-letter amino acid codes. The broken line denotes the genes which are the same order, but not shown in here.

表 2 黄脸油葫芦线粒体基因组碱基组成情况

Tab. 2 Nucleotide composition within the *Teleogryllus emma* mitochondrial genome

	T	C	A	G	Number	A+T	AT-skew ^c	GC-skew ^d
All gene	32.6	17	40.5	9.8	15660	73.1	0.11	-0.27
rRNA genes	42.3	7.6	31.5	18.5	2105	73.8	-0.15	0.42
tRNA genes	37.8	9.8	37	15.4	1456	74.8	-0.01	0.22
D-Loop	32.2	16.1	41.7	10	940	73.9	0.13	-0.23
Protein genes	41.5	13.8	31.2	13.6	11124	72.7	-0.14	-0.01
All codons								
1st	36.5	13.3	31.7	18.5	3708	68.2	-0.07	0.16
2st	44.3	17.3	23.8	14.6	3708	68.1	-0.30	-0.08
3st	43.7	10.7	37.8	7.8	3708	81.5	-0.07	-0.16
Genes on H-strand ^a								
1st	30.8	17.7	33.7	17.9	2265	64.5	0.04	0.01
2st	41.3	21.1	25.2	12.4	2265	66.5	-0.24	-0.26
3st	38.1	13.5	44.4	4	2265	82.5	0.08	-0.54
Total	36.7	17.4	34.4	11.4	6795	71.1	-0.03	-0.21
Genes on L-strand ^b								
1st	45.5	6.5	28.5	19.5	1426	74	-0.23	0.5
2st	49	11.4	21.5	18.1	1426	70.5	-0.39	0.23
3st	52.7	6.2	27.4	13.8	1426	80.1	-0.32	0.38
Total	49.1	8	25.8	17.1	4278	74.9	-0.31	0.36

a: *COI*, *COII*, *COIII*, *ATP6*, *ATP8*, *Cyt b*, *ND2*, *ND3* and *ND6*; b: *ND1*, *ND4*, *ND4L* and *ND5*; c: [A-T]/[A+T];

d: [G-C]/[G+C].

3 讨论

3.1 直翅目线粒体基因排列次序

黄脸油葫芦线粒体基因组最主要的特征是在昆虫中通常所认为的 tRNA 基因簇 ARNESF(Flook et al,1995)发生了重排，形成了 ARESNF 排列。这是一种在直翅目和昆虫纲以前从未报道的新的基因

排列方式。形成这种排列方式的原因是由于 *tRNA^{Glu}*，*tRNA^{Ser(AGN)}*和 *tRNA^{Asn}* 这三个基因所对应的 DNA 片段发生了倒置。关于 DNA 片段倒置产生的机理，目前最有说服力的是 Mark 和 Nick 的观点，即双链断裂和重新结合模型。在 mtDNA 上两相邻位点可能同时断裂，断裂产生的 mtDNA 双链片段在相应酶的作用下反向接回断裂位点，致使基因倒

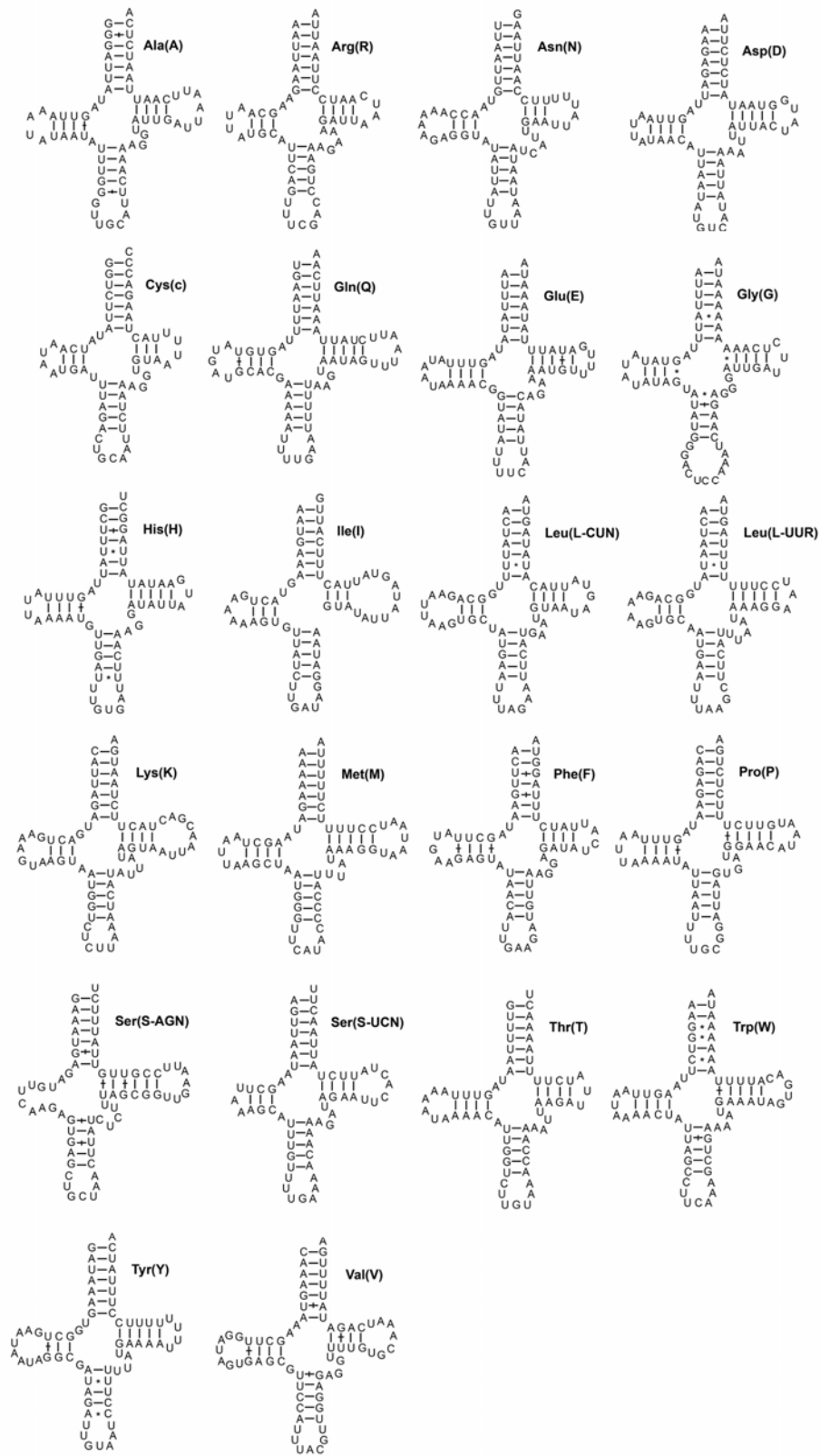


图 3 黄脸油葫芦线粒体 22 个 tRNA 的二级结构

Fig. 3 Inferred secondary structure of tRNAs of *Teleogryllus emma* mitochondrial genome

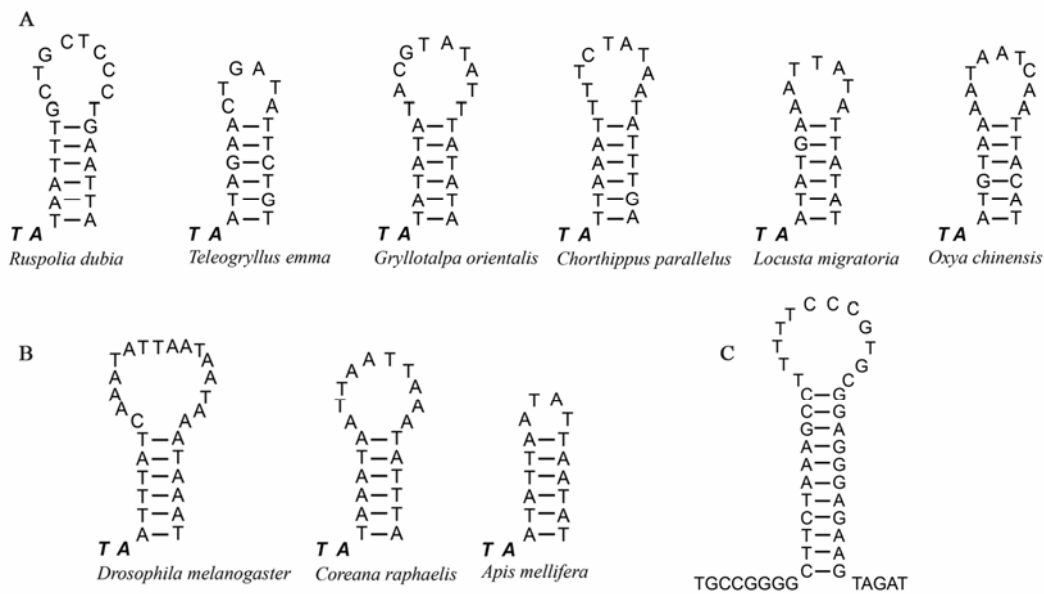


图 4 昆虫 A+T 富集区和人类控制区存在的保守二级结构

Fig. 4 The conserved secondary structure of A+T rich regions in insects and control region of human mitochondrial genome

A: 6 种直翅目昆虫 A+T 富集区的二级结构 (The secondary structure of the A+T rich regions in six species of Orthoptera); B: 双翅目、鳞翅目和膜翅目昆虫 A+T 富集区的二级结构 (The secondary structure of the A+T rich regions in Diptera, Lepidoptera and Hymenoptera); C: 人类的线粒体基因组轻链复制起点 (O_L) 附近的二级结构 (The secondary structure at O_L of human mitochondrial DNA) (Hixson et al., 1986).

置 (Zhong et al., 2005)。

线粒体基因顺序是研究生物较高阶元系统发育的一种非常有用的信息 (Brooe & Brown, 1998), 从图 2 不难看出, 从 tRNA^{Lys}(K) 和 tRNA^{Asp}(D) 的排列顺序上支持直翅目分为两个亚目: 蝗亚目 (Caelifera) 和螽亚目 (Ensifera)。而这样的分类系统也是为学者们广为支持的。与其他螽亚目种类不同的是黄脸油葫芦出现的 ARESNF 基因排列, 如果这种排列方式在所有的蟋蟀总科都一致的话, 就可以推断蟋蟀总科是螽亚目其他类群的姐妹群; 甚至可以像 Ragge 认为的那样, 蟋蟀类可以从螽亚目中划分出来, 成为独立的类群 (Zheng et al., 1999)。

3.2 密码子变异分析

黄脸油葫芦线粒体的 *CO I* 基因的起始区以及与其 *tRNA^{Tyr}* 相邻区域找不到典型的起始密码子, 通过与直翅目其它种的比对, 发现与非洲飞蝗 *CO I* 的起始密码 (ATTA) 处于同一位置的 CTTA, 这是一个从未被报道过的起始密码, 所以我们选择了一个不常见但是很有可能是起始密码的 TTA (Lavrov et al., 2000; Nardi et al., 2001) 作为黄脸油葫芦的 *CO I* 的起始密码子。*CO I* 起始密码子的多样性在昆虫中相当普遍。以前的多篇文献都对 *CO I* 的这

种特殊情况进行过讨论, 如双翅目昆虫中发现的特殊起始密码子 TCG (Junqueira et al., 2004) 和六联起始密码子 (ATTTAA) (Beard et al., 1993), 也有提出四联起始密码子的 (ATAA、TTAA 和 ATTA) (Wilson et al., 2000)。在直翅目中, 除了东方蝼蛄 *CO I* 基因的起始密码子为标准的 ATG (Kim et al., 2005) 外, 非洲飞蝗具有特殊四联起始密码子 ATTA (Flook et al., 1995), 摩门螽斯 (Fenn et al., 2007) 的起始密码则为 CCG, 疑钩额螽 (Zhou et al., 2007) 则与黄脸油葫芦相同, 皆为 TTA。但是对于 *CO I* 基因起始密码子多样性的原因, 至今尚不清楚。

3.3 A+T 富集区的重复序列、调控序列与保守的二级结构

直翅目昆虫中非洲飞蝗 A+T 富集区有重复两次的重复序列 (Flook et al., 1995)。黄脸油葫芦 A+T 富集区和与之相邻的 12S *rRNA* 基因组成的区域内, 也有两段重复序列 Rpt1(187bp) 和 Rpt2(188bp)。以前曾有人认为这些重复序列的结构是保守的, 并且与线粒体的复制过程相关 (Zhang et al., 1995, 1997)。但是从已经公布的数据以及我们实验室一些尚未发表的数据来看, 这样的重复结构似乎与线粒体基因组的复制过程关系不大。因为在同一目中一些物种的线粒体 DNA 的 A+T 富集区不含有这样

的重复结构,典型的例子如疑钩额蚤(Zhou et al, 2007)。所以这类重复结构可能是由复制滑移造成的,而包含复制信息的区域可能是位于重复区之外的部分。

A+T 富集区包含与线粒体基因组复制相关的信息,这是一个被普遍接受的观点。目前对该区域的关注也主要集中于所包含的复制相关的调控信息。现在公认的是昆虫的线粒体 N 链的复制起点(O_N)位于控制区内, J 链的复制则是在轻链的复制完成 97% 之后才开始,即认为 O_J (J 链的复制起点)位于距离 O_N 97% mtDNA 长度的位置上(Goddard & Wolstenholme, 1978, 1980; Saito et al, 2005)。但关于 O_J 的识别及与之相关的调控元件,至今尚无定论。

Saito et al (2005) 以果蝇为主要研究对象,同时也对 10 个目的 20 种昆虫的 A+T 富集区进行了研究。根据实验结果,作者认为 O_N 位于控制区内,同时在 O_N 的上游有一段 Poly(T) 结构,可能与复制起点的识别有关。同时 Saito et al (2005) 根据“97%”这个原则,推断 O_J 也位于控制区内,并且在相应的区域也发现了 Poly(T) 结构。由于果蝇的控制区的结构与其它类群明显不同(Zhang et al, 1995),所以该实验结果在其它类群的适用性还有待进一步的证明。Saito et al (2005) 也认为像非洲飞蝗这样没有典型 PolyT 结构的类群,可能会有一段序列来行使 Poly(T) 的职能。

黄脸油葫芦的 A+T 富集区找不到典型的 Poly(T) 结构,同时现有的直翅目昆虫线粒体基因组中也没有发现类似的结构。而由于疑钩额蚤的控制区只有 70bp (Zhou et al, 2007),所以据此推测,如果轻链的复制是从控制区开始的,并且在 O_N 附近要有一些调控元件的话,那么这 70 bp 的区域内,必须包括全部的信息。通过 RNAdraw1.10 (Hofacker et al, 1995) 预测发现疑钩额蚤的控制区会形成一个 5' 端为“TA”的 6bp 的茎环结构。之后通过对直翅目的其他 6 个种——非洲飞蝗, 东方蝼蛄, 沙漠蝗

(*Schistocerca gregaria*) (Zhang et al, 1995), 草地蝗 (*Chorthippus parallelus*) (Zhang et al, 1995), 中华稻蝗 (*Oxya chinensis*) (Zhang & Huang, 2008) 以及黄脸油葫芦的控制区(对于有重复区的物种,使用了不包括重复区的序列)进行了二级结构的预测,发现这样的结构同样存在于上述的 6 个物种中(图 4A)。而同样的结构也存在于全变态类的韩国灰纹蝶 (*Coreana raphaelis*) (Kim et al, 2006), 意大利蜜蜂 (*Apis mellifera*) (Crozier & Crozier, 1993), 以及黑腹果蝇 (*Drosophila melanogaster*) (GenBank 序列登陆号: NC_001709) 中(图 4B), 但是在弹尾目的 *Tetrodontophora bielanensis* (Nardi et al, 2001) 和 *Gomphiocephalus hodgsoni* (GenBank 序列登陆号: NC_005438) 中没有发现相似的结构。这些二级结构除了在 5' 端都有“TA”结构外,在序列的组成上完全不同。在脊椎动物的线粒体控制区内,也具有类似的与 O_L 相连二级结构,被认为是参与了轻链的复制及转录过程(图 4C) (Tapper & Clayton, 1981; Hixson et al, 1986; Ray & Densmore, 2002, 2003)。同时在昆虫纲的大部分物种中, A+T 富集区中都含有相当数量的回文结构 (palindromes) 和反向重复序列 (inverted repeats, IRs), 这些结构被认为是与 mtDNA 的复制起始相关联 (Arunkumar & Nagaraju, 2006)。所以我们推测这样的结构可能参与了轻链的复制过程。而由于昆虫纲早期分化的类群弹尾目的两个物种没有发现这样的茎环结构,所以推测这样的结构可能是在昆虫纲的有翅亚纲中开始出现的,或在昆虫纲进化的早期就存在而在目前在弹尾目中丢失了。

致谢: 感谢陕西师范大学生命科学学院卢慧麓博士和刘念博士在引物设计和实验过程中给与的支持和帮助。

参考文献:

- Anderson S, Bankier AT, Barrell BG, Bruijn MHL, Coulson AR, Drouin J. 1981. Sequence and organization of the human mitochondrial genome[J]. *Nature*, **290**(5806): 457-465.
- Arunkumar KP, Nagaraju J. 2006. Unusually long palindromes are abundant in mitochondrial control regions of insects and nematodes[J]. *PLoS ONE*, **1**: e110.
- Beard CB, Hamm DM, Collins FH. 1993. The mitochondrial genome of the mosquito *Anopheles gambiae*: DNA sequence, genome organization, and comparisons with mitochondrial sequences of other insects[J]. *Insect Mol Biol*, **2**(2): 103-124.
- 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.
- Brehm A, Harris DJ, Hernandez M, Cabrera V, Larruga J, Pinto F, Gonzalez AM. 2001. Structure and evolution of the mitochondrial DNA complete control region in the *Drosophila subobscura* subgroup[J]. *Insect Mol Biol*, **10**(6): 573-578.
- Crozier RH, Crozier YC. 1993. The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization[J]. *Genetics*, **133**(1): 97-117.
- Fenn JD, Cameron SL, Whiting MF. 2007. The complete mitochondrial genome sequence of the Mormon cricket *Anabrus simplex*: Tettigoniidae: Orthoptera and an analysis of control region variability[J]. *Insect Mol Biol*, **16**(2): 239-252.
- Flook PK, Rowell CHF, Gellissen G. 1995. The sequence, organization, and evolution of the *Locusta migratoria* mitochondrial genome[J]. *J Mol Evol*, **41**(6): 928-941.
- Goddard JM, Wolstenholme DR. 1978. Origin and direction of replication in mitochondrial DNA molecules from *Drosophila melanogaster*[J]. *Proc Natl Acad Sci U S A*, **75**(8): 3886-3890.
- Goddard JM, Wolstenholme DR. 1980. Origin and direction of replication in mitochondrial DNA molecules from the genus *Drosophila*[J]. *Nucleic Acids Res*, **8**(4): 741-757.
- Hixson JE, Wong TW, Clayton DA. 1986. Both the conserved stem-loop and divergent 5'-flanking sequences are required for initiation at the human mitochondrial origin of light-strand DNA replication[J]. *J Biol Chem*, **261**(5): 2384-2390.
- Hofacker IL, Stadler PF, Bonhoeffer S, Fontana W. 1995. The Vienna RNA package: A comprehensive package for RNA secondary structure calculation and analysis for Windows 95/NT 3.51. <ftp.itc.univie.ac.at>.
- Junqueira ACM, Lessinger AC, Torresa TT, Silvab FR, Vettorec AL, Arrudad P, Espin AMLA. 2004. The mitochondrial genome of the blowfly *Chrysomya chloropyga* Diptera: Calliphoridae[J]. *Gene*, **339**: 7-15.
- Kim I, Cha SY, Yoon MH, Hwang JS, Lee SM, Sohn HD, Jin BR. 2005. The complete nucleotide sequence and gene organization of the mitochondrial genome of the oriental mole cricket, *Gryllotalpa orientalis* Orthoptera: Gryllotalpidae[J]. *Gene*, **353**(2): 155-168.
- Kumar S, Tamura K, Nei M. 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment[J]. *Brief Bioinform*, **5**(2): 150-163.
- Lavrov DV, Boore JL, Brown WM. 2000. The complete mitochondrial DNA sequence of the horseshoe crab *Limulus polyphemus*[J]. *Mol Biol Evol*, **17**(5): 813-824.
- Liu N, Hu J, Huang Y. 2006. Amplification of grasshoppers complete mitochondrial genomes using long PCR[J]. *Chin J Zool*, **41**: 61-65.[刘念, 胡靖, 黄原. 2006. 应用长 PCR 扩增蝗虫线粒体全基因组. 动物学杂志, **41**: 61-65.]
- 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.
- Lu Ch, Liu YQ, Liao ShY, Li B, Xiang ZH, Han H, Wang XG. 2002. Complete sequence determination and analysis of *Bombyx mori* mitochondrial genome[J]. *Journal of Agricultural Biotechnology*, **10**(2): 163-170.[鲁成, 刘运强, 廖顺尧等. 2002. 家蚕线粒体基因组全序列测定与分析. 农业生物技术学报, **10**(2): 163-170.]
- Nardi F, Carapelli A, Fanciulli PP, Dallai R, Frati F. 2001. The complete mitochondrial DNA sequence of the basal hexapod *Tetrodontophora bielanensis*: Evidence for heteroplasmy and tRNA translocations[J]. *Mol Biol Evol*, **18**(7): 1293-1304.
- Ojala D, Montoya J, Attardi G. 1981. tRNA punctuation model of RNA processing in human mitochondria[J]. *Nature*, **290**(5806): 470-474.
- Ray DA, Densmore L. 2002. The crocodilian mitochondrial control region: general structure, conserved sequences, and evolutionary implications[J]. *J Exp Zool*, **294**(4): 334-345.
- Ray DA, Densmore LD. 2003. Repetitive sequences in the crocodilian mitochondrial control region: poly-A sequences and heteroplasmic Rychlik tandem repeats[J]. *Mol Biol Evol*, **20**(6): 1006-1013.
- Rychlik W. 2007. OLIGO 7 primer analysis software[J]. *Methods Mol Biol*, **402**: 35-60.
- Saito S, Tamura K, Aotsuka T. 2005. Replication origin of mitochondrial DNA in insects[J]. *Genetics*, **171**(4): 1695-1705.
- Simon C, Frati F, Bekenbach A, Crespi B, Liu H, Flook P. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a complication of conserved polymerase chain reaction primers[J]. *Ann Entomol Soc Am*, **87**: 651-701.
- Takashi Y, Yohichi W, Kimitsuna W. 1991. A novel clover leaf structure found in mammalian mitochondrial tRNASer[J]. *Nucl Acids Res*, **19**: 6101-6105.
- Tapper DP, Clayton DA. 1981. Mechanism of replication of human mitochondrial DNA. Localization of the 5' ends of nascent daughter strands[J]. *J Biol Chem*, **256**(10): 5109-5115.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools[J]. *Nucleic Acids Res*, **25**(24): 4876-82.
- Tsujino FK, Osemura A, Inohira K, Hara TO, tsuka1 YF, Obara MK, Matsuura ET. 2002. Evolution of the A+T-rich region of mitochondrial DNA in the melanogaster species subgroup of *Drosophila*[J]. *J Mol Evol*, **55**(5): 573-83.
- Watanabe Y, Watanabe K. 1994. Higher order structure of bovine mitochondrial tRNASerUGA chemical modification and computer modeling[J]. *Nucl Acids Res*, **22**: 5378-5384.
- Wilson K, Cahill V, Ballment E, Benzie J. 2000. The complete sequence of the mitochondrial genome of the crustacean *Penaeus monodon*: are malacostracan crustaceans more closely related to insects than to branchiopods? [J]. *Mol Biol Evol*, **17**(6): 863-74.
- Wolstenholme DR. 1992. Animal mitochondrial DNA: structure and evolution[J]. *Int Rev Cytol* **141**: 173-216.
- Yokobori S, Paabo S. 1995. Transfer RNA editing in land snail mitochondria[J]. *Proc Natl Acad Sci U S A*, **92**(22): 10432-5.
- Zhang CY, Huang Y. 2008. Complete mitochondrial genome of *Oxya chinensis* (Orthoptera, Acridoidea)[J]. *Acta Biochim Biophys Sin*, **40**(1): 7-18.
- Zheng LY, Gui H. 1999. Insect Classification[M]. Nanjing: Nanjing Normal University Press, 256-257.[郑乐怡, 归鸿. 1999. 昆虫分类. 南京: 南京师范大学出版社, 256-257.]
- Zhong J, Li G, Liu ZQ, Li QW, Wang YQ. 2005. Gene rearrangement of mitochondrial genome in the vertebrate[J]. *Acta Genetica Sinica*, **32**(3): 322-330.[钟婧, 李光, 刘忠权, 李庆伟, 王义权. 2005. 脊椎动物线粒体 DNA 的基因重排. 遗传学报, **32** (3): 322-330.]
- Zhou Z, Huang Y, Shi FM. 2007. The mitochondrial genome of *Ruspolia dubia* Orthoptera: Conocephalidae contains a short A+T-rich region of 70 bp in length[J]. *Genome*, **50**(9): 855-66.