

家蚕地方品种线粒体基因组 A+T 丰富区的序列及分子进化分析

陈丽媛^{1, 2} 赵巧玲² 沈兴家² 张志芳² 唐顺明² 徐安英² 张国政² 郭锡杰^{1, 2}

(¹江苏科技大学, 镇江 212003; ²中国农业科学院蚕业研究所, 农业部家蚕生物技术重点开放实验室, 镇江 212018)

摘要 为了研究家蚕线粒体基因组 A+T 丰富区的结构和家蚕品种的进化, 用 PCR 方法扩增了 12 个家蚕 (*Bombyx mori*) 地方品种线粒体基因组 A+T 丰富区及其侧翼序列, 分离纯化后克隆到 pMD18 T 载体进行测序。序列分析表明, 克隆片段长度约 1.1 kb, 基因排列顺序与 C108 线粒体相同, 依次为 12S rRNA 基因 3' 端、A+T 丰富区、tRNA^{Met}、tRNA^{Ile}、tRNA^{Gln} 和 ND2 基因 5' 端, 在 tRNA^{Gln} 和 ND2 基因之间有 47 bp 的非编码区。以日本野桑蚕 (*Bombyx mandarina*) 为外群, 用 Phy lip 软件包构建了基于 12 个家蚕品种线粒体基因组 A+T 丰富区序列的 NJ 进化树。结果显示, 甘肃种单独聚为一群, 其进化早于其它 11 个品种聚成的类群, 说明甘肃种是供试家蚕品种中进化最早的品种。这一结果在分子水平上为黄河流域是家蚕品种的发祥地之一提供了证据, 也进一步支持了家蚕品种的中国起源说。对 A+T 丰富区及其侧翼基因的结构分析表明, 家蚕线粒体 12S rRNA 和 ND2 基因都十分保守, 14 个品种统计只分别发生 1 个和 2 个碱基转换; A+T 丰富区中 (A+T) 比例高达 94.9% 以上, 第 27 nt 开始有 1 个 T 串结构, 长度为 16~19 bp 不等; 同时还根据 3 个 tRNA 基因的核苷酸序列推定了其二级结构。

关键词 家蚕; 线粒体 DNA; A+T 丰富区; 分子进化; 起源

中图分类号 S881.2 Q78 **文献标识码** A **文章编号** 0257-4799(2007)01-0005-09

Nucleotide Sequences in the A+T rich Regions of Mitochondrial DNA From Local Races of *Bombyx mori* and Their Molecular Evolution

CHEN LiYuan^{1, 2} ZHAO Qiao-Ling² SHEN Xing-Jia^{2*} ZHANG Zhi-Fang²

TANG Shun-Ming² XU An-Ying² ZHANG Guo-Zheng² GUO Xi-Jie^{1, 2}

(¹Jiangsu University of Science and Technology, Zhenjiang Jiangsu 212003 China; ²The Key Laboratory of Silkworm Biotechnology, Ministry of Agriculture, Sericultural Research Institute, Chinese Academy of Agricultural Sciences, Zhenjiang Jiangsu 212018 China)

Abstract To study the structure of A+T rich region of mitochondrial DNA (mtDNA) of silkworm, *Bombyx mori*, and evolution of silkworm race, the A+T rich regions of mtDNA from 12 local silkworm races were amplified by polymerase chain reaction (PCR) method and cloned into pMD18 T vector after purification for sequencing. Sequence analysis revealed that the cloned fragments were about 1.1 kb in length with the same gene order as C108 mtDNA: 3' end portion of 12S rRNA, A+T rich region, tRNA^{Met}, tRNA^{Ile}, tRNA^{Gln} and 5' end portion of ND2 successively. Between tRNA^{Gln} and ND2 there was a 47 bp non-coding region. Molecular

phylogenetic analysis was carried out and a NJ tree was constructed based on the nucleotide sequences of A+T rich regions of mtDNA from 12 local silkworm races with Phy lip software package. The results showed that Gansu race formed one group by itself and its evolution was earlier than the other group formed by the other races, strongly suggesting that Gansu race is the earliest evolved *Bombyx*.

收稿日期: 2006-11-03

资助项目: 国家重点基础研究发展计划“973”项目 (编号 2005CB121000); 江苏省自然科学基金前期预研重大项目 (编号 BK2004206)。

作者简介: 陈丽媛 (1983-), 女, 山东, 硕士研究生。

E-mail: liyuan3391@126.com

通讯作者: 沈兴家, 研究员, 博士生导师。

Tel: 0511-5616659 E-mail: shenxj@sricas.com

mori race in the study. And it did therefore not only provide an evidence in the molecular level for the theory that the Yellow River valley was at least one of the origin of *Bombyx mori* but also support the theory of China origination. The 12S rRNA and ND2 genes of *B. mori* mtDNA are highly conservative and only one and two base substitutions were occurred respectively in the 14 races examined. In the A+T rich region, the A+T content was extremely high to over 94.9% and there was a T stretch ranging from 16 to 19 bps starting at the 27 nt. The secondary structures of three rRNAs were deduced based on their nucleotide sequences as well.

Key words *Bombyx mori*; Mitochondrial DNA; A+T rich region; Molecular evolution; Origination

后生动物线粒体 DNA (mtDNA) 是一个闭环环状的双链分子, 分子量为 15 ~ 20 kb^[1]。mtDNA 具有分子量小、便于实验操作、变异速度快和很少重组 (母性遗传) 等特点, 被广泛用于种群遗传和物种分子进化分析^[2-9]。

动物 mtDNA 控制区 (D-环) 因富含腺嘌呤 (A) 和胸腺嘧啶 (T), 被称为 A+T 丰富区。A+T 丰富区是脊椎动物线粒体基因组序列和长度变异最大的区域, 它位于 rRNA^{Pro} 和 rRNA^{Phe} 之间, 包含 H 链复制起始区 (O_H)、保守序列节段 (CSB)、L 链启动子 (LSP)、H 链启动子 (HSP)、终止结合序列 (TAS)^[6,7]。昆虫线粒体 A+T 丰富区的位置与脊椎动物不同, 如果蝇 (*Drosophila melanogaster*; *D. yakuba*) 位于小 rRNA 和 rRNA^{Lys} 之间, 而且果蝇线粒体 A+T 丰富区中也存在保守序列^[8-13]。虽然昆虫中这些保守序列不同于脊椎动物, 但是它们对 mtDNA 的复制起着同样重要的作用。果蝇 (*D. melanogaster*; *D. immigrans*; *D. virilis*)、家蚕 (*Bombyx mori*) 等完全变态昆虫线粒体 A+T 丰富区控制 mtDNA 的复制, 它们的复制原点 (O_R) 的位置互不相同, 但是都在紧挨 A+T 丰富区 T 串 (thymine nucleotides stretch; T stretch) 下游的数个核苷酸内, T 串与昆虫 mtDNA 复制原点的识别有关^[14]。

研究家蚕线粒体基因组 A+T 丰富区的结构, 不仅有利于阐明其复制机制, 而且有助于阐明家蚕品种的起源和进化关系。利用 mtDNA 蛋白编码基因或 rRNA 基因等分析蚕类分子进化已经有许多报道^[15-18]。本试验对 12 个家蚕地方品种线粒体基因组 A+T 丰富区进行了克隆和测序, 并以日本野桑蚕 (*Bombyx mandarina*) 为外群线粒体, 对包含 C108 和青熟 (Aojuku) 在内的 14 个家蚕品种 mtDNA 的 A+T 丰富区序列的进化关系进行了分析。

1 材料与方法

1.1 材料及主要试剂

试验用的家蚕品种由中国农业科学院蚕业研究所保存。家蚕品种 C108、青熟和日本野桑蚕线粒体基因组相应序列引自 GenBank (表 1)。宿主菌 *E. coli* DH10B 由本实验室保存。克隆载体 pMD18-T, Taq DNA 聚合酶、T4 DNA 连接酶为 Takara 公司产品, 其它主要试剂为 GibcoBRL 公司产品。核苷酸引物合成、测序均委托上海生物工程有限公司完成。

1.2 mtDNA 抽提

以刚孵化的蚁蚕为材料, 参照文献 [19] 的方法并作改进提取 mtDNA, 溶于适量 pH 8.0 的 1×TE, -20 °C 保存备用。

1.3 PCR 引物设计及 A+T 丰富区序列的扩增

通过分析已发表的家蚕 mtDNA 控制区侧翼序列 (GenBank accession No.: AB070264 AB083339 AY048187), 设计 1 对 PCR 引物: P1, 5'-AAGGATCGTATAAACCGCAACTGCTGGCA-3'; P2, 5'-GATCTFAGATGAGTTTGAAATTAGGGGGAT-3'。引物 P1、P2 分别对应于 C108 mtDNA 第 9 672 ~ 9 692 nt 和互补链 10 741 ~ 10 761 nt。

分别以各家蚕品种 mtDNA 为模板, 用上述引物进行 A+T 丰富区及其侧翼序列的 PCR 扩增。反应体系 50 μL, 94 °C 变性 4 min 后, 按如下条件循环 35 次: 94 °C 30 s, 55 °C 1 min, 72 °C 1 min, 30 s, 最后 72 °C 延伸 7 min。

1.4 PCR 产物的纯化

PCR 产物进行凝胶电泳, 按照文献 [20] 的方法用 glassm 分离目的 DNA 片段, -20 °C 保存备用。

1.5 mtDNA 控制区及其侧翼片段的克隆

在 T4 DNA 连接酶的作用下, 将纯化后包含 A+T 丰富区及其侧翼序列的片段连接到 pMD18-T 载体上。连接产物转化感受态的 *E. coli* DH10B LB 培养基 37 °C 培养 10 ~ 12 h, 挑取数个单菌落分别接种到 3 mL 液体 LB 培养基中, 37 °C 培养过夜, 提取

重组质粒 DNA, 用 *HindIII* 酶切后凝胶电泳鉴定, 将正确的重组质粒从两个方向分别进行测序。

表 1 试验用家蚕品种的来源和系统

Table 1 The source and system of silkworm races *Bombyx mori* used in the experiment

家蚕品种名称 Names of silkworm races	编号 ¹⁾ Serial No	来源 原产地 Source or origin	系统及主要特性 System and basic characteristics	GenBank 登录号 Accession No
中 108 C 108	1	日本 Japan	中系, 二化性四眠 Chinese bivoltine	AB070264
三四 Sansi	2	中国(产地不详) China (unknown)	中系, 有滞育多化 Chinese polyvoltine	EF157997
大草 Daisou	3	日本 Japan	中系, 二化性四眠 Chinese bivoltine	EF157991
邯郸种 Handanzhong	4	中国河北 Hebei China	中系, 二化性四眠 Chinese bivoltine	EF157993
迈索尔 Mysore	5	印度 India	热带, 多化性四眠 Tropical polyvoltine	EF157995
大造 Dazao	6	中国广东、广西 Guangdong and Guangxi China	热带, 二化性四眠 Tropical bivoltine	EF157992
萨尼斯 Sanisi	7	俄罗斯 Russia	欧系, 一化性四眠 European monovoltine	EF157996
余杭二化 Yuhang bivoltine	8	中国浙江 Zhejiang China	中系, 二化性四眠 Chinese bivoltine	EF158000
沿河 1 号 Yanhe 1	9	中国贵州 Guizhou China	中系, 一化性三眠 Chinese monovoltine	EF157999
新昌 12 号 Xinchang 12	10	中国浙江 Zhejiang China	中系, 一化性四眠 Chinese monovoltine	EF157998
安康 4 号 Ankang 4	11	中国陕西 Shaanxi China	中系, 一化性三眠 Chinese monovoltine	EF392866
甘肃种 Gansu zhong	12	中国甘肃 Gansu China	中系, 一化性三眠 Chinese monovoltine	EF158001
兰 5 Lan 5	13	中国浙江 Zhejiang China	中系, 多化性四眠 Chinese polyvoltine	EF157994
青熟 Aojuku	14	日本 Japan	日系, 二化性四眠 Japanes bivoltine	AB083339
日本野桑蚕 Japanese <i>Bombyx mandarina</i>	15	日本 Japan	染色体 n=27 Chromosome n= 27	AB070263 / NC003395

¹⁾表 2 的编号相同。
¹⁾The serial numbers are the same in the Table 2.

1.6 序列分析

测序结果用 Blast 软件在 NCBI 等数据库进行搜索, 并与已发表的 C 108 等家蚕品种 mDNA 序列进行比较, 用 DNA star 5.0 软件包中的 EdiSeq 软件及目测进行序列拼接^[21]。进化分析以 GenBank 公

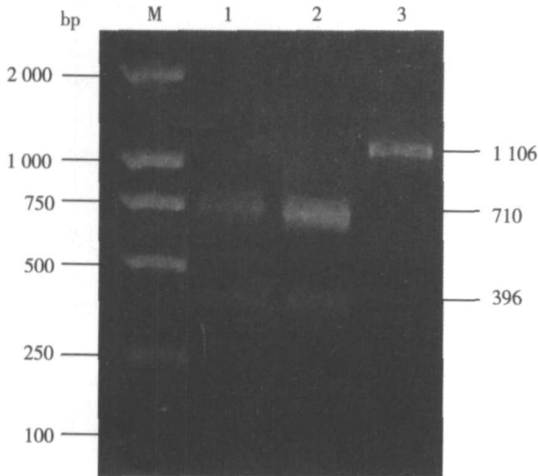
布的日本野桑蚕作外群, 先用 ClustaX 1.83 软件排列 14 个家蚕品种 mDNA 控制区序列^[22]; 再用 Phy lip3.66 软件包分析, 采用邻近法 (NJ) 构建进化树, 遗传距离以二元法 (Kimura 2-parameter) 计算, 进化树采用 Bootstrapping 法估算 (100 个重复)^[23];

最后由 TreeView 软件阅读^[24]。

2 结果与分析

2.1 家蚕 mDNA (A+T) 丰富区及其侧翼片段的扩增与纯化

PCR 产物的凝胶电泳结果显示, 所有扩增反应均获得 1 个特异性片段, 其大小约为 1.1 kb (图 1)。经凝胶电泳分离后, 按照上述方法进行纯化。



M. DL 2 000 marker 1 2 分别为安康 4 号和甘肃种 mDNA PCR 扩增片段的 *Hind* III 酶切片段
3 甘肃种 mDNA PCR 扩增片段

M. DL 2 000 marker 1 and 2 were *Hind* III digestion fragments of PCR products from mDNA of Ankan 4 and Gansu Zhong respectively

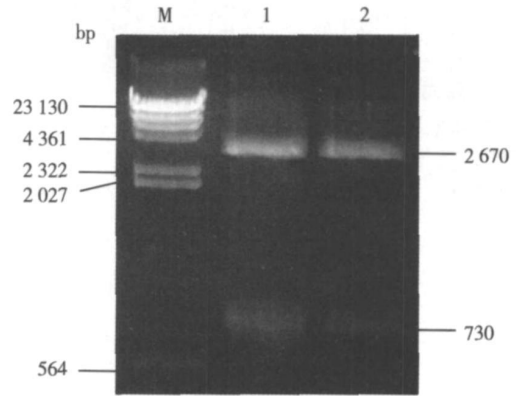
3. PCR fragment amplified from Gansu Zhong mDNA

1 家蚕 mDNA (A+T) 丰富区及其侧翼片段的 PCR 扩增

Fig 1 Amplification of A+T rich region and its flank from mDNA s of *Bombyx mori* by PCR

2.2 家蚕 mDNA (A+T) 丰富区及其侧翼片段的克隆

将纯化后的 PCR 产物直接克隆到 pMD18-T 载体上, 转化感受态 *E. coli* DH10B, 每个家蚕品种获得数个重组质粒。用 *Hind* III 酶切重组质粒后进行凝胶电泳鉴定。由于在载体和扩增片段 702~707 bp 之间各有 1 个 *Hind* III 酶切位点, 当用 *Hind* III 酶切重组质粒时, 正向连接获得约 0.73 kb 和 2.67 kb 2 个片段 (图 2), 反向连接则获得 0.42 kb 和 3.38 kb 2 个片段 (图略)。每个供试品种选择 1 个重组质粒测序。



M. λ DNA *Hind* III marker 1 2 分别为含安康 4 号和甘肃种 mDNA 片段的重组质粒 *Hind* III 酶切片段

M. λ DNA *Hind* III marker 1 and 2 were *Hind* III digestion fragments of the recombinated plasmids containing mDNA fragment amplified from mDNA of Ankan 4 and Gansu Zhong respectively

图 2 家蚕 mDNA (A+T) 丰富区及其侧翼片段克隆的酶切鉴定

Fig 2 Identification of recombinated plasmids containing the A+T rich region and its flank of mDNA from *Bombyx mori* by *Hind* III digestion

2.3 家蚕 mDNA 控制区及其侧翼片段的序列分析

序列测定结果, 克隆的家蚕 mDNA 片段长度为 1 088~1 091 bp 经 Blast 搜索以及与 C108 mDNA 序列比对, 这些片段包含的基因依次为 12S rRNA 基因 3' 端部分、A+T 丰富区 (D-loop)、rRNA^{Met}、rRNA^{Lk}、rRNA^{Gln} 和 ND2 基因 5' 端部分, 其中 rRNA^{Gln} 和 ND2 基因之间有一段 47 bp 的非编码区。基因排列顺序与 C108 mDNA 相同。图 3 为家蚕品种安康 4 号克隆片段序列, 1~189 bp 为 12S rRNA 基因, 190~683 bp 为 A+T 丰富区, 684~751 bp 为 rRNA^{Met} (AUG) 基因, 750~815 bp 为 rRNA^{Lk} (AUC) 基因, 813~881 bp 互补链为 rRNA^{Gln} (CAA) 基因, 882~928 为非编码序列, 929~1 090 bp 为 ND2 基因 5' 端序列。

用 MegAlign 和 ClustalX 1.83 软件对 12 个家蚕品种的克隆片段序列及 GenBank 公布的 C108 和 Aojiku 相应序列分析, 家蚕品种间核苷酸序列的同源性非常高, 均在 99% 以上, 其中: C108 兰 5 和青熟三者完全一致; 沿河 1 号和大造与 C108 相比, 只在控制区的 T 串和 A 串 (Adenine 串) 长度上有 1~

2个碱基的差异。在 14个品种中,共检测到 19个碱基替换(15个转换,4个颠换)和 1个碱基缺失,碱基替换发生比例为 0.124%,主要分布在控制区部分(表 2)。

GTATAACCGCAACTGCTGGCACAAAATTTGTTATTAATTTAAATATTACTAAATCCTAATTACTTAAATT 70
TTTAATTCCTAATTACTACAAATAAATATTAATTATTATTTAAATAAATTAATAAATAACACTAAAATTTA 140
TATGTAAAATAAATTTATAATAAATTTTAAACCATAAATAATTTTATTATTTAATGTAATTTTTTTTAC 210
ATAGATTTTTTTTTTTTTTTTTTATATTAATTTATTATTAATTTATTATTATTATTAATTTGAATATTTG 280
ATTTAATATTTTTTTTATTAATAAATAAATCAATAAATGATTAATTAATAATAAATAAATTAATATTTAATGATT 350
ATTTAATATTTAAATTTAAATATTAATTGATTAATTAATATAAATTTATAAATTTTTAATATTTCTCTTA 420
TTTTTTTCTTATAATATTAAGTTTAAATATAAATCAATATTCAACCTATAATATTCATTAAAAATAAAA 490
AAAAATTAATATAATTAATTAATTTTTTAATAATTTATTATATATATATATATTAATTATATAAAT 560
AATTTATTATATATAAATTTATATTAATAAATTAATAAATTTAATATATATATATATATAAGTATTATT 630
TATTTAAATTAATAACAAACCATTGTTAATTTTTTTTCATTAAAAAAAAAATTAAAAATAAGCTAATA 700
AAAGCTTTTGGGCTCATACCTCAAATATAAAGGAAATACCTTTTTTTTAAATAAAGTGCCTGAATAAAG 770
ATTATCTGATAGAATAAATTAAGTAGAAATATCTACCTTTATTATATTTTATAGAATTAAACTATATCT 840
AATAATATCAAAAATTATTGTGCATTTTACACTAAAATATAATTTAAATAATTAAAAATAAGAATTATATA 910
ATTTCTATTAAATATATTTATATTTTTTTTTTATTTTAACAAATAATAATTCAAATAAATATTTTTTTTAT 980
TTATTTTATTTTTAGAACATTAATTTCTATTTCCTTCAAACTCATGATTGGATGTTGAATTGGATTAGA 1 050
AATTAAC TTATTAAGATTTATCCCCCTAATTTCAAAC TCA 1 090

下划线表示 T串, 阴影表示 A串, 方框为 TATA 框。
The underlined part and shaded part indicates a t stretch and a stretch of the d loop respectively.
The boxed part indicates a putative tata box of ND2 gene promoter

图 3 家蚕品种安康 4号线粒体基因组 A+T丰富区核苷酸序列
Fig 3 Nucleotide sequence of A+T rich region of m itochondrial DNA from Ankang 4 silkworm, *Bombyx mori*

表 2 家蚕线粒体基因组 A+T丰富区及其侧翼序列碱基替换与分布
Table 2 Base pair substitutions and their distribution in the A+T rich region of m itochondrial DNAs from silkworms *Bombyx mori*

编号 Number	碱基替换的位置 / nt Site of base pair substitution																		
	78	233	236	273	281	286	330	347	369	379	437	576	630	676	719	735	870	988	1 023
1	C	T	T	A	A	A	A	T	A	G	T	A	T	A	T	G	T	C	C
2													C				...		
3			A												C				
4						G	G							G					
5											C								
6																			
7									G										
8		C																	
9																			
10								G		T									
11				G	G													T	
12	T							G				G			A				T

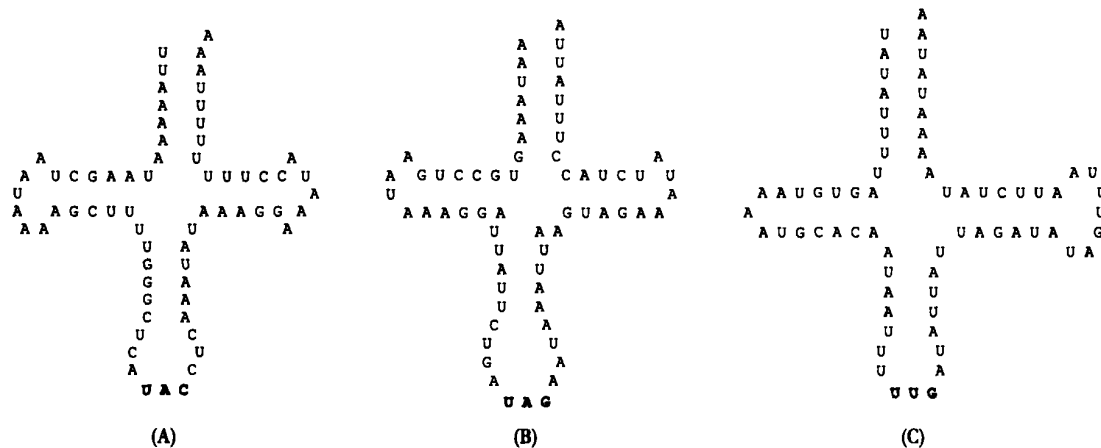
“...”表示碱基缺失 兰 5和青熟 与 C108一致。
“...” indicates the absence of nucleotide. Nucleotide sequences of Lan 5 and Aojuku are identical with that of C108.
©1994-2014 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

2.3.1 12S rRNA 基因的位置 无脊椎动物线粒体基因组 12S rRNA 和 16S rRNA 基因的位置不同于脊椎动物线粒体基因组。家蚕 mDNA 的 12S rRNA 和 16S rRNA 基因的位置与果蝇线粒体基因组相同, 位于 $rRNA^{Leu}$ 与 (A+T) 丰富区之间, 中间有 $rRNA^{Val}$ 相隔。本试验克隆片段中包含 12S rRNA 基因的 3' 端部分 189 个核苷酸, 在 14 个家蚕品种 (含 C108 和 Aojiku) 中有 13 个品种完全一致, 只有甘肃种在第 78 nt 处发生了 G-T 转换, 说明 12S rRNA 基因在家蚕品种间是非常保守的。同时, 12S rRNA 基因 3' 端序列中 (A+T) 的含量高达 87.3%, (G+C) 只占 12.70%, 与脊椎动物线粒体 rRNA 的特性相似。

2.3.2 A+T 丰富区的结构 动物 mDNA 控制区因富含 A 和 T 被称为 A+T 丰富区, 它是线粒体基因组序列和长度变异最大的区域。但是在 14 个家蚕品种间 A+T 丰富区仍然相当保守, 其核苷酸序列的同源性均在 98.2% 以上, 其中 C108、兰 5 和青

熟 3 个品种完全相同。该区段 (A+T) 比例高达 94.9% 以上, 而 (G+C) 含量不足 5.1%。供试的 14 个品种在该区段第 27 nt 开始均有 1 个 T 串结构, 长度为 16~19 bp 不等; 末端有 1 个 11~12 bp 的 A 串结构。分析表明该区段序列共发生 13 个碱基替换, 其中 10 个转换、3 个颠换, 品种间的分布 0~4 个不等, 平均每个品种 1 个左右。家蚕品种间 A+T 区序列的差异主要来自 T 串和 A 串长度的差异。

2.3.3 rRNA 基因及其二级结构 克隆片段中, 紧挨 A+T 丰富区 3' 端的是 3 个 rRNA 基因: $rRNA^{Met}$ (AUG)、 $rRNA^{Ile}$ (AUC) 和 $rRNA^{Gln}$ (CAA) (互补链), 基因之间有 2~3 bp 的重叠。线粒体 rRNA 也能形成典型的三叶草形结构。根据安康 4 号线粒体 rRNA 基因的核苷酸序列, 推测了其编码的 rRNA 二级结构, 如图 4 所示的 3 个 rRNA 的氨基酸接受臂碱基配对正确, 而其它各臂碱基对存在错配现象。



粗体字母表示反密码子

The block letters indicate the anticodons

图 4 推定的家蚕品种安康 4 号 3 个 rRNA 的二级结构

Fig 4 The putative secondary structures of three rRNA s from mDNA of *Bombyx mori* Ankang 4

2.3.4 ND2 基因的保守性 线粒体中 ND2 基因是高度保守的基因^[18], 对 14 个家蚕品种线粒体基因组 ND2 基因 5' 端序列的分析结果, 12 个品种完全一致, 只有安康 4 号和甘肃种分别在基因的 58 nt 和 93 nt 处发生了 1 个 C-T 转换。在 ND2 基因 5' 侧翼有一段 47 bp 的非编码序列甚至比 ND2 基因更保守, 不仅 14 个家蚕品种完全相同, 而且与日本野桑

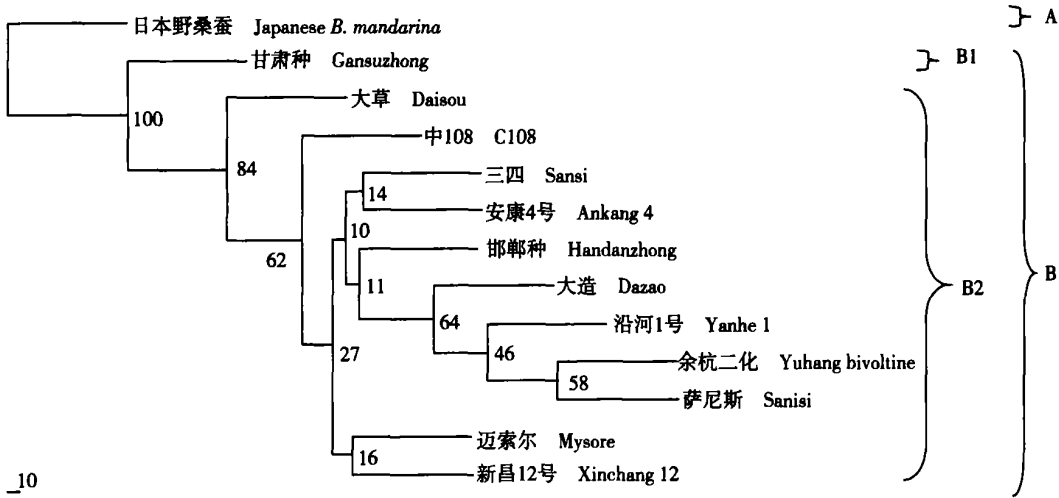
蚕也相同。在 47 bp 序列中包含 1 个原核启动子的特征基序 TATAAT, 位于 ND2 基因编码区上游 -22~-18 nt 处, 推测该段序列可能是 ND2 基因的启动子。

2.4 家蚕 A+T 丰富区序列的进化分析

用 Phy lip3.66 软件邻近法 (NJ) 构建进化树, 以二元法计算遗传距离, 采用 Bootstrapping 法对进化树

(100 个重复) 进行评估 (图 5)。在进化树中, 日本野桑蚕单独聚为一群 (A), 14 个家蚕品种聚为一群 (B)。在 B 群中, 甘肃种单独聚为一个亚群 (B1), 其分化早于其它 13 个品种组成的亚群 (B2), 而且,

亚群分支节点上的支持率高达 100%, 说明这种进化的可能性较大。用 Phy lip3. 66 软件最大似然法 (M axim um L kelihood) 构建进化树在类群和亚群的聚类上有相似的结果 (数据略)。



兰 5 和青熟 A+T 区序列与 C 108 一致。节点数据为 Bootstraping 法对 100 棵树评估的支持率。

e A+T rich regions of mDNAs from Lan 5 and Aojuku strains are identical with that of C108. The statistical support for each node on the tree was evaluated by using bootstrapping method (100 replications) with Phy lip3. 66 software package

图 5 用邻近法构建的基于 14 个家蚕品种线粒体基因组 A+T 丰富区序列的进化树

Fig 5 NJ tree based on sequences of A+T rich regions of m itochondrial DNAs from 14 *Bombyx mori* strains

3 讨论

3.1 进化分析序列的选择

动物线粒体基因组共编码 13 个蛋白质 (或亚基)、22 个 rRNA、蛋白质基因中细胞色素 b 基因 (Cytb)、细胞色素 C 氧化酶 3 个亚基的基因 (COX1、COX2、COX3) 最保守, 同源性最高, 不同物种间可达 80% 左右; 其次保守的是 rRNA 基因; 而 ATP 合酶 2 个亚基的基因 (ATP 8、ATP 6)、NADH 氧化还原酶 7 个亚基的基因 (ND 1、ND 2、ND 3、ND 4、ND 4L、ND 5、ND 6) 的变异较大^[25]。因此, Cytb、COX3 和 rRNA 等基因的核苷酸或氨基酸序列常被用于物种进化分析, 而这些保守码基因在同一物种的不同品系间更加保守。本试验中 14 个家蚕品种线粒体 12S rRNA 基因 3' 端 189 bp 和 ND 2 基因 5' 端 162 bp 序列比较结果, 有 12 个品种的序列同源性为 100%, 只有甘肃种和安康 4 号在 2 个基因中各有 1 个碱基的替换, 而且 ND 2 基因中的碱基替换为同义突变, 并不影响氨基酸序列, 说

明这 2 个基因的序列不适用于家蚕种内地方品种 (品系) 间的进化关系分析。鉴此, 本试验选择了碱基替换率相对较高的线粒体 A+T 丰富区序列作为进化分析的基础。

3.2 分析方法和外群的选择

用于分子进化分析的软件很多, 每种软件都有其特点和不足。本试验在 ClustaK 进行序列排列后选用 Phy lip 软件包进行分析。进化树的构建方法主要分为两类: 一是独立元素法 (Discrete character methods), 包括最大简约性法 (Maximum Parsimony methods) 和最大似然法 (Maximum Likelihood methods); 二是距离依靠法 (Distance methods), 包括除权配对法 (UPGMAM) 和邻位相连法 (NJ)。NJ 法是一种常用的算法, 它构建的进化树相对准确且计算快捷, 缺点是序列上的所有位点都被同等对待, 而且分析的序列进化距离不能太大^[23]。因此, 外群的选

择对构建的进化树的可靠性有很大的影响。

已有的研究表明,在进化距离上家蚕与野桑蚕比家蚕与其它昆虫近^[16-26],而家蚕与中国野桑蚕的进化距离最近,与日本野桑蚕的进化距离相对较远^[27]。由于家蚕与中国野桑蚕的进化距离过近,部分基因序列甚至比家蚕品种间还要近,不适合做家蚕地方品种进化分析的外群,所以本试验选用日本野桑蚕为外群。

3.3 家蚕线粒体基因组 A+T 丰富区的结构及其进化关系

家蚕虽然线粒体基因组 A+T 区中, A+T 的含量高达 94.9 以上,而 G+C 的含量不足 5.1%。虽然 A+T 区是物种间线粒体基因组序列和长度变异最大的区域,但是它在家蚕种内仍然相当保守,14 个品种共检测到 13 个碱基替换(10 个转换 3 个颠换),核苷酸序列的同源性均在 98.2% 以上,有的品种甚至完全相同。家蚕线粒体基因组 A+T 区有 1 个 16~19 bp 的 T 串结构,与 mtDNA 复制原点的识别有关,而复制原点位于 T 串结构互补链下游的 7 个核苷酸内^[14]。

“家蚕起源于中国”的学说已经成为学界的共识^[28-29],最近几年的分子生物学研究从分子水平上为“家蚕起源于中国古野桑蚕”提供了有力的证据^[17-18-27]。然而,对于家蚕起源的地域性和化性分化却一直存在着争论。日本学者吉武根据对家蚕不同地理品系的生化分析和 Vavilov 的起源中心说,提出了“中国一化性品种为起源”的学说;而我国学者蒋猷龙根据我国气候的演变和历史文物的考证,提出了“中国多化性多中心起源”学说,即家蚕由中国黄河、长江和珠江流域等地的多化性及热带多化性品种进化而来。目前学术界对于家蚕起源比较一致的观点是“从多化(性)到一化(性)”。华北野桑蚕为三化性的事实是对多化性起源的一个有力的证据^[29]。但是,对于家蚕品种的地域起源问题,学术界仍然存在家蚕起源黄河流域的“一中心”学说和多地域起源的“多中心”学说。

甘肃种为我国西部地区黄河上游甘肃省的地方品种,蚕茧为天然橘红色,独特的色彩使其易于其它品种相区别,继代饲养中不容易被混杂,品种相对较纯。本试验利用 Phyip 软件,以家蚕线粒体基因组中变异最大的 A+T 丰富区核苷酸序列为基础,采用 NJ 法构建了家蚕地方品种的进化树。分析结果

显示,在包含中国地方品种、日本品种、欧洲品种和热带多化性品种的全部试验品种中,甘肃种的进化早于试验的其它 13 个品种。说明甘肃种是本试验研究的家蚕品种中进化最早的品种。这一结果在分子水平上为黄河流域是家蚕品种的发祥地之一提供了证据,也进一步支持了家蚕品种的中国起源说。

虽然本试验中也使用了贵州的地方品种沿河 1 号,但并未如以往的研究^[17]那样成为独立的组群。然而两者与“多中心”起源学说并不矛盾。同一个家蚕品种在不同的分析中之所以出现不同的结果,可能有多方面的原因:一是分析所用的核苷酸序列不同;二是外群(参照)选择上的差异;三是还可能与品种的复杂有关。家蚕的驯化至少已经有 5000 年以上的历史,由于数千年中地球气候的变化和家蚕在长期的人工饲养、选择以及传播(如古丝绸之路)过程中,品种变化非常复杂,从而给家蚕品种间进化分析带来一定的难度。正因为如此,就需要开展更多和更深入的研究,通过大量的分析比较,才能获得比较真实地重现家蚕进化历史的分子树,阐明家蚕的地域起源和品种进化。

参考文献 (References)

- [1] Wolstenholme D R. Animal mitochondrial DNA: structure and evolution[J]. *Int Rev Cytol* 1992; 141: 173-216
- [2] Wolstenholme D R, Clary D O. Sequence evolution of drosophila mitochondrial DNA[J]. *Genetics* 1985; 109: 725-744
- [3] Sankoff D, Leduc G, Antoine N *et al*. Gene order comparisons for phylogenetic inference: evolution of the mitochondrial genome[J]. *Proc Natl Acad Sci USA* 1992; 89: 6575-6579
- [4] Kim I, Bae J S, Sohn H D *et al*. Genetic homogeneity in the domestic silkworm *Bombyx mori* and phylogenetic relationship between *B. mori* and the wild silkworm *B. mandarina* using mitochondrial CO I gene sequences[J]. *Int J Indust Entomol* 2000; 1: 9-17
- [5] Yagi T, Katoh T, Chichvarkhin A *et al*. Molecular phylogeny of butterflies *Pamassius glacialis* and *P. stubbendorffii* at various localities in East Asia[J]. *Genes Genet Syst* 2001; 76: 229-234
- [6] Saccone G, Attinonelli M, Sbisà E. Structural elements highly preserved during the evolution of the D-loop containing region in vertebrate mitochondrial DNA[J]. *J Mol Evol* 1987; 26(3): 205-211
- [7] Gray M W, Burger G, Lang B F. Mitochondrial evolution[J]. *Science* 1999; 283: 1467-1481
- [8] Clary D O, Wolstenholme D R. Drosophila mitochondrial DNA: conserved sequences in the A+T rich region and supporting evidence for a secondary structure model of the small ribosomal RNA

- [J]. *J Mol Evol* 1987 25(2): 116-125
- [9] Monforte A, Barrio E, Labarre A. Characterization of the length polymorphism in the A+T rich region of the *Drosophila obscura* group species[J]. *J Mol Evol* 1993 36(3): 214-223
- [10] Lewis D L, Farr C L, Farquhar A L, *et al*. Sequence organization and evolution of the A+T region of *Drosophila melanogaster* mitochondrial DNA[J]. *Mol Biol Evol* 1994 11(3): 523-538
- [11] Inohira K, Hara T, Matsuura E T. Nucleotide sequence divergence in the A+T-rich region of mitochondrial DNA in *Drosophila simulans* and *Drosophila auritiana*[J]. *Mol Biol Evol* 1997 14(8): 814-822
- [12] Brehm A, Harris D J, Hernandez M, *et al*. Structure and evolution of the mitochondrial DNA complete control region in the *Drosophila subobscura* subgroup[J]. *Insect Mol Biol* 2001 10(6): 573-578
- [13] Tsujino E, Kosmura A, Inohira K, *et al*. Evolution of the A+T rich region of mitochondrial DNA in the melanogaster species subgroup of *Drosophila*[J]. *J Mol Evol* 2002 55(5): 573-583
- [14] Saito S, Tanura K, Aotsuka T. Replication origin of mitochondrial DNA in insects[J]. *Genetics* 2005 171(4): 1695-1705
- [15] 魏兆军, 赵巧玲, 张志芳, 等. 蓖麻蚕线粒体基因组细胞色素氧化酶亚基III的序列及其分子进化分析[J]. *昆虫学报*, 2002 45(2): 193-197
- [16] Xing Jia Shen, Qiao Ling Zhao, Zhi Fang Zhang, *et al*. Cytochrome subunit b and NADH oxidoreductase subunit I genes of mitochondrial genome from the eri silkworm, *Samia cynthia ricini* and phylogenetic analysis[J]. *International Journal of Wild Silk worm and Silk* 2003 8 51-56
- [17] Ailing Li, Qiaoling Zhao, Zhifang Zhang, *et al*. Molecular phylogeny of the domesticated silkworm, *Bombyx mori* based on the sequences of mitochondrial cytochrome b genes[J]. *Journal of Genetics* 2005 84(2): 101-106
- [18] Anunkumar K, R Metta M, Nagaraju J. Molecular phylogeny of silkmoths reveals the origin of domesticated silkmoth, *Bombyx mori* from Chinese *Bombyx mandarina* and paternal inheritance of *Antheraea pernyi* mitochondrial DNA[J]. *Mol Phylogenet Evol* 2006 40(2): 419-27
- [19] Tanura K, Aotsuka T. Rapid isolation of animal mitochondrial DNA by the alkaline lysis procedure[J]. *Biochemical genetics* 1988 26(11/12): 815-819
- [20] 沈兴家, 赵巧玲, 张志芳, 等. 家蚕蛹体线粒体基因组 *EωRI* 2.0 kb 片段的克隆及序列分析[J]. *蚕业科学*, 2000 26(4): 234-238
- [21] Burland T G. DNASTAR's lasergene sequence analysis software[J]. *Methods Mol Biol* 2003 12 71-91
- [22] Jeanmougin F, Thompson J D, Gouy M, *et al*. Multiple sequence alignment with Clustal X[J]. *Trends Biochem Sci* 1998 23 403-405
- [23] Felsenstein J. PHYLIP: Phylogenetic inference package, version 3.6 [M]. [2007-08-07], <http://evolution.genetics.washington.edu/phylip.html>
- [24] Page R D M. TREEVIEW, version 1.6.0 [M]. [2007-08-07], <http://taxonomy.zoology.gla.ac.uk/rod/treeview.html>
- [25] 张方, 米志勇. 动物线粒体 DNA 的分子生物学研究进展[J]. *生物工程进展*, 1998 18(3): 25-31
- [26] 沈兴家, 赵巧玲, 张志芳, 等. 蓖麻蚕线粒体基因组中 *nd1* 及其侧翼 rRNA 基因的克隆与结构分析[J]. *蚕业科学*, 2002 28(4): 289-293
- [27] 李爱玲, 徐安英, 沈兴家, 等. 家蚕、野桑蚕线粒体 *Cyt b* 基因片段序列分析及分子进化研究[J]. *蚕业科学*, 2004 30(1): 80-84
- [28] 中国农业科学院蚕业研究所. 家蚕遗传育种学[M]. 北京: 科学出版社, 1981 237-249
- [29] 中国农业科学院蚕业研究所. 中国养蚕学[M]. 上海: 上海科学技术出版社, 1991 80-87