

山羊线粒体 DNA D-loop 区部分序列变异位点分析

武艳平^{1,2}, 关伟军¹, 赵倩君¹, 何晓红¹, 浦亚斌¹, 霍俊宏², 敖红¹, 李奎¹, 马月辉^{1*}

(1. 中国农业科学院北京畜牧兽医研究所, 北京 100193;

2. 江西农业科学院畜牧兽医研究所, 南昌 330200)

摘要: 截取了山羊线粒体 DNA D-loop 高变区 453 bp 的部分片段(线粒体全序列的 15 735~16 187 bp 处), 共计 1 182 条序列进行了多重比对, 并进行系统建树和网络关系分析, 结果这些序列明显的分成了 4 个支系(A、B、C、D), 分别含有 1 001、134、24 和 13 条序列, 进一步比对分析, 发现这 4 个支系有各自特有的碱基变异特点, A、B、C 和 D 分别有 5、6、18 和 7 个特有的碱基变异位点, 除了各个支系特有的变异外, 还发现有 2 个支系共有的变异位点。

关键词: 山羊; 线粒体 DNA; D-loop; 变异

中图分类号: S826.2

文献标识码: A

文章编号: 0366-6964(2008)08-1137-05

Analysis on Partial Sequence of Mitochondrial DNA D-loop Region Variation Position of Goat

WU Yan-ping^{1,2}, GUAN Wei-jun¹, ZHAO Qian-jun¹, HE Xiao-hong¹, PU Ya-bin¹,
HUO Jun-hong², AO Hong¹, LI Kui¹, MA Yue-hui^{1*}

(1. Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing 100193, China; 2. Institute of Animal Science, Jiangxi Academy of Agricultural Sciences, Nanchang 330200, China)

Abstract: In this study, we aligned and analyzed 453 bp segment of mitochondrial DNA D-loop hypervariable region (Positions 15 735-16 187 bp of complete mitochondrial genome) from 1 182 sequences, neighbor-joining (NJ) tree and network showed that the sequences were divided into four distinct mtDNA lineages A-D, and included 1 001, 134, 24 and 13 sequences, respectively. After the alignment of the sequences, nucleotide position of A, B, C and D lineage specific were acquired, and included 5, 6, 18 and 7 variation positions, respectively, moreover, 2 lineages common variation positions were found.

Key words: goat; mitochondrial DNA; D-loop; variation

我国有丰富的品种资源^[1], 然而, 随着外来品种的不断引进以及杂交改良、人工授精和胚胎移植等工作的深入展开, 我国古老的地方品种受到严酷的竞争和排挤, 以至造成地方品种的群体数量急剧下降, 有的甚至处于濒临灭绝的境地^[2], 遗传多样性降低。线粒体 DNA (Mitochondrial DNA, mtDNA) 是唯一的核外遗传物质, 由于线粒体的高拷贝数、缺

乏重组以及单亲及母系遗传等特点而使其成为遗传结构和系统发育研究的良好素材, 线粒体 DNA 的遗传分析成为理解进化的重要工具, 山羊线粒体 DNA 是存在于胞质细胞器线粒体中的一个闭合环状、有 16 640 个碱基的分子^[3], 在每个细胞中存在成百上千的拷贝。mtDNA 基因排列紧凑, 几乎没有间隔序列和内含子, 编码 13 种线粒体氧化磷酸化

收稿日期: 2007-08-31

基金项目: 科技支撑计划(2006BAD13B08); 国家科技基础条件平台项目(2005DKA21101); 中国农业科学院北京畜牧兽医研究所“家畜种质资源研究与创新”项目(ywf-td-1); 绵山羊核心种质构建“重要基因发掘研究”项目(ywf-qn-8)资助

作者简介: 武艳平(1976-), 女, 山西武乡人, 博士, 主要从事畜禽资源分子评价的研究, E-mail: wuyanping0902@tom.com

* 通讯作者: 马月辉, 研究员, E-mail: yuehui.ma@263.net

呼吸链组成酶所必需的多肽、12SrRNA、16SrRNA 和 22 个 tRNA。而其中长 1 212 个碱基的 mtDNA D-环控制区(D-Loop)是 mtDNA 基因组中进化速率最高、最具多态的区域,尤其在 高变区,由于 D 环受自然选择影响较大,存在巨大的变异,所以一般用于种群间的系统进化分析。目前,我国利用线粒体研究家养动物的起源进化及遗传变异已取得一些进展^[4-8]。根据线粒体 DNA 研究,目前家养山羊有 5 (A、B、C、D 和 E) 个支系(lineages)。Luikart 等^[9]通过对 88 个品种的线粒体 D 环高变区测序分析,得出山羊主要有 3 个母系起源,其中 A 最普遍,存在于所有的品种中,C 最少,只存在于蒙古、瑞士和斯洛文尼亚。后来,Sultana 等^[10]分析了 13 个巴基斯坦本地山羊品种的线粒体 D 环全序列和细胞色素 b 基因序列,除了发现上面 3 个支系外,还发现了一个新的类群(D 支系),Chen 等对中国 18 个地方品种的线粒体 D 环高变区分析,也发现了这个稀有支系,E 支系仅在印度山羊中发现^[11-12]。本研究通过分析来自不同国家不同品种的 1 182 条序列,来揭示 A、B、C 和 D 这 4 个支系的特有的碱基变异,为山羊系统发育及遗传变异的深入研究提供基础和便利。

1 材料与方法

1.1 序列的获得

根据关于 mtDNA D-环控制区的高变区的报道,利用现有的网络资源,从 GenBank 数据库获得所需序列,其中有 Chen 等^[11]报道的 18 个中国山羊品种的序列 345 条(DQ089106-DQ089450),Joshi 等^[12]报道的印度山羊 323 条 (AY155674-AY155695, AY155697-AY155791, AY155834-AY156039),Sultana 等^[10]报道的巴基斯坦山羊 38 条 (AB110552-AB110589),Luikart 等^[9]报道的包括覆盖欧洲、亚洲、非洲和中东的 44 个国家的 331 条序列(AJ317533-AJ317863);另外还有笔者前期已经测序完成的 12 个中国山羊品种的 145 条序列(EU035991-EU036135),共计 1 182 条序列。

1.2 序列的截取及统计分析

选取变异较大的 453 bp 的高变区部分片段(线粒体全序列的 15 735~16 187 bp 处)。用 BioEdit7.0 软件^[13]对所有序列进行多重比对,并进行人工校对,以确保比对的准确性,MEGA3.0 软件^[14]用于系统进化分析及进化树的构建。用 Network 4.2^[15]进行各支系网络关系分析。

2 结果

2.1 系统进化树及网络关系分析

我们对 1 182 条序列在 BioEdit7.0 软件进行比对后,用 MEGA 3.0 以 Kimura 2-parameter 模型,建立 NJ 树,并对拓扑图进行了自展检验(bootstrap),重复抽样次数为 1 000 次。所建的系统进化树见图 1,这 10 个山羊品种很明显的分成了 4 个支系 A、B、C 和 D。用 Network 4.2 进行网络关系分析也证实了这一点。其中 A、B、C 和 D 支系分别含有 1 001、134、24 和 13 条序列。

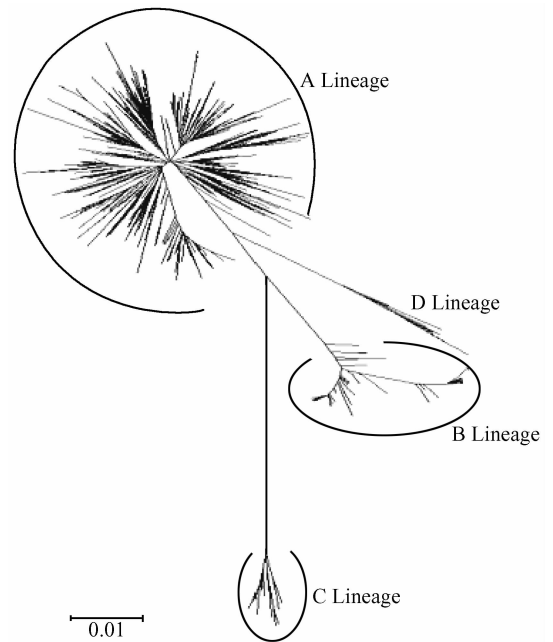


图 1 1 182 条序列构建的 NJ 树

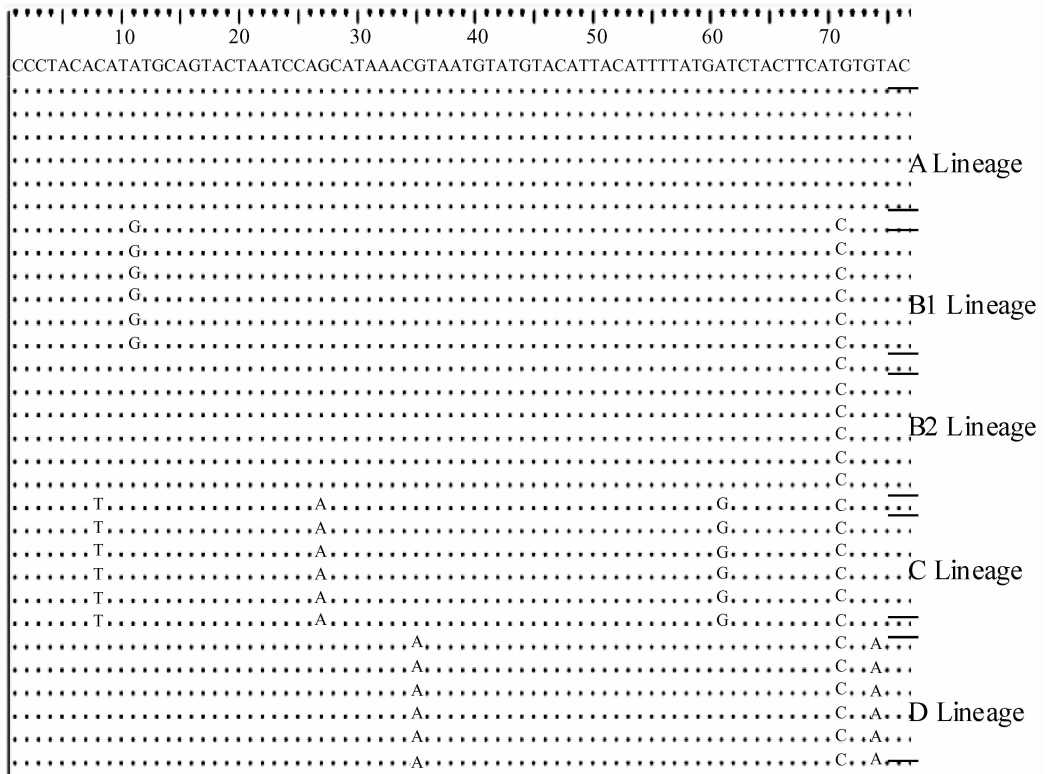
Fig. 1 A neighbour-joining tree for 1 182 sequences

2.2 各支系变异分析

对 A、B、C 和 D 4 个支系利用 BioEdit 软件,以已经发表的山羊线粒体全序列(AF533441)15 735~16 187 片段为参考序列进行比对,4 个支系所特有的变异位点,可以清晰的显示出来,如图 2 是随机选取了这 4 个支系的部分序列。我们共发现能区分 4 个支系的 48 个位点,每个支系有自己特有的碱基变异位点,其中 A 支系的特有变异位点为:71、270、283 和 291,B 支系特有变异位点为:157、229、241、246、301 和 347,C 支系的特有变异位点为:8、27、61、86、126、130、134、173、248、258、267、286、290、301、318、348、383 和 414,D 支系特有变异位点为:35、74、137、153、175、242 和 313。这些位点均表示

在我们截取的 453 bp 长度的位置,在线粒体全序列的位置及具体的碱基替换情况见表 1,所有的变异都是转换。另外,以前有报道^[10]B 支系是由 2 个呈星状辐射的单倍群所构成。所以,为了更清楚的了解这 2 个单倍群的碱基变异特点,我们把这 2 个单倍群进行了分类比对,分别称为 B1 和 B2,从图 2 可以看出,这 2 个单倍群有各自不同的特点,如在 11

和 97 位点,当然,它们也拥有 B 支系所共有的碱基变异特征。除了各个支系特有的变异外,我们还发现有几个支系共有的变异,其中 A 和 D 共有的是 107、239 和 274,A 和 B1 共有的是 237 和 247,B2 和 C 共有的是 128、238 和 275,C 和 D 则在 302 位点。



* . 表示相同碱基 * . The same bases

图 2 截取 4 个支系部分序列用 BioEdit7.0 软件比对结果

Fig. 2 The result of 4 Lineages partial sequences alignment by BioEdit7.0

3 讨论

线粒体 DNA 由于其自身的特点而广泛的应用于山羊的起源进化及系统发育研究,尤其线粒体 D-loop 区,随着研究的深入,越来越多的序列提交到 GenBank 数据库,主要分成 5 个支系,其中 A 支系占的比例最大,依次为 B、C 和 D。为了进一步了解各个支系的碱基变异特点,本研究选取了来自不同国家的 1 182 条序列进行分析,发现 A、B、C 和 D 这 4 个支系有各自特有的碱基变异特点。利用本研究得出的各个支系特有的碱基变异位点,就可以通过直接测定这些特有变异位点来大概区分各个支系,

而省去了测定整个要研究序列进行聚类分析来区分支系的繁琐。在我们分析的各个支系中,C 支系的特有变异位点最多,A 最少,也就说明 A 支系是最普遍的支系,而 C 则是稀有支系。Sultana 等^[10]用 PCR-RFLP 分析巴基斯坦山羊线粒体 D-loop 全序列时得出:377 位点是 B、C、D 特有的,653 位点是 B 特有的,及相当于我们分析的 71、347 位点。通过序列比对,并没有发现不同国家间有特异性的变异,表明山羊品种间分化并不明显,表现出较弱的地理结构,这与以前 Luikart 等^[9]报道的是一致的,在我国李祥龙等^[16-17]研究了国内几个地方山羊品种的线粒体 DNA 多态性,也显示了我国地方山羊品种较低

的分化程度。张红平等^[18]通过对中国 16 个地方品种的山羊线粒体 D-loop 区的分子等级差异分析,品种内的方差组分则占了 75.78%,也表明了山羊品种间弱的地理结构。造成这样的原因可能是古代山

羊经过不同区间的大量运输,从而表现出有家养山羊不明显的地理分化,山羊这种弱的地理结构也正反映了人类迁徙和商贸活动的历史。

表 1 4 个支系的变异位点
Table 1 Variation positions of 4 Lineages

| | | | | | | | | | | | | | | | | |
|--------------------------|----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| mtDNA 全序列位置 | | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| The position of complete | | 7 | 7 | 7 | 7 | 7 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 |
| mtDNA sequence | | 4 | 4 | 6 | 7 | 9 | 0 | 0 | 2 | 3 | 4 | 6 | 6 | 6 | 6 | 7 |
| | | 3 | 6 | 2 | 0 | 6 | 6 | 9 | 1 | 2 | 2 | 1 | 3 | 5 | 9 | 2 |
| 本研究截取序列位置 | | | | | | | | | | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| The position of the | | | 1 | 2 | 3 | 6 | 7 | 7 | 8 | 9 | 0 | 2 | 2 | 3 | 3 | 3 |
| sequences in our data | | 8 | 1 | 7 | 5 | 1 | 1 | 4 | 6 | 7 | 7 | 6 | 8 | 0 | 4 | 7 |
| A Lineage | | C | A | G | G | A | T | G | T | C | G | A | C | A | T | A |
| | B1 | C | G | G | G | A | C | G | T | C | A | A | C | A | T | A |
| B Lineage | B2 | C | A | G | G | A | C | G | T | T | A | A | T | A | T | A |
| C Lineage | | T | A | A | G | G | C | G | C | C | A | G | T | G | C | A |
| D Lineage | | C | A | G | A | A | C | A | T | C | G | A | C | A | T | G |
| | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| mtDNA 全序列位置 | | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 6 |
| The position of complete | | 8 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 0 |
| mtDNA sequence | | 9 | 0 | 6 | 6 | 6 | 7 | 7 | 7 | 7 | 7 | 7 | 8 | 8 | 8 | 9 |
| | | 2 | 8 | 4 | 6 | 8 | 2 | 3 | 4 | 5 | 6 | 7 | 1 | 2 | 3 | 3 |
| 本研究截取序列位置 | | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| The position of the | | 5 | 7 | 2 | 3 | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 5 |
| sequences in our data | | 7 | 3 | 9 | 1 | 3 | 7 | 8 | 9 | 0 | 1 | 2 | 6 | 7 | 8 | 7 |
| A Lineage | | C | A | T | T | T | G | C | T | C | C | T | G | G | C | T |
| | B1 | T | A | C | T | C | G | C | C | C | T | T | A | G | C | T |
| B Lineage | B2 | T | A | C | C | C | A | T | C | C | T | T | A | A | C | T |
| C Lineage | | C | G | T | T | C | A | T | C | C | C | T | G | A | T | C |
| D Lineage | | C | A | T | T | C | A | C | T | T | C | C | G | A | C | T |
| | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| mtDNA 全序列位置 | | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 |
| The position of complete | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| mtDNA sequence | | 0 | 0 | 1 | 1 | 2 | 2 | 2 | 3 | 3 | 4 | 4 | 5 | 8 | 8 | 1 |
| | | 5 | 9 | 0 | 8 | 1 | 5 | 6 | 6 | 7 | 2 | 8 | 3 | 2 | 3 | 8 |
| 本研究截取序列位置 | | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 4 |
| The position of the | | 7 | 7 | 7 | 8 | 8 | 9 | 9 | 0 | 0 | 0 | 1 | 1 | 4 | 4 | 8 |
| sequences in our data | | 0 | 4 | 5 | 3 | 6 | 0 | 1 | 1 | 2 | 7 | 3 | 8 | 7 | 8 | 3 |
| A Lineage | | C | T | T | G | A | A | G | C | A | T | A | G | T | T | G |
| | B1 | T | C | T | A | A | A | A | T | A | T | A | G | C | T | G |
| B Lineage | B2 | T | C | C | A | A | A | A | T | A | T | A | G | C | T | G |
| C Lineage | | T | C | C | A | G | G | A | C | G | C | A | A | T | C | A |
| D Lineage | | T | T | T | A | A | A | A | C | G | T | G | G | T | T | G |

参考文献:

- [1] 马月辉,徐桂芳,王端云,等. 中国畜禽遗传资源信息动态研究[J]. 中国农业科学, 2002, 35(5):89-92.
- [2] 马月辉,吴常信. 畜禽遗传资源受威胁程度评价[J]. 家畜生态, 2001, 22(2):8-13.
- [3] FELIGINE M, PARMA P. The complete nucleotide sequence of goat(*capra hircus*) mitochondrial genome [J]. DNA Sequence, 2003, 14:199-203.
- [4] 李齐发,李隐侠,赵兴波,等. 牦牛线粒体 DNA 细胞色素 b 基因序列测定及其起源、分类地位研究[J]. 畜牧兽医学报, 2006, 37(11):118-123.
- [5] 赖松家,刘延鑫,李学伟,等. 四川黄牛品种线粒体 DNA 遗传多样性研究[J]. 畜牧兽医学报, 2005, 36(9):887-892.
- [6] 刘益平,曹少先,傅泽红,等. 中国部分山羊品种线粒体 DNA D-loop 序列遗传多样性分析[J]. 畜牧兽医学报, 2007, 38(4):313-320.
- [7] 张汤杰,李慧芳,陈宽维,等. 利用线粒体 D-loop 区分析家鸭品种遗传多态性与系统进化[J]. 畜牧兽医学报, 2007, 38(11):1168-1175.
- [8] GUO J, DU L X, MA Y H, et al. A novel maternal lineage revealed in sheep (*Ovis aries*) [J]. Animal Genetics, 2005, 36: 331-336.
- [9] LUIKART G, GIELLY L, EXCOFFIER L, et al. Multiple maternal origins and weak phylogeographic structure in domestic goats [J]. Proceedings of the National Academy of Sciences of the United States of America, 2001, 98:5 927-5 932.
- [10] SULTANA S, MANNEN H, TSUJI S. Mitochondrial DNA diversity of Pakistani goats [J]. Animal Genetics, 2003, 34: 417-421.
- [11] CHEN S Y, SU Y H, WU S F, et al. Mitochondrial diversity and phylogeographic structure of Chinese domestic goats [J]. Molecular Phylogenetics and Evolution, 2005, 37:804-814.
- [12] JOSHI M B, ROUT P K, MANDAL A K, et al. Phylogeography and origin of Indian domestic goats [J]. Molecular Biology and Evolution, 2004, 21: 454-462.
- [13] HALL T A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT [J]. Nucleic Acids Symposium Series, 1999, 41:95-98.
- [14] KUMAR S, TAMURA K, NEI M. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment [J]. Briefings in Bioinformatics, 2004, 5:150-163.
- [15] BANDELT H, FORSTER P, ROHL A. Median joining networks for inferring intraspecific phylogenies [J]. Molecular Biology and Evolution, 1999, 16: 37-48.
- [16] 李祥龙,张亚平,陈圣偶,等. 山羊品种间线粒体 DNA 限制性片段长度多态性研究[J]. 动物学研究, 1997, 18(4):421-428.
- [17] 李祥龙,张亚平,陈圣偶,等. 山羊 mtDNA 多态性及其起源分化研究[J]. 畜牧兽医学报, 1999, 30(4): 313-319.
- [18] 张红平,李利,向德,等. 四川 7 个地方山羊品种(类群)mtDNA 遗传多样性研究[J]. 畜牧兽医学报, 2007, 38(12):1300-1305.