

# 家驴基因组高度重复 DNA 序列的 分离及其文库的构建

赵英杰 吕 品 周祥山 张元兴

华东理工大学生物工程学院/生物反应器工程国家重点实验室, 上海 200237

**摘要** 为研究家驴基因组重复序列,通过2次羟基磷灰石柱层析,分离纯化家驴 Cot 0-Cot 0.1 之间的基因组高度重复 DNA 序列,用于构建家驴基因组 Cot 文库,并向 GenBank 提交了 18 条重复序列,其中 4 条含有 LINEs 序列片段,4 条含有 SINEs 序列片段,3 条含有卫星序列片段,4 条含有简单重复序列片段,1 条含有 LTR-逆转录转座子片段,2 条未知功能的重复序列。

**关键词** 高度重复序列; 分离; 纯化; 羟基磷灰石色谱; Cot 文库; 家驴

**中图分类号** Q 959 **文献标识码** A **文章编号** 1000-2421(2011)05-0563-05

真核生物基因组中存在大量 DNA 重复序列<sup>[1]</sup>,主要包括散布重复序列(interspersed repeats)和串联重复序列(tandem repeats)<sup>[2]</sup>。其中散布重复序列又可分为 DNA 转座子(DNA transposons)、LTR 型反转录转座子(LTR retrotransposons)、长散布重复序列元件(long interspersed repetitive elements, LINEs)和短散布重复序列元件(short interspersed repetitive elements, SINEs) 4 种<sup>[3]</sup>,而串联重复序列则主要包括卫星(satellites)、微卫星(microsatellites)和小卫星(minisatellites)序列<sup>[4]</sup>。Daniel 等<sup>[5]</sup>在研究 DNA 复性动力学的基础上,首次构建了高粱(*Sorghum bicolor*)的 Cot 文库,根据不同 Cot 值成功实现基因组高度、中度重复序列和低拷贝序列的分离和大规模测序。这种快速高效基于 Cot 分析所实现的对各种重复序列的克隆与测序,简称 CBCS(Cot-based cloning and sequencing)。

家驴(*Equus asinus*)隶属于奇蹄目(Perrisodactyla)、马科(Equidae)、马属(*Equus*)<sup>[6]</sup>,具有很高的役用、肉用、奶用和药用价值<sup>[7-9]</sup>,因此,对家驴重复序列的研究有重要的意义。目前,国内外还没有家驴 DNA 重复序列的相关研究报道,笔者分离纯化家驴的高度重复序列并构建了家驴的高度重复序列 Cot 文库,以期为进一步进行驴物种鉴定以及进化

学研究打下基础。

## 1 材料与方法

### 1.1 试验材料

家驴血液采自山东省东阿县无棣驴站。超声波破碎机购自宁波新芝生物科技股份有限公司, Sephadex G-50 购自 Amersham biosciences 公司, 羟基磷灰石购自 Bio-Rad 公司, *Taq* 酶及 Mung Bean Nuclease 购自 TaKaRa 公司, PCR Purification Kit 购自 QIAGEN 公司, pGEM-T 载体及 JM109 感受态细胞购自 Promega 公司。NanoDrop ND-1000 Spectrophotometer 购自 Thermo Fisher Scientific 公司。

### 1.2 家驴血液 DNA 提取与剪切

DNA 提取采用文献<sup>[10]</sup>记述的方法进行。通过调整超声波发生器的功率、振幅与时间,进行 DNA 的超声波剪切,直到 DNA 片段在 500~1 000 bp 左右。为防止超声产热对样品所造成的破坏,应将样品放在冰水浴中。

### 1.3 高度重复序列样品的制备

将超声剪切的 DNA 片段溶解于 1 mL 0.12 mol/L 磷酸盐缓冲液(PBS 中含 1 mmol/L 的 EDTA)中,煮沸 10 min。取出煮沸的 DNA 样品,置于 60 °C 水浴中复性到 Cot 值为 0.1 时,迅速取出样品

收稿日期: 2010-06-20

基金项目: 国家自然科学基金项目(30600809)

赵英杰, 硕士, 研究方向: 中药分子鉴定. E-mail: yangmu\_114@hotmail.com

通讯作者: 周祥山, 博士, 教授, 研究方向: 应用生物技术. E-mail: xszhou@ecust.edu.cn

于干冰乙醇中冷冻,使复性反应终止。将冰冻的复性 DNA 样品融化,注入到 Sephadex G-50 层析柱内脱盐,用 25 mmol/L 氯化钠洗脱(洗脱速率为 1 mL/min),收集 DNA 样品。将收集的 DNA 样品注入到用 2 倍柱体积缓冲液平衡的羟基磷灰石层析柱内,分别用 0.12、0.48 mol/L PBS 洗脱单链和双链 DNA(层析柱温度为 60 °C),分别收集 0.12、0.48 mol/L PBS 洗脱液的洗脱峰,即单链和双链 DNA。将收集的 0.48 mol/L PBS 洗脱液(含双链 DNA)注入到 Sephadex G-50 层析柱内脱盐,用 25 mmol/L 氯化钠洗脱(速率为 1 mL/min),收集 DNA 样品。将上一步收集的 DNA 样品 100 °C 煮沸 10 min,迅速取出样品于干冰乙醇中冷冻,此时复性的 Cot 值为 0。让冷冻样品融化,并将其注入到羟基磷灰石层析柱内,重复上述洗脱操作,分别收集单链和双链 DNA。

#### 1.4 高度重复序列单链 DNA 的双链化补平

在微量离心管中配制反应液:高度重复序列单链 DNA 4  $\mu\text{g}$ ,  $2\times\text{Ex Taq}^{\text{TM}}$  Buffer 20  $\mu\text{L}$ , 6 bp 随机引物 0.4 ng, 加灭菌蒸馏水补足至 80  $\mu\text{L}$ 。设置 PCR 反应程序为 98 °C 6 min, 30 °C 10 min, 72 °C 7 min, 将单链 DNA 补平成双链。用 PCR 纯化试剂盒纯化所得 DNA。

#### 1.5 DNA 样品的平头化及加 A 反应

1) 平头化反应。取纯化所得 DNA 4  $\mu\text{g}$ ,  $10\times\text{Mung Bean Buffer}$  20  $\mu\text{L}$ ,  $\text{Mung Bean Nuclease}$  50 U, 加灭菌蒸馏水补足至 200  $\mu\text{L}$ 。37 °C 平头化反应 10 min。用 PCR 纯化试剂盒对反应终止的 DNA 进行纯化。

2) 加 A 反应。取平头化 DNA 样品 2  $\mu\text{g}$ ,  $2\times\text{Ex Taq}^{\text{TM}}$  Buffer 10  $\mu\text{L}$ , 加灭菌蒸馏水至 40  $\mu\text{L}$ 。72 °C 反应 10 min 进行加 A 反应。用 PCR 纯化试剂盒纯化所得 DNA。

#### 1.6 连接与转化

按照试剂盒说明书上记述的方法将本文“1.5”纯化的 DNA 片段连接到 pGEM-T 载体上,转化至 JM109 感受态细胞。

#### 1.7 克隆的挑取与保存

用无菌牙签将平板上的单菌落挑至 96 孔板中,每个菌落 1 个孔,每孔加入 150  $\mu\text{L}$  含 12.5  $\mu\text{g}/\text{mL}$  氯霉素的 LB 液体培养基,用封口膜封口,放置 37 °C 恒温培养箱中培养过夜。次日,加入甘油至终体积分数为 20%。将孔板中的单克隆过夜培养物平均

分配到 2 张 96 孔微孔板中。将 96 孔微孔板进行系统编号后,1 份置于 -80 °C 超低温冰箱中,1 份置于 -20 °C 冰箱中保存。

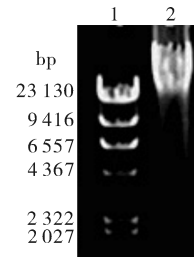
#### 1.8 菌落 PCR 鉴定及测序分析

随机挑选单克隆菌株作为菌液模板,于 PCR 管中配制反应液:菌液模板 2  $\mu\text{L}$ , 上游 M13 引物(10  $\mu\text{mol}/\text{L}$ ) 0.3  $\mu\text{L}$ , 下游 M13 引物(10  $\mu\text{mol}/\text{L}$ ) 0.3  $\mu\text{L}$ ,  $2\times\text{Taq Premix}$  7.5  $\mu\text{L}$ , 加灭菌蒸馏水至 15  $\mu\text{L}$ 。设置 PCR 反应程序为 95 °C 5 min, 94 °C 30 s, 55 °C 30 s, 72 °C 1 min, 30 个循环, 72 °C 7 min。反应完成后,电泳检测 DNA 片段大小。将菌落 PCR 鉴定中含有阳性条带的菌株取样送到上海英骏生物技术有限公司用 3730 测序仪进行序列测定,测得的序列在 RepeatMasker 数据库上进行序列比对分析,完成后将序列提交 GenBank。

## 2 结果与分析

#### 2.1 家驴基因组 DNA 的提取及超声剪切

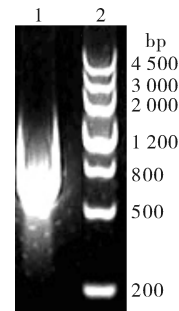
图 1 表明,血液基因组的 DNA 片段比较完整,由 NanoDrop ND-1000 全波长紫外/可见光扫描分光光度计检测,  $A_{260}/A_{280} = 1.91$ 。超声剪切后, DNA 片段大小主要集中在 500~1 000 bp(图 2)。



1.  $\lambda\text{HindIII}$ ; 2. 家驴基因组 DNA Genomic DNA of *E. asinus*.

图 1 家驴基因组 DNA 电泳图

Fig. 1 Electrophoresis of genomic DNA of *E. asinus*



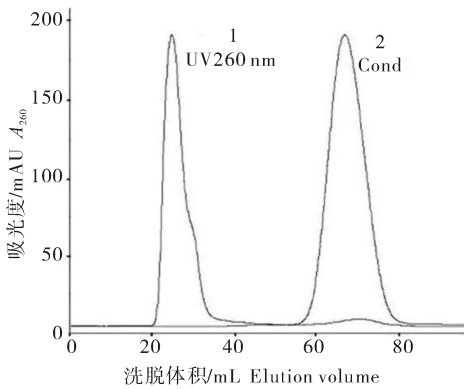
1. 超声剪切 DNA Sonicated DNA of *E. asinus*; 2. Marker III.

图 2 超声剪切 DNA 电泳图

Fig. 2 Electrophoresis of sonicated DNA of *E. asinus*

## 2.2 高度重复序列的制备

在羟基磷灰石柱层析之前,对核酸样品进行了 Sephadex G-50 柱层析以除盐(图 3),一方面可以去除 EDTA 对羟基磷灰石的破坏,另一方面也可以满足 DNA 在低盐浓度下与羟基磷灰石结合的要求。在图 3 中,洗脱峰 1 是 260 nm 紫外光的吸收峰,为核酸的洗脱峰,洗脱峰 2 是电导率的吸收峰,为盐离子的洗脱峰。图 3 表明,核酸样品中的盐离子已被



1. UV-260 吸收峰  $A_{260}$ ; 2. 电导率吸收峰 Conductivity.

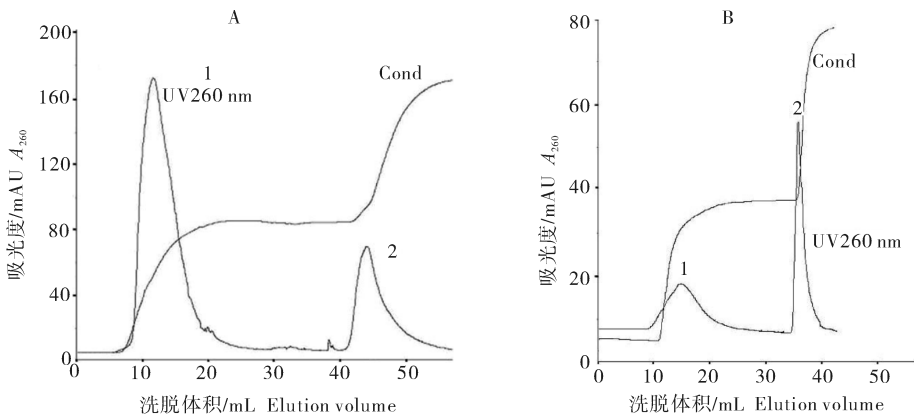
图 3 Sephadex G-50 柱层析除盐

Fig. 3 Desalting DNA sample using Sephadex G-50 column by gel filtration

较好地去除。

本试验采用羟基磷灰石柱层析分离单链和双链 DNA 洗脱时,在同一 pH 值(pH 6.8)下,用 2 种不同盐离子浓度的磷酸缓冲液洗脱,收集到 2 个吸收峰(图 4),低盐浓度下先洗出来的峰 1 为单链 DNA 洗脱峰,高盐浓度下后洗脱出来的峰 2 为双链 DNA 洗脱峰。

在图 4-A 中,第 1 次羟基磷灰石柱层析时,洗脱峰 1 包含了在  $Cot=0.1$  的复性条件下驴基因组中那些没有复性的仍为单链的单拷贝、低拷贝和中度重复基因组 DNA,洗脱峰 2 包含了在  $Cot=0.1$  的复性条件下驴基因组中那些已经复性的、双链高度重复 DNA 和链内折叠 DNA。为了分离高度重复 DNA 和链内折叠 DNA,将图 4-A 中洗脱峰 2 收集下来,脱盐,解链,并进行第 2 次羟基磷灰石柱层析,结果见图 4-B。在图 4-B 中,洗脱峰 1 为  $Cot=0$  到  $Cot=0.1$  之间没有复性的仍为单链的基因组高度重复 DNA,洗脱峰 2 是在瞬间能够完成复性的双链的链内折叠 DNA。图 4-B 中,洗脱峰 1 为目的基因组高度重复 DNA,是单链 DNA,因此有必要对其进行互补链的合成。



A. 第 1 次 First; B. 第 2 次 Second;

A-1. 单链 DNA 洗脱峰,含低拷贝和中度重复 DNA ssDNAs,including lowly repetitive DNA and moderately repetitive DNA;

A-2. 双链 DNA 洗脱峰,含高度重复和链内折叠 DNA dsDNAs,including highly repetitive DNA and foldback DNA;

B-1. 单链 DNA 洗脱峰,含高度重复 DNA ssDNAs,including highly repetitive DNA;

B-2. 双链 DNA 洗脱峰,含链内折叠 DNA dsDNAs,including foldback DNA.

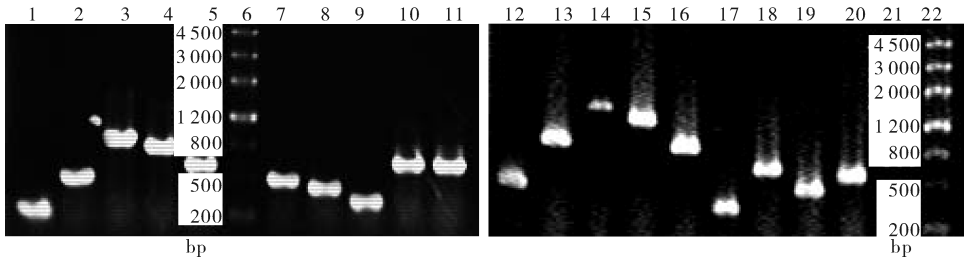
图 4 羟基磷灰石柱层析,  $Cot0.1$  DNA 的分离

Fig. 4 Hydroxyapatite chromatography, separation of  $Cot0.1$  DNA

## 2.3 菌落 PCR 鉴定

将羟基磷灰石层析后所得目的基因组高度重复 DNA 进行互补链合成,补平并加 A 后连接到 pGEM-T 载体上,转化感受态大肠杆菌 JM109,通

过氨苄抗性和蓝白斑筛选,随机挑取了 960 个阳性克隆。利用 PCR 检测出 684 个含有插入片段的克隆(如图 5),1~5、7~20 泳道为随机挑选的单克隆菌株 PCR 产物,其中 21 泳道为空载质粒 PCR 结



6, 22. Marker III; 1~5, 7~20. 随机挑选的单克隆菌株 PCR 产物 Inserted DNA fragments in highly repetitive Cot library by PCR; 21. 对照 CK.

图 5 菌落 PCR 鉴定高度重复序列 Cot 文库克隆插入片段

Fig. 5 Identification of the inserted DNA fragments in highly repetitive Cot library by PCR

果, 没有得到非特异性条带, 插入载体的目的片段主要集中在 500~1 000 bp 之间。这与超声剪切的片段大小相吻合, 表明本试验比较成功。

将序列提交 GenBank 并在 RepeatMasker 数据库上进行比对分析, 该数据库基于 Smith-Waterman-Gotoh 算法, 专门适用于分散重复序列如 LINE、SINE、MIR 等的搜索和确定。比对结果汇总如下, 在所提交的 18 条随机高度重复序列 Cot 文库插入片段中: 有 4 条含有 LINE 序列片段 (GenBank 序列号: FI763399, FI763400, FI763404,

FI763408); 有 4 条含有 SINE 序列片段 (GenBank 序列号: FI763405, FI763406, FI763409, FI763416); 有 3 条含有卫星序列片段 (GenBank 序列号: FI763401, FI763403, FI763410); 有 4 条含有简单重复序列 (GenBank 序列号: FI763402, FI763412, FI763413, FI763415); 有 1 条含有 LTR-逆转录转座子序列片段 (GenBank 序列号: FI763407), 2 条未知功能的重复序列 (GenBank 序列号: FI763411, FI763414), 结果分析见表 1。

表 1 Cot 文库序列结果分析

Table 1 Analysis of repeat sequence screened from the Cot library

序列 Query sequence	长度/bp Total length	匹配的重复序列 Matching repeat	重复序列分类 Repeat class	比对应分值 Score	偏差/% Divergence
FI763399	423	L1-1_EC	LINE/L1	1 344	1.3
FI763400	401	L1-3_EC	LINE/L1	1 204	2.1
FI763401	491	ES22	卫星序列 Satellite	817	16.1
FI763402	554	(CCCTAA) <i>n</i>	简单重复序列 Simple repeat	787	6.9
FI763403	279	ES22	卫星序列 Satellite	1 343	16.2
FI763404	424	L1-1_EC	LINE/L1	462	5.0
FI763405	596	ERE3	SINE/tRNA	903	21.6
FI763406	450	ERE3B	SINE/tRNA	363	14.9
FI763407	635	ERV1-3N-EC_I-int	LTR/ERV1	612	14.2
FI763408	308	L1-1_EC	LINE/L1	2 633	2.3
FI763409	591	ERE1B	SINE/tRNA	1 072	15.9
FI763410	563	SAT_EC	卫星序列 Satellite	973	14.9
FI763411	417	N/A	N/A <sup>1)</sup>	N/A	N/A
FI763412	483	(CA) <i>n</i>	简单重复序列 Simple repeat	333	12.5
FI763413	471	(TTAGGG) <i>n</i>	简单重复序列 Simple repeat	1 260	0
FI763414	294	N/A	N/A	N/A	N/A
FI763415	644	(TCCA) <i>n</i>	简单重复序列 Simple repeat	588	2.9
FI763416	182	MIRb	SINE/MIR	338	24.2

1) N/A: 没有匹配序列 No matches.

### 3 讨论

采用羟基磷灰石柱层析法分离基因组高度重复序列, 为了得到较高的准确性和良好的重复性, 控制 DNA 复性条件和洗脱条件是必要的。在进行复性反应时, 一般复性缓冲液 PBS 的浓度为 0.12 mol/L, DNA 片段大小为 200~1 500 bp, 复性温度

为 60 °C。洗脱时, 所用的磷酸盐缓冲液 (PBS) 由等物质的量之比的磷酸氢二钠和磷酸二氢钠配制, pH 6.8。用 0.12 mol/L 的 PBS 洗脱单链 DNA, 用 0.48 mol/L 的 PBS 洗脱双链 DNA。洗脱前, PBS 要预热, 并且在层析过程中, 羟基磷灰石柱置于 60 °C 下保温进行操作, 通常的做法是用夹套层析柱, 循环水维持柱温。

传统分离高度重复序列的方法通常采用1次羟基磷灰石色谱,即分离 Cot 0.1 DNA,而 Cot 0.1 DNA 包含了高度重复 DNA 以及链内折叠 DNA,这种做法无法除去链内折叠 DNA。本试验用了2次羟基磷灰石色谱,收集 Cot=0 和 Cot=0.1 之间的 DNA,这样就可去除链内折叠 DNA 对高度重复序列的影响。从图4可以看出,链内折叠 DNA 在 Cot 0.1 DNA 中所占比例比较大,会对后续基因组高度重复序列研究造成影响。

对通过羟基磷灰石柱层析收集得到的高度重复 DNA 为单链 DNA 进行互补链合成是有必要的,本试验所采用的方法是随机引物合成法。随后,将高度重复 DNA 纯化,连接到 pGEM-T 载体上,转化感受态大肠杆菌 JM109 并测序分析,成功建立起了驴高度重复序列 Cot 文库。对其中 18 条序列进行研究表明,文库插入片段主要集中在 500~1 000 bp 之间,插入片段除了 2 个未知功能的重复序列外,其他均为高度重复序列,说明文库质量较好。更深入的研究可对文库序列进行大规模测序,以深入了解家驴高度重复序列的组成、分布以及作用等。对这些高度重复序列的深入分析将有助于驴的物种鉴定以及进化学研究。

### 参 考 文 献

[1] BRITTEN R J, KOHNE D E. Repeated sequences in DNA:

hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms [J]. *Science*, 1968, 161: 529-540.

[2] WICKSTEAD B, ERSFELD K, GULL K. Repetitive elements in genomes of parasitic protozoa [J]. *Microbiol Mol Biol Rev*, 2003, 67: 360-375.

[3] LEWIS B. *Genes VIII* [M]. New Jersey: Pearson Prentice Hall, 2004: 70-123.

[4] WICKER T, ROBERTSON J S, SCHULZE S R, et al. The repetitive landscape of the chicken genome [J]. *Genome Res*, 2005, 15: 126-136.

[5] PETERSON D G, SCHULZE S R, SCIARA E B, et al. Integration of Cot analysis, DNA cloning, and high-throughput sequencing facilitates genome characterization and gene discovery [J]. *Genome Res*, 2002, 12: 795-807.

[6] 谭杰邦. 哺乳动物分类名录[M]. 北京: 中国医药科技出版社, 1992: 390-393.

[7] POLIDORI P, VINCENZETTI S, CAVALLUCCI C, et al. Quality of donkey meat and carcass characteristics [J]. *Meat Sci*, 2008, 80: 1222-1224.

[8] VINCENZETTI S, POLIDORI P, MARIANI P, et al. Donkey's milk protein fractions characterization [J]. *Food Chem*, 2008, 106: 640-649.

[9] 李时珍. 本草纲目[M]. 北京: 人民卫生出版社, 1982: 2780-2785.

[10] SAMBROOK J, RUSSELL D W. *Molecular cloning laboratory manual* [M]. New York: Cold Spring Harbor Laboratory Press, 2002: 463-470.

## Separation of highly repetitive DNA for construction HR Cot library of *Equus asinus*

ZHAO Ying-jie LÜ Pin ZHOU Xiang-shan ZHANG Yuan-xing

*School of Bioengineering/State Key Laboratory of Bioreactor Engineering,  
East China University of Science and Technology, Shanghai 200237, China*

**Abstract** In order to investigate the highly repetitive (HR) DNA sequences among *Equus asinus* genome, Cot0-Cot0.1 DNA of *E. asinus* genome was separated and purified through two steps of hydroxypatite chromatography. Then the HR Cot library of *E. asinus* was constructed by gene cloning and sequencing. After the analysis and screening of the library, 18 repetitive elements were identified and submitted to GenBank, including 4 LINEs, 4 SINEs, 3 satellites, 4 simple repeats, 1 LTR retrotransposon and 2 unknowns.

**Key words** highly repetitive DNA; separation; purification; hydroxypatite chromatography; Cot library; *Equus asinus*

(责任编辑:边书京)