

华中农业大学硕士学位论文

基于 CAPSS 的糜子 BAC 克隆在参考基因组上定位

**Mapping broomcorn millet BAC clones to the reference
genome based on CAPSS**



目录

摘要.....	i
Abstract	ii
缩略词表.....	i
1 前言.....	1
1.1 基因组测序技术的发展	1
1.1.1 Roche 454 测序平台	2
1.1.2 ABI SOLiD 测序平台	3
1.1.3 Solexa 测序平台	4
1.1.3 PacBio 测序平台	4
1.1.4 Oxford Nanopore 测序平台	6
1.2 全基因组测序策略的发展	7
1.3 全基因组辅助拼装技术	9
1.4 基因组文库	10
1.4.1 基因组文库的类型	10
1.4.2 基因组文库的应用	11
1.5 糜子的简介	14
1.6 研究目的和意义.....	15
1.7 研究策略	15
2 材料与方法	17
2.1 实验材料	17
2.2 实验仪器与试剂.....	17
2.3 数据分析工具.....	18
2.4 BAC 混合池 NGS 文库构建	19
2.4.1 BAC 二级混合池构建	19
2.4.2 BAC 混合池 DNA 制备.....	20
2.4.3 Illumina 测序文库构建.....	21
2.4.4 实时荧光定量 PCR 测定相对浓度.....	25
2.5 BAC 末端序列解析方法	26

2.6 BAC 末端序列重复序列注释	29
2.7 BAC 克隆在陇糜 4 号基因组中的定位	30
2.8 基于 JBrowse 展示 BAC 信息.....	30
2.9 Sanger 法验证 BAC 末端序列	31
2.10 鉴定叶绿体和线粒体 BAC 克隆.....	32
3 结果和分析	33
3.1 二级混合池测序文库构建	33
3.2 测序数据预处理.....	34
3.3 糜子 BAC 短末端序列解析	37
3.4 二级混合池 NGS 数据组装与长末端序列解析.....	38
3.5 糜子 BAC 末端序列中重复序列注释.....	39
3.6 糜子 BAC 末端序列与陇糜 4 号基因组比对及定位.....	40
3.7 Sanger 测序验证 BAC 定位	44
3.8 细胞器 BAC 克隆鉴定	44
4 讨论.....	46
4.1 大肠杆菌基因组污染.....	46
4.2 质量控制和污染去除是数据分析前的重要步骤.....	47
4.3 末端序列解析过程中软件的选择.....	47
4.4 BAC 末端序列定位中的问题	48
4.5 关于细胞器基因组.....	49
参考文献.....	50
附录 1 试剂配方	56
附录 2 BAC 质粒制备(用于 BAC 末端测序).....	58
附录 3 二级混合池统计.....	59
附录 4 用于 NGS 建库的引物序列	64
致谢.....	68

摘要

糜子是中国传统“五谷”之一，抗旱耐贫瘠，是干旱半干旱地区的重要作物。目前已有两个糜子品种的参考基因组发表。这两个基因组主要是以全基因组鸟枪法测序策略结合 Illumina 测序技术、PacBio 测序技术、HiC 技术和图谱技术完成的，其组装质量较好。BAC 文库作为一种基因组资源，在基因组完善、功能基因研究、基因组比较等方面起着重要作用。本研究的主要目的是将糜子 BAC 文库与参考基因组进行整合，以便最大程度地利用 BAC 文库资源。

本研究以糜子 BAC 文库中的 9216 个克隆为材料，利用 CAPSS (克隆矩阵混合池鸟枪测序) 策略和 Illumina 测序技术，经过两种途径解析 BAC 末端序列，最终将 8262 个 BAC 克隆定位到糜子参考基因组上，成功率为 89.65%。其中通过双末端定位的克隆有 5391 个，通过单末端定位的克隆有 2871 个。随机挑拣 55 个 BAC 克隆通过 Sanger 法验证定位的准确性，其中 54 个 BAC 定位结果与其一致。通过 BAC 双末端定位结果，估计文库的平均插入片段大小为 123.48 kb。通过将 BAC 末端序列与细胞器基因组进行比对，发现了 14 个来自叶绿体基因组的 BAC 克隆，但没有发现属于线粒体基因组的 BAC 克隆。相对于 Sanger 测序法，本研究中所使用的方法极大地降低了获取 BAC 末端序列的成本，缩短了实验周期。BAC 文库与基因组的整合可为获取包含特定功能基因的 BAC 克隆提供便利，为基因组序列校正和完善提供可靠保障，为分子育种和基因组功能研究奠定基础。

关键词：糜子；BAC 文库；克隆矩阵混合池测序；BAC 末端序列；定位

Abstract

Broomcorn millet (*Panicum miliaceum*) is one of the traditional five grains in China. It is drought-resistant and barren-tolerant and is an important crop in arid and semi-arid regions. At present, reference genomes of two broomcorn millet varieties have been published, which are accomplished by whole-genome shotgun sequencing strategy combined with Illumina sequencing technology, PacBio sequencing technology, HiC technology and map technology. The assembly qualities are high. BAC library, as a genomic resource, is of great use in genome sequence improvement, functional gene research and comparative genomics field. The purpose of this research was to integrate BAC library into broomcorn millet genome to maximize the use of BAC library resource.

In this study, 9216 clones from the broomcorn millet BAC library were chosen as a material set. Using CAPSS (Clone-Array Pooled Shotgun Sequencing) strategy and Illumina sequencing technology, 8262 BAC clones were mapped onto the broomcorn millet reference genome, of which 5391 were mapped by paired-end sequences and 2871 by single-end sequence. In order to assess the accuracy, we randomly picked up 55 BAC clones and sequenced by Sanger method. The loci of 54 BAC clones mapping by Sanger sequences were identical to the former. The average of BAC inserted fragment lengths is 123.48 kb, according to an estimate by the result of BAC paired-end mapping. By aligning the BAC end sequences with the organelle genomes, 14 BAC clones from chloroplast genome and no clones from mitochondria genome were identified. Compared to the Sanger sequencing method, our method used in this study can greatly reduce the cost of acquiring BAC end sequences and shorten the period of the experiment. The integration between the BAC library and the genome will facilitate the acquisition of BAC clones with specific functional genes, provide a reliable basis for the correction and improvement of genome sequences, and form the basis of molecular breeding and genomic function research.

Key words: broomcorn millet; BAC library; CAPSS; BES; mapping

缩略词表

缩写形式	英文全称	中文意译
BAC	Bacterial Artificial Chromosome	细菌人工染色体
BES	BAC End Sequence	BAC 末端序列
CAPSS	Clone Array Pooled Shotgun Sequencing	克隆矩阵混合池鸟枪测序法
Cm	Chloramphenicol	氯霉素
EST	Expressed Sequence Tag	表达序列标签
FISH	Fluorescence <i>In Situ</i> Hybridization	荧光原位杂交
Hi-C	High-throughput Chromosome conformation capture	高通量染色体构象捕获
LINE	Long Interspersed Nuclear Elements	长散在重复序列
LTR	Long Terminal Repeats	长末端重复
MTP	Minimal Tiling Path	最小重叠路径
NGS	Next Generation Sequencing	下一代测序技术
PCR	Polymerase Chain Reaction	聚合酶链式反应
PFGE	Pulsed Field Gel Electrophoresis	脉冲场凝胶电泳
RFLP	Restriction Fragment Length Polymorphism	限制性片段长度多态性
SINE	Short Interspersed Nuclear Elements	短散在重复序列
SMRT	Single-Molecule Real-Time	单分子实时测序技术
SNP	Single Nucleotide Polymorphism	单核苷酸多态性
SOLiD	Sequencing by Oligonucleotide Ligation and Detection	寡核苷酸连接和检测测序
SR	Short Read	短序列
SSLP	Simple Sequence Length Polymorphism	简单序列长度多态性
VES	Vector End Sequence	载体末端序列
ZMW	Zero-Mode Waveguide	零模波导孔

1.5 糜子的简介

糜子 (*Panicum miliaceum*), 又称黍、稷, 属于禾本科一年生草本植物, 在中国已有 1 万多年的栽培历史, 是中国传统“五谷”之一。糜子的生长周期较短, 抗旱, 耐贫瘠, 是干旱半干旱地区的重要作物 (Yang et al 2012)。虽然糜子生长在土壤贫瘠的地区, 其对水资源的利用率是极高的。随着全球水资源和人口危机的日益加剧, 糜子高效的水分利用可为人类未来的粮食安全提供保障。

糜子是异源四倍体植物, 由 18 对染色体组成, 基因组大小约 887.8Mb (Shi et al 2019)。目前, 已有中国农业大学赖锦盛课题组和中国科学院上海植物逆境生物学研究中心张衡课题组分别公布了两个不同品种的糜子基因组序列 (表 1-1)。

表 1-1 已公布的两个糜子参考基因组信息

Table 1-1 The genome information for two broom millet varieties

类型	育成品种(陇糜 4 号)	本地品种
No. NCBI BioProject	PRJNA429322	PRJNA431363
No. NCBI Assembly	GCA_002895445.2	GCA_003046395.1
Tect. type	Illumina、PacBio、HiC、 BioNano	Illumina、PacBio、HiC、 binmap
Total contigs	1308 (838.9 Mb)	5,541 (838 Mb)
N50 contigs	2.55 Mb	396 kb
Total scaffolds	475 (848.4 Mb)	1309 (855 Mb)
N50 scaffolds	48.26 Mb	46.66 Mb
GC content	46.8%	46.8%
Numbers of genes	63,671	65,573
Publication	(Shi et al 2019)	(Zou et al 2019)

1.6 研究目的和意义

BAC 文库作为一种基因组资源，对于基因组序列完善，功能基因研究、比较基因组学等方面具有重要意义。本实验室陈波构建了一个糜子 BAC 文库，为 BAC 文库的应用研究奠定了基础。

本研究主要目的是将该 BAC 文库中的克隆定位到已发表的糜子参考基因组上。在定位过程中，基于克隆矩阵混合池测序策略和 Illumina 测序技术，从 BAC 混合池测序序列中解析出 BAC 末端序列，通过解析的末端序列，将 BAC 克隆定位参考基因组上。以往，为了将 BAC 克隆定位到基因组上，常利用 Sanger 测序法对逐个 BAC 克隆进行双末端测序，再根据双末端序列确定 BAC 在基因组上的位置。这种方法适用于少量 BAC 的定位。本研究采用的方法可同时将几千个 BAC 克隆定位到基因组上，这极大的降低了成本，缩短了实验周期。此外，BAC 文库与基因组的整合可为后续的基因组序列校正和完善提供可靠保障，同时为分子育种和基因组功能研究奠定基础。

1.7 研究策略

克隆矩阵混合池鸟枪法最早由 Cai 等提出并应用到恒河猴基因组测序中 (Milosavljevic et al 2005)，其主要思想是：BAC 克隆首先被排列成二维矩阵的形式，分别将每一行或每一列所有克隆混合在一起，构建亚克隆文库并测序。通过比对行池序列和列池序列将序列定位到交叉位点上，然后组装出每个 BAC 的序列，再通过人工校正形成参考基因组序列 (Cai et al 2001, Csuros et al 2003)。本研究主要是以糜子 BAC 文库为材料，通过克隆矩阵混合池测序策略解析 BAC 末端序列，同时利用末端序列将 BAC 克隆定位到陇糜 4 号参考基因组上。为了方便理解，以图 1-4 为例描述整个研究思路：

(1) 首先将 BAC 克隆排列成一个矩阵的形式，我们称这个矩阵的所有克隆集合为超级混合池。矩阵的每一行或每一列的克隆集合称为二级混合池，其中每一行的克隆集合称为行池，每一列中的克隆集合称为列池。在图 1-4 中，由 64 个克隆组成一个 8 X 8 的矩阵，其中包含 8 个列池和 8 个行池，共 16 个二级混合池，每个二级混合池中均有 8 个 BAC 克隆。

(2) 分别对 16 个混合池进行 BAC DNA 提取、随机打断并构建 Illumina 测序文库，在文库构建过程中对这些文库添加不同的 index 序列，以便进行多文库混样测序，节约测序成本。

(3) 原始下机数据需要先进行质量控制和污染去除，获得有效的测序序列，再通过 index 序列将数据拆分到每个对应的二级混合池中。为了兼顾 BAC 末端序列的数量和长度，我们通过短末端和长末端两种途径来解析每个混合池的末端序列集合。随后分别对每个行与列池末端序列集合求交集，得到交叉点上 BAC 的末端序列。为了排除超级混合池中重复 BAC 的影响，使用了一种差集法移除假阳性序列(Pan et al 2016)。

(4) 将解析的 BAC 末端序列比对到参考基因组上，通过 BAC 克隆长度和双末端方向特征，确定 BAC 克隆在基因组上的唯一位置。最后通过 JBrowse 展示 BAC 克隆在基因组上的位置信息。

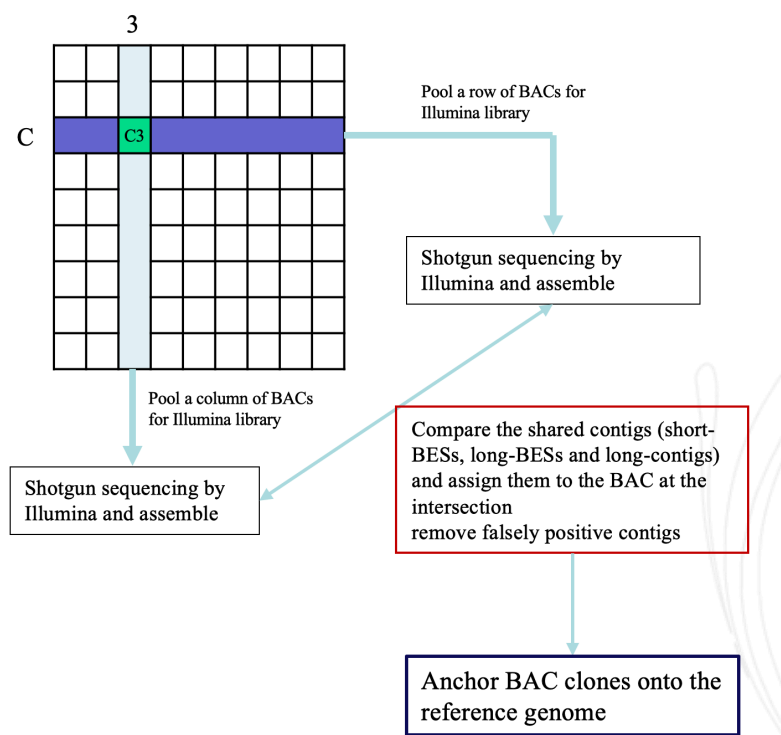


图 1-4 基于克隆矩阵混合池的末端序列解析策略

Fig. 1-4 The strategy of BAC end sequence location based on CAPSS

2 材料与方法

2.1 实验材料

构建 BAC 文库的主要过程包括：取暗室培养的黄化苗，随后进行液氮研磨，分离出细胞核。加等体积 1% 低熔点琼脂糖混匀后制成细胞核包埋块 (plug)。Plug 用蛋白酶 K 处理 48 h 后，再用 PMSF (苯甲基磺酰氟) 去除蛋白酶 K，然后存放于 4℃ 的 TE 中备用。使用限制性内切酶 *Hind*III 对基因组 DNA 进行部分酶解，经脉冲凝胶电泳筛选，回收合适大小的 DNA 片段并与 pIndigoBAC536-S 载体连接，连接产物转化到大肠杆菌菌株 DH10B T1 Phage-Resistant 细胞后进行蓝白斑筛选，白色菌落捡入 384 孔板中。该文库共包含 76,032 (198 X 384) 个克隆，平均插入片段为 132 kb，空载率小于 1%，约覆盖基因组的 10.6 倍 (图 2-1)。该文库长期保存在本实验室 -80℃ 超低温冰箱中，以供研究使用。

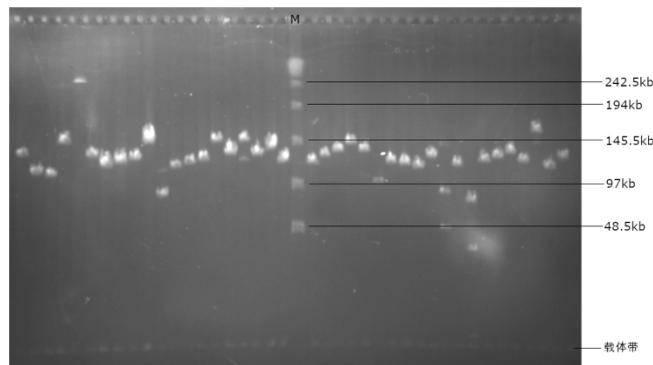


图 2-1 糜子插入片段检测 (陈波 20160529)

Fig. 2-1 The inserted size detection of broomcorn millet BAC library (Chen 20160529)

图为随机挑选的 40 个克隆的插入片段检测图片。Maker 为 λ Ladder。

The figure is inserted fragment detection image of 40 randomly selected clones. The maker used is λ Ladder.

2.2 实验仪器与试剂

实验主要仪器：高速冷冻离心机 (BECKMAN)；超声波破碎仪 (BioRuptor)；恒温培养箱 (上海智诚)；超微量紫外分光光度计 (Merinton SMA4000)；凝胶成像系统 (Carestream)；超低温冰箱 (SANYO)；PCR 仪 (BioRad)。

实验主要试剂：Epicentre Plasmid-Safe™ ATP-Dependent DNase，Thermo Fermentas T4 Polynucleotide Kinase，Thermo Fermentas T4 DNA polymerase，TaKaRa *Taq* DNA polymerase，Thermo Fermentas T4 DNA ligase，TOYOBO KOD FX，生工 25mM dNTP mixture，AXYGEN AxyPrep Easy-96 Plasmid kit，DNA 分子量标准购自北京全式金生物技术有限公司，MagBead DNA Purification Kit 购自康为世纪生物科技有限公司，提取 BAC 末端质粒的 S1、S2、S3 试剂由本实验室自行配制，测序文库构建过程所使用的长引物均在金唯智公司合成，定量使用的短引物在武汉擎科合成。

2.3 数据分析工具

(1) 数据预处理

数据质量查看工具 FastQC-v0.11.5 (Andrews 2010)

<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

数据质控工具 Trimmomatic-v0.36 (Bolger et al 2014)

<http://www.usadellab.org/cms/?page=trimmomatic>

(2) 序列比对软件

通用序列比对软件 BLAST-v2.9.0 (Altschul et al 1990)

<ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+>

序列比对软件 BLAT-v35 (Kent 2002)

<http://genome.ucsc.edu/>

长序列比对软件 Minimap2-v2.9 (Li 2016)

<https://lh3.github.io/minimap2>

短序列比对软件 Bowtie2-v2.3.3.1 (Langmead 2013)

<https://github.com/BenLangmead/bowtie2>

(3) 序列组装软件

短序列，单细胞数据组装软件 SPAdes-v3.13.0 (Bankevich et al 2012)

<http://bioinf.spbau.ru/spades>

scaffold 组装软件 Scarpa-v0.241 (Donmez and Brudno 2013)

<http://compbio.cs.toronto.edu/scarpa>

(4) 重复序列注释工具

重复序列在线注释工具 RepeatMasker-v4.0.8 和 RepeatModeler-v1.0.11 (Smit et al 1996)

<http://www.repeatmasker.org>

(5) 文件转换和数据统计工具

Samtools-v1.5 (Sam et al 2014)

<http://samtools.sourceforge.net>

Bedtools-v2.25.0 (Quinlan and Hall 2010, Quinlan 2014)

<https://bedtools.readthedocs.io/en/latest/>

BBTools-v38.32 (Bushnell 2014)

<https://jgi.doe.gov/data-and-tools/bbtools/>

seqkit-v0.9.0 (Shen et al 2016)

<https://bioinf.shenwei.me/seqkit/>

(6) 结果展示

基于网页的轻量级数据展示工具 JBrowse-v1.16.3 (Skinner et al 2009)

<http://jbrowse.org>

(7) 编程环境

Linux Bash <https://www.gnu.org/software/bash/>

GCC 4.4.7 <https://gcc.gnu.org/>

Jdk 1.8.0_45 <http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>

Python 3.6.3 <https://www.python.org>

Python 模块: HTSeq https://htseq.readthedocs.io/en/release_0.11.1/

2.4 BAC 混合池 NGS 文库构建

2.4.1 BAC 二级混合池构建

本实验同时构建行池和列池，因此需要将原始板复制两份。在超净工作台中，用灌板仪向 384 孔板中灌注 LB 培养基 (12.5 $\mu\text{g/mL}$ Cm)，每孔灌注 60 μL ，4 $^{\circ}\text{C}$ 保存备用。

把糜子 BAC 文库编号为 25-48 的 COPY I 从-80 $^{\circ}\text{C}$ 中取出化冻，总共 24 块 384 板。化冻前用无菌面巾纸擦净板面的霜层，将 384 板平铺在超净工作台中，板上覆盖一层面巾纸，打开超净台风机，微风吹 1h 左右。完全化冻后，若板内盖覆着有

冷凝水，需在无菌条件下用灭菌的滤纸擦去，避免因积水流动导致孔之间的污染。在超净工作台中，将 384 复制叉蘸取 75%酒精后在火焰上灼烧一遍，再蘸取 95%酒精烧两次，待复制叉冷却后，蘸取母板菌液，按相同方向依次接菌至复制板中，并在复制板上做好标记，与母版一一对应。每板 384 板复制后，需要按上述流程将 384 复制叉灼烧灭菌。复制板复制完成后，置于 37℃恒温培养箱中培养 24 h。

将 24 块 384 板按 6 行 4 列排列构成一个 96 x 96 的矩阵，将这个矩阵的每一行、每一列菌液混合在一起，共混成 96 个行池和 96 个列池，将这些混合池菌液按序分装到 96 孔深孔板中，每孔分装 1 ml 菌液。行池和列池各分装到 4 个 96 孔深孔板中，出口用 BF-400 膜封住。

2.4.2 BAC 混合池 DNA 制备

采用 AXYGEN AxyPrep Easy-96 Plasmid kit 制备 BAC 混合池 DNA，具体方法如下：

- (1) 将装有混合池菌液的 96 孔深孔板置于离心机中，1500 x g 离心 5 min；
- (2) 弃去 BF-400 膜，将 96 孔板迅速倒置弃去培养上清，然后将 96 孔深孔板倒扣在滤纸上轻轻拍打以丢弃残留的培养上清；
- (3) 向每孔加入 300 μ l Buffer S1 (加有 RNase A)，用不干胶片封口，用涡旋仪重悬细菌。重悬必须充分，不应留有小的细菌块；
- (4) 向每孔加入 300 μ l Buffer S2，用新的不干胶封口，温和地上下翻转 8-10 次以混合均匀。该步骤不宜超过 5 min；
- (5) 向每孔加入 300 μ l Buffer NP，用新的不干胶封口，温和地上下翻转 8-10 次使其混合均匀，冰上放置 15 min；
- (6) 将 96 孔滤板放置在洁净的 96 孔深孔板上，用排枪从一侧伸入上一步骤的 96 孔深孔板中，吸取所有的溶液并转移到 96 孔滤板中；
- (7) 同时将 96 孔滤板和 96 孔深孔板放入离心机内，2000 x g 离心 5 min；
- (8) 丢弃 96 孔滤板，向 96 孔深孔板每孔加入 600 μ l 异丙醇，用新的硅胶片封口，上下翻转 8-10 次以便充分混匀；
- (9) 将 96 孔深孔板放入离心机中，3000 x g 离心 20 min。将 96 孔深孔板迅速倒置于废液槽中，弃尽上清，然后倒扣在滤纸上 2-3 min；

(10) 向每孔加入 600 μL 新配置的 70%乙醇，用硅胶片封口，上下翻转 4-5 次。3000 x g 离心 3 min。弃去上清，倒扣在洁净的滤纸上 2-3 min。在离心机低于 50 x g 短暂倒置离心；

(11) 将 96 孔深孔板放置在超净工作台风干 (约 10 min)；

(12) 向每孔中加入 20 μL ddH₂O，用新的硅胶片封口，轻轻旋转以便溶解 DNA，4°C放置，备用；

(13) 取 1 μL BAC 质粒溶解液，用微量分光光度计测定其浓度和纯度。

2.4.3 Illumina 测序文库构建

对检测合格的样品，首先使用核酸外切酶 (Epicentre Plasmid-Safe™ ATP-Dependent DNase) 去除大肠杆菌基因组 DNA 污染，然后采用超声波将 BAC 质粒破碎成 500 bp 左右的片段，随后对这些片段进行末端修复、3'末端添加碱基 'A'、5'端添加磷酸基团，再将 DNA 片段与 3'端带有碱基 'T' 的特殊接头连接，通过磁珠纯化的方法选择性回收连接接头后的 DNA 片段，最后再进行 PCR 扩增和产物纯化。具体步骤如下：

(1) 核酸外切酶酶解线性 DNA

在 96 孔 PCR 板配置酶解体系 (90 μL)：BAC 混合池 DNA (来自 4 mL 菌液)，10 X Buffer for Dnase (9 μL)，Dnase (1 μL)，ATP (3.6 μL)，余下用灭菌 ddH₂O 补足。

表 2-1 线性 DNA 酶解体系

Table 2-1 liner DNA digestion system

试剂名称	体积 (μL)
10X buffer	9
10 U/ μL Epicentre DNase	1
25 mmol/L ATP	3.6
ddH ₂ O	16.4
DNA	60
Total	90

体系混匀后，短暂离心，使所有组分收集到管底，随后将 PCR 板置于 PCR 仪中，37°C酶切 16 h，最后 70°C变性 30 min。

(2) 超声波破碎 DNA

分别将 PCR 板每孔中的体系转移到 1.5 mL AxyGen 平底 EP 管中, 再加入 10 μ L 10 X EDTA, 保证 EDTA 终浓度为 1 mmol/L, Tris 终浓度为 33 mmol/L (由上一步中的 buffer 提供)。用枪头轻轻吸打混匀, 冰上放置。利用 BioRuptor 超声波破碎仪将大片段 DNA 随机打断, 工作条件设置为: Level=HIGH, ON=30 sec, OFF=90 sec, Cycle=20, 水槽温度保持在 4°C。样品需对称放置, 每次最多破碎 6 个样品。将破碎完成的样品及时放置在 4°C 冰箱中。

(3) DNA 修复

将破碎的 DNA 转移到洁净的 AxyGen PCR 对应的孔中, 按表 2-2 配置反应体系。用 12 道移液枪轻轻吸打混均体系, 短暂离心使组分聚集到管底, 随后将 PCR 板置于 PCR 仪中, 反应程序设置为 12°C 反应 15 min, 37°C 反应 15 min, 然后 72°C, 20 min 终止反应。

表 2-2 DNA 修复体系

Table 2-2 DNA repair system

试剂名称	体积 (μ l)
10X Tango buffer	7
25 mmol/L dNTPs	0.28
100 mmol/L ATP	0.7
T4 DNA pol.	1.5
Taq DNA pol.	1
T4 Polynucleotide kinase	1
ddH ₂ O	8.52
Template DNA	50
Total	70

(4) 连接测序接头

在连接测序接头前, 需现制连接接头。通过将合成的两条 DNA 序列经过逐渐降温形成 12 bp 配对的 Y 字形接头。按表 2-3 配制接头 Mix, 混匀后, 置于 PCR 仪中, 设置程序为 95°C, 10 sec, 再以 0.1 °C/sec 的速度下降到 12°C。

表 2-3 测序接头制备体系

Table 2-3 Sequencing adapter preparation system

100 μ mol/L Adapter1	100 μ mol/L Adapter2	Oligo nucl. Buffer	ddH ₂ O
15 μ l	15 μ l	10 μ l	60 μ l

接头制备完成后，按表 2-4 配制连接体系，混匀后，短暂离心，随后将 PCR 板至于 PCR 仪中，20℃反应 15 min。反应结束后，及时取出，冰上放置。

表 2-4 接头连接体系

Table 2-4 Adapter ligation system

反应体系	体积 (μ l)
10X T4 DNA ligase buffer	2
T4 DNA ligase	1
PEG 4000 (50%)	9
Adapter Mix	2.5
ddH ₂ O	5.5
Template DNA	70
Total	90

(5) DNA 片段的选择性回收

使用康为世纪磁珠法 DNA 纯化回收试剂盒选择性回收连接产物中的 DNA，具体操作步骤如下：

- 1) 涡旋振荡 CMPure 30 sec，使其混合均匀；
- 2) 向 PCR 板连接反应液中加入 90 μ L 混合均匀的 CMPure，用移液器吸打混合均匀，然后室温静置 5 min；
- 3) 将 PCR 板放在磁力架上 5 min，使磁珠完全吸附到磁力架上，此时溶液澄清；
- 4) 用移液器吸去上清液，期间避免接触到结合目的 DNA 的磁珠；
- 5) 保持 PCR 板固定在磁力架上，向每孔中加入 250 μ L 新配置的 80%乙醇，静置 2 min，待悬浮的磁珠完全吸附后，用移液器吸去乙醇；
- 6) 重复步骤 5；
- 7) 用 10 μ L 枪头小心吸去 PCR 板每孔底部的积液，后室温放置 10 min，使乙醇完全挥发；

8) 将 PCR 板从磁力架上取下, 向每孔加入 20 μL ddH₂O, 涡旋震荡使磁珠完全悬浮于洗脱液中, 室温静置 5 min;

9) 短暂离心, 将 PCR 板放回到磁力架上 5 min, 待溶液澄清后, 将 20 μL 洗脱液转移到一个新的 AxyGen PCR 板中, 冰上放置。

(6) PCR 富集

通过 PCR 富集并对每个二级混合池文库标记不同的 7 bp index 标签。每个二级混合池构建成一个测序文库, 将 96 个二级混合池的测序文库混合上机测序。Index 设计遵循下列标准: (1) Index 序列碱基数组成相对均匀, 相同碱基不超过两个且尽量避免相同的碱基连续出现; (2) 不同的 index 序列之间至少有两个碱基的差异以免造成后期序列拆分错误。使用洁净的 AxyGen PCR 板按表 2-5 制备反应体系, 需要特别注意 DNA 样品和 Index Primer 按顺序排列, 以免造成数据混乱。体系制备完成后, 用排枪上下吸打混匀, 短暂离心, 然后将 PCR 板至于 PCR 仪中。反应程序设置为预变性 94°C, 2 min; 一个循环: 98°C 变性 10 sec, 60°C 退火 30 sec 和 72°C 延伸 30 sec, 12 个循环; 72°C 终末延伸 5 min; 最后 12°C。

表 2-5 PCR 富集反应体系

Table 2-5 PCR Amplification system

试剂	体积 (μl)
2X KOD buffer	25
2 mmol/L dNTPs	10
10 $\mu\text{mol/L}$ P5+ADA	1
10 $\mu\text{mol/L}$ Index Primer	1
1 U/ μl KOD DNA pol.	1
Template DNA	12
Total	50

(7) PCR 产物纯化

通过磁珠纯化 PCR 产物, 去除反应体系中的酶, 引物, dNTP 等杂质, 获得最终的混合池文库, 具体步骤如下:

1) 涡旋振荡 CMPure 30 sec, 使其混合均匀;

- 2) 向 PCR 板连接反应液中加入 50 μL 混合均匀的 CMPure，用移液器吸打混合均匀，然后室温静置 5 min；
- 3) 将 PCR 板放在磁力架上 5 min，使磁珠完全吸附到磁力架上，此时溶液澄清；
- 4) 用移液器吸去上清液，期间避免接触到结合目的 DNA 的磁珠；
- 5) 保持 PCR 板固定在磁力架上，向每孔中加入 250 μL 新配置的 80%乙醇，静置 2 min，待悬浮的磁珠完全吸附后，用移液器吸去乙醇；
- 6) 重复步骤 5；
- 7) 用 10 μL 枪头小心吸去 PCR 板每孔底部的积液，后室温放置 10 min，使乙醇完全挥发；
- 8) 将 PCR 板从磁力架上取下，向每孔加入 20 μL ddH₂O，涡旋震荡使磁珠完全悬浮于洗脱液中，室温静置 5 min；
- 9) 短暂离心，将 PCR 板放回到磁力架上 5 min，待溶液澄清后，将 20 μL 洗脱液转移到一个新的 AxyGen PCR 板中，4℃保存。

2.4.4 实时荧光定量 PCR 测定相对浓度

为了将含有不同 Index 的样品等量混合，需要对每个 DNA 文库进行精确定量。本实验使用 Real-time PCR 方法对样品进行相对定量。荧光定量采用 TOYOBO 的定量试剂盒 (QuantiFast® SYBR® Green PCR Kit)。反应体系见表 2-6：

表 2-6 实时荧光定量 PCR 反应体系

Table 2-6 The reaction system for Real-time quantitative PCR

试剂	体积 (μL)
2 x SYBR	6
10 $\mu\text{mol/L}$ HiQ-R	0.5
10 $\mu\text{mol/L}$ HiQ-L	0.5
ddH ₂ O	1
1/40000X Diluted DNA	4
Total	12

加样后上机检测。扩增程序包括 95°C 酶激活 5 min; 一个循环: 95°C 变性 5 sec, 60°C 退火 5 sec 和 70°C 延伸 30 sec, 40 个循环; 溶解曲线 65°C-95°C, 5 sec/step。DNA 的相对浓度以 $2^{-\Delta Ct}$ 的方法来计算, 将其中浓度最低的一组数据的浓度设为 1。

将具有不同 Index 的样品等量混合在一起, 用移液器吸打均匀, 送至公司检测和测序。

2.5 BAC 末端序列解析方法

本研究主要策略是构建二级混合矩阵, 分别对每个行列池建库并测序, 利用行池和列池纵横交叉性质, 对混合池 Illumina HiSeq 2000 测序序列进行分析, 解析出交叉点 BAC 末端序列, 通过与基因组数据比对, 尽可能准确地将 BAC 克隆定位到糜子基因组上。我们自主设计了数据分析流程(图 2-1), 具体如下:

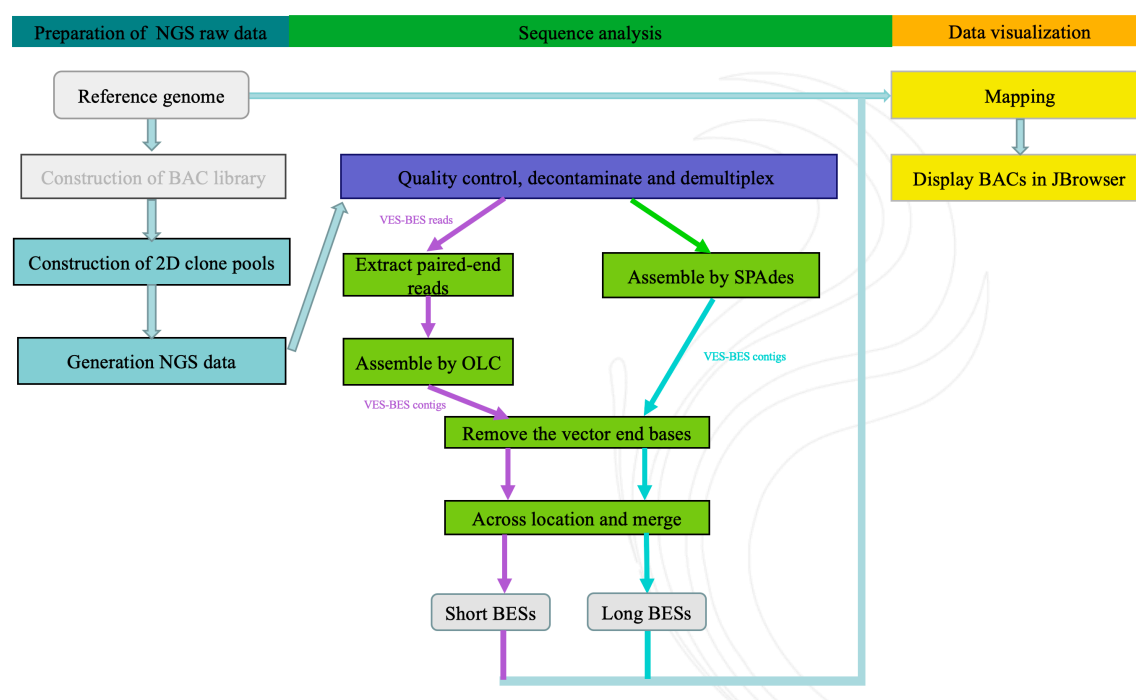


图 2-1 BAC 末端序列分析流程图

Fig. 2-1 The pipeline of BAC end sequence analysis

(1) 原始数据预处理

在 DNA 文库构建过程中, 我们采用超声波破碎 DNA, 使用康为世纪 CMPure 磁珠两次纯化 DNA。这些步骤将导致 DNA 插入片段长度的方差较大, 片段大小不集中。当插入片段长度小于单端测序长度时, 在测序序列 3' 端将包含部分接头序列。在进行后续分析前, 去除这些接头序列是必须的。此外, 测序数据质量受测序平台

的影响，使得测序碱基质量不同，在后续分析中应过滤掉质量较低的碱基序列。对原始测序数据用 Trimmomatic 去除测序接头、剪切末端低质量碱基和过滤平均质量低的 reads，设置参数为：LEADING:20 (截去 5'端 Q 值小于 20 的碱基)，TRAILING:20 (截去 3'端 Q 值小于 20 的碱基)，ILLUMINACLIP:adpater.fasta:2:18:7:2:true (adpater.fasta 文件中包含两条测序接头序列)，SLIDINGWINDOW:5:23，AVGQUAL:23 (过滤平均 Q 值小于 23 的序列)，MINLEN:70 (过滤长度小于 70 bp 的序列)。最后使用 FastQC 查看数据预处理效果。

BAC 质粒提取过程中，大肠杆菌基因组 DNA 与 BAC 难以完全分开，故在 Illumina 测序文库构建前，通过 Epicentre ATP-dependence DNase 水解线性 DNA 分子，能有效降低大肠杆菌基因组污染比例，但不能完全去除污染。使用 Bowtie2 将质量控制后的测序数据比对到大肠杆菌 DH10B 参考基因组上，通过 un-conc 参数保留未比对的有效混合池 BAC 序列，其他参数为默认值。

在构建每个二级混合池文库时，添加了不同的 7-bp index 序列，以便将多个混合池文库混合成一个文库，有利于节约测序成本。我们依据这些 index 信息，利用 BBDemux/demuxbyname.sh 将前面处理的数据拆分到各个混合池中，其中 prefixmode 设置为 f。

(2) VES-BES reads 筛选

BAC 克隆序列由 BAC 载体序列和插入片段序列组成。BAC 载体序列是已知的，而插入片段的序列是未知的。BAC 载体序列通过 *HindIII* 酶切位点与插入片段序列相连。如图 2-2 所示，将与 *HindIII* 相邻的那段载体序列称为载体末端序列 (Vector End Sequence, VES)，将与 VES 通过 *HindIII* 相邻的那段插入片段序列称为 BAC 末端序列 (BAC end Sequence, BES)。BAC 插入片段大小一般在 100 kb-200 kb，两端的序列均称为 BES。为了区分这两个 BAC 末端序列，通常将与 *LacZ* 基因上游相连的 BES 称为正向末端序列 (Forward BES)，将与 *lacZ* 下游相连的 BES 称为反向末端序列 (Reverse BES)。为了快速筛选出与 VES 重叠的 paired-end reads，首先使用载体末端 12 bp 作为种子，通过全局比对算法判断混合池的每条 reads 是否包含这一种子，其中 edit=1，如果种子在 reads 中出现，则进一步使用载体末端 200 bp 用半全局比对算法确认 paired-end reads 是否与末端重叠。筛选这些特殊的 reads 是整个分析流程中最耗时的部分。为了减少计算时间，在实际中将实现这一核心功能的

Python 代码 Cython 化并编译成 C 扩展库，并在 Python 中调用。通过 Cython 化可以将 Python 代码运算效率大幅提升。数据筛选通过 Python 脚本 `filter_reads_from_NGS.py` 实现。

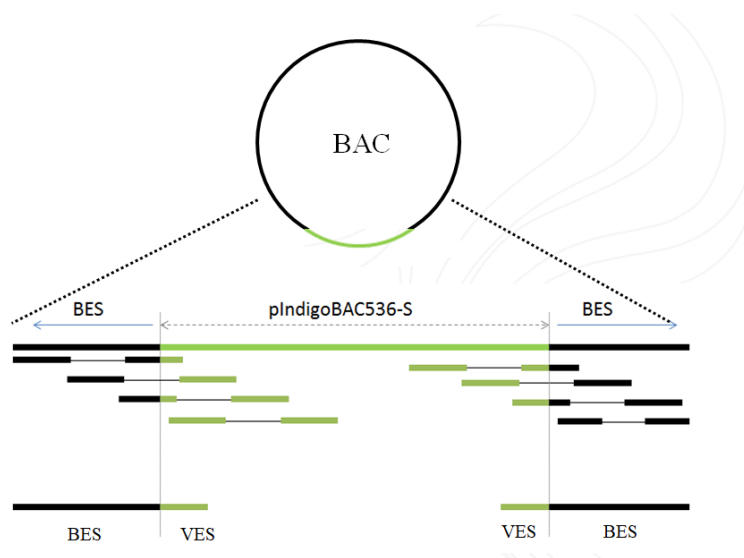


图 2-2 BAC 质粒示意图

Fig. 2-2 The schematic of BAC plasmid

(3) 序列组装与剪切

为了解析出更多更长的末端序列，我们通过两种途径来解析：

第一种途径是：首先从混合池中筛选出包含 VES 的 paired-end reads，为了延长 reads，采用基于 OLC 算法的 Cap3 对这些 reads 进行组装，获得 consensus。然后利用配对 reads 的信息，用 Scarpa 软件结合 Bowtie2 比对结果将 consensus 进一步组装成更长的 scaffold。最终获得的 scaffold 序列不仅包含 BES，还包含 VES，为了后续解析的方便，将 scaffold 序列中 VES 剪切掉，保留以 *HindIII* 酶切位点 ‘AAGCTT’ 开始的 BES。因受文库插入片段的限制，通过该途径解析到的末端序列相对较短，因此称为短的 BAC 末端序列，简称为短末端序列。

第二种途径是：先组装混合池测序数据，再提取末端序列。利用 SPAdes 依次组装每个二级混合池预处理后的数据，获得每个混合池 contigs 数据，kmer 设为 31、51、71、91。将组装结果与载体序列比对，找出与载体序列重叠的 contigs，经过截取获得需要的 BAC 末端序列，截取后的 BAC 末端序列均以 *HindIII* 酶切位点 ‘AAGCTT’ 开始，这有利于后续的 BAC 末端序列定孔。因通过这种途径获得的

BAC 末端序列远远长于第一种途径获得的，所以将这些序列称为长的 BAC 末端序列，简称为长末端序列，以便区分。

(4) BAC 末端序列交叉定孔

在前一步结果中，获得每个二级混合池中所有的 BES。利用 blast 分别将每一个行池和每一个列池中的 BES 逐一进行比对，比对参数 `perc_identity` 设为 99。提取满足比对限制条件的成对末端序列，这些成对序列则是位于该行池和列池交叉点上 BAC 的末端序列。但当实验构建的超级混合池较大，而物种基因组较小时，在超级混合池中必定存在两个具有相同末端序列的 BAC，若它们在不同的行列中，比对包含这 BAC 的两个行池和列池，将会在四个位点出现这两个 BAC 的 BES，其中有两个属于假阳性末端序列，为了剔除假阳性 BES，我们采用如下两种解决方法：第一种是，设定二级混合池内的 BES 之间不重叠。在构建的 96 X 96 超级混合池中，每个二级混合池均包含 96 个 BAC 克隆，这相对于糜子基因组来说覆盖度是极低的，这些 BAC 具有相同的 BES 的概率非常小，可以忽略不计。因此，我们逐一判断行列池交叉法比对结果中具有唯一 BES 的位点中的 BES 在行列池中的唯一性。第二种是，在初步的行池与列池两个集合交集的基础上，先分别计算交叉位点所对应行池与列池上的末端序列绝对补集，再对两个绝对补集求交集，最后求前一个交集与后一个交集的差集，获得最终的每个交叉点上的 BAC 末端序列(潘永龙 2014)。在末端解析过程中，短的 BAC 末端序列和长的 BAC 末端序列解析过程是独立完成的，正向与反向 BAC 末端序列解析过程也是独立完成的。利用 Cap3 将解析获得的来自行池与列池的成对序列合并成单条序列。

2.6 BAC 末端序列重复序列注释

为了评估末端序列中重复序列种类和含量，以及更方便与参考序列比对，我们对解析出的 BAC 末端序列使用从头注释重复元件的方法。首先以长末端序列为数据源使用 RepeatModeler 通过 NCBI blast 训练一个重复序列数据库，然后依据该重复序列数据库使用 RepeatMasker 分别注释短末端序列和长末端序列，将重复序列部分用小写字母表示，方便后续的比对。

2.7 BAC 克隆在陇糜 4 号基因组中的定位

为了确定 BAC 克隆在参考基因组上的具体位置，方便 BAC 文库的利用，首先，使用 blastn 分别将重复序列注释前后的短末端和长末端序列比对到陇糜 4 号基因组上。其中在比对短末端序列时，blastn 参数 qcov_hsp_perc 设为 99，perc_identity 设为 99，outfmt 设为 6。在比对长末端序列时，blastn 参数 culling_limit 设为 1，perc_identity 设为 99。在比对重复序列屏蔽的序列时，添加限制参数 lcase_masking。本研究使用的 BAC 文库，并不是用陇糜 4 号构建的，存在基因组结构差异，因此在长序列比对结果中，我们以长末端序列 5'端的比对结果为准。最后从比对结果中解析出 BAC 在参考基因组上的位置，解析规则如下：若 BAC 的正向与反向末端序列都被解析出且能比对到基因组上，则判断 BAC 的两个末端序列是否均比对到同一染色体上且方向朝内，两端距离在 250 kb 以内，提取满足条件的克隆；若 BAC 仅解析出单个末端序列且比对到参考基因组唯一的位置，也作为有效的定位。为了将更多的 BAC 克隆定位，我们结合上面的规则整合了短末端序列和长末端序列的比对结果。

2.8 基于 JBrowse 展示 BAC 信息

JBrowse 是由 GMOD (Generic Modal Organism Database) 开发的一个功能齐全的基因组浏览工具。它是基于 javascript 和 HTML 构建，能够很好的与浏览器兼容，方便内嵌到网站中。它可快速、平滑地滚动和缩放，能较快浏览基因组数据，支持 GFF3、BED、FASTA、Wiggle、BAM、VCF 等众多数据格式。在本研究中，我们将 JBrowse 配置在实验室工作站上，通过 Apache 服务，可以方便的在局域网内的电脑上查看数据。

为了在 JBrowse 中展示 BAC 末端定位信息，我们整理了 BAC 末端序列定位信息，将其转换成 GFF3 格式。在 GFF3 格式中，如果 BAC 是双端定位，则每个 BAC 记录由三行信息组成，第一行是 BAC 在基因组上的起点和终点位置，第二、三行分别是两端的 BAC 末端在基因组上的位置，它们通过最后一列的"Parent"和"ID"字段相关联(图 2-4)；如果 BAC 仅通过单端定位，则每个 BAC 记录由一行信息组成，这行信息是 BAC 单侧末端在基因组上的起点和终点位置。

6	GResource	BES	4968299	5109892	.	.	.	ID=25A01;Name=25A01
6	GResource	BES_end	4968299	5008417	.	+	.	Parent=25A01
6	GResource	BES_end	5109493	5109892	.	-	.	Parent=25A01

图 2-4 GFF3 格式的 BAC 位置注释信息

Fig. 2-4 The information of BAC location in GFF3 format file

每个 BAC 记录由三行信息组成，第一行是 BAC 在基因组上的起点和终点位置，第二、三行分别是两端的 BAC 末端在基因组上的位置，它们通过最后一列的 Parent 和 ID 字段相关联。

Each BAC record consists of three lines, the first line records the starting and ending positions of BAC on the genome, and the second and third lines record the positions of the two BAC end sequences on the genome, which are associated with the Parent and ID fields in the last column.

2.9 Sanger 法验证 BAC 末端序列

从糜子 BAC 文库 384 板中随机选取克隆进行双末端测序。碱裂解法提取 BAC 质粒。测序反应使用 BigDye v3.1 (Applied Biosystems) 12 μ L 体系(见表 2-1)。双末端测序反应引物分别为 BAC-F: 5'-aacgacggccagtgaattg-3' , BAC-R: 5'-gataacaatttcacacagg-3'。PCR 反应程序为: 96°C, 5 min; (96°C, 10 sec; 55 °C, 5 sec; 60°C, 3 min) x 75 cycles; 10°C, 5 min。反应完成后对产物进行纯化。1) 每个样品加入 100 μ L 无水乙醇: 2.5 mol/L 乙酸钠 (20: 1), 混匀后 4000 x g 离心 30 min; 2) 弃上清后加入 150 μ L 70%乙醇, 4000 x g 离心 15 min; 3) 重复(2)一次, 弃上清后风干 30 min。上样前加 8 μ L 甲酰胺 95°C 变性 5 min, 4°C 冷却 5 min。后在 ABI 3730 DNA analyzer 上进行毛细管电泳检测。测序结果用 Applied Biosystems Sequence Scanner v1.0 Software 提取, 并去掉低质量碱基。

表 2-1 BAC 末端测序反应体系

Table 2-1 Reaction system for BAC end sequencing

Total	12 μ L
5 \times BigDye Sequencing Buffer	4.3 μ L
BigDye	0.7 μ L
160 μ mol/L dNTP	0.5 μ L
2 μ mol/L Primer	1.5 μ L
BAC template	5 μ L

2.10 鉴定叶绿体和线粒体 BAC 克隆

高等植物基因组不仅包含细胞核基因组，还包含叶绿体和线粒体的基因组。糜子叶绿体基因组长139,929 bp (Cao et al 2017)，与BAC插入片段大小非常接近。本研究将解析出的BAC末端序列与糜子叶绿体基因组进行比对，以一致性大于99%为临界值，鉴定潜在的叶绿体BAC克隆。植物线粒体基因组大小一般在200 kb–750 kb。糜子线粒体基因组尚未报道。与糜子进化关系较近的高粱线粒体基因组长468,628 bp (Saski et al 2007, Paterson et al 2009)，玉米线粒体基因组长569,630 bp (Jiao et al 2017)。将解析出的BAC末端序列与NCBI/Genome数据中所有禾本科线粒体基因组序列 (NC_008331、NC_007982、NC_036024、NC_031164、NC_029816、NC_022714、NC_022666、NC_013816、NC_007886、NC_011033、NC_008362、NC_008360、NC_008332、NC_008333) 进行比对，以一致性99%为临界值，鉴定文库中潜在的线粒体BAC克隆。

3 结果和分析

3.1 二级混合池测序文库构建

为了高效率、低成本获取 BAC 克隆的末端序列，我们设计了一套基于克隆矩阵混合池鸟枪测序策略定位 BAC 克隆的分析流程。

为了验证该流程的可行性，我们选择了糜子 BAC 文库中 24 块 384 板，包含 9216 个克隆，覆盖糜子基因组的 1.28 倍。将这 24 块 384 板按 6 行 4 列排列构成一个的超级混合池 (图 3-1)。该超级混合池包含 96 个行池和 96 个列池，这些行池和列池均称为二级池，每一个二级池均包含 96 个 BAC 克隆。在这 24 块 384 板构成的超级混合池中，我们共构建了 192 个二级池。然后分别提取混合池 BAC DNA 并对其进行 Illumina 测序文库构建，建库过程中对不同的混合池添加不同的 index 序列，最后等量混合文库。我们将 96 个行池测序文库混合成测序文库 X，将 96 个列池测序文库混合成列池测序文库 Y。

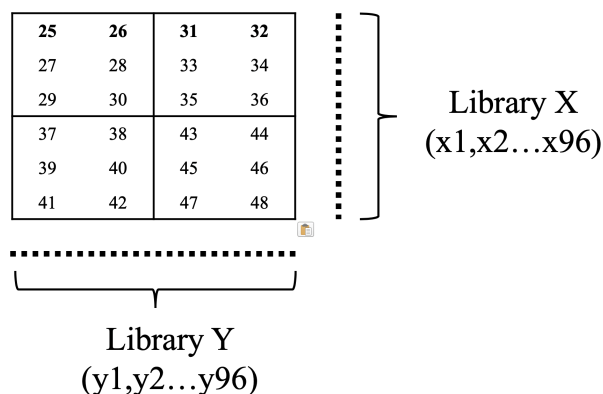


图 3-1 24 块 384 板在超级混合池中的布局

Fig. 3-1 The layout of 24 384-plates in the superpool

使用 Agilent 2100 生物分析仪分析两个文库片段长度范围以及质量。结果如图 3-2 所示，A、B 分别显示了测序文库 X 和 Y 的建库片段大小范围和浓度。测序文库 X 平均长度为 491 bp，DNA 浓度为 836 pg/μL，文库 Y 平均长度为 443 bp，浓度为 1370 pg/μL。文库峰单一，片段大小合适，符合上机测序要求。



表 3-5 糜子 BAC 末端序列与糜子基因组序列比对情况

Table 3-5 Anchoring the Broom millet clones onto the reference sequences through BESs

类别	短末端序列	长末端序列	长短整合
BAC end sequences			
Clones in superpool	9216	9216	-
Clones with successful BESs	9126	7725	-
with paired BESs	7790	3890	-
with single-end BESs	1336	3835	-
only forward	921	2189	-
only reverse	415	1646	-
Total successful BESs	17811 (96.63%)	12721(66.17%)	-
Alignment			
Aligned BESs with repeats unmasked	14862 (83.44%)	12120 (95.28%)	-
Single hit BESs	12971	12120	-
Multi-hit BESs	1891	-	-
Aligned BESs with repeats masked	7760	10626	-
Single hit BESs	7295	10626	-
Multi-hit BESs	465	-	-
Anchoring to reference sequences			
Clones anchored to single sites	5795	6973	8262
Clones anchored with single BES	2507	3907	2871
Clones anchored with paired BESs	3288	3066	5391

表 3-6 BAC 在糜子参考基因组染色体上的定位情况

Table 3-6 The location result of BAC on broomcorn millet each chromosome

染色体	长度(Mb)	gap 数	BAC 数目	覆盖长度(Mb)	覆盖 gap 数
chr 1	69.18	85	439	35.25	34
chr 2	61.15	59	379	31.04	19
chr 3	57.97	50	394	31.03	20
chr 4	56.29	34	359	27.90	14
chr 5	54.13	52	361	29.86	19
chr 6	52.84	46	358	28.24	16
chr 7	51.23	67	286	23.67	27
chr 8	48.26	29	339	26.49	8
chr 9	45.11	70	240	20.96	21
chr10	44.65	53	308	25.44	28
chr11	43.18	30	267	21.90	14
chr12	42.47	30	254	20.79	11
chr13	40.72	50	261	21.11	17
chr14	38.49	32	269	20.23	10
chr15	34.36	34	213	18.14	8
chr16	33.61	45	212	16.72	15
chr17	32.99	25	226	18.26	10
chr18	32.24	38	199	15.32	17
unplaced	9.48	0	27	-	0
Total	848.47	829	5391	432.47	308

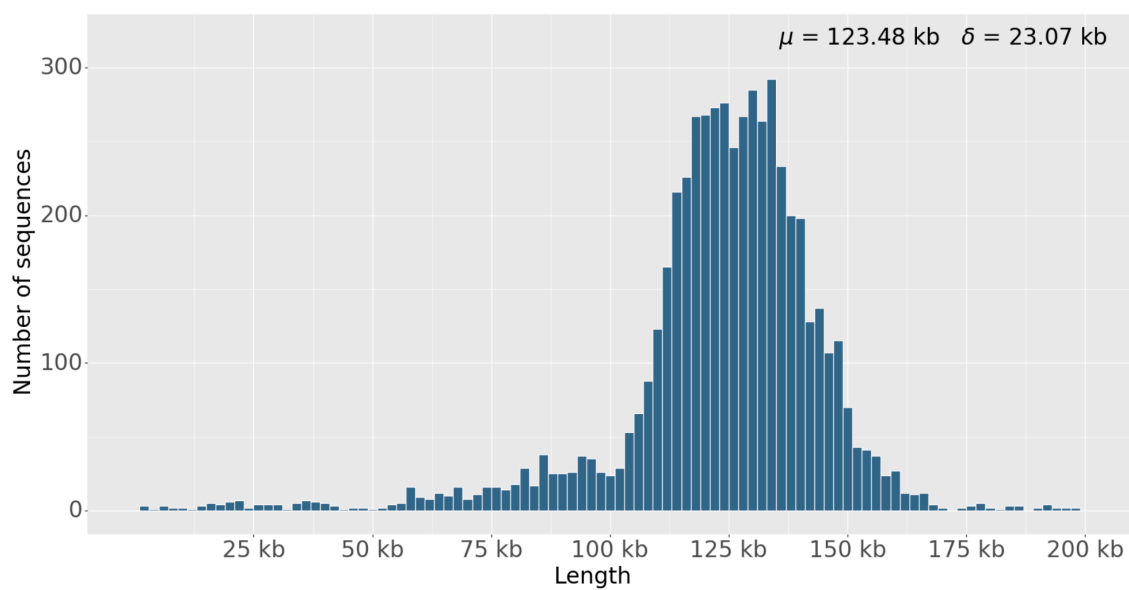


图 3-9 双末端定位的 BAC 长度分布统计

Fig. 3-9 The statistic of located BAC lengths by double BESs



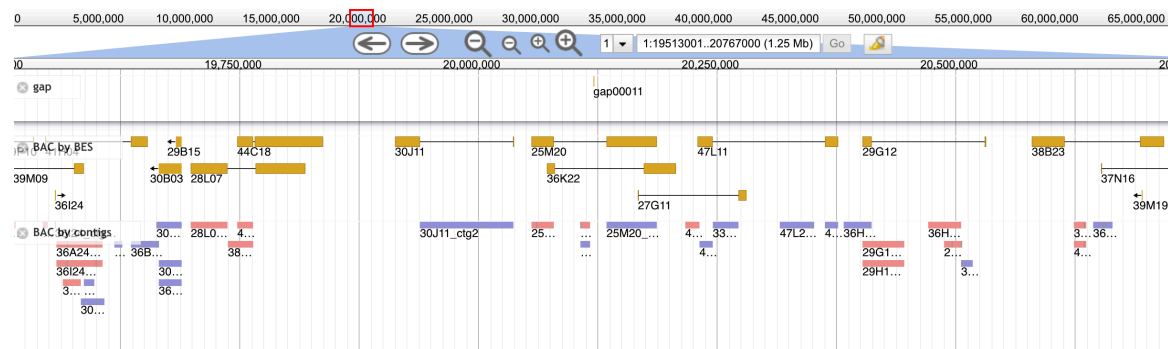


图 3-10 糜子 BAC 位置在 JBrowse 中的展示结果

Fig 3-10 The presentation of broomcorn millet BAC location in JBrowse

图为 1 号染色体局部放大后的效果。展示界面由三栏组成，从上至下分别是 gap，末端定位的 BAC 和 contigs。

The figure shows partial of chromosome 1. The display consists of three parts, including gap, BAC by BESs, contigs respectively.

参考文献

1. 程远芳. 基于BAC文库指纹特征的油菜物理图谱构建及其测序组装. [硕士学位论文]. 武汉: 华中农业大学图书馆, 2018
2. 窦怀乾, 李仰平, 吕佳, 窦锦壮, 李语丽, 王师. 基于WGP物理作图法的虾夷扇贝 fosmid克隆混池策略及解码率研究. 中国海洋大学学报, 2019, 42(2):44–51
3. 郭刚刚, 董国清, 周进, 达瓦顿珠, 袁兴淼, 张京. 大麦BAC文库三维混合池的构建与HvGW2筛选. 中国农科学, 2013, 46(1):9–17
4. 潘永龙. 两个水稻参考基因组序列组装质量的评估和基于混合池进行全基因组测序的方法研究. [博士学位论文]. 武汉: 华中农业大学图书馆, 2014
5. 谢明树, 孙建, 陈松林, 邵长伟. 半滑舌鳎含sox9基因BAC克隆的筛选及BAC-FISH定位. 渔业科学进展, 2014, 35(1):34–40
6. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, George RA, Lewis SE, Richards S, Ashburner M, Henderson SN, Sutton GG, Wortman JR, Yandell MD, Zhang Q *et al.* The genome sequence of *Drosophila melanogaster*. *Science*, 2000, 287(5461):2185–2195
7. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*, 1990, 215(3):403–410
8. Andrews S. FastQC: a quality control tool for high throughput sequence data. *Bioinformatics*, 2010, 33(19):3137–3149
9. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin A V., Sirotkin A V., Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*, 2012, 19(5):455–477
10. Begum D, Fiandt M, Jendrisak J. An improved approach for construction and analysis of BAC libraries using the CopyControl™ pCC1BAC vector containing an inducible high-copy origin. *Epicentre*, 2002, 9(2):9–11
11. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 2014, 30(15):2114–2120
12. Brown T. *Genomes*. 2nd edition. Oxford: Wiley-Liss, 2002
13. Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO, Shendure J. Chromosome-

- scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat Biotechnol*, 2013, 31(12):1119
14. Bushnell B. BBMap: a fast, accurate, splice-aware aligner. *Joint Genome Institute*, 2014
 15. Cai WW, Chen R, Gibbs RA, Bradley A. A clone-array pooled shotgun strategy for sequencing large genomes. *Genome Res*, 2001, 11(10):1619–1623
 16. Cao X, Wang J, Wang H, Liu S, Chen L, Tian X, Qin H, Wang L, Na X, Qiao Z. The complete chloroplast genome of *Panicum miliaceum*. *Mitochondrial DNA Part B Resour*, 2017, 2(1):43–45
 17. Craig Venter J. A life decoded. *Penguin Books*, 2007
 18. Csuros M, Li B, Milosavljevic A. Clone-array pooled shotgun mapping and sequencing: design and analysis of experiments. *Genome Inf*, 2003, 14:186–195
 19. Donmez N, Brudno M. SCARPA: Scaffolding reads with practical algorithms. *Bioinformatics*, 2013, 29(4):428–434
 20. Du B, Zhang X, Fang S, Wang D. Construction and characterization of a lipotes vexillifer genomic DNA BAC library. *Zoolog Sci*, 2007, 24(4):401–427
 21. Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, Scherer S, Scott G, Steffen D, Worley KC, Burch PE, Okwuonu G, Hines S, Lewis L, Deramo C, Delgado O, Dugan-Rocha S, Miner G, Morgan M, Hawes A *et al*. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature*, 2004, 428(6982):493–520
 22. Gilbert W, Maxam A. The nucleotide sequence of the lac operator. *Proc Natl Acad Sci*, 1973, 70(12):3581–3604
 23. Gualberto JM, Newton KJ. Plant mitochondrial genomes: dynamics and mechanisms of mutation. *Annu Rev Plant Biol*, 2017, 68(1):225–252
 24. van Heesch S, Kloosterman WP, Lansu N, Ruzius FP, Levandowsky E, Lee CC, Zhou S, Goldstein S, Schwartz DC, Harkins TT, Guryev V, Cuppen E. Improving mammalian genome scaffolding using large insert mate-pair next-generation sequencing. *BMC Genomics*, 2013, 14(1):1–11
 25. Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, Campbell MS, Stein JC, Wei X, Chin CS, Guill K, Regulski M, Kumari S, Olson A, Gent J, Schneider KL, Wolfgruber TK, May MR, Springer NM *et al*. Improved maize reference genome with single-molecule technologies. *Nature*, 2017, 546(7659):524

26. Kaplan N, Dekker J. High-throughput genome scaffolding from in vivo DNA interaction frequency. *Nat Biotechnol*, 2013, 31(12):1143
27. Kaul S, Koo HL, Jenkins J, Rizzo M, Rooney T, Tallon LJ, Feldblyum T, Nierman W, Benito MI, Lin X, Town CD, Venter JC, Fraser CM, Tabata S, Nakamura Y, Kaneko T, Sato S, Asamizu E, Kato T *et al*. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, 2000, 408(6814):796–815
28. Kent WJ. BLAT - The BLAST-like alignment tool. *Genome Res*, 2002, 12(4):656–664
29. Lander ES. Initial impact of the sequencing of the human genome. *Nature*, 2011, 470(7333):187–197
30. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, Fitzhugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, Levine R, McEwan P *et al*. Initial sequencing and analysis of the human genome. *Nature*, 2001, 409(6822):860–921
31. Langmead. Bowtie2. *Nat Methods*, 2013
32. Lapitan NLV, Brown SE, Kennard W, Stephens JL, Knudson DL. FISH physical mapping with barley BAC clones. *Plant J*, 1997, 11(1):149–156
33. Lee H-R, Eom E-M, Lim Y-P, Bang J-W, Lee D-H. Construction of a garlic BAC library and chromosomal assignment of BAC clones using the FISH technique. *Genome*, 2003, 46(3):514–520
34. Leggett RM, Clark MD. A world of opportunities with nanopore sequencing. *J Exp Bot*, 2017, 68(20):5419–5429
35. Li H. Minimap and miniasm: Fast mapping and de novo assembly for noisy long sequences. *Bioinformatics*, 2016, 32(14):2103–2110
36. Ling HQ, Ma B, Shi X, Liu H, Dong L, Sun H, Cao Y, Gao Q, Zheng S, Li Y, Yu Y, Du H, Qi M, Li Y, Lu H, Yu H, Cui Y, Wang N, Chen C *et al*. Genome sequence of the progenitor of wheat A subgenome *Triticum urartu*. *Nature*, 2018, 557(7705):424
37. Liu W, Zhao Y, Liu Z, Zhang Y, Lian Z, Li N. Construction of a 7-fold BAC library and cytogenetic mapping of 10 genes in the giant panda (*Ailuropoda melanoleuca*). *BMC Genomics*, 2006, 7:294
38. Lu H, Giordano F, Ning Z. Oxford Nanopore MinION sequencing and genome assembly. *Genomics, Proteomics Bioinforma*, 2016, 14(5):265–279

39. Marie-Nelly H, Marbouty M, Cournac A, Flot JF, Liti G, Parodi DP, Syan S, Guillén N, Margeot A, Zimmer C, Koszul R. High-quality genome (re)assembly using chromosomal contact data. *Nat Commun*, 2014, 5:5695
40. Maxam AM, Gilbert W. A new method for sequencing DNA. *Proc Natl Acad Sci*, 1977, 74(2):560–574
41. Milosavljevic A, Harris RA, Sodergren EJ, Jackson AR, Kalafus KJ, Hodgson A, Cree A, Dai W, Csuros M, Zhu B, de Jong PJ, Weinstock GM, Gibbs RA. Pooled genomic indexing of rhesus macaque. *Genome Res*, 2005, 15(2):292–301
42. Myers G. Whole-genome DNA sequencing. *Comput Sci Eng*, 1999, 1(3):33–43
43. Oeveren J, Ruiter M, Jesse T, Poel H, Tang J, Yalcin F, Janssen A, Volpin H, Stormo KE, Bogden R, Eijk MJT, Prins M. Sequence-based physical mapping of complex genomes by whole genome profiling. *Genome Res*, 2011, 21(4):618–625
44. Pan Y, Wang X, Liu L, Wang H, Luo M. Whole genome mapping with feature sets from high-throughput sequencing data. *PLoS One*, 2016, 11(9):1–17
45. Park MH, Lee HJ, Bok J, Kim CH, Hong ST, Park C, Kimm KC, Oh BS, Lee JY. Korean BAC library construction and characterization of *HLA-DRA*, *HLA-DRB3*. *BMB Rep*, 2011, 39(4):418–425
46. Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U *et al*. The *Sorghum bicolor* genome and the diversification of grasses. *Nature*, 2009, 457(7229):551–6
47. Quinlan AR. BEDTools: The Swiss-Army tool for genome feature analysis. *Curr Protoc Bioinforma*, 2014, 47:11–22
48. Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 2010, 26(6):841–852
49. Rhoads A, Au KF. PacBio sequencing and its applications. *Genomics, Proteomics Bioinforma*, 2015, 13(5):278–289
50. Romanov MN, Koriabine M, Nefedov M, de Jong PJ, Ryder OA. Construction of a California condor BAC library and first-generation chicken-condor comparative physical map as an endangered species conservation genomics resource. *Genomics*, 2006, 88(6):711–738
51. Sabehi G, Béjà O. Preparation of BAC libraries from marine microbial populations. *Methods in Enzymology*. 2013, 111–122

52. Sam T, Working S. The sequence alignment and map format specification. *Bioinformatics*, 2014, 1(1):1–16
53. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci*, 1977, 74(12):5463–5487
54. Saski C, Lee SB, Fjellheim S, Guda C, Jansen RK, Luo H, Tomkins J, Rognli OA, Daniell H, Clarke JL. Complete chloroplast genome sequences of *Hordeum vulgare*, *Sorghum bicolor* and *Agrostis stolonifera*, and comparative analyses with other grass genomes. *Theor Appl Genet*, 2007, 115(4):571–590
55. Shen W, Le S, Li Y, Hu F. SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One*, 2016, 11(10):1–16
56. Shendure J, Balasubramanian S, Church GM, Gilbert W, Rogers J, Schloss JA, Waterston RH. DNA sequencing at 40: Past, present and future. *Nature*, 2017, 550(7676):345–353
57. Shi J, Ma X, Zhang J, Zhou Y, Liu M, Huang L, Sun S, Zhang X, Gao X, Zhan W, Li P, Wang L, Lu P, Zhao H, Song W, Lai J. Chromosome conformation capture resolved near complete genome assembly of broomcorn millet. *Nat Commun*, 2019, 10:1–9
58. Skinner ME, Uzilov A V., Stein LD, Mungall CJ, Holmes IH. JBrowse: A next-generation genome browser. *Genome Res*, 2009, 19(9):1630–1648
59. Smit A, Hubley R, Green P. RepeatMasker Open-3.0. *RepeatMasker*, 1996
60. Song X, Goicoechea JL, Ammiraju JSS, Luo M, He R, Lin J, Lee SJ, Sisneros N, Watts T, Kudrna DA, Golser W, Ashley E, Collura K, Braidotti M, Yu Y, Matzkin LM, McAllister BF, Markow TA, Wing RA. The 19 genomes of *Drosophila*: A BAC library resource for genus-wide and genome-scale comparative evolutionary research. *Genetics*, 2011, 187(4):1023–1037
61. Tian X. The rice mitochondrial genomes and their variations. *PLANT Physiol*, 2005, 140(2):401–410
62. Wallis JW, Aerts J, Groenen MAM, Crooljman RPMA, Layman D, Graves TA, Scheer DE, Kremitzki C, Fedele MJ, Mudd NK, Cardenas M, Higginbotham J, Carter J, McGrane R, Gaige T, Mead K, Walker J, Albracht D, Davito J *et al.* A physical map of the chicken genome. *Nature*, 2004, 432(7018):761–774
63. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, Antonarakis SE, Attwood J, Baertsch R,

- Bailey J, Barlow K, Beck S, Berry E, Birren B, Bloom T *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature*, 2002, 420(6915):520–562
64. Wei X, Xu Z, Wang G, Hou J, Ma X, Liu H, Liu J, Chen B, Luo M, Xie B, Li R, Ruan J, Liu X. PBACode: A random-barcode-based high-throughput approach for BAC paired-end sequencing and physical clone mapping. *Nucleic Acids Res*, 2017, 45(7):1–17
65. Wu R, Kaiser AD. Structure and base sequence in the cohesive ends of bacteriophage lambda DNA. *J Mol Biol*, 1968, 35(3):523
66. Xu HL, Qian YP, Nie WH, Chi JX, Yang FT, Su B. Construction, characterization and chromosomal mapping of bacterial artificial chromosome (BAC) library of Yunnan snub-nosed monkey (*Rhinopithecus bieti*). *Chromosom Res*, 2004, 12(3):251–262
67. Yang X, Wan Z, Perry L, Lu H, Wang Q, Zhao C, Li J, Xie F, Yu J, Cui T, Wang T, Li M, Ge Q. Early millet use in northern China. *Proc Natl Acad Sci*, 2012, 109(10):3726–3730
68. Zeng CJ, Pan HJ, Gong S Bin, Yu JQ, Wan QH, Fang SG. Giant panda BAC library construction and assembly of a 650-kb contig spanning major histocompatibility complex class II region. *BMC Genomics*, 2007, 8:1–15
69. Zhang J, Chen L-L, Xing F, Kudrna DA, Yao W, Copetti D, Mu T, Li W, Song J-M, Xie W, Lee S, Talag J, Shao L, An Y, Zhang C-L, Ouyang Y, Sun S, Jiao W-B, Lv F *et al.* Extensive sequence divergence between the reference genomes of two elite indica rice varieties Zhenshan 97 and Minghui 63. *Proc Natl Acad Sci*, 2016, 113(35):5163–5171
70. Zhao G, Zou C, Li K, Wang K, Li T, Gao L, Zhang X, Wang H, Yang Z, Liu X, Jiang W, Mao L, Kong X, Jiao Y, Jia J. The *Aegilops tauschii* genome reveals multiple impacts of transposons. *Nat Plants*, 2017, 3(12):946–955
71. Zhao R, Miao H, Song W, Chen C, Zhang H. Identification of sesame (*Sesamum indicum* L.) chromosomes using the BAC-FISH system. *Plant Biol*, 2018, 20(1):85–92
72. Zou C, Li L, Miki D, Li D, Tang Q, Xiao L, Rajput S, Deng P, Peng L, Jia W, Huang R, Zhang M, Sun Y, Hu J, Fu X, Schnable PS, Chang Y, Li F, Zhang H *et al.* The genome of broomcorn millet. *Nat Commun*, 2019, 10(1):1–12

附录 1 试剂配方

1.1 LB 液体培养基 (1L)

胰化蛋白胨	10 g
酵母提取物	5 g
NaCl	10g

若配置固体培养基，则每 1L 液体培养基中加入 15g Agar。121℃灭菌 20min。

1.2 冰冻培养基

配制每升冰冻培养基，应在 800 mL 去离子水中加入

胰化蛋白胨	10 g
酵母提取物	5 g
NaCl	10 g
K ₂ HPO ₄	8.215 g
KH ₂ PO ₄ ·3H ₂ O	1.795 g
柠檬酸钠	1.795 g
(NH ₄) ₂ SO ₄	0.9 g
甘油	44 mL

摇动容器直至溶质溶解。用去离子水定容至 1 L。在 15 psi (1.05 kg/cm²) 高压下蒸汽灭菌 20 min。使用之前，在冰冻培养基中加入无菌 MgCl₂ 溶液，使得 Mg²⁺终浓度为 0.4 mmol/L。

1.3 碱裂解 S1, S2, S3 溶液 (制备质粒)

碱裂解 P1 溶液 (制备质粒)

10 mmol/L EDTA (pH8.0)

S1 溶液可以一次配制 1 L，在 15 psi (1.05 kg/cm²) 高压下蒸汽灭菌 15 min，保存于 4℃。

碱裂解 S2 溶液 (制备质粒)

0.2 mmol/L NaOH

1% (m/V) SDS

S2 溶液要现用现配制，室温下使用。

碱裂解 S3 溶液 (制备质粒)

5 mol/L 乙酸钾	60.0 mL
-------------	---------

冰乙酸	11.5 mL
-----	---------

H ₂ O	28.5 mL
------------------	---------

所配制的溶液中钾的浓度为 3 mol/L，乙酸根的浓度为 5 mol/L。保存于 4℃，用时置于冰浴中。

附录 2 BAC 质粒制备 (用于 BAC 末端测序)

- (1) 2 mL 离心管收集菌体, 离心机 12000 g, 5 min (倒掉上清, 瞬时离心, 用枪吸净培养基)。
- (2) 依次加 S1, S2, S3 溶液 (加完 S1 溶液后, 需要用涡旋震荡仪充分悬浮, 并且加 P2 和 S3 之间勿超过 5 min, 同时需要摇晃 6-8 次)。
- (3) 放置-20°C, 10 min, 离心 4°C, 15000 g, 15 min。
- (4) 取上清 700 μ L, 加 24:1 (氯仿: 异戊醇) 按照 1:1 加量, 通过摇床转速 150 r/min, 时间为 20 min。
- (5) 离心 4°C, 12000 g, 15 min, 吸取上清液 600 μ L 到新的 1.5 mL EP 管内。
- (6) 加入异丙醇 420 μ L (体积是上清液体 0.7 倍), 加完后摇晃 6-8 次。
- (7) 离心 15°C, 12000 g, 15 min, 倒掉上清液体。
- (8) 加入 70%无水乙醇 500 μ L, 15°C, 离心 12000 g, 5 min, 倒上清液体。
- (9) 瞬时离心, 用枪头吸净酒精, 通风橱吹干 15-20 min, 加 15 μ L 灭菌水, 4°C 保存。

附录 3 二级混合池统计

Pool	reads	Total_bases	coverage	Assembl contigs					
				Num	Max (bp)	N50	LN50 (bp)	N90	LN90 (bp)
X01	1,793,860	507,200,587	40	11074	105206	212	22153	2372	372
X02	2,296,911	649,889,916	51	10899	200505	204	22465	2219	368
X03	2,655,766	751,641,400	59	14484	167059	208	21533	4175	268
X04	5,007,381	1,415,799,899	112	18919	147106	197	22389	7747	256
X05	2,226,696	629,306,620	50	12230	130874	218	20062	3133	308
X06	1,870,970	528,491,316	42	19490	102290	498	7615	7669	251
X07	2,537,085	717,665,596	57	15937	134187	215	20698	5337	262
X08	1,459,631	413,377,992	33	16623	111000	470	7695	6332	264
X09	1,285,582	363,593,353	29	22933	77459	1125	2944	11918	251
X10	2,189,416	621,289,133	49	22487	94821	576	5283	9703	245
X11	1,761,138	498,127,492	39	20927	112364	637	6052	9382	254
X12	1,259,092	355,865,314	28	9515	96613	499	5651	4195	383
X13	1,666,337	470,440,570	37	19099	103975	542	7758	7500	253
X14	2,078,801	588,689,742	46	15359	119971	270	15745	4661	267
X15	1,847,720	522,855,840	41	12287	123179	234	18792	3142	326
X16	1,817,716	514,348,817	41	10314	130796	184	23431	1931	457
X17	2,142,634	606,264,745	48	5463	135033	164	25231	1062	1329
X18	2,128,517	600,646,252	47	13770	115262	362	11535	3823	333
X19	2,030,802	575,113,253	45	11215	134737	254	15385	2752	376
X20	1,073,716	303,698,436	24	16984	90238	763	4674	6980	267
X21	1,444,452	407,754,027	32	23129	74288	1237	2885	11428	252
X22	1,892,852	534,581,432	42	12223	134860	200	22668	2953	322
X23	1,815,702	513,798,898	41	19694	117771	494	6906	8044	258
X24	2,671,571	759,695,799	60	16763	73697	1537	2104	9378	276
X25	3,937,544	1,115,006,973	88	14361	139961	165	30031	4358	268
X26	2,513,414	712,519,367	56	12054	127757	197	22761	2942	327
X27	1,797,745	508,396,214	40	17851	125594	386	9891	6235	264
X28	2,142,977	608,307,065	48	10222	134438	250	15193	2528	445
X29	1,538,986	434,727,477	34	16684	114324	329	10881	5835	262
X30	2,517,321	712,432,290	56	12327	177347	152	30111	2956	296
X31	1,839,233	516,927,241	41	16577	126699	313	13779	5165	265
X32	1,200,095	337,803,965	27	17685	118132	654	5889	7488	264
X33	1,619,581	455,623,333	36	15610	120235	237	16968	5504	263
X34	1,291,194	363,145,129	29	13385	110576	332	11093	4445	281
X35	1,549,299	435,840,882	34	14715	177347	184	23736	4289	269
X36	1,947,404	549,077,776	43	16587	180505	188	22755	5628	264
X37	2,192,085	620,999,923	49	12574	131918	239	18209	3181	324
X38	2,614,604	741,158,253	58	6449	186675	196	18770	1580	1049

X39	2,062,900	584,238,034	46	18148	126245	404	7909	6436	259
X40	1,223,586	346,358,608	27	9173	86237	637	4772	4333	457
X41	1,468,270	414,510,470	33	13452	133439	289	11186	4236	327
X42	2,372,729	670,115,415	53	14251	141886	290	13399	4668	296
X43	2,391,132	676,757,230	53	13398	100008	233	20155	3550	321
X44	1,306,965	369,692,570	29	16533	88057	835	4056	7238	271
X45	1,198,634	338,387,362	27	18594	83551	801	3010	9371	259
X46	1,251,567	354,476,752	28	17126	146121	713	5210	7187	268
X47	1,443,491	408,396,651	32	17173	77883	943	3608	8269	270
X48	2,109,150	598,034,088	47	13595	190689	256	17017	3859	293
X49	1,166,303	330,272,628	26	16115	101627	896	3833	7006	287
X50	2,693,610	762,728,989	60	15299	151220	280	11872	4685	290
X51	3,879,737	1,095,640,173	86	16872	150220	169	28117	5768	264
X52	2,274,745	643,300,205	51	12556	140304	249	16724	3149	321
X53	3,549,281	1,003,132,228	79	12311	146407	148	33288	3212	298
X54	2,702,378	764,506,762	60	16531	137221	278	12128	6153	270
X55	1,909,876	540,275,804	43	22249	146377	585	5481	9997	247
X56	1,737,784	492,314,931	39	20092	113074	532	6128	8321	254
X57	1,798,679	508,984,840	40	21522	106708	595	6110	9298	251
X58	2,202,163	622,361,415	49	15879	125982	238	20348	5672	266
X59	1,885,606	533,746,202	42	17260	119805	400	8992	7140	261
X60	1,959,371	553,266,393	44	18664	119006	343	11186	7032	258
X61	1,650,762	467,617,939	37	12030	140543	270	14489	3095	360
X62	2,138,625	606,064,703	48	16611	97179	346	9212	5658	268
X63	1,770,067	500,868,884	40	10611	115583	229	18920	2376	402
X64	1,872,535	529,952,662	42	13317	97683	234	19134	3400	297
X65	2,179,543	614,939,401	49	9723	149245	165	27346	1754	412
X66	2,122,312	600,061,661	47	12742	116033	237	19137	3055	319
X67	1,586,394	448,555,330	35	14693	136775	273	12377	4994	292
X68	1,630,964	460,841,766	36	17370	118781	278	12034	6060	262
X69	1,793,903	507,970,284	40	12172	140231	281	11737	4023	367
X70	2,458,197	696,145,926	55	12389	143042	216	17472	3667	314
X71	2,243,696	635,126,169	50	16562	131854	280	12466	6106	273
X72	2,011,424	567,960,654	45	11494	141353	188	22392	2969	323
X73	2,060,920	584,060,714	46	21402	135770	496	6239	9080	250
X74	1,902,345	539,003,924	43	14177	119232	291	11708	4232	317
X75	2,682,828	760,705,653	60	11122	139559	188	23716	2234	388
X76	2,621,000	742,367,449	59	12677	117619	275	12622	3679	346
X77	2,866,386	812,123,628	64	12072	142072	201	20145	2872	339
X78	1,997,761	566,592,820	45	11288	133108	264	11551	3206	400
X79	1,221,719	346,196,763	27	7458	102865	319	10763	2646	668
X80	1,454,760	412,287,864	33	19288	111314	516	5484	8064	262
X81	1,427,335	404,713,771	32	17197	121071	430	5993	6834	265

基于 CAPSS 的糜子 BAC 克隆在参考基因组上定位

X82	2,185,715	620,585,861	49	11773	137270	220	12893	3915	328
X83	2,021,355	573,278,001	45	10147	126526	208	16323	2626	361
X84	2,563,829	727,239,407	57	14830	137905	249	14127	5666	271
X85	1,125,481	318,110,352	25	6126	138792	276	11606	1868	1049
X86	1,343,719	380,235,336	30	12730	127317	257	11996	4131	309
X87	2,428,921	685,921,374	54	7533	120096	190	20911	1651	433
X88	2,967,078	840,193,635	66	14079	147047	204	20804	4773	270
X89	1,709,161	483,482,267	38	12191	166625	314	13204	3696	322
X90	2,764,403	782,017,837	62	15557	127161	168	25046	5776	264
X91	2,995,541	847,593,349	67	16297	124593	245	16810	5759	262
X92	1,582,815	447,837,333	35	16965	114151	412	6155	7351	265
X93	1,575,175	442,857,191	35	12960	118345	238	16324	4214	283
X94	1,208,282	342,538,374	27	12193	115844	474	6885	4333	394
X95	2,205,471	624,149,754	49	15990	137168	286	9933	6451	266
X96	1,439,584	407,663,795	32	17859	141854	492	6714	7117	262
Y01	4,143,729	1,194,088,526	94	20418	174676	280	8290	11163	260
Y02	4,142,056	1,194,381,844	94	20297	149911	214	14674	11530	256
Y03	2,814,285	811,975,788	64	22142	147157	320	7415	13447	262
Y04	3,660,226	1,056,378,139	83	24349	136276	303	8083	15442	261
Y05	2,350,325	678,012,946	54	19927	124004	241	11877	11517	263
Y06	4,115,671	1,187,072,392	94	31850	131725	742	1889	20705	258
Y07	2,464,498	710,971,285	56	20046	122060	255	12261	11070	257
Y08	2,506,163	723,828,090	57	23396	143060	276	10717	14037	258
Y09	2,025,922	584,736,916	46	15121	191488	186	17025	7420	268
Y10	4,214,770	1,217,251,278	96	18306	157769	206	16108	9721	262
Y11	2,002,523	577,676,810	46	12910	114048	166	18534	5893	265
Y12	2,673,491	771,950,880	61	15546	106892	206	15587	7727	261
Y13	2,524,234	727,700,500	57	17003	120697	231	13862	9059	270
Y14	4,304,899	1,241,999,389	98	26347	152080	362	6787	16516	261
Y15	3,616,210	1,042,559,150	82	35187	140410	1404	1016	24334	259
Y16	2,221,250	640,035,947	51	16707	132812	251	12132	8525	273
Y17	2,150,836	620,313,769	49	19001	127292	274	10531	10643	264
Y18	2,784,059	802,549,466	63	27492	151097	368	6643	17894	261
Y19	2,006,218	579,505,890	46	14400	134479	180	16628	7240	266
Y20	1,975,572	569,609,773	45	17276	137827	193	17712	8368	258
Y21	1,700,183	490,604,598	39	14133	132055	215	13555	6240	275
Y22	2,829,944	816,988,648	64	12848	130083	153	21315	5616	264
Y23	3,034,591	875,490,567	69	11699	134851	139	27050	4608	265
Y24	3,961,103	1,143,733,410	90	13369	129492	141	20113	5900	255
Y25	9,479,738	2,736,074,385	216	35377	130408	873	2402	23652	257
Y26	5,953,517	1,717,066,281	136	41943	180390	2483	833	29347	258
Y27	3,419,267	986,069,172	78	21765	141541	226	13045	12970	262
Y28	3,802,833	1,096,818,685	87	19862	146317	221	14087	11143	260

Y29	2,861,848	825,492,005	65	27268	140128	521	9706	12589	268
Y30	3,882,597	1,120,505,985	88	31403	121581	879	2204	19998	261
Y31	2,895,265	835,355,471	66	23613	139672	315	9018	14697	259
Y32	2,353,553	679,027,331	54	15280	126740	180	17934	7547	263
Y33	2,773,277	800,247,281	63	12012	133132	130	24998	4780	266
Y34	2,724,077	785,429,967	62	21141	135765	340	7017	12043	261
Y35	3,968,365	1,145,675,309	90	18271	165734	207	16417	10120	265
Y36	4,044,031	1,166,235,257	92	16533	151772	157	18639	8173	256
Y37	5,565,154	1,606,422,607	127	27917	137770	514	3329	17934	257
Y38	6,670,265	1,925,923,520	152	37590	134388	1876	1609	24369	266
Y39	2,930,487	846,519,081	67	18119	138222	209	14885	9732	261
Y40	3,266,160	942,433,570	74	25924	133101	546	10393	10493	277
Y41	4,115,182	1,187,021,324	94	39619	140123	831	5421	19621	255
Y42	3,184,582	918,943,015	73	19290	131316	228	13329	10521	258
Y43	4,168,323	1,202,141,235	95	24288	142078	308	7880	15129	261
Y44	3,003,091	866,791,933	68	21050	117935	288	9847	11958	263
Y45	2,213,442	638,475,977	50	14891	115737	188	16545	6943	263
Y46	3,034,446	875,249,638	69	18855	141902	191	16218	10298	259
Y47	3,695,285	1,065,415,172	84	14564	142143	136	24632	7064	261
Y48	3,482,826	1,005,141,991	79	16236	142646	170	18172	8414	257
Y49	3,142,073	907,130,289	72	16048	158371	310	8380	7378	271
Y50	4,948,332	1,425,739,712	113	29975	131806	774	2634	19217	263
Y51	3,126,727	902,388,316	71	15388	192583	217	15765	7779	273
Y52	2,207,871	637,024,584	50	15230	128787	214	14722	7531	270
Y53	2,657,792	766,945,288	61	16259	129357	222	13921	8006	272
Y54	2,733,290	788,654,987	62	15976	137943	201	17545	8040	274
Y55	1,562,143	450,777,244	36	13519	143213	186	15347	6540	265
Y56	3,349,350	964,852,809	76	28127	149364	983	2110	16976	268
Y57	3,588,393	1,035,316,451	82	18971	142940	215	13303	10423	260
Y58	4,952,802	1,428,259,746	113	29754	162194	581	2816	19271	260
Y59	3,551,615	1,022,863,736	81	21537	124749	265	8932	12602	260
Y60	9,592,708	2,766,706,099	218	25627	131154	378	6684	15560	256
Y61	3,175,319	914,214,325	72	25562	125127	460	4069	15466	258
Y62	3,219,240	929,651,742	73	26373	147334	512	3596	17277	264
Y63	2,293,930	660,268,475	52	17828	108150	234	12588	8886	258
Y64	4,298,927	1,238,465,197	98	30352	127100	466	13066	13435	260
Y65	2,211,029	636,817,292	50	19412	138447	282	9565	10454	264
Y66	2,605,845	751,608,152	59	18369	129266	215	11452	10387	263
Y67	3,200,809	923,979,845	73	26910	160573	370	5495	17412	262
Y68	2,415,801	696,276,629	55	17710	126242	207	16793	9782	263
Y69	2,133,589	615,375,798	49	15891	119518	190	16151	8516	265
Y70	2,569,170	739,758,094	58	21813	115309	292	8729	12266	252
Y71	2,555,672	736,019,047	58	16365	124027	210	13362	8787	261

基于 CAPSS 的糜子 BAC 克隆在参考基因组上定位

Y72	4,069,319	1,175,114,780	93	15892	159534	190	16490	7772	269
Y73	2,388,962	688,945,787	54	17316	118458	295	9212	8432	265
Y74	2,184,407	630,106,420	50	29794	119741	1041	2604	16550	266
Y75	2,705,633	780,412,508	62	16520	135825	191	16390	8006	273
Y76	22,945,498	6,620,232,821	522	63213	179044	3818	691	46189	258
Y77	2,514,768	724,521,191	57	24350	134578	372	6500	15266	263
Y78	2,017,301	581,574,758	46	21521	133943	283	8595	13126	261
Y79	2,557,598	736,469,690	58	30719	118565	896	1992	19839	259
Y80	2,323,301	670,153,319	53	23012	143310	332	6535	14252	262
Y81	1,854,734	534,071,301	42	17011	126281	244	12377	9340	265
Y82	2,706,266	778,965,205	61	17148	151252	172	17883	9326	263
Y83	3,157,758	910,567,055	72	20253	120082	262	10555	11581	257
Y84	2,462,564	709,577,296	56	13491	130752	155	20678	6423	274
Y85	2,673,421	770,611,611	61	30125	127306	545	2956	19393	256
Y86	2,337,476	674,303,409	53	18865	143217	291	9555	10294	265
Y87	2,628,312	757,471,733	60	18269	120461	322	9009	9573	265
Y88	1,968,236	567,355,363	45	15419	117965	199	15660	7966	267
Y89	5,315,991	1,531,627,522	121	28970	130250	650	3105	18414	261
Y90	3,584,318	1,032,542,704	81	24247	148560	267	9575	14846	254
Y91	599,906	173,092,771	14	14806	83876	653	3863	7454	260
Y92	1,788,766	514,244,039	41	15366	129431	182	18403	7773	266
Y93	3,245,110	935,435,757	74	19906	231579	194	16859	11731	261
Y94	2,341,752	675,340,586	53	16525	140402	176	19218	8476	263
Y95	3,132,534	902,837,084	71	20161	144060	194	13952	11418	258
Y96	4,631,475	1,336,010,594	105	19080	147074	162	21113	11014	260

附录 4 用于 NGS 建库的引物序列

Notes: Phos: 磷酸化修饰; *: 硫代修饰

引物名称	序列 5'-3'
Adpater1	/5Phos/ GATCGGAAGAGCACACGTCTGAACTCCAGT*C
Adpater2	ACACTCTTTCCCTACACGACGCTCTTCCGATC*T
P5+ADA	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC*T
I-001	CAAGCAGAAGACGGCATACGAGAT CCTGCG AGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-002	CAAGCAGAAGACGGCATACGAGAT TCAGAG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-003	CAAGCAGAAGACGGCATACGAGAT ACCTAGG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-004	CAAGCAGAAGACGGCATACGAGAT TTGATCC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-005	CAAGCAGAAGACGGCATACGAGAT TCTCCAT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-006	CAAGCAGAAGACGGCATACGAGAT CATCAG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-007	CAAGCAGAAGACGGCATACGAGAT TTCCAGC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-008	CAAGCAGAAGACGGCATACGAGAT AGTTGGT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-009	CAAGCAGAAGACGGCATACGAGAT GTACCGG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-010	CAAGCAGAAGACGGCATACGAGAT CGGAGTT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-011	CAAGCAGAAGACGGCATACGAGAT TGATAGT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-012	CAAGCAGAAGACGGCATACGAGAT CAGGTCG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-013	CAAGCAGAAGACGGCATACGAGAT CGCATTA GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-014	CAAGCAGAAGACGGCATACGAGAT GGTACCT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-015	CAAGCAGAAGACGGCATACGAGAT GGACGCA GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-016	CAAGCAGAAGACGGCATACGAGAT GAGATTC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-017	CAAGCAGAAGACGGCATACGAGAT GAGCATG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-018	CAAGCAGAAGACGGCATACGAGAT GTTGCGT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-019	CAAGCAGAAGACGGCATACGAGAT CCAATGC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-020	CAAGCAGAAGACGGCATACGAGAT CGAGATC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-021	CAAGCAGAAGACGGCATACGAGAT CATATTG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-022	CAAGCAGAAGACGGCATACGAGAT GACGTCAG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-023	CAAGCAGAAGACGGCATACGAGAT GTAATTG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-024	CAAGCAGAAGACGGCATACGAGAT CCTATCT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-025	CAAGCAGAAGACGGCATACGAGAT CAATCGG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-026	CAAGCAGAAGACGGCATACGAGAT GCGGCAT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-027	CAAGCAGAAGACGGCATACGAGAT AGTACTG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-028	CAAGCAGAAGACGGCATACGAGAT TACTATT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-029	CAAGCAGAAGACGGCATACGAGAT CCGGATG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-030	CAAGCAGAAGACGGCATACGAGAT ACCATGA GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-031	CAAGCAGAAGACGGCATACGAGAT CGGTTCT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-032	CAAGCAGAAGACGGCATACGAGAT TATTCCA GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-033	CAAGCAGAAGACGGCATACGAGAT CCTCTGT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-034	CAAGCAGAAGACGGCATACGAGAT AGGTATT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-035	CAAGCAGAAGACGGCATACGAGAT GCATTCC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-036	CAAGCAGAAGACGGCATACGAGAT TTGAATT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-037	CAAGCAGAAGACGGCATACGAGAT CTGCGCG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-038	CAAGCAGAAGACGGCATACGAGAT AGACCTT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-039	CAAGCAGAAGACGGCATACGAGAT GTCCAGT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-040	CAAGCAGAAGACGGCATACGAGAT ACCTGCT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-041	CAAGCAGAAGACGGCATACGAGAT CCGGTAC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-042	CAAGCAGAAGACGGCATACGAGAT CTTGACC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-043	CAAGCAGAAGACGGCATACGAGAT CATCATT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-044	CAAGCAGAAGACGGCATACGAGAT TCTAGTT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T

I-045	CAAGCAGAAGACGGCATACGAGAT GCCATAG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-046	CAAGCAGAAGACGGCATACGAGAT ACCGTCG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-047	CAAGCAGAAGACGGCATACGAGAT CTTGGT TGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-048	CAAGCAGAAGACGGCATACGAGAT TACGCCG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-049	CAAGCAGAAGACGGCATACGAGAT GGACTGC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-050	CAAGCAGAAGACGGCATACGAGAT GCGCGAG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-051	CAAGCAGAAGACGGCATACGAGAT GTTCGAG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-052	CAAGCAGAAGACGGCATACGAGAT CATACGT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-053	CAAGCAGAAGACGGCATACGAGAT TCAGTAT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-054	CAAGCAGAAGACGGCATACGAGAT CTAAGTA GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-055	CAAGCAGAAGACGGCATACGAGAT TTAGCTT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-056	CAAGCAGAAGACGGCATACGAGAT CGCCGTC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-057	CAAGCAGAAGACGGCATACGAGAT GTCTTCT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-058	CAAGCAGAAGACGGCATACGAGAT GCCGGAC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-059	CAAGCAGAAGACGGCATACGAGAT AAGCTGA GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-060	CAAGCAGAAGACGGCATACGAGAT GCGCTCT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-061	CAAGCAGAAGACGGCATACGAGAT CGTAGGC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-062	CAAGCAGAAGACGGCATACGAGAT ATGATTA GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-063	CAAGCAGAAGACGGCATACGAGAT GCAGGTT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-064	CAAGCAGAAGACGGCATACGAGAT AATCGTC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-065	CAAGCAGAAGACGGCATACGAGAT CGGCCTA GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-066	CAAGCAGAAGACGGCATACGAGAT CTATGCC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-067	CAAGCAGAAGACGGCATACGAGAT GAGTTAA GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-068	CAAGCAGAAGACGGCATACGAGAT TAGACTA GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-069	CAAGCAGAAGACGGCATACGAGAT TCATGCA GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-070	CAAGCAGAAGACGGCATACGAGAT GCTTATT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-071	CAAGCAGAAGACGGCATACGAGAT CAAGGCT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-072	CAAGCAGAAGACGGCATACGAGAT AGGTTGG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-073	CAAGCAGAAGACGGCATACGAGAT CTTCTGC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-074	CAAGCAGAAGACGGCATACGAGAT TAATTCT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-075	CAAGCAGAAGACGGCATACGAGAT GATGCTG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-076	CAAGCAGAAGACGGCATACGAGAT CTAGAGG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-077	CAAGCAGAAGACGGCATACGAGAT TATCCGG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-078	CAAGCAGAAGACGGCATACGAGAT AGGGGCG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-079	CAAGCAGAAGACGGCATACGAGAT GGTCGTT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-080	CAAGCAGAAGACGGCATACGAGAT GGAAGTA GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-081	CAAGCAGAAGACGGCATACGAGAT ATTGCCA GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-082	CAAGCAGAAGACGGCATACGAGAT GATTAGC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-083	CAAGCAGAAGACGGCATACGAGAT AGAAGTC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-084	CAAGCAGAAGACGGCATACGAGAT ATAGTAC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-085	CAAGCAGAAGACGGCATACGAGAT GATCTCG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-086	CAAGCAGAAGACGGCATACGAGAT GGCTGCG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-087	CAAGCAGAAGACGGCATACGAGAT GCCTCTC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-088	CAAGCAGAAGACGGCATACGAGAT CTGAAGC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-089	CAAGCAGAAGACGGCATACGAGAT TGGATAT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-090	CAAGCAGAAGACGGCATACGAGAT ATTAAGG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-091	CAAGCAGAAGACGGCATACGAGAT GTTGGCG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-092	CAAGCAGAAGACGGCATACGAGAT GCGAGTA GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-093	CAAGCAGAAGACGGCATACGAGAT TGGGCGT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-094	CAAGCAGAAGACGGCATACGAGAT TTACTTA GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-095	CAAGCAGAAGACGGCATACGAGAT AGAGGAT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-096	CAAGCAGAAGACGGCATACGAGAT GACCGCC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T

I-097	CAAGCAGAAGACGGCATACGAGAT ATCATAT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-098	CAAGCAGAAGACGGCATACGAGAT GCGTACCG TGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-099	CAAGCAGAAGACGGCATACGAGAT CGTTCTT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-100	CAAGCAGAAGACGGCATACGAGAT TCAACGGG TGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-101	CAAGCAGAAGACGGCATACGAGAT GATTGCT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-102	CAAGCAGAAGACGGCATACGAGAT ACGGAGCG TGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-103	CAAGCAGAAGACGGCATACGAGAT AGAGTTG TGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-104	CAAGCAGAAGACGGCATACGAGAT GGAGTCT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-105	CAAGCAGAAGACGGCATACGAGAT AAGATAG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-106	CAAGCAGAAGACGGCATACGAGAT ACTGCTT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-107	CAAGCAGAAGACGGCATACGAGAT CGTTAGGG TGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-108	CAAGCAGAAGACGGCATACGAGAT TTATGAT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-109	CAAGCAGAAGACGGCATACGAGAT GTTATCA GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-110	CAAGCAGAAGACGGCATACGAGAT TAGGAGA GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-111	CAAGCAGAAGACGGCATACGAGAT TGCTCTG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-112	CAAGCAGAAGACGGCATACGAGAT AATAGCG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-113	CAAGCAGAAGACGGCATACGAGAT GACGATC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-114	CAAGCAGAAGACGGCATACGAGAT CTGCCGT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-115	CAAGCAGAAGACGGCATACGAGAT ATCGGTC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-116	CAAGCAGAAGACGGCATACGAGAT CGCAAGT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-117	CAAGCAGAAGACGGCATACGAGAT GGCTTAC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-118	CAAGCAGAAGACGGCATACGAGAT GCATGGC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-119	CAAGCAGAAGACGGCATACGAGAT ACGTCCG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-120	CAAGCAGAAGACGGCATACGAGAT TACTCGC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-121	CAAGCAGAAGACGGCATACGAGAT TCCGCGT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-122	CAAGCAGAAGACGGCATACGAGAT ACGCTTC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-123	CAAGCAGAAGACGGCATACGAGAT AGTTACCG TGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-124	CAAGCAGAAGACGGCATACGAGAT CTATATT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-125	CAAGCAGAAGACGGCATACGAGAT GGCGAGG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-126	CAAGCAGAAGACGGCATACGAGAT CTATTAA GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-127	CAAGCAGAAGACGGCATACGAGAT GCTTCGG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-128	CAAGCAGAAGACGGCATACGAGAT ATAACGA GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-129	CAAGCAGAAGACGGCATACGAGAT CAGAGCC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-130	CAAGCAGAAGACGGCATACGAGAT TTCCAAG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-131	CAAGCAGAAGACGGCATACGAGAT AGCTTCA GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-132	CAAGCAGAAGACGGCATACGAGAT GTCAACC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-133	CAAGCAGAAGACGGCATACGAGAT ATGGCGG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-134	CAAGCAGAAGACGGCATACGAGAT CTCTGGT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-135	CAAGCAGAAGACGGCATACGAGAT TCTACGC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-136	CAAGCAGAAGACGGCATACGAGAT GCGAAGG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-137	CAAGCAGAAGACGGCATACGAGAT AGGCCAG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-138	CAAGCAGAAGACGGCATACGAGAT CCTTGCG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-139	CAAGCAGAAGACGGCATACGAGAT TCTCTCA GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-140	CAAGCAGAAGACGGCATACGAGAT GAATCTT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-141	CAAGCAGAAGACGGCATACGAGAT AACCGAT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-142	CAAGCAGAAGACGGCATACGAGAT GGCTCGT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-143	CAAGCAGAAGACGGCATACGAGAT GTATCCA GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-144	CAAGCAGAAGACGGCATACGAGAT CTGGAAT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-145	CAAGCAGAAGACGGCATACGAGAT AATGAGT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-146	CAAGCAGAAGACGGCATACGAGAT ATAAGCT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-147	CAAGCAGAAGACGGCATACGAGAT AGATCGC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-148	CAAGCAGAAGACGGCATACGAGAT ACTGGAG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T

I-149	CAAGCAGAAGACGGCATACGAGAT TAAGTTC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-150	CAAGCAGAAGACGGCATACGAGAT CTCGCT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-151	CAAGCAGAAGACGGCATACGAGAT ATCTGAG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-152	CAAGCAGAAGACGGCATACGAGAT TCGTCTT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-153	CAAGCAGAAGACGGCATACGAGAT GTCAGTT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-154	CAAGCAGAAGACGGCATACGAGAT GTTACTC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-155	CAAGCAGAAGACGGCATACGAGAT GTCTATG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-156	CAAGCAGAAGACGGCATACGAGAT CCGTAGT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-157	CAAGCAGAAGACGGCATACGAGAT GCTAGCC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-158	CAAGCAGAAGACGGCATACGAGAT CTCATGG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-159	CAAGCAGAAGACGGCATACGAGAT AACCAGC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-160	CAAGCAGAAGACGGCATACGAGAT CTACTCT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-161	CAAGCAGAAGACGGCATACGAGAT GTTGATA GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-162	CAAGCAGAAGACGGCATACGAGAT GTACGTC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-163	CAAGCAGAAGACGGCATACGAGAT GAAGTAG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-164	CAAGCAGAAGACGGCATACGAGAT CCGACCA GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-165	CAAGCAGAAGACGGCATACGAGAT ATGGTCT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-166	CAAGCAGAAGACGGCATACGAGAT GTTCGGA GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-167	CAAGCAGAAGACGGCATACGAGAT CAACCTC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-168	CAAGCAGAAGACGGCATACGAGAT ACTCATA GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-169	CAAGCAGAAGACGGCATACGAGAT AACGAAG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-170	CAAGCAGAAGACGGCATACGAGAT CGGACGG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-171	CAAGCAGAAGACGGCATACGAGAT TAACGTT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-172	CAAGCAGAAGACGGCATACGAGAT AATATGC GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-173	CAAGCAGAAGACGGCATACGAGAT TGCGGTT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-174	CAAGCAGAAGACGGCATACGAGAT TATGGAT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-175	CAAGCAGAAGACGGCATACGAGAT CGATGAG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-176	CAAGCAGAAGACGGCATACGAGAT AACGCGA GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-177	CAAGCAGAAGACGGCATACGAGAT GGTCAAG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-178	CAAGCAGAAGACGGCATACGAGAT CAGCTAT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-179	CAAGCAGAAGACGGCATACGAGAT AGGAACG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-180	CAAGCAGAAGACGGCATACGAGAT TGAGGCG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-181	CAAGCAGAAGACGGCATACGAGAT TCGATTG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-182	CAAGCAGAAGACGGCATACGAGAT CGAGCGT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-183	CAAGCAGAAGACGGCATACGAGAT TTGCCTG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-184	CAAGCAGAAGACGGCATACGAGAT AGTCTCT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-185	CAAGCAGAAGACGGCATACGAGAT ACCAATT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-186	CAAGCAGAAGACGGCATACGAGAT ATTCTTG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-187	CAAGCAGAAGACGGCATACGAGAT ATCTCTT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-188	CAAGCAGAAGACGGCATACGAGAT TAGTCAG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-189	CAAGCAGAAGACGGCATACGAGAT CAGGCTT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-190	CAAGCAGAAGACGGCATACGAGAT TTCGTTG GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-191	CAAGCAGAAGACGGCATACGAGAT CCAACCT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
I-192	CAAGCAGAAGACGGCATACGAGAT TGCGTTT GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T