AFLP 基因组捕获技术：一种通用的低阶元的系统发育分析方法

AFLP Genome Capture Technique: A General Method for Shallow Scale Phylogenetic Analysis

李佳璇，梁丹，张鹏*

¹ 中山大学生命科学学院，广州 510006

*通讯作者邮箱: zhangp35@mail.sysu.edu.cn

引用格式: 李佳璇，梁丹，张鹏. (2021). AFLP基因组捕获技术：一种通用的低阶元的系统发育分析方法.

Bio-101 e1010598. Doi: 10.21769/BioProtoc.1010598.

How to cite: Li, J. X., Liang, D. and Zhang, P. (2021). AFLP Genome Capture Technique: A General Method for Shallow Scale Phylogenetic Analysis. *Bio-101* e1010598. Doi: 10.21769/BioProtoc.1010598.(in Chinese)

摘要：AFLP 基因组捕获技术是在扩增性片段长度多态性 (Amplified Fragment Length Polymorphism, AFLP) 的实验基础上，将 AFLP 片段制备成捕获探针，从目标样本的基因组文库中富集特定区域，并进行高通量测序的一种技术。该技术能够在属内种间水平上获取大量的匿名基因序列，或在种下水平上获取大量的 SNP 位点，以用于系统发育关系重建，群体遗传结构等分析。该技术的主要流程包括: (1) AFLP 探针制备；(2) 基因组文库构建；(3) 杂交捕获及高通量测序。后续的数据分析流程简单易学，且对计算资源的要求较低，主要为序列组装和提取目标直系同源序列或 SNP 位点。该技术的优势是: (1) 整个过程不依赖任何参考基因组资源，适合非模式生物的研究；(2) 自制捕获探针，省时，大大降低实验成本；(3) 对目标样本的 DNA 质量要求较低，高度降解的 DNA 样本也可被用于分析。

关键词：目标序列捕获，自制探针，系统发育学，群体遗传学

材料与试剂

1. 0.2 ml 和 1.5 ml 微量离心管 (BBI Life Sciences, 常温储存)
2. 200 μ l 八联排管 (BBI Life Sciences, 常温储存)

3. 各种型号的枪头 (10 μ l, 200 μ l, 1 ml, BBI Life Sciences, 常温储存)
4. 基因组 DNA 提取试剂盒 (北京天根生化科技有限公司, 常温储存, catalog number: DP304-03)
5. 无水乙醇 (分析纯) (广州化学试剂厂, 常温储存)
6. AMPure XP beads (Beckman Coulter, 4 °C 储存, catalog number: A63881)
7. Dynabeads MyOne Streptavidin C1 (Life Technologies, 4 °C 储存, catalog number: 65001)
8. FastDigest MluI (Thermo Fisher Scientific, -20 °C 储存, catalog number: FD0564)
9. FastDigest SbfI (Thermo Fisher Scientific, -20 °C 储存, catalog number: FD1194)
10. T4 DNA ligase (5 U/ μ l) (Thermo Fisher Scientific, -20 °C 储存, catalog number: EL0014)
11. Human Cot-1 DNA (1 μ g/ μ l) (Thermo Fisher Scientific, -20 °C 储存, catalog number: 15279011)
12. NEBNext Ultra DNA Library Prep Kit for Illumina (New England Biolabs, -20 °C 储存, catalog number: E7645L)
13. NEBNext dsDNA Fragmentase (New England Biolabs, -20 °C 储存, catalog number: M0348L)
14. TransTaq DNA Polymerase High Fidelity (HiFi) (北京全式金生物技术有限公司, -20 °C 储存, catalog number: AP131-03)
15. 100 mM ATP (上海生工生物工程股份有限公司, -20 °C 储存, catalog number: A600311)
16. 10 mM dNTPs (上海生工, -20 °C 储存, catalog number: A610056)
17. 1X TE buffer (上海生工, 4 °C 储存, catalog number: B548106)
18. 20X SSPE buffer (上海生工, 4 °C 储存, catalog number: B548111)
19. 50X Denhardt's Solution (上海生工, -20 °C 储存, catalog number: B548209)
20. 10 % SDS Solution (上海生工, 常温储存, catalog number: B548118)
21. 20X SSC buffer (上海生工, 4 °C 储存, catalog number: B548109)
22. 5 M NaCl (上海生工, 4 °C 储存, catalog number: B548121)
23. 1 M Tris-HCl Solution, pH 8.0 (上海生工, 4 °C 储存, catalog number: B548127)
24. 0.5 M EDTA, pH8.0 (上海生工, 4 °C 储存, catalog number: B300599)

25. 吐温 20 (上海生工, 常温储存, catalog number: A600560)
26. 矿物油 (上海生工, 常温储存, catalog number: A630217)
27. 琼脂糖 (Invitrogen, 常温储存, catalog number: 75510-019)
28. 双蒸水 (广州誉维生物科技仪器有限公司 Unique 超纯水机制)
29. 吸附缓冲液 (见配方, 4 °C 储存)
30. 洗涤缓冲液 1 (见配方, 4 °C 储存)
31. 洗涤缓冲液 2 (见配方, 4 °C 储存)

仪器设备

1. 0.5-5 µl, 2-20 µl, 20-200 µl, 100-1000µl 移液器 (Thermo Fisher Scientific F2 系列)
2. Gene Pro PCR 扩增仪 (杭州博日科技股份技术有限公司, model: TC-E-96G)
3. Vortex Genie2 漩涡混合器 (Scientific Industries, model: G560E)
4. Cubee 迷你离心机 (广州美津生物技术有限公司, model: aqbd-i)
5. 振荡型恒温金属浴 (杭州奥盛仪器有限公司, model: MS-100)
6. DynaMag-2 Magnet 磁力架 (Thermo Fisher Scientific, model: 12321D)
7. 超微量分光光度仪 (Thermo Fisher Scientific, model: NanoDrop2000)
8. 通用电泳仪 (北京百晶生物技术有限公司, model: BC-power6001)
9. 高压蒸汽灭菌锅 (Kagoshima Seisakusyo Inc., model: SX-500)
10. 服务器 (Linux 系统, 32 G 以上的内存, 500 G 的存储)

软件

1. Seqtk_demultiplex https://github.com/jameslz/fastx-utils/blob/master/seqtk_demultiplex/
2. SPAdes <https://github.com/ablab/spades/>
3. CD-HIT <https://github.com/weizhongli/cdhit/>
4. BLAST 2.6.0 <ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/2.6.0/>
5. Python 2.7.9 <https://www.python.org/ftp/python/2.7.9/>
6. Biopython 1.68 <https://pypi.org/project/biopython/1.68/>
7. PASTA <https://github.com/smirarab/pasta>

8. FastTree <http://www.microbesonline.org/fasttree/#Install>
9. Phylopypruner <https://gitlab.com/fethalen/phylopypruner/-/wikis/Installation>
10. RAxML <https://github.com/stamatak/standard-RAxML>
11. ASTRAL 5.7.4 <https://github.com/smirarab/ASTRAL>

实验步骤

AFLP 基因组捕获技术的实验流程,可分为三个步骤:AFLP 探针制备,基因组文库构建以及杂交捕获,其工作流程如图 1 所示。

图 1. AFLP 基因组捕获技术的工作流程图

一、AFLP 探针制备 (总耗时: 约 7 h)

读者需要挑选一个与研究类群相关的探针制备物种, 并将它的基因组 DNA 作为起始模板, 按照以下四个步骤完成 AFLP 探针的制备 (见图 1): (1) 限制性内切酶消化; (2) 接头连接和片段筛选; (3) 预扩增; (4) 选择性扩增。其中, 限制性内切酶消化和选择性扩增是影响 AFLP 探针制备结果的关键步骤, 因为它们决定了 AFLP 探针的长度范围和数量等。需要指出的是, 探针制备物种的基因组 DNA 质量对 AFLP 探针制备的实验结果至关重要。在利用限制性内切酶消化基因组 DNA 的过程中, 如果基因组 DNA 的质量较差 (如 DNA 无主带或 DNA 有主带, 但同时存在明显的弥散条带), 大量的限制性酶切识别位点会因 DNA 降解而丢失, 这将导致酶切片段总数的减少, 从而导致 AFLP 探针数量的减少。因此, 建议读者优先挑选基因组 DNA 质量较好的样本作为探针制备物种, 选择标准为: 在基因组 DNA 的电泳检测结果中, 基因组 DNA 条带的长度大于 20 Kb, 且无明显的弥散条带。

1. 限制性内切酶消化 (耗时: 约 1 h)

1.1 取 1 μg 高分子量的基因组 DNA (>10Kb) 于新的 0.2 ml 离心管中。

1.2 按下表建立酶切反应体系 (表 1)。

表 1. 酶切反应体系

成分	体积 (μl)	终浓度 (20 μl)
FastDigest buffer (10X)	2.0	1X
FastDigest MluI	1.0	
FastDigest SbfI	1.0	
DNA (1 μg) + 灭菌双蒸水	16.0	
总体积	20	

1.3 将配制好的酶切反应体系充分涡旋混匀, 简短离心, 37 °C 孵育 20 min。

1.4 将上一步溶液转移至新的 1.5 ml 离心管中, 使用 AMPure XP 磁珠纯化酶切产物。

注: 纯化开始前将 AMPure XP 磁珠平衡至室温, 然后涡旋 5-10 min, 使其充分悬浮。

1.5 向 20 μl 酶切反应液中加入 36 μl (体积比 1:1.8) AMPure XP 磁珠, 用移液器吸

打混匀至少 10 次，室温孵育 5 min。

1.6 将含磁珠的离心管置于磁力架上，室温孵育 5 min，待溶液变澄清。

1.7 小心地移除上清液，注意动作要轻柔，不要扰动磁珠。

1.8 保持离心管置于磁力架上，向离心管底部轻柔地加入 200 μ l 80 %新鲜配制的乙醇溶液 (注意不要扰动磁珠)，室温孵育 30 s，小心地移除上清液。

1.9 重复操作步骤 1.8。

1.10 使用量程 < 20 μ l 的移液器移除管底残留的乙醇，注意不要扰动磁珠。

1.11 将离心管开盖置于磁力架上，室温孵育 5 min，自然晾干。

1.12 将离心管从磁力架上取下，向离心管中加入 52 μ l 0.1X TE 缓冲液，并使用移液器吸打混匀，使磁珠充分悬浮，室温孵育 2 min。

1.13 将离心管置于磁力架上，室温孵育 5 min，待溶液变澄清。

1.14 小心地吸取上清液 (酶切纯化产物) 至新的 0.2 ml 离心管中。

1.15 取 1 μ l 酶切纯化产物，使用 1.2 %的琼脂糖凝胶电泳进行检测。

注：在酶切纯化产物的电泳检测结果中，如果酶切后的基因组 DNA 与酶切反应前的基因组 DNA 相比，无主带，且呈现出明显的弥散条带，则证明酶切实验成功。

1.16 立刻进行下一步操作或将产物保存于 -20 °C (建议保存时间不超过 7 d)。

2. 接头连接和片段筛选 (耗时：约 1.5 h)

2.1 在新的 0.2 ml 离心管中，按下表建立接头连接反应体系 (表 2)。

表 2. 接头连接反应体系

成分	体积 (μl)	终浓度 (60 μl)
T4 DNA ligase buffer (10X)	6.0	1X
ATP (100 mM)	0.6	1 mM
Mlul adapter (10 μM)	1.0	0.167 μM
Sbfl adapter (10 μM)	1.0	0.167 μM
酶切纯化产物	50.4	
总体积	59	

注: Mlul adapter 和 Sbfl adapter 的序列和制备方法见附录和溶液配方。

2.2 将配制好的接头连接反应体系充分涡旋混匀，并简短离心，45 °C 孵育 5 min。

2.3 迅速将含有连接反应液的离心管置于冰上，向离心管中加入 1 μl T4 DNA ligase (5U/μl)，充分涡旋混匀，并简短离心，25 °C 孵育 30 min。

2.4 将 60 μl 连接产物转移至新的 1.5 ml 离心管中，使用 AMPure XP 磁珠进行片段筛选。

注: 片段筛选开始前需要涡旋 AMPure XP 磁珠，使其充分悬浮。

2.5 向离心管中加入 15 μl AMPure XP 磁珠，用移液器吸打混匀至少 10 次，室温孵育 5 min。

2.6 将含磁珠的离心管置于磁力架上，室温孵育 5 min，待溶液变澄清。

2.7 小心地吸取上清液至新的 1.5 ml 离心管中，注意动作要轻柔，不要吸到磁珠。

注: 保留上清液，丢弃磁珠，目的是去除大片段 DNA (> 5 Kb)。

2.8 向离心管中加入 20 μl AMPure XP 磁珠，用移液器吸打混匀至少 10 次，室温孵育 5 min。

2.9 将含磁珠的离心管置于磁力架上，室温孵育 5 min，待溶液变澄清。

2.10 小心地移除上清液，注意动作要轻柔，不要扰动磁珠。

注: 保留磁珠，丢弃上清液，目的是去除小片段 DNA (< 200 bp)，并确保连接产物的长度范围在 200-5000 bp 之间。

2.11 保持离心管置于磁力架上，向离心管底部轻柔地加入 200 μ l 80 %新鲜配制的乙醇溶液 (注意不要扰动磁珠)，室温孵育 30 s，小心地移除上清液。

2.12 重复操作步骤 2.11。

2.13 将离心管开盖置于磁力架上，室温孵育 5 min，自然晾干。

2.14 将离心管从磁力架上取下，向离心管中加入 30 μ l 0.1X TE 缓冲液，并使用移液器吸打混匀，使磁珠充分悬浮，室温孵育 2 min。

2.15 将离心管置于磁力架上，室温孵育 5 min，待溶液变澄清。

2.16 小心地吸取上清液 (片段筛选产物) 至新的 0.2 ml 离心管中。通过 NanoDrop2000 分光光度计测量浓度，并使用 0.1X TE 缓冲液稀释片段筛选产物，使其终浓度达到 1 ng/ μ l。

2.17 建议立刻进行下一步的操作，剩余产物可保存于-20 $^{\circ}$ C (建议保存时间不超过 1 年)。

3. 预扩增 (耗时：约 2 h)

3.1 在新的 0.2 ml 离心管中，按下表建立预扩增反应体系 (表 3)。

表 3. 预扩增反应体系

成分	体积 (μ l)	终浓度 (25 μ l)
TransTaq HiFi buffer (10X)	2.5	1X
dNTPs (10 mM)	0.5	200 μ M
TransTaq HiFi DNA Polymerase (5 U/ μ l)	0.25	0.05 U/ μ l
Mlul-F (10 μ M)	0.5	0.2 μ M
Sbfl-R (10 μ M)	0.5	0.2 μ M
稀释后的片段筛选产物 (1 ng/ μ l)	5.0	
灭菌双蒸水	15.75	
总体积	25	

注：Mlul-F 和 Sbfl-R 的序列见附录。

3.2 将配制好的预扩增反应体系充分涡旋混匀，并简短离心，随后将离心管放入 PCR 仪中进行预扩增，反应程序见下表 (表 4)。

表 4. 预扩增反应程序

循环步骤	温度	时间	循环数
预变性	94 °C	4 min	1
变性	94 °C	30 s	25
复性	56 °C	1 min	
延伸	72 °C	1 min	
最终延伸	72 °C	10 min	1
保存	4 °C	∞	

3.3 取 2 µl 预扩增产物，使用 1 % 的琼脂糖凝胶电泳进行检测。

注：在预扩增产物的电泳检测结果中，如果在 200-5000 bp 的范围内出现弥散条带，则证明预扩增实验成功。

3.4 向离心管中加入 10 倍体积的 0.1X TE 缓冲液，以作下一步选择性扩增的模板。

3.5 建议立刻进行下一步操作，剩余产物保存于 -20 °C (建议保存时间不超过 1 年)。

4. 选择性扩增 (耗时：约 2.5 h)

本方法提供了两条上游的选择性扩增引物和两条下游的选择性扩增引物，共四对组合 (引物序列见附录，组合见表 5)。每一条引物的 3' 末端都添加了单个选择性核苷酸 (见图 1 的 AFLP 探针制备)，目的是为了控制 AFLP 片段数目。例如，使用表 5 中任意一对引物组合进行选择性的扩增，理论上能够获得已回收酶切片段数目的 1/16 ($1/4 \times 1/4$)。此外，两条下游的选择性扩增引物的 5' 末端都添加了生物素标记 (biotin)，目的是通过 PCR 扩增的方法，将生物素标记添加到 AFLP 片段中，以获得可直接用于杂交捕获实验的 AFLP 探针 (见图 1 的 AFLP 探针制备)。考虑到每对引物组合扩增出的 AFLP 探针不同，读者可以根据自身对目标基因数目的需求量仅使用其中某一组探针，或者将不同组的探针混合在一起使用。根据本方法在臭蛙属 (隶属于两栖纲，蛙科) 中的测试结果 (Li *et al.*, 2019)，当混合使用四组 AFLP 探针 (表 5) 进行序列捕获时，最终能够在属内种间水平上得到 511 个直系同源序列组 (orthologous groups, OGs)。这 511 个 OGs 的长度分布范围为 203-7779 bp，与 AFLP 探针的长度范围相近 (200-5000 bp)。其中有一些 OG 的长度大于 5000

bp，这证明了该方法不仅可以捕获目标区域，还可以捕获目标区域两侧的侧翼序列 (flanking sequence)，获得这些额外的侧翼序列可能会增强系统发育分析的信号。本方法提供的引物具有通用性，适用于不同的生物类群。若读者需要更多的选择性扩增引物，可参考上述设计思路自行设计。值得一提的是，选择性核苷酸的数目与研究对象的基因组大小有关，若基因组小于 5 Gb，建议在上游和下游引物中添加单个选择性核苷酸 (与本方法提供的引物设计思路一致)，若基因组大于 5 Gb，建议在上游或下游引物中添加两个及两个以上的选择性核苷酸。

表 5. 选择性扩增引物组合信息表

组合	1	2	3	4
上游选择性扩增引物	Mlul-F-SA	Mlul-F-SA	Mlul-F-ST	Mlul-F-ST
下游选择性扩增引物	Sbfl-R-SC	Sbfl-R-SG	Sbfl-R-SC	Sbfl-R-SG

4.1 在新的 0.2 ml 离心管中，按下表建立选择性扩增反应体系 (表 6)。

表 6. 选择性扩增反应体系

成分	体积 (μl)	终浓度 (25 μl)
TransTaq HiFi buffer (10X)	2.5	1X
dNTPs (10 mM)	0.5	200 μM
TransTaq HiFi DNA Polymerase (5U/μl)	0.25	0.05 U/μl
上游选择性扩增引物 (10 μM)	0.5	0.2 μM
下游选择性扩增引物 (10 μM)	0.5	0.2 μM
稀释后的预扩增产物	5.0	
灭菌双蒸水	15.75	
总体积	25	

表 7. 选择性扩增反应程序

循环步骤	温度	时间	循环数
预变性	94 °C	4 min	1
变性	94 °C	30 s	13
复性	65 °C (每经历一个循环, 降 0.7 °C)	60 s	
延伸	72 °C	60 s	
变性	94 °C	30 s	12
复性	56 °C	60 s	
延伸	72 °C	60 s	
最终延伸	72 °C	10 min	1
保存	4 °C	∞	

- 4.2 将配制好的选择性扩增反应体系充分涡旋混匀, 并简短离心, 随后将离心管放入 PCR 仪中进行选择性扩增, 反应程序见上表 (表 7)。
- 4.3 反应结束后, 将不同组的选择性扩增产物分别转移至新的 1.5 ml 离心管中。
- 4.4 使用 AMPure XP 磁珠纯化选择性扩增产物, 最后使用 20 μ l 0.1X TE 缓冲液进行洗脱。
- 4.5 取 1 μ l 纯化的选择性扩增产物 (AFLP 探针), 使用 1 %的琼脂糖凝胶电泳进行检测。如图 2 所示, 每一对引物组合都能扩增出长度不同的 AFLP 片段, 同时, 不同引物组合间呈现出的 AFLP 带型规律也各不相同, 则证明了 AFLP 探针制备实验成功。
- 4.6 通过 NanoDrop2000 分光光度计测量浓度, 并向离心管中加入 0.1X TE 缓冲液稀释 AFLP 探针, 使其最终浓度达到 50 ng/ μ l, 以便用于后续的杂交捕获实验。剩余的探针溶液可保存于 -20 °C (建议保存时间不超过 1 年)。

图 2. AFLP 探针的电泳结果示意图。泳道 1-4 代表了四对选择性引物组合的扩增结果。

二、基因组文库构建 (总耗时：约 5 h)

本方法对高通量测序平台的测序读长没有严格要求，所以它适用于所有的二代测序平台，如 Illumina, Ion Torrent, BGISEQ 等。读者可根据自身对测序平台的熟悉程度，购买与测序平台相匹配的基因组文库构建试剂盒进行文库构建。本方法以 NEBNext Ultra DNA Library Prep Kit for Illumina 试剂盒为例，简要地介绍了 Illumina 基因组文库构建的方法。为了保证基因组文库具有较高的质量和良好的多样性，建议尽量选择 DNA 质量好 (> 4 Kb) 的样本用于文库构建，DNA 初始投入量在 200 ng-1 µg, 1 µg 最优。

1. DNA 片段化 (耗时：约 1 h)

该步骤的目的是为了获得符合 Illumina 测序平台的要求的，具有一定长度范围的 DNA 片段。以 HiSeq X ten 测序平台为例，该平台期望的插入 DNA 片段大小在 200-400 bp 之间，所以片段化后的 DNA 片段最好集中在这个区域。片段化的方法有超声处理，酶处理等，对于新手而言，推荐使用后者，因为该过程不依赖特殊仪器 (超声波 DNA 打断仪)，且入手简单。本方法以 NEBNext DNA 双链片段化酶为例，介绍了 DNA 片段化的方法，具体操作步骤和注意事项可参阅产品说明书：

<https://international.neb.com/protocols/0001/01/01/digestion-with-nebnext-dsdna-fragmentase-m0348>

1.1 取 1 μ g 基因组 DNA 于新的 0.2 ml 离心管中。

1.2 按下表建立片段化反应体系 (表 8)。

表 8. 片段化反应体系

成分	体积 (μ l)	终浓度 (20 μ l)
dsDNA fragmentase buffer (10X)	2.0	1X
dsDNA fragmentase	1.0	
DNA (1 μ g) + 灭菌双蒸水	17.0	
总体积	20	

注：片段化体系的准备工作在冰上完成。

1.3 将配制好的片段化反应体系充分涡旋混匀，并简短离心，37 °C 孵育 10-25 min。

注：建议读者做时间梯度预实验 (如：10 min, 12min, 14 min) 来确定最优的片段化时间。判断标准是：在片段化产物的电泳检测结果中，与片段化反应前的基因组 DNA 相比，片段化后的 DNA 片段主要集中在 200-400 bp 的范围内。

1.4 迅速将离心管转移至冰上，随后向离心管中加入 2 μ l 0.5 M EDTA，充分涡旋混匀，并简短离心，65 °C 孵育 15 min，使片段化酶彻底失活。

1.5 取 2 μ l 片段化产物，使用 2 % 的琼脂糖凝胶电泳进行检测。在片段化的电泳检测结果中，如果片段化后的产物存在主带或片段大小不在期望的范围内 (200-400 bp)，则需要调整片段化时间，重新进行实验。

1.6 使用 AMPure XP 磁珠纯化片段化产物，最后使用 20 μ l 灭菌双蒸水进行洗脱。

2. 补平加 A 和接头连接 (耗时：约 2.5 h)

这些步骤的主要目的是为了在 DNA 片段的两端添加 Illumina 测序接头。此外，为了区分不同的样本，测序接头上还应包含样本特异性标签 (barcode 或 index)。具体操作步骤请参阅文库构建试剂盒说明书

<https://international.neb.com/protocols/2015/01/14/protocol-for-use-with-nebnext-ffpe-dna-repair-mix-m6630-and-ultra-dna-library-prep-kit-for-illumina-e7370>

3. 片段筛选及 PCR 扩增 (耗时：约 1.5 h)

3.1 片段筛选

该步骤的目的是为了特异性地回收 300-500 bp 的 DNA 片段 (包括测序接头的长度), 去除过大或过小的双链或单链 DNA 片段。实验操作与 AFLP 探针制备过程中的片段筛选步骤一致, 但磁珠用量不同, 对于回收 300-500 bp 范围内的 DNA 片段, 可参阅试剂盒说明书。

3.2 PCR 扩增

PCR 扩增循环数与 DNA 初始投入量相关。为了防止过度地扩增基因组文库, 使其保持良好的多样性, 当 DNA 初始投入量为 1 µg, 并采用片段化酶处理的方法构建基因组文库时, 推荐使用 6-8 个 PCR 循环。

三、杂交捕获 (总耗时: 约 41 h)

在系统发育研究中, 捕获样本往往与探针制备样本不同。为了从捕获样本的基因组文库中杂交捕获到与探针同源的基因组区域, 特别是从亲缘关系较远的样本中, 通常需要根据捕获样本与探针制备样本之间的序列相似度, 对杂交体系中的盐离子浓度, 杂交反应温度等条件进行优化和调整。在本实验方法中, 杂交反应体系中盐离子的终浓度为 5X, 且杂交反应的温度呈梯度下降 (Touch-down), 即杂交温度从 65 °C 起始, 每隔一段时间下降 3 °C, 最终下降至 50 °C。在此条件下, 与探针的核酸序列相似度在 65 % 以上的单链 DNA 文库将被捕获。这种反应条件适合于大多数的系统发育研究。

1. 杂交 (耗时: 约 37 h)

杂交反应体系共 50 µl, 其中包括 18.5 µl 的杂交缓冲液体系和 31.5 µl 的探针文库混合体系, 在杂交反应开始前, 两个子体系需要进行不同的预处理 (具体操作见下文)。为了降低实验成本, 提高实验效率, 本方法建议使用混合文库进行杂交实验。在准备混合文库时, 读者首先需要根据样本间亲缘关系的远近, 基因组文库的质量 (指浓度和片段范围) 和文库特异性标签对捕获样本进行分组, 将亲缘关系相近, 基因组文库质量相近且不具有相同特异性标签的样本分为一组, 每组的样本数量为 2-7 个, 4 个最优 (Portik *et al.* 2016)。随后将同一组的基因组文库等质量混合, 以用于后续的杂交实验。

1.1 在新的 0.2 ml 离心管中, 按下表建立杂交缓冲液体系 (表 9)。

表 9. 杂交缓冲液体系

成分	体积 (μl)	终浓度 (50 μl)
20X SSPE	12.5	5X
0.5 M EDTA	0.5	0.005 M
50X Denhardt's	5.0	5X
10 % SDS	0.5	0.1 %
总体积	18.5	

注: 20X SSPE 在使用前需要充分涡旋混匀, 10 % SDS 需先置于室温充分溶解。

- 1.2 将杂交缓冲液体系充分涡旋混匀, 并简短离心。然后向杂交缓冲液体系中加入 12 μl 矿物油 (确保封盖反应液表面), 并将离心管放入 PCR 仪中, 65 °C 孵育待用。
- 1.3 在新的 0.2 ml 离心管中, 按下表建立探针文库混合体系 (表 10)。

表 10. 探针文库混合体系

成分	体积 (μl)	终浓度 (50 μl)
500 ng 混合 DNA 文库	23.5	10 ng/μl
200 ng AFLP 混合探针 (共四组, 每组 50 ng)	4.0	4 ng/μl
Human Cot-1 DNA (1 μg/μl)	2.5	0.05 μg/μl
BO1.P5.F (100 μM)	0.5	1 μM
BO3.P7.part1.F (100 μM)	0.5	1 μM
BO5.P7.part.F (100 μM)	0.5	1 μM
总体积	31.5	

注: (a) Human Cot-1 DNA 是人的胎盘 DNA, 长度主要为 50-300 bp, 并且富含重复的 DNA 序列, 在杂交反应中的作用是阻断非特异性杂交。(b) 寡核苷酸阻碍物 BO1-5 (序列见附录) 的作用是防止 DNA 文库的接头发生交联, 避免非特异性杂交。

- 1.4 将探针文库混合体系充分涡旋混匀, 简短离心, 然后向探针文库混合体系中加入 15 μl 矿物油 (确保封盖反应液表面), 并将离心管放入 PCR 仪中, 95 °C 孵育 5 min。
- 1.5 使用量程 50-200 μl 的移液器, 迅速将探针文库混合体系 (包括矿物油) 转移至

65 °C 预热的杂交缓冲液体系中，吸打混匀 5-10 次，迅速盖紧离心管管盖进行杂交。杂交反应程序为: 65 °C 6 h, 62 °C 6 h, 59 °C 6 h, 56 °C 6 h, 53 °C 6 h, 50 °C 6 h。

注：杂交开始后，每 6-12 h 检查离心管是否渗漏。

1.6 杂交反应结束后，杂交产物可继续 50 °C 孵育数小时，直至准备好下一步所需的链霉亲和素磁珠。

2. 链霉亲和素磁珠吸附 (耗时：约 1 h)

链霉亲和素磁珠的用量与探针投入量存在对应关系，磁珠用量过多会造成浪费，过少可能会影响吸附效果。以 Dynabeads Myone streptavidin C1 磁珠为例，1 μ l (50 μ g) 磁珠可以吸附约 200 ng 的 DNA 片段。根据该对应关系，当杂交反应的 AFLP 探针投入量为 200 ng 时，建议使用 3 μ l 磁珠进行吸附。增加磁珠用量的原因是在吸附反应开始前，本实验方法会对链霉亲和素磁珠进行洗涤，部分磁力较弱的珠子将被丢弃，这有助于提高捕获的效率。

2.1 取 $n \times 3 \mu$ l (n = 杂交反应数目， n 不大于 20) 链霉亲和素磁珠于新的 1.5 ml 离心管中。

注：链霉亲和素磁珠在使用前需要涡旋 3-5 min，使其充分悬浮。

2.2 将含磁珠的离心管置于磁力架上，室温孵育 2 min，待溶液变澄清。然后小心地移除上清液，注意不要扰动磁珠。

2.3 将离心管从磁力架上取下，向离心管中加入 200 μ l 吸附缓冲液 (2 M NaCl, 10 mM Tris-HCl, pH 8.0, 1 mM EDTA, pH 8.0)，涡旋 5-10 s 后，将离心管置于磁力架上，室温孵育 2 min，待溶液变澄清。然后小心地移除上清液，注意不要扰动磁珠。

2.4 重复操作步骤 2.3 两次。

2.5 将离心管从磁力架上取下，向离心管中加入 $n \times 20 \mu$ l 吸附缓冲液和 $n \times 0.1 \mu$ l 10 % 吐温 20，充分涡旋混匀，使磁珠悬浮。

2.6 取 n 个新的 1.5 ml 离心管，并向管中加入 180 μ l 吸附缓冲液和 20 μ l 上一步重悬的磁珠，充分涡旋混匀，50 °C (最终的杂交温度) 孵育 2 min。

2.7 小心地吸取杂交产物 (不包括矿物油) 至 50 °C 预热的含有磁珠的吸附缓冲液

中。然后将离心管置于振荡型恒温金属浴中,在 50 °C, 1500 × g 的条件下, 孵育 30 min。

3. 洗涤 (耗时: 约 1.5 h)

该步骤的主要目的是利用不同盐离子浓度的缓冲液对杂交产物进行洗涤, 以去除非目标捕获 (non-target) 和非特异性杂交的单链 DNA 文库, 从而提高杂交捕获的效率。在本实验方法中, 洗涤条件先宽松后严格, 首先使用高离子浓度的洗涤缓冲液 1 (1X SSC, 0.1 % SDS) 对杂交产物进行洗涤, 以去除非目标捕获的单链 DNA 文库, 然后使用 50 °C (最终杂交温度) 预热的低离子浓度的洗涤缓冲液 2 (0.1X SSC, 0.1 % SDS) 对杂交产物进行洗涤, 以去除非特异性杂交的单链 DNA 文库。

3.1 吸附反应结束后, 将离心管从振荡型恒温金属浴中取下, 并简短离心。然后将离心管置于磁力架上, 室温孵育 2 min, 待溶液变澄清, 彻底地移除上清液, 注意不要扰动磁珠。

注: 该步骤结束后, 50 °C 孵育洗涤缓冲液 2。

3.2 将离心管从磁力架上取下, 向离心管中加入 200 µl 洗涤缓冲液 1, 并使用移液器吸打混匀反应液, 使磁珠充分悬浮 (动作要轻柔, 以免产生大量气泡), 室温孵育 10 min。然后将离心管置于磁力架上, 室温孵育 2 min, 待溶液变澄清, 彻底地移除上清液, 注意不要扰动磁珠。

3.3 重复操作步骤 3.2 一次。

3.4 将离心管从磁力架上取下, 向离心管中加入 200 µl 50 °C 洗涤缓冲液 2, 并使用移液器吸打混匀反应液, 使磁珠充分悬浮 (动作要轻柔, 以免产生大量气泡), 50 °C 孵育 10 min。然后将离心管置于磁力架上, 室温孵育 2 min, 待溶液变澄清, 彻底地移除上清液, 注意不要扰动磁珠。

3.5 重复操作步骤 3.4 两次。

3.6 将离心管从磁力架上取下, 向离心管中加入 30 µl 0.1X TE 缓冲液, 并使用移液器吸打混匀, 使磁珠充分悬浮。

3.7 建议立刻进行下一步操作, 剩余的杂交产物可保存于 -20 °C (建议保存时间不超过 1 年)。

4. PCR 扩增及高通量测序 (耗时: 约 1.5 h)

该步骤的目的是通过 PCR 扩增的方式将杂交捕获到的单链 DNA 文库变为双链 DNA 文库。对于探针和捕获文库的投入量分别为 200 ng 和 500 ng 的杂交反应，本方法推荐 PCR 扩增的模板投入量为 3 μ l，扩增循环数为 10-12。PCR 扩增反应体系的其余成分（如引物和 dNTPs 等）以及扩增反应程序，均与基因组文库制备流程中的 PCR 扩增步骤一致，详细信息可参阅试剂盒说明书。需要注意的是，在添加 PCR 扩增的模板时，务必保证带有磁珠的杂交产物溶液是充分悬浮的状态。

4.1 PCR 扩增结束后，取 1 μ l 扩增产物，使用 2.0 % 的琼脂糖凝胶电泳进行检测。

4.2 使用 AMPure XP 磁珠对扩增产物进行纯化，最后使用 25 μ l 0.1X TE 缓冲液进行洗脱。

4.3 通过 NanoDrop2000 分光光度计测量纯化产物的浓度。然后将不同杂交反应的纯化的扩增产物等质量混合，并在 2 % 的琼脂糖凝胶中割胶回收 300-500 bp 范围内的 DNA 片段。

注：混合扩增产物时需要保证样本间的特异性标签是不重复的。

4.4 将胶回收产物送往公司进行 Illumina 高通量测序。

结果与分析

AFLP 基因组捕获技术可以产生匿名基因序列数据和 SNP 数据。其中，匿名基因序列数据可以用于属内种间水平上的系统发育分析，而 SNP 数据可以用于种群水平上的系统地理学分析。读者需要根据自身的研究需求进行数据的提取。目前，SNP 数据提取与分析的流程已在现有文献中被详细描述 (McKenna *et al.* 2010; Depristo *et al.* 2011; Van der Auwera *et al.* 2013)，读者可参考上述文献对 SNP 数据进行提取与分析。本文将重点介绍从 AFLP 捕获数据中获取匿名基因序列数据的生物信息学分析流程，其主要流程可概括为：数据分选与组装，直系同源序列的鉴定与提取，序列比对，数据质量控制以及系统发育分析。

一、数据分选与组装

1. 使用 Trimmomatic (v 0.39) 对测序得到的原始数据 (raw data) 进行质量控制和筛选，去除 reads 中的测序接头序列，去除平均碱基质量值低于 20 的 reads，去除 reads 首端和末端碱基质量小于 3 或 N 的碱基，得到高质量的测序数据

(clean reads)。

以处理双端测序数据为例:

```
$ java -jar trimmomatic-0.39.jar PE RawReads_R1.fq.gz RawReads_R2.fq.gz RawReads_R1_paired.fq.gz RawReads_R1_unpaired.fq.gz RawReads_R2_paired.fq.gz RawReads_R2_unpaired.fq.gz ILLUMINA:TruSeq3-PE:fa:2:30:10:2:keepBothReads LEADING:3 AVGQUAL:20 TRAILING:3 MINLEN:36
```

2. 依据各个样本的特异性标签 (barcode), 使用 fastq-multx 将 clean reads 分选到对应的实验样本中。

以分选双端测序数据为例:

```
$ fastq-multx -B barcode.txt -b CleanReads_R1_paired.fq.gz CleanReads_R2_paired.fq.gz -o %_R1_clean.fq %_R2_clean.fq
```

注:“-B”用于提供特异性标签文件, %指的是输出文件名与 *barcode* 文件中的样本名称相对应。详细使用方法可参阅 <https://github.com/brwnj/fastq-multx>。

3. 使用基因组拼接软件 SPAdes 对每个样本的 clean data 进行从头组装 (de novo assembly)。

以组装样本 A 的 clean reads 为例:

```
$ python spades.py -t 4 -m 20 -1 taxaA_R1_clean.fq -2 taxaA_R2_clean.fq -o taxaA_outdir
```

注:“-t”用于设定线程数,“-m”用于设定内存大小 (Gb),“-o”用于指定输出文件夹的路径和名称。

4. 使用 cd-hit-est 软件去除 contigs 中的冗余序列。

以处理样本 A 的 contigs 为例:

```
$ cd-hit-est -i taxaA_contig.fasta -o taxaA_contig095.fasta -c 0.95 -n 10 -M 16000 -T 4
```

注:“-c”用于设定冗余序列相似度阈值,“-M”用于设定内存大小 (Mb),“-T”用于设定线程数。

二、直系同源序列的鉴定与提取

使用同一套 AFLP 探针从不同样本的基因组文库中进行序列捕获时，不同样本间被捕获到的序列具有同源性。所以，本方法直接采取双向 BLASTN 策略从所有样本的 contigs 中鉴定直系同源序列组 (orthologous groups, OGs) (如图 3 中的步骤一)。做法是：首先从所有样本拼接的 contigs 中选择拼接总长度最长的 contigs 作为参考序列，然后将其他样本的 contigs 分别与参考序列进行双向 BLASTN。若 1 号样本的 A 序列在参考序列的第一个 BLAST Hit 是 B 序列，反过来参考序列的 B 序列在 1 号样本中的第一个 BLAST Hit 还是 A 序列，则 A 序列与 B 序列为 1 号样本与参考序列的直系同源序列。考虑到在一个 OG 中，样本间 contigs 的长度有所差异，这使它们很难进行序列比对。为了解决这个问题，本方法提出了一种修剪 OG 的策略，即在原始 OG 中寻找一个最佳区域，以降低比对的难度 (如图 3 中的步骤二)。上述所有过程可通过本实验室开发的“SearchOGs.py”脚本实现 (下载地址：<https://figshare.com/s/5a4eee383e2dc9afba53>)。

```
$ python SearchOG.py -c ./contigDir -r ./contigDir/reference.fasta -e 1E-10 -L 200
-S 0.7 -B 2 -C 0.5 -N 2 -o outdir
```

注：“-c”用于提供包含所有样本 contigs 文件在内的文件夹路径，“-r”用于提供参考序列文件的路径，“-e”用于设置 BLAST 比对的 E-value 阈值，“-L”用于设置 BLAST Hit 长度阈值，“-S”用于设置 BLAST Hit 的序列相似度阈值，“-B”用于设置最佳比对区域中上下限包含样本数目的阈值，“-C”用于设置最佳比对区域中修剪后的 contigs 长度与参考序列长度比例的阈值，“-N”用于设置直系同源序列组中所包含样本数目的阈值，“-o”用于指定输出文件夹的路径和名称。运行环境要求和参数详解请参阅该脚本的说明文件。

图 3. 鉴定直系同源序列组的流程图

三、序列比对

常用的多序列比对的软件有 Muscle (<http://www.drive5.com/muscle/>), MAFFT (<https://mafft.cbrc.jp/alignment/software/>), Bali-Phy (<http://www.bali-phy.org/>), SATé (<http://phylo.bio.ku.edu/software/sate/>), PASTA (<https://github.com/smirarab/pasta>)等。考虑到大多数被 AFLP 探针捕获的匿名基因组序列都是快速进化的非编码序列，因此，

为了提高序列比对的准确性，推荐使用针对高变序列比对专门设计的程序如 **SATé** 和 **PASTA**。

例如，使用 **PASTA** 处理一个 **OG** 文件：

```
$ python run_pasta.py -i OG1.fasta
```

注：**PASTA** 具体使用方法请参阅 <https://github.com/smirarab/pasta/blob/master/pasta-doc/pasta-tutorial.md#step-2-inspecting-the-output-of-pasta>

四、数据质量控制

为了进一步提高数据质量，确保所鉴定的直系同源基因数据集的准确性，本方法对每个 **alignment** 构建单基因树。通过分析基因树中各分枝的长度和判断基因树的拓扑结构，将枝长异常的序列，旁系同源序列以及污染序列从数据集中剔除（如图 4）。以臭蛙属为例，本方法能够从 11 个臭蛙属物种和 2 个外类群物种中鉴定出 511 个 **OGs**，共 4 142 条序列，碱基缺失率为 35.6%（包含外类群物种在内）。经过数据质量控制之后，有 11 条不合格的序列被剔除，碱基缺失率变为 35.9%。虽然数据集的碱基缺失率有所增加，但剔除枝长异常的序列，旁系同源序列以及污染序列将大大提高最终数据集的可靠性。

图 4. 基于单基因树构建的方法去除长枝，旁系同源序列以及污染序列的示意图 (Kapli et al. 2020)

以处理一个 **OG** 为例：

1. 单基因树构建：

```
$ FastTree -gtr -nt OG1_align.fasta > OG1_align.tree
```

2. 借助 Phylopypruner 去除长枝（含有长枝的基因树拓扑结构可参照图 3 中的拓

扑结构 A)，例如，删除某一枝长大于树中所有分枝标准差 3 倍的序列：

```
$ phylopypruner --trim-lb 3 OG1_align.fasta OG1_align.tree
```

注：*Phylopypruner* 具体使用方法请参阅

<https://gitlab.com/fethalen/phylopypruner/-/wikis/methods#paralogy-pruning->

3. 去除旁系同源序列或污染序列，其标准是观察单基因树拓扑结构中是否存在聚类关系怪异的分支，可参照图 4 中基因树拓扑结构 B 和 C 同正确的基因树拓扑结构之间的差异。

4. 使用 PASTA 对数据质量控制后的 alignment 再次进行比对。

```
$ python run_pasta.py -i OG1_realign.fasta
```

5. 使用 Gblocks 剔除比对模糊的区域。

```
$ Gblock OG1_realign.fasta -b5=a
```

注：*Gblock* 软件的详细参数介绍可参阅 http://molevol.cmima.csic.es/castresan/a/Gblocks/Gblocks_documentation.html

五、系统发育分析

1. 串联树构建

将所有的 OG 比对矩阵进行串联处理，生成一个大型的数据矩阵集，并使用 RAXML 构建 ML 树：

```
$ raxmlHPC-PTHREADS-AVX -T 4 -f a -x 12347 -p 12845 -# 500 -m GTRGAMMA -s CombineOGs.fasta -n CombineOGs_MLtree_500bootstrap
```

注：“-f”用于选择 RAXML 运算的算法，a 表示执行快速 *Bootstrap* 分析并搜索最佳得分的 ML 树，“-x”用于指定一个随机数作为随机种子，以启动快速 *Bootstrap* 算法，“-#”用于提供 *Bootstrap* 的次数，“-m”用于指定核苷酸或氨基酸的替代模型。RAXML 参数的详细使用说明请参阅 <https://cme.h-its.org/exelixis/resource/download/NewManual.pdf>

2. 物种树构建

- 2.1 使用 RAXML 构建每一个 OG 的单基因树，以处理一个 OG 为例：

```
$ raxmlHPC-PTHREADS-AVX -T 4 -f a -x 12347 -p 12845 -# 100 -m GTRGAMMA -s OG1_realign.fasta -n OG1_MLtree_100bootstrap
```

- 2.2 使用 ASTRAL 构建物种树：

```
$ java -jar astral.5.7.4.jar -i OG_best_trees.tre -b OG_bs_files -o OG_as-  
tral_result
```

注: *ASTRAL* 参数的详细使用说明请参阅 <https://github.com/smirarab/ASTRAL/blob/master/astral-tutorial.md>

溶液配方

1. 10X 接头淬火缓冲液配方 (表 11)

表 11. 10X 接头淬火缓冲液体系

成分	体积 (μl)	终浓度 (1 ml)
NaCl (5 M)	100	500 mM
Tris-HCl, pH 8.0 (1 M)	10	10 mM
EDTA, pH 8.0 (0.5 M)	2	1 mM
灭菌双蒸水	888	
总体积	1000	

2. 10 μM Mlul adapter 制备方法

在 0.2 ml 离心管中, 建立双链 Mlul 接头淬火体系 (表 12)。

表 12. 双链 Mlul 接头淬火体系

成分	体积 (μl)	终浓度 (100 μl)
Mlul-adapter-F (100 μM)	10	10 μM
Mlul-adapter-R (100 μM)	10	10 μM
10X 淬火缓冲液	10	1X
灭菌双蒸水	70	
总体积	100	

充分涡旋混匀, 利用下述条件进行淬火反应, 形成局部互补的 DNA 双链接头:
95 °C 5 min; 95 °C 降到 12 °C, 每秒 0.1 °C; 保持在 12 °C。反应结束后, 将退火后的接头溶液放置于-20 °C 保存

3. 10 μ M SbfI adapter 制备方法

在 0.2 ml 离心管中，建立双链 SbfI 接头淬火的体系 (表 13)。

表 13. 双链 SbfI 接头淬火体系

成分	体积 (μ l)	终浓度 (100 μ l)
SbfI-adapter-F (100 μ M)	10	10 μ M
SbfI-adapter-R (100 μ M)	10	10 μ M
10X 淬火缓冲液	10	1X
灭菌双蒸水	70	
总体积	100	

充分涡旋混匀，利用下述条件进行淬火反应，形成局部互补的 DNA 双链接头：
95 $^{\circ}$ C 5 min；95 $^{\circ}$ C 降到 12 $^{\circ}$ C，每秒 0.1 $^{\circ}$ C；保持在 12 $^{\circ}$ C。反应结束后，将退火后的接头溶液放置于-20 $^{\circ}$ C 保存。

4. 吸附缓冲液配方 (表 14)

表 14. 吸附缓冲液体系

成分	体积 (μ l)	终浓度 (1 ml)
NaCl (5 M)	400	2 M
Tris-HCl, pH 8.0 (1 M)	10	10 mM
EDTA, pH 8.0 (0.5 M)	2	1 mM
灭菌双蒸水	588	
总体积	1000	

5.洗涤缓冲液 1 (表 15)

表 15. 洗涤缓冲液 1 体系

成分	体积 (μl)	终浓度 (1 ml)
20X SSC	50	1X
10 % SDS	10	0.1 %
灭菌双蒸水	940	
总体积	1000	

6.洗涤缓冲液 2 (表 16)

表 16. 洗涤缓冲液 2 体系

成分	体积 (μl)	终浓度 (1 ml)
20X SSC	5	0.1X
10 % SDS	10	0.1 %
灭菌双蒸水	985	
总体积	1000	

致谢

实验方案摘自发表的文章 Jia-Xuan Li, Zhao-Chi Zeng, Ying-Yong Wang, Dan Liang, Peng Zhang*. (2019). [Sequence capture using AFLP-generated baits: A cost effective method for high-throughput phylogenetic and phylogeographic analysis.](#) Ecology and Evolution, 9(10), 5925-5937.

竞争性利益声明

作者声明没有利益冲突

参考文献

1. DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C. and Daly, M. J. (2011). [A framework for variation discovery and genotyping using next - generation DNA sequencing data.](#) *Nature Genetics* 43(5), 491–498.
2. Li, C., Hofreiter, M., Straube, N., Corrigan, S. and Naylor, G. J. (2013). [Capturing protein-coding genes across highly divergent species.](#) *BioTechniques* 54(6), 321-326.
3. Li, J., Zeng, Z., Wang, Y., Liang, D., Zhang, P. (2019). [Sequence capture using AFLP-generated baits: A cost effective method for high-throughput phylogenetic and phylogeographic analysis.](#) *Ecology and Evolution* 9(10): 5925-5937.
4. Kapli, P., Yang, Z. H. and Telford, M. J. (2020). [Phylogenetic tree building in the genomic age.](#) *Nature Review Genetics* 21(7): 428-444.
5. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A. and DePristo, M. a. (2010). [The genome analysis tool - kit: A MapReduce framework for analyzing next - generation DNA sequencing data.](#) *Genome Research* 20(9), 1297–1303.
6. Portik, D. M., Smith, L. L., Bi, K. (2016). [An evaluation of transcriptome-based exon capture for frog phylogenomics across multiple scales of divergence \(Class: Amphibia, Order: Anura\).](#) *Molecular Ecology Resources* 16(5):1069-83.
7. Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy - Moonshine, A. and DePristo, M. A. (2013). [From fastQ data to high - confidence variant calls: The genome analysis toolkit best practices pipeline.](#) *Current Protocols in Bioinformatics* 43, 11.10.1–11.10.33.

附录

附表 1. 引物及寡核苷酸序列信息

类型	名称	序列 (5'-3')
Mlul adapter	Mlul-adapter-F	AGTCGTCTGACAGTTACCCGATA-CACCA*A
	Mlul-adapter-R	Phosphate-CGCGTTGGTGTATCGATCTG-GACTAGCCTCTAC
Sbfl adapter	Sbfl-adapter-F	GGGTGTGACAGAAGCGAAGCCGAA-GAGTCCTGC*A
	Sbfl-adapter-R	Phosphate-GGACTCTTCGGCCATCTTAGGTGTGGTTTG
预扩增引物	Mlul-F	AGTCGTCTGACAGTTACC
	Sbfl-R	CAAACCACACCTAAGATG
选择性扩增引物	Mlul-F-SA	GTTACCCGATACACCAACGCGTA
	Mlul-F-ST	GTTACCCGATACACCAACGCGTT
	Sbfl-R-SC	Biotin-GATGGCCGAAGAGTCCTGCAGGC
	Sbfl-R-SG	Biotin-GATGGCCGAAGAGTCCTGCAGGG
寡核苷酸阻碍物	BO1.P5.F	AATGATACGGCGACCACCGAGATCTACA CTCTTTCCCTACACGACGCTCTTCCGATC T-Phosphate
	BO3.P7.part1.F	AGATCGGAAGAGCACACGTCTGAACTCC AGTCAC-Phosphate
	BO5.P7.part2.F	ATCTCGTATGCCGTCTTCTGCTTG- Phosphate

注：“*”表示乙酰化修饰；“Phosphate”表示磷酸化修饰；“Biotin”表示生物素修饰。

附表 2. AFLP 基因组捕获实验中常见问题处理

步骤	常见问题	可能原因	解决方法
AFLP 探针 制备	酶切反应结束后， DNA 依然有主带	酶切时间短	适当延长酶切反应时间 (可尝试 1-2 h)
		酶切失败	检查酶切反应体系，反应温度是否出现问题，并重复酶切实验
	没有预扩增产物	没有扩增模板	检查连接体系接头浓度或片段筛选过程是否出现问题。重新从酶切实验开始，以产生新的预扩增模板
		扩增失败	检查 PCR 扩增体系，重新进行预扩增
	有 AFLP 片段，但 长度多态性很差 或 AFLP 片段范围 偏小 (如<500bp)	基因组 DNA 质量较差	重新提取高质量的基因组 DNA
	没有 AFLP 片段	扩增失败	检查 PCR 扩增体系，重新进行选择 性扩增
杂交 捕获	没有杂交富集产物	杂交失败	1) 检查下游选择性扩增引物是否 带有生物素修饰; 2) 检查杂交体 系，杂交温度是否出现问题
		扩增失败	检查 PCR 扩增体系，重新进行扩增