

跨物种靶基因富集技术实验方案

Protocol for Cross-species Target-gene Enrichment

黄俊满, 袁昊, 李晨虹*

海洋动物系统分类与进化上海高校重点实验室, 上海海洋大学, 上海, 201306

*通讯作者邮箱: chli@shou.edu.cn

引用格式: 黄俊满, 袁昊, 李晨虹. (2021). 跨物种靶基因富集技术实验方案. *Bio-101* e1010606. Doi: 10.21769/BioProtoc. 1010606.

How to cite: Huang, J. M., Yuan, H. and Li, C. H. (2021). Protocol for Cross-species Target-gene Enrichment. *Bio-101* e1010606. Doi: 10.21769/BioProtoc. 1010606. (in Chinese)

摘要: 本实验方案介绍了一种跨物种靶基因富集的方法, 可以获得与探针相似度低至 60-70% 的目标基因序列。该方法能用于获得成千上万个单拷贝核基因编码序列, 用于系统发育基因组学 (phylogenomics) 分析; 也可以用于其它非模式生物功能基因的富集。实验方案包括三个实验方法和相关数据分析方法。1."MagNA 磁珠清洗法纯化 DNA 产物"描述了通过自配 MagNA 清洗磁珠和缓冲液, 可以几乎无成本地纯化 DNA, 用于靶基因文库构建和富集的各个步骤, 也可用于一般的 DNA 纯化需求; 2."文库构建"介绍了如何构建用于靶基因富集的 Illumina 文库; 3."靶基因富集"是关于跨物种靶基因富集实验过程的详细步骤。另外, 本方案还介绍了以单拷贝外显子为目标基因富集数据的组装和分析方法。本方案尽量选用实验室自配溶液和降低成本的设计, 并尽可能地详尽描述实验步骤, 旨在为相关科研工作者开展系统发育基因组学研究或其它方向的研究提供一种可操作、高效、低成本的解决方案。

关键词: 目标基因捕获, 外显子, 文库构建, 高通量测序, 数据组装, 系统发育基因组学

研究背景

传统分子系统学研究往往采用一个或几个基因标记, 通过设计 PCR 引物, 扩增目标基因, 然后克隆、测序。这种方法耗时费力, 一次只能获得一个样本一个或几个基因的序

列。随着测序技术的提高和测序成本的不断下降，越来越多的动物系统学和进化研究开始使用基因组水平的数据。但是目前采用全基因组测序的方法成本太高，特别是当研究涉及物种较多或者每个物种有多个样品的情况下。因此，越来越多的研究者通过使用简化基因组或靶基因富集等方法，获得基因组的部分同源序列用于系统学研究。

简化基因组依靠限制性内切酶位点来选择同源序列，此方法用于遗传距离较远的物种往往不能保证序列的同源性，还可能会导致大量数据的缺失。而通过锚定超保守序列的基因富集方法，如 **ultraconserved elements (UCEs)** 或 **anchored hybrid enrichment (AHE)** 等方法获得的大部分序列是超保守的序列和非编码的侧翼序列。超保守序列提供的系统发育信息不足，而侧翼非编码序列与编码序列相比无论是在序列比对和适用分子进化模型等方面都还不成熟，不太适合用于系统学分析。另外，无论是 **UCEs** 还是 **AHE** 可用的标记数量有限。本方案介绍一种跨物种靶基因富集技术，通过改变杂交条件，降低杂交和清洗温度等步骤获得相似度低至 60-70% 的目标基因序列。因此，该方法比 **UCEs** 或 **AHE** 更具优势，可以用于富集不同物种的编码基因，能更好地用于系统发育基因组学的研究；也可以用于获得非模式物种的目标基因序列，用于基因功能进化的研究。

本实验方案结合自主开发用于系统发育基因组学研究的分子标记，已成功应用于桑科植物 (Yao *et al.*, 2013)、蚂蚁 (Stroeher *et al.*, 2013)、绦虫 (Huan *et al.*, 2016)、线虫、纤毛虫，鱼类 (Song *et al.*, 2017, Yin *et al.*, 2019)、爬行类和哺乳类等生物类群中。本实验方案尽量采用自配溶液和降低成本的设计，可以一次性获得一百多个样品的几千个基因序列。每个样品建库、富集和测序的总成本可以控制在 300-400 元。

材料与试剂

一、"MagNA 磁珠清洗法纯化 DNA"试剂清单

1. 5% Sera-Mag Magnetic Speed-beads (Fishersci, catalog number: 09-981-123, 4 °C 储存)
2. PEG-8000 (生工, catalog number: A600433-0500, 常温储存)
3. 5 M NaCl (生工, catalog number: A100241, 常温存储, 自配至对应浓度)
4. 1 M Tris-Cl (pH = 8.0) (生工, catalog number: B548127-0500, 常温储存)
5. 0.5 M EDTA (pH = 8.0) (Invitrogen, catalog number: 15575-020, 常温存储)

二、"靶基因富集文库构建"试剂清单

1. Buffer Tango (10x) (Thermo, catalog number: BY5, -20 °C 储存)
2. dNTPs (10 mM each) (Invitrogen, catalog number: 18427088, -20 °C 储存)
3. ATP (100 mM) (Thermo, catalog number: R0441, -20 °C 储存)
4. T4 polynucleotide kinase (10 U/μl) (Thermo, catalog number: EK0031, -20 °C 储存)
5. T4 DNA polymerase (5 U/μl) (Thermo, catalog number: EP0061, -20 °C 储存)
6. T4 DNA ligase (5 U/μl) (Invitrogen, catalog number: 15224041, -20 °C 储存)
7. T4 DNA ligase buffer (10x) (T4 DNA ligase 试剂盒配套)
8. PEG-4000 (50%) (T4 DNA ligase 试剂盒配套)
9. Bsm polymerase, large fragment (8 U/μl) (Thermo, catalog number: EP0691, -20 °C 储存)
10. Bsm buffer (10x) (Bsm polymerase, large fragment (8U/μl) 试剂盒配套)
11. KAPA HiFi Hotstart Ready Mix (2x) (KAPABIOSYSTEMS, catalog number: KK2602, -20 °C 储存)

三、"靶基因富集"试剂清单

1. UltraPure™ SSPE, 20x (Invitrogen, catalog number: 15591043, 常温储存)
2. UltraPure™ 0.5M EDTA, pH 8.0 (Invitrogen, catalog number: 15575020, 常温储存)
3. Denhardt's Solution (50x) (Invitrogen, catalog number: 750018, -5 °C ~-30 °C 储存)
4. 10% SDS 溶液 (生工, catalog number: B548118-0100, 常温储存)
5. SUPERase•In™ RNase Inhibitor (20 U/μl) (Invitrogen, catalog number: AM2696, -20 °C 储存)
6. DEPC 处理水 (生工, catalog number: B501005-0500, 常温储存)
7. RNA baits (Arbor Bioscience, -20 °C 储存)

注: RNA baits 是一段带有生物素标记、与目标基因互补配对的 RNA 序列。通过杂

交反应与目标基因结合，其特有的生物素能与链霉亲和素磁珠特异结合，根据此特性清洗去除非目标序列以达到富集效果。*RNA baits* 可以根据目标基因进行设计，例如通过比较模式生物基因组寻找单拷贝核基因编码序列，设计 *RNA baits* (Li et al., 2012)。

8. Human Cot-1 DNA (Invitrogen, catalog number: 15279011, -20 °C 储存)
9. BO1 (封阻片段 1, 序列见"溶液配制")
10. BO3 (封阻片段 3, 序列见"溶液配制")
11. Dynabeads™ MyOne™ Streptavidin C1 (Invitrogen, catalog number: 65002, 2-8 °C 储存)
12. Binding Buffer (成分见"溶液配制")
13. Tween-20 (Amresco, catalog number: 0777-1L, 常温储存)
14. Wash Buffer2 (成分见"溶液配制")
15. Wash Buffer1 (成分见"溶液配制")
16. 1x TE buffer (生工, catalog number: B548106, 常温储存)
17. KAPA HiFi Hotstart Ready Mix (2x) (KAPABIOSYSTEMS, catalog number: KK2602, -20 °C 储存)
18. 20x SSC 缓冲液 (生工, catalog number: B548109-0200, 常温储存)
19. 1 M Tris-HCl 溶液, pH 7.5, 无菌 (生工, catalog number: B548124-0500, 常温储存)
20. HPLC 超纯水 (TEDIA, catalog number: WS2211-001, 常温储存)

四、其他耗材

1. 0.2 ml 薄壁 PCR 管 (Axygen, catalog number: PCR-02-L-C)
2. 各种型号的带滤芯枪头 (10 µl, 100 µl, 200 µl, 1000 µl, Axxygen)
3. 0.2 ml 96 孔 PCR 板盒 (淘宝)
4. 自制磁板 (超强磁条 (淘宝) 嵌套于 0.2 ml 96 孔 PCR 板盒)

仪器设备

1. PCR 仪 (Eppendorf AG, 22331 Hamburg)

2. 杂交仪 (UVP HB-1000 Hybridizer)
3. 0.1-2.5 µl, 0.5-10 µl, 2-20 µl, 10-100 µl, 20-200 µl, 100-1000 µl 移液枪 (Eppendorf 系列)
4. 漩涡混匀仪 (上海沪析实业有限公司, Vortex-2)
5. 微型掌上离心机 (上海沪析实业有限公司, HL-6KS)
6. Nanodrop 3300 超微量荧光分光光度计 (Thermo)
7. 破碎仪 (Covaris M220)
8. 高压蒸汽灭菌锅 (上海博讯实业有限公司)
9. 服务器 (Linux 系统, 32 G 以上的内存, 500 G 的存储)
10. 电泳仪 (BIO-RAD, PowerPac-Basic)

分析软件

1. Perl v5.18 or higher (<https://www.perl.org/get.html>)
2. trim_galore v0.4.1 or higher (<https://github.com/FelixKrueger/TrimGalore>)
3. cutadapt v1.2.1 or higher (<https://github.com/marcelm/cutadapt>)
4. USEARCH 10.0.240 or higher (<http://www.drive5.com/usearch/download.html>)
5. SGA v0.10.15 or higher (<https://github.com/jts/sga>)
6. Exonerate v2.2.0 or higher (<https://www.ebi.ac.uk/about/vertebrate-genomics/software/exonerate>)
7. Assexon (<https://github.com/yhadevol/Assexon>)
8. Mafft v7.294b or higher (<https://mafft.cbrc.jp/alignment/software/>)
9. RAxML (<https://github.com/stamatak/standard-RAxML>)
10. ASTRAL (<https://github.com/smirarab/ASTRAL>)
11. Figtree (<http://tree.bio.ed.ac.uk/software/figtree/>)

实验步骤

本实验方案包括三个部分：1. MagNA 磁珠清洗法纯化 DNA 产物；2. 文库构建；3. 靶基因富集。

一、MagNA 磁珠清洗法纯化 DNA 产物

1. 将 30 µl 的 MagNA 磁珠加入到 0.2 ml PCR 空管, 将 PCR 管放在磁板上静置片刻, 待磁珠吸附到管壁后, 去除上清液 (可丢弃吸附上的磁性不强磁珠)。

2. 以 1:0.9 的比例添加 DNA 样品与 MagNA 缓冲液 (见溶液配制), 随后涡旋振荡 PCR 管片刻, 使其充分混匀。

注: 1:0.9 的比例是根据经验值而定, 可最大化减少试剂的使用量、减少短片段的丢失, 建议读者在配制 MagNA 缓冲液后, 做浓度梯度预实验来确定最优的比例。

3. 常温下将 PCR 管置于 96 孔板上 5 min (使磁珠吸附 DNA), 用微型离心机离心 5 s。

4. 将 PCR 管转移至磁板上 5 min, 直到磁珠吸附到管壁上。随后去除上清液, 注意不要吸到磁珠。

5. 继续将 PCR 管置于磁板上, 添加 186 μ l 的 70%乙醇, 用于清洗残留的 MagNA 缓冲液。静置 1 min, 随后去除上清液。

注: 不要吹打溶液, 无需混匀磁珠。

6. 重复步骤 5 一次。

7. 将 PCR 管盖敞开, 室温静置 5 min, 使残留的乙醇彻底挥发。

8. 添加 20 μ l DEPC 处理水或 1x TE 缓冲液, 将 PCR 管从磁板转移到 96 孔板上, 用移液枪缓慢吹打溶液 3-5 次, 使其充分混匀, 室温静置 1 min (释放 DNA), 随后离心 5 s。

9. 将 PCR 管放回磁板上, 静置 1 min, 随后将上清液转移至新管中。

注: 在下面建库和富集过程中, 往往不需洗脱 DNA, 而是保留有 DNA 的磁珠, 此时省略第 8-9 步。

二、文库构建

1. DNA 片段化

- 1.1 取 0.3-1 μ g DNA 样品 (DNA 片段大于 100 bp), 用 Covaris M220 破碎仪将 DNA 破碎成 100-500 bp 片段 (破碎仪可参考如下设置: Average Incident Power: 15 Watt, Peak Incident Power: 75 Watt, Duty Factor: 20%, Cycles/Burst: 200, Duration: 360 s; 片段化 DNA 样品所需体积 270 μ l)。

- 1.2 分别取 1 μ l 片段化产物和 1 μ l 初始 DNA, 使用 1%琼脂糖凝胶电泳进行检测。

- 1.3 取 130 μ l 片段化 DNA, 正对照 (用于做正对照的样品可以是一般 PCR 产物, 如 300 bp 左右的 PCR 产物) 和负对照 (30 μ l 超纯水) 作对比实验, 用 MagNA

磁珠清洗法将片段化 DNA 吸附到磁珠上，随后去除上清液。

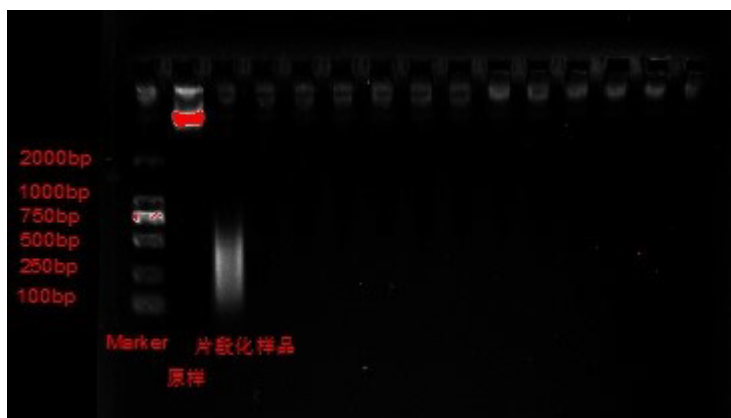


图 1. 破碎后的 DNA 示意图 图上分别为 DNA 样品和片段化 DNA。

2. 平末端修复

2.1 在冰盒上配制"平末端修复"反应溶液。将 20 μ l 的反应溶液加入每个样品 (上一步保留的磁珠)，并轻轻吹打 3-5 次混匀。

成分	体积 (μ l)	终浓度 (20 μ l 体系)
Buffer Tango (10x)	2	1x
dNTPs (10 mM each)	0.2	100 μ M each
ATP (100 mM)	0.2	1 mM
T4 polynucleotide kinase (10 U/ μ l)	1	0.5 U/ μ l
T4 DNA polymerase (5 U/ μ l)	0.4	0.1 U/ μ l
DEPC 处理水	16.2	
总体积	20	

注：建议在冰盒上进行操作

2.2 设置"平末端修复"程序如下：

步骤	循环阶段	温度	时间	循环数
1	平末端	25 $^{\circ}$ C	15 min	1
2	DNA 5'端磷酸化修饰	12 $^{\circ}$ C	5 min	1

注：反应完成后，将 PCR 管离心 5 s 以免管盖内壁附着液滴，可以减少污染。

2.3 利用 MagNA 磁珠清洗法进行纯化 (无需添加新的磁珠)，最后去除上清，保留磁珠。(仅做到磁珠清洗法的第 7 步)

注：反应完成后需立即纯化，纯化产物可以 4 °C 过夜保存。

3. 连接测序接头

3.1 在冰盒上配制"连接测序接头"反应溶液：

成分	体积 (μl)	终浓度 (40 μl 体系)
T4 DNA ligase buffer (10x)	4	1x
PEG-4000 (50%)	4	5%
Adapter mix P5 (50 μM each)	1	1.25 μM each
Adapter mix P7 (50 μM each)	1	1.25 μM each
T4 DNA ligase (5 U/μl)	1	0.125 U/μl
DEPC 处理水	29	
总体积	40	

3.2 将 40 μl 反应溶液加入上一步样品 (上一步保留的磁珠)中，轻轻吹打 3-5 次混匀。

3.3 设置"连接测序接头"程序如下：

步骤	循环阶段	温度	时间	循环数
1	连接测序接头	22 °C	30 min	1

注：反应完成后，将 PCR 管离心 5 s。

3.4 利用 MagNA 磁珠清洗法进行纯化 (无需添加新的磁珠)，最后去除上清，保留干燥的磁珠。(仅做到磁珠清洗法的第 7 步)

注：纯化产物可以 4 °C 过夜保存

4. 缺刻补齐

4.1 在冰盒上配制"缺刻补齐"反应溶液：

成分	体积 (μl)	终浓度 (40 μl 体系)
Bsm buffer (10x)	4	1x
dNTPs (10 mM each)	1	250 μM each
Bsm polymerase, large fragment (8 U/μl)	1.5	0.3 U/μl
DEPC 处理水	33.5	
总体积	40	

4.2 将 40 μl 反应溶液加入到上一步的样品 (上一步保留的磁珠) 中, 轻轻吹打 3-5 次混匀。

4.3 设置"缺刻补齐"程序如下:

步骤	循环阶段	温度	时间	循环数
1	补齐缺刻	37 °C	20 min	1

注: 反应完成后, 将 PCR 管离心 5 s。

4.4 利用 MagNA 磁珠清洗法进行纯化 (无需添加新的磁珠), 去除上清, 将 35 μl 的 1x TE 缓冲液加入到每个样品中, 保留磁珠在 PCR 管中, 在管盖上标注"样品名+lib"。此时靶基因待富集文库已构建完成, 可放于-20 °C 环境下短期储存一个月。

5. 预杂交 PCR 扩增在冰盒上配制"预杂交 PCR 扩增"反应溶液:

成分	体积 (μl)	终浓度 (25 μl 体系)
KAPA HiFi taq Ready Mix (2x)	12.5	1x
Primer IS7 (10 μM)	0.75	0.3 μM
Primer IS8 (10 μM)	0.75	0.3 μM
总体积	14	

5.2 将 14 μl 上述反应液加入到 0.2ml PCR 空管中, 再分别加入 11 μl 第四步获得

的样品 "lib"文库 (切记磁珠与溶液混匀后再吸取)。将样品与试剂吹打 3-5 次，充分混匀。设置"预杂交 PCR"程序如下：

步骤	循环阶段	温度	时间	循环数
1	预变性	98 °C	45 s	1
2	变性	98 °C	15 s	12-18 (具体
	退火	60 °C	30 s	循环数根据
	延伸	72 °C	45 s	所添加样品的量而定)
3	延伸	72 °C	1 min	1
4	降温	4 °C	5 min	1
5	保存	12 °C	∞	1

5.3 利用 MagNA 磁珠清洗法进行纯化 (需要添加新的 MagNA 磁珠！)，然后将 25 μ l 的 1x TE 缓冲液加入到每个样品管里，将上清液转移至新的 0.2 ml PCR 空管中，在管盖上标注"样品名+PreH"，可放于-20 °C 环境下短期储存一个月。

5.4 分别取 1 μ l "PreH"产物、1 μ l 正对照原液 (300 bp)，使用 1%的琼脂糖凝胶电泳进行检测。若文库构建成功，则正对照的"PreH"产物长度比正对照原液稍长，且负对照的"PreH"产物中无条带。

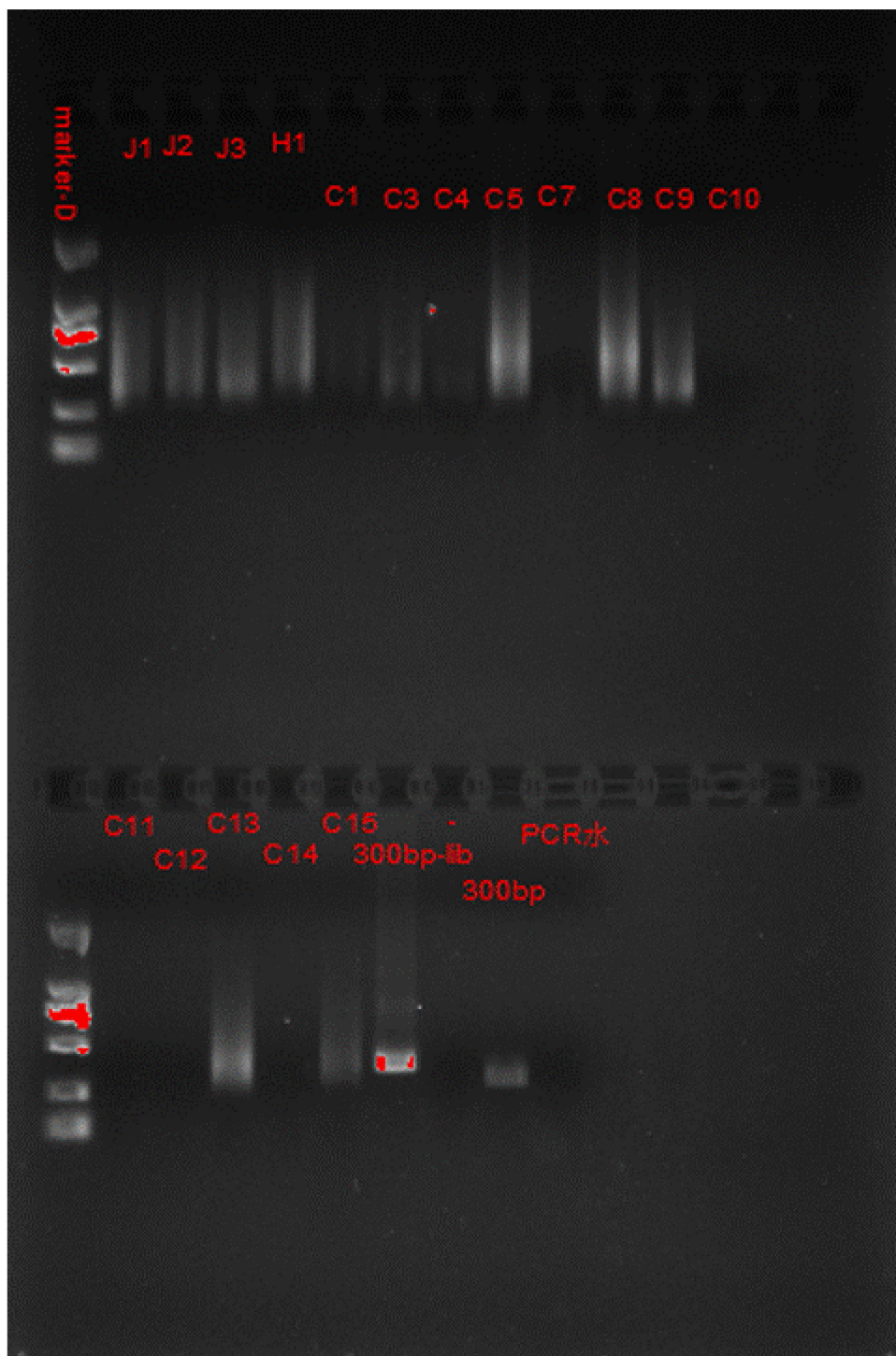


图 2. 靶基因待富集文库构建结果电泳图 "J1"- "C15"共 14 个实验样品, "300 bp-lib"为正对照, "-"为负对照, "300 bp"为建库前的正对照 DNA 原液。Marker-D 的片段大小从下至上分别为 100 bp, 250 bp, 500 bp, 750 bp, 1 k, 2 k。

三、靶基因富集

1. 杂交

1.1 设置"靶基因富集"程序:

步骤	循环阶段	温度	时间	循环数
1	变性解链	95 °C	5 min	1
2	预退火杂交	65 °C	5 min	1
3	退火杂交	65 °C	240 min	1
4	退火杂交	60 °C	240 min	1
5	退火杂交	55 °C	240 min	1
6	退火杂交	50 °C	240 min	1
	保存	50 °C	∞	

注: 共计 16 h 左右。(通过降低退火温度、设定退火梯度温度, 可杂交结合相似度低至 60-70%的序列)

1.2 在室温下配制 Hyb Baits Mix 溶液 (杂交反应体系与 RNA 酶抑制剂), 反应体系如下:

成分	体积 (μl)
HYB 20x SSPE	7
HYB 0.5 M EDTA (pH = 8.0)	0.3
HYB 50x Denhardt's	3
HYB 10% SDS	0.3
SUPERase•In (20U/μl)	1
RNA Baits (MYselect)	0.2
DEPC 处理水	4.8
总体积	16.6

注: 各试剂添加完后, 用移液枪轻轻地吹打混匀, 随后用微型离心机离心 5 s, 将溶液集中到底部。

1.3 在室温下配制 Lib Mix 溶液 (封阻简单重复序列与接头测序序列), Lib Mix 反应体系如下:

成分	体积 (μl)
Block Human Cot1 (1 μg/μl)	2.5
BO1.P5.F (200 μM)	0.5
BO3.P7.part1.R (200 μM)	0.5
总体积	3.5

各试剂添加完后，用移液枪轻轻地吹打混匀，随后离心 5 s，将溶液集中到底部。

取 0.2 ml 薄壁 PCR 空管，在管盖上标注样品名。将 3 μl 的 Lib Mix 溶液加入到每个 PCR 管中，然后再分别加入 11 μl 样品"PreH"产物 (含有至少 100 ng DNA)，用移液枪轻轻地吹打混匀，随后用微型离心机离心 5 s，将溶液集中到底部。

- 1.4 将混有 Lib Mix 和样品文库的 PCR 管放至 PCR 仪上，并启动反应程序。此时进入程序第一步，对 DNA 文库进行 95 °C 高温变性解链处理。
- 1.5 当程序到达第二步 (65 °C)，将含有 Hyb Baits Mixs 的管子载入 PCR 仪，预热 Hyb Baits Mix 5 min，期间混有 Lib Mix 和样品文库的 PCR 管始终置于 PCR 仪上。
- 1.6 5 min 后，PCR 仪上的管子不取出，继续维持 65 °C，将 16 μl 的 Hyb Baits Mix 转移到混有 Lib Mix 和样品的 PCR 管中，用移液枪轻轻吹打混匀 3-5 次。盖紧 PCR 管。
- 1.7 维持反应体系在恒温体系下，直至程序结束。

注：程序结束后，样品仍可在 PCR 仪上 50 °C 条件下放置数小时。

2. Dynabeads 链霉亲和素磁珠吸附与清洗

- 2.1 加 nx 10 μl Dynabeads 链霉亲和素磁珠到 0.2 ml PCR 新管中，每个 PCR 管中添加的磁珠不得超过 100 μl，方便后续步骤清洗。
- 2.2 将 PCR 管静置于磁板上片刻，待磁珠吸附到管壁之后，去除上清液。
- 2.3 加 100 μl 的 Binding 缓冲液于 PCR 管中清洗磁珠。用移液枪轻轻吹打混匀。继续将 PCR 管静置于磁板上 2 min，待磁珠吸附到管壁之后，去除上清液 (包含尚未被吸附到管壁上的少许磁珠)。

2.4 重复步骤 2.3 两次。

2.5 将 nx 20 µl Binding 缓冲液、1 µl 10%的 Tween 加入到步骤 2.4 的 PCR 管 (此时仅有磁珠), 轻轻吹打混匀, 将磁珠再次悬浮。

2.6 先将 180 µl 的 Binding 缓冲液加入到新的 0.2 ml PCR 空管中, 再加入 20 µl 步骤 2.5 的磁珠悬浮液。

2.7 将步骤 2.6 配制好的 200 µl 磁珠溶液放置在 PCR 仪 50 °C 加热 5 min。

2.8 将"靶基因富集"步骤 1.7 中的杂交产物加入到磁珠悬浮液中, 用移液枪轻轻吹打混匀。随后将 PCR 管放入杂交仪中, 50 °C 反应 30 min。同时在 PCR 仪上预热 3x nx 190 µl Wash Buffer 2 溶液 (提前配制 Wash buffer 1, 以 Wash buffer 1 为原料配制 Wash buffer 2), 保持 50 °C 10 min。

2.9 待步骤 2.8 杂交结束后, 用微型离心机离心 5 s, 将 PCR 管转移至磁板上 (磁板提前在杂交仪中预热), 并在杂交仪中静置片刻 (不超过 2 min), 待磁珠吸附到管壁后, 去除上清液。

2.10 将带有磁珠的 PCR 管放回 PCR 仪上 (程序维持在 50 °C)。添加 186 µl 50 °C 的 Wash Buffer 2 到 PCR 管中, 轻轻地将磁珠吹打混匀。继续在 50 °C PCR 仪上反应 10 min。待 10 min 反应结束后, 将 PCR 管静置于磁板上片刻 (不超过 2 min), 待磁珠吸附到管壁后, 去除上清液。

2.11 重复步骤 2.10 两次。

2.12 在室温下将 100 µl 的 1x TE 缓冲液加入到 PCR 管中 (用以清洗残留的 Wash Buffer 2)。将 PCR 管静置于磁板上 1 min, 待磁珠吸附到管壁后, 去除上清液。

2.13 将 35 µl 的 1x TE 缓冲液加入 PCR 管中, 在管盖上标注"样品名+1stcap", 将其放置在 -20 °C 短期保存 (溶液中依然保留有磁珠)。

注: 清洗步骤时, 减少 PCR 管暴露在室温的时间, 尽量保持反应溶液在 50 °C。

步骤 2.9 应尽量移除干净上清液。

3. 第一次杂交 PCR (为第二次杂交作准备)

3.1 在冰盒上配制 PCR 扩增反应溶液:

成分			体积 (µl)	终浓度 (25 µl 体系)
KAPA	HiFi	HotStart	12.5	1x

Ready Mix (2x)		
P5 引物 IS7 (10 μ M)	0.75	0.3 μ M
P7 引物 IS8 (10 μ M)	0.75	0.3 μ M
总体积	14	

将 14 μ l 上述所配制的 PCR 扩增反应溶液加入到新的 0.2 ml PCR 空管中，随后添加 11 μ l 第一次杂交 1stcap 产物 (确保磁珠与溶液混匀后再吸取)，在管盖上标注样品名。用移液枪轻轻吹打 3-5 下。用以下 PCR 反应程序进行扩增：

步骤	循环阶段	温度	时间	循环数
1	预变性	98 °C	45 s	1
2	变性	98 °C	15 s	12-14 (具体循环数根据所添加样品的量而定)
	退火	60 °C	30 s	
	延伸	72 °C	45 s	
	延伸	72 °C	1 min	
3	延伸	72 °C	1 min	1
4	降温	4 °C	5 min	1
5	保存	12 °C	∞	1

3.2 用 MagNA 磁珠法清洗 PCR 扩增产物 (需要添加新的 MagNA 磁珠!)。清洗结束后，用 25 μ l 的 1x TE 缓冲液进行洗脱，将上清液转移至新的 200 μ l PCR 空管中，在管盖上标注"样品名+preH2"

3.3 取 1 μ l 的 PCR 产物稀释 10 倍后，用 Nanodrop3300 分光光度计进行浓度的测定，产物稀释后浓度约在 1 ng/ μ l 左右。

4. 用步骤 3 的 PCR 产物再做一遍杂交富集，即二次杂交 (重复步骤 1 和 2)。第二次杂交并清洗后的产物命名为"样品名+2ndcap"。

5. Indexing PCR

5.1 在冰盒上配制 PCR 扩增反应溶液如下：

成分	体积 (μl)	终浓度 (25 μl 体系)
KAPA HiFi HotStart Ready Mix (2x)	12.5	1x
P5 引物 IS4 (10 μM)	0.5	0.2 μM
总体积	13	

将 13 μl 上述所配制的 PCR 扩增反应溶液加入新的 0.2 ml PCR 空管，添加 0.5 μl P7 index 扩增引物 (见溶液配制)，随后添加 11.5 μl 的第二次杂交 2ndcap 产物 (确保磁珠与溶液混匀后再吸取)。吹打混匀并离心 5 s。用以下 PCR 程序进行扩增：

步骤	循环阶段	温度	时间	循环数
1	预变性	98 °C	45 s	1
2	变性	98 °C	15 s	12-16 (具体循环数根据所添加样品的量而定)
	退火	60 °C	30 s	
	延伸	72 °C	45 s	
	延伸	72 °C	1 min	
4	降温	4 °C	5 min	1
5s	保存	12 °C	∞	1

5.2 用 MagNA 磁珠法清洗 PCR 扩增产物 (需要添加新的 MagNA 磁珠!)。清洗结束后，用 25 μl 的 1x TE 缓冲液进行洗脱，将上清液转移至新的 200 μl PCR 空管中，在管盖上标注"样品名+Ind"

5.3 分别取 1 μl PCR 产物，使用 1%的琼脂糖凝胶电泳进行检测。

6. 混样及送测

6.1 取 1 μl PCR 产物稀释 10 倍，用 Nanodrop3300 超微量荧光分光光度计进行浓度测定。确保样品稀释后浓度均在 0.1-0.9 ng/μl 范围之内。

6.2 将所有样品等摩尔量混合。

6.3 再次用 Nanodrop 3300 超微量荧光分光光度计对混合后产物进行浓度测定。终产物应有 20 μ l 体积，且浓度应在 2-50 nM (0.5-13 ng/ μ l) 之间。当满足上述条件后，即可送测。以 Illumina Hiseq PE150 测序平台为例。

数据分析

利用 (Yuan *et al.*, 2019) 发表的外显子富集数据组装流程进行富集数据的组装 (Assexon, <https://github.com/yhadevol/Assexon>)。

此富集数据组装流程共分为 3 步：数据预处理，数据组装，数据后期处理。

一、数据预处理

1. 批量解压二代测序得到的原始数据

```
$ perl gunzip_Files.pl \
--gzip raw_reads \
--gunzipped gunzipped_raw_reads
```

2. 去除接头及低质量的序列

```
$ perl trim_adaptor.pl \
--raw_reads gunzipped_raw_reads \
--trimmed trimmed
```

3. 矫正编码基因的阅读框，并翻译成氨基酸序列

```
$ perl predict_frames.pl \
--baits reference.fas \
--cds reference.onehitCDSmakers.column1.txt \
--ref_prot reference.pep.fas
```

或者手动矫正编码基因阅读框

二、数据组装

1. 检查软件是否已装齐

```
$ perl assemble.pl --check_depends
```

2. 序列组装

```
$ perl assemble.pl \
--trimmed trimmed \
--query reference.aa.fas \
```

```
--queryn reference.dna.fas \
--db reference.genome.fas \
--dbtype nucleo \
--ref_name reference_name \
--outdir assemble_result
```

三、数据后期处理

1. 添加已知基因组物种的直系同源序列，并与组装的数据合并

```
$ perl get_orthologues.pl \
--query reference.dna.fas \
--querydb reference.genome.fas \
--subdb "species1.genome.fas|species2.genome.fas" \
--subname "species1 species2" \
--outdir orthologs
$ perl merge_loci.pl \
--indir "assemble_result/nf orthologs/nf" \
--outdir merged_nf \
--min_seq 3
```

2. 删除不需要的样品序列信息

```
$ perl pick_taxa.pl \
--indir merged_nf \
--outdir merged_nf_deselected \
--deselected_taxa "species2"
```

3. 序列比对

```
$ perl mafft_aln.pl \
--dna_unaligned merged_nf \
--dna_aligned merged_nf_aligned
```

4. 过滤比对质量差的序列

```
$ perl filter.pl \
--indir merged_nf_aligned \
--filtered merged_nf_filtered \
--ref_taxa "reference_name"
```

5. 统计序列信息

```
$ perl statistics.pl \
--coding_aligned merged_nf_filtered
```

四、基因串联构建系统树

1. 串联基因

```
$ perl concat_loci.pl \
--indir merged_nf_filtered \
--outfile concat
```

2. 基于最大似然法构建基因树

```
$ raxmlHPC \
-m GTRGAMMA \
-p 12345 \
-q dna12_3.partition.txt \
-s concat.phy \
-n genetree
```

五、基于多物种溯祖模型的物种树构建

1. 为每个基因构建一个基因树

```
$ perl construct_tree.pl \
--indir merged_nf_filtered
```

2. 利用 Linux 的"cat"命令将步骤 1 生成的基因树合并到一个文件里

```
$ cat ./merged_nf_filtered/* > catree.fas
```

3. 利用 ASTRAL 构建物种树

```
$ java -jar astral.jar -l catree.fas -o species.tre
```

六、寻找 SNP 位点

1. 生成一致序列 (若参考序列已有基因组, 则可省略此步, 直接用基因组数据)

```
$ perl consensus.pl \
--indir merged_nf_filtered \
--outfile consensus.fas
```

2. 利用 GATK 软件寻找 SNP 位点

```
$ gatk.sh consensus_reference num_of_thread sample1 sample2 sample3 ...
```

3. 将 vcf 文件转格式，用于 BEAST, STRUCTURE 等分析

```
$ perl vcftosnps.pl \
--vcf species.vcf
```

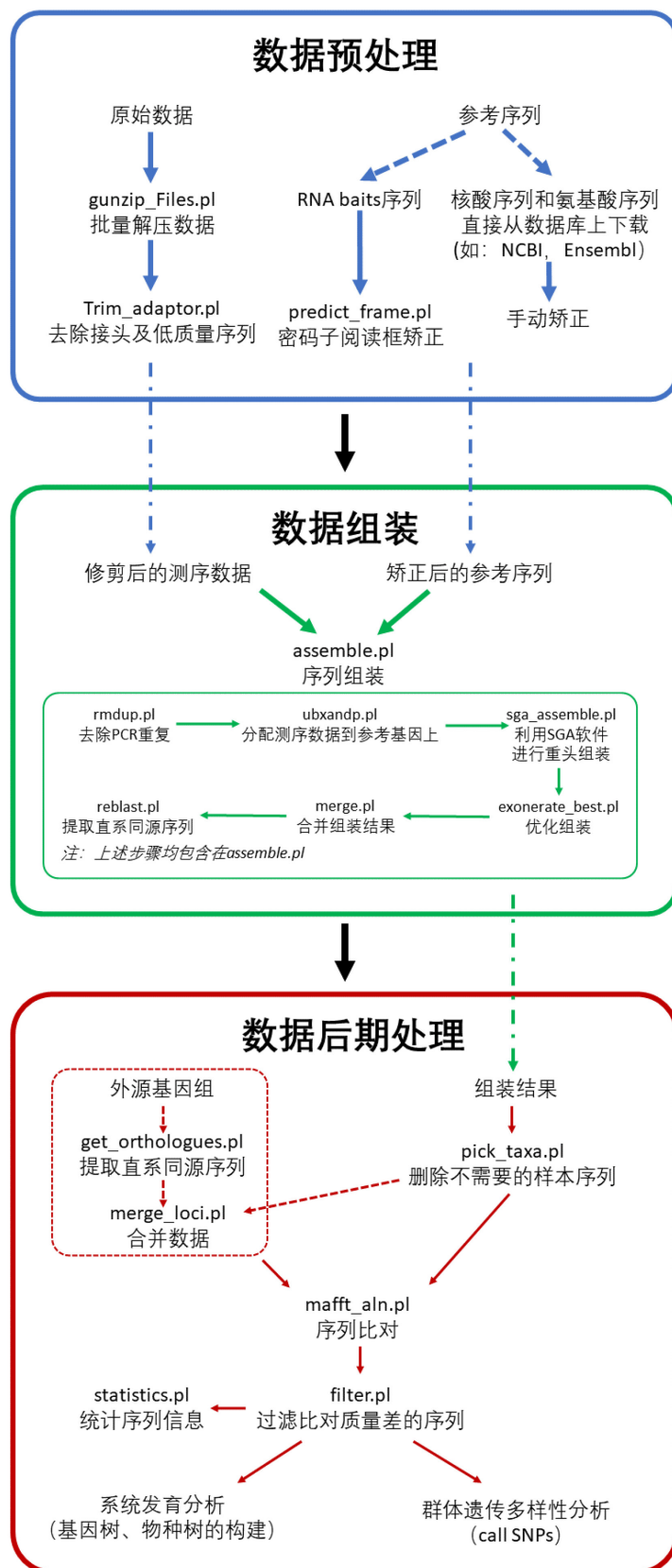


图 3. 数据分析流程图

溶液配方

一、MagNA 磁珠法清洗 DNA 溶液配方

1. 100 ml 的 MagNA beads

成分	体积	终 浓 度
5% Sera-Mag Magnetic Speed-beads (FisherSci, catalog number: 09-981-123, 4 °C 储存)	2 ml	0.1%
PEG8000 (生工, catalog number: A600433-0500, 常温储存)	18 g	18%
5 M NaCl (生工, catalog number: A100241, 常温存储, 自配至对应浓度)	50 ml	2.5 M
1 M Tris-Cl, pH 8.0 (生工, catalog number: B548127-0500)	1 ml	10 mM
0.5 M EDTA, pH 8.0 (Invitrogen, catalog number: 15575-038)	200 µl	1 mM
加超纯水补齐至 100 ml		

2. 100 ml 的 MagNA 缓冲液

成分	体积	终 浓 度
PEG8000 (生工, catalog number: A600433-0500, 常温储存)	18 g	18%
5 M NaCl (生工, catalog number: A100241, 常温存储, 自配至对应浓度)	50 ml	2.5 M
1 M Tris-Cl, pH 8.0 (生工, catalog number: B548127-0500)	1 ml	10 mM
0.5 M EDTA, pH 8.0 (Invitrogen, catalog number: 15575-038)	200 µl	1 mM

加超纯水补齐至 100 ml

二、文库构建溶液配方

1. Adapter Mix

配制下述试剂需在不同 PCR 管中进行

1.1 寡核苷酸杂交缓冲液 (10x):

成分	体积 (μl)	终浓度 (10 ml 体系)
NaCl (5 M)	1 ml	500 mM
Tris-Cl, pH 8.0 (1 M)	100 μl	10 mM
EDTA, pH 8.0 (0.5 M)	20 μl	1 mM
H ₂ O	8.88 ml	

1.2 Adapter mix P5 (100 μM):

成分	体积 (μl)	终浓度 (100 μl 体系)
IS1_adapter_P5.F (500 μM)	20	100 μM
IS3_adapter_P5+P7.R (500 μM)	20	100 μM
寡核苷酸杂交缓冲液 (10x)	10	1x
H ₂ O	50	

1.3 Adapter mix P7 (100 μM):

成分	体积 (μl)	终浓度 (100 μl 体系)
IS2_adapter_P7.F (500 μM)	20	100 μM
IS3_adapter_P5+P7.R (500 μM)	20	100 μM
寡核苷酸杂交缓冲液 (10x)	10	1x

H₂O

50

1.4 将步骤 1.2, 1.3 的溶液分别振荡混匀, 利用下述程序进行退火反应, 形成局部互补的双链接头序列

步骤	循环阶段	温度	时间	循 环 数
1	变性解链	95 °C	10 s	1
2	退火	从 95 °C 梯度降温至 12 °C, 降温速率为 0.1 °C/s		1

三、靶基因富集溶液配方

1. Binding Buffer

成分	体积 (总体积 200 ml)
5 M NaCl (生工, catalog number: A100241, 常温存储, 自配至对应浓度)	40 ml
1 M Tris-HCL (pH = 7.5) (生工, catalog number: B548124-0500)	2 ml
0.5 M EDTA (pH = 8) (Invitrogen, catalog number: 15575-038)	0.4 ml
H ₂ O (TEDIA, catalog number: WS2211-001)	157.6 ml

2. Wash Buffer1

成分	体积 (总体积 50 ml)
20x SSC (生工, catalog number: B548109-0200)	2.5 ml
10% SDS (生工, catalog number: B548118-0100)	500 µl
0.5 M EDTA (Invitrogen, catalog number: 15575-038)	100 µl

H ₂ O (TEDIA, catalog number: WS2211-001)	46.9 ml
--	---------

3. Wash Buffer2

成分	体积 (总体积 50 ml)
Wash Buffer 1	5 ml
10% SDS (生工, catalog number: B548118-0100)	0.5 ml
H ₂ O (TEDIA, catalog number: WS2211-001)	44.5 ml

序列信息

1. 接头与引物

Name	Sequences
IS1_adapter.P5	a*c*a*c*TCTTCCCTACACGACGCTCTTCCg*a*t*c*t
IS2_adapter.P7	g*t*g*a*CTGGAGTTCAGACGTGTGCTCTTCCg*a*t*c*t
IS3_adapter.P5+	
P7	a*g*a*t*CGGAa*g*a*g*c AATGATACGGCGACCAACGAGATCTACACTCTTCCCTA
IS4_indPCR.P5	CACGACGCTCTT
IS7_short_amp.P	
5	ACACTCTTCCCTACACGAC
IS8_short_amp.P	
7	GTGACTGGAGTTCAGACGTGT
Index_primer_8n	CAAGCAGAAGACGGCATACGAGATxxxxxxxxGTGACTGG
t_XXX	AGTTCAGACGTGT

注: *代表的是 *PTO* 键 (硫代修饰键)

2. BLOCK

类型	序列
BO1.P5.F	AATGATACGGCGACCAACGAGATCTACACTCTTCCCTACAC GACGCTCTTCCGATCT-Pho

BO3.P7.pa AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC-Pho
rt1.F

3. P7 Index Primer 8nt 信息

ID	P7 Index name	P7 primer seq
1	index_8nt_1	CAAGCAGAAGACGGCATACGAGATcggttggtGTGACTGGA GTTTCAGACGTGT
2	index_8nt_2	CAAGCAGAAGACGGCATACGAGATttctggttGTGACTGGAG TTCAGACGTGT
3	index_8nt_3	CAAGCAGAAGACGGCATACGAGATgcatggttGTGACTGGA GTTTCAGACGTGT
4	index_8nt_4	CAAGCAGAAGACGGCATACGAGATgttcggttGTGACTGGA GTTTCAGACGTGT
5	index_8nt_5	CAAGCAGAAGACGGCATACGAGATtggcggttGTGACTGGA GTTTCAGACGTGT
6	index_8nt_6	CAAGCAGAAGACGGCATACGAGATaaccggttGTGACTGGA GTTTCAGACGTGT
7	index_8nt_7	CAAGCAGAAGACGGCATACGAGATtctaggttGTGACTGGA GTTTCAGACGTGT
8	index_8nt_8	CAAGCAGAAGACGGCATACGAGATctgaggttGTGACTGGA GTTTCAGACGTGT
9	index_8nt_9	CAAGCAGAAGACGGCATACGAGATggcaggttGTGACTGGA GTTTCAGACGTGT
10	index_8nt_10	CAAGCAGAAGACGGCATACGAGATacttcggttGTGACTGGA GTTTCAGACGTGT
11	index_8nt_11	CAAGCAGAAGACGGCATACGAGATtagtcggttGTGACTGGA GTTTCAGACGTGT
12	index_8nt_12	CAAGCAGAAGACGGCATACGAGATctatcggttGTGACTGGA GTTTCAGACGTGT

13	index_8nt_13	CAAGCAGAAGACGGCATAACGAGATgatgcggttGTGACTGGA GTTTCAGACGTGT
14	index_8nt_16	CAAGCAGAAGACGGCATAACGAGATgcgccggttGTGACTGGA GTTTCAGACGTGT
15	index_8nt_68	CAAGCAGAAGACGGCATAACGAGATcggttggtGTGACTGGA GTTTCAGACGTGT
16	index_8nt_89	CAAGCAGAAGACGGCATAACGAGATaagttcgtGTGACTGGA GTTTCAGACGTGT
17	index_8nt_128	CAAGCAGAAGACGGCATAACGAGATccggttctGTGACTGGA GTTTCAGACGTGT
18	index_8nt_129	CAAGCAGAAGACGGCATAACGAGATgtcggttctGTGACTGGA GTTTCAGACGTGT
19	index_8nt_179	CAAGCAGAAGACGGCATAACGAGATtcagtactGTGACTGGA GTTTCAGACGTGT
20	index_8nt_184	CAAGCAGAAGACGGCATAACGAGATgttcaactGTGACTGGA GTTTCAGACGTGT
21	index_8nt_200	CAAGCAGAAGACGGCATAACGAGATatacctatGTGACTGGA GTTTCAGACGTGT
22	index_8nt_208	CAAGCAGAAGACGGCATAACGAGATtgcttgatGTGACTGGA GTTTCAGACGTGT
23	index_8nt_220	CAAGCAGAAGACGGCATAACGAGATgcagtcacGTGACTGGA GTTTCAGACGTGT
24	index_8nt_229	CAAGCAGAAGACGGCATAACGAGATtcgataatGTGACTGGA GTTTCAGACGTGT
25	index_8nt_233	CAAGCAGAAGACGGCATAACGAGATtaatcaatGTGACTGGA GTTTCAGACGTGT
26	index_8nt_236	CAAGCAGAAGACGGCATAACGAGATagagggttgGTGACTGG AGTTCAGACGTGT
27	index_8nt_237	CAAGCAGAAGACGGCATAACGAGATcctcggttgGTGACTGGA GTTTCAGACGTGT
28	index_8nt_254	CAAGCAGAAGACGGCATAACGAGATctcaattgGTGACTGGA GTTTCAGACGTGT

29	index_8nt_256	CAAGCAGAAGACGGCATAACGAGATaaggtctgGTGACTGGA GTTTCAGACGTGT
30	index_8nt_288	CAAGCAGAAGACGGCATAACGAGATctgggttgGTGACTGGA GTTTCAGACGTGT
31	index_8nt_298	CAAGCAGAAGACGGCATAACGAGATccagctggGTGACTGG AGTTCAGACGTGT
32	index_8nt_301	CAAGCAGAAGACGGCATAACGAGATgactatggGTGACTGGA GTTTCAGACGTGT
33	index_8nt_302	CAAGCAGAAGACGGCATAACGAGATctatatggGTGACTGGA GTTTCAGACGTGT
34	index_8nt_303	CAAGCAGAAGACGGCATAACGAGATtaagatggGTGACTGG AGTTCAGACGTGT
35	index_8nt_309	CAAGCAGAAGACGGCATAACGAGATaggatcggGTGACTGG AGTTCAGACGTGT
36	index_8nt_316	CAAGCAGAAGACGGCATAACGAGATgtcgccggGTGACTGG AGTTCAGACGTGT
37	index_8nt_327	CAAGCAGAAGACGGCATAACGAGATcgtcgaggGTGACTGG AGTTCAGACGTGT
38	index_8nt_332	CAAGCAGAAGACGGCATAACGAGATttggcaggGTGACTGGA GTTTCAGACGTGT
39	index_8nt_335	CAAGCAGAAGACGGCATAACGAGATattcaaggGTGACTGGA GTTTCAGACGTGT
40	index_8nt_337	CAAGCAGAAGACGGCATAACGAGATaatattcgGTGACTGGA GTTTCAGACGTGT
41	index_8nt_340	CAAGCAGAAGACGGCATAACGAGATaggcgtcgGTGACTGG AGTTCAGACGTGT
42	index_8nt_347	CAAGCAGAAGACGGCATAACGAGATtccgatcgGTGACTGGA GTTTCAGACGTGT
43	index_8nt_362	CAAGCAGAAGACGGCATAACGAGATgagtcgcgGTGACTGG AGTTCAGACGTGT
44	index_8nt_389	CAAGCAGAAGACGGCATAACGAGATccgcaacgGTGACTGG AGTTCAGACGTGT

45	index_8nt_398	CAAGCAGAAGACGGCATAACGAGATatgaatagGTGACTGG AGTTCAGACGTGT
46	index_8nt_407	CAAGCAGAAGACGGCATAACGAGATacctcgagGTGACTGG AGTTCAGACGTGT
47	index_8nt_420	CAAGCAGAAGACGGCATAACGAGATatgctcagGTGACTGGA GTTTCAGACGTGT
48	index_8nt_427	CAAGCAGAAGACGGCATAACGAGATttatccagGTGACTGGA GTTTCAGACGTGT
49	index_8nt_432	CAAGCAGAAGACGGCATAACGAGATtaacgaagGTGACTGG AGTTCAGACGTGT
50	index_8nt_434	CAAGCAGAAGACGGCATAACGAGATgactcaagGTGACTGG AGTTCAGACGTGT
51	index_8nt_438	CAAGCAGAAGACGGCATAACGAGATaacggttcGTGACTGGA GTTTCAGACGTGT
52	index_8nt_439	CAAGCAGAAGACGGCATAACGAGATcagcggttcGTGACTGGA GTTTCAGACGTGT
53	index_8nt_440	CAAGCAGAAGACGGCATAACGAGATatacggttcGTGACTGGA GTTTCAGACGTGT
54	index_8nt_455	CAAGCAGAAGACGGCATAACGAGATcgtacgttcGTGACTGGA GTTTCAGACGTGT
55	index_8nt_475	CAAGCAGAAGACGGCATAACGAGATttatcatcGTGACTGGA GTTTCAGACGTGT
56	index_8nt_476	CAAGCAGAAGACGGCATAACGAGATgaggcatcGTGACTGG AGTTCAGACGTGT
57	index_8nt_497	CAAGCAGAAGACGGCATAACGAGATaacctggcGTGACTGG AGTTCAGACGTGT
58	index_8nt_502	CAAGCAGAAGACGGCATAACGAGATagttaggcGTGACTGGA GTTTCAGACGTGT
59	index_8nt_509	CAAGCAGAAGACGGCATAACGAGATcgatatcgcGTGACTGGA GTTTCAGACGTGT
60	index_8nt_517	CAAGCAGAAGACGGCATAACGAGATgctaacgcGTGACTGG AGTTCAGACGTGT

61	index_8nt_	CAAGCAGAAGACGGCATACTAGAGATagcagagcGTGACTGG
522		AGTTCAGACGTGT
62	index_8nt_	CAAGCAGAAGACGGCATACTAGAGATtgaccagcGTGACTGG
524		AGTTCAGACGTGT
63	index_8nt_	CAAGCAGAAGACGGCATACTAGAGATggagaagcGTGACTGG
526		AGTTCAGACGTGT
64	index_8nt_	CAAGCAGAAGACGGCATACTAGAGATccaattccGTGACTGGA
530		GTTTCAGACGTGT
65	index_8nt_	CAAGCAGAAGACGGCATACTAGAGATttcaagccGTGACTGGA
549		GTTTCAGACGTGT
66	index_8nt_	CAAGCAGAAGACGGCATACTAGAGATttggtaccGTGACTGGA
551		GTTTCAGACGTGT
67	index_8nt_	CAAGCAGAAGACGGCATACTAGAGATattctaccGTGACTGGA
553		GTTTCAGACGTGT
68	index_8nt_	CAAGCAGAAGACGGCATACTAGAGATcaactaccGTGACTGG
554		AGTTCAGACGTGT
69	index_8nt_	CAAGCAGAAGACGGCATACTAGAGATcgccgaccGTGACTGG
559		AGTTCAGACGTGT
70	index_8nt_	CAAGCAGAAGACGGCATACTAGAGATgactaaccGTGACTGG
560		AGTTCAGACGTGT
71	index_8nt_	CAAGCAGAAGACGGCATACTAGAGATgctgaaccGTGACTGG
561		AGTTCAGACGTGT
72	index_8nt_	CAAGCAGAAGACGGCATACTAGAGATgccaagacGTGACTGG
582		AGTTCAGACGTGT
73	index_8nt_	CAAGCAGAAGACGGCATACTAGAGATaacttaacGTGACTGGA
583		GTTTCAGACGTGT
74	index_8nt_	CAAGCAGAAGACGGCATACTAGAGATtattgaacGTGACTGGA
585		GTTTCAGACGTGT
75	index_8nt_	CAAGCAGAAGACGGCATACTAGAGATgttgcaacGTGACTGGA
589		GTTTCAGACGTGT
76	index_8nt_	CAAGCAGAAGACGGCATACTAGAGATggctgctaGTGACTGGA
604		GTTTCAGACGTGT

77	index_8nt_	CAAGCAGAAGACGGCATAACGAGATcctaactaGTGACTGGA
610		G TTCAGACGTGT
78	index_8nt_	CAAGCAGAAGACGGCATAACGAGATagcattgaGTGACTGGA
620		G TTCAGACGTGT
79	index_8nt_	CAAGCAGAAGACGGCATAACGAGATactactgaGTGACTGGA
625		G TTCAGACGTGT
80	index_8nt_	CAAGCAGAAGACGGCATAACGAGATcttcatgaGTGACTGGA
627		G TTCAGACGTGT
81	index_8nt_	CAAGCAGAAGACGGCATAACGAGATtccaacgaGTGACTGG
647		A GTTCAGACGTGT
82	index_8nt_	CAAGCAGAAGACGGCATAACGAGATcgaatagaGTGACTGG
650		A GTTCAGACGTGT
83	index_8nt_	CAAGCAGAAGACGGCATAACGAGATtgcgtccaGTGACTGGA
678		G TTCAGACGTGT
84	index_8nt_	CAAGCAGAAGACGGCATAACGAGATgtactccaGTGACTGGA
679		G TTCAGACGTGT
85	index_8nt_	CAAGCAGAAGACGGCATAACGAGATcgtcgccaGTGACTGG
682		A GTTCAGACGTGT
86	index_8nt_	CAAGCAGAAGACGGCATAACGAGATtaacgccaGTGACTGG
683		A GTTCAGACGTGT
87	index_8nt_	CAAGCAGAAGACGGCATAACGAGATaattaccaGTGACTGGA
684		G TTCAGACGTGT
88	index_8nt_	CAAGCAGAAGACGGCATAACGAGATtcataccaGTGACTGGA
685		G TTCAGACGTGT
89	index_8nt_	CAAGCAGAAGACGGCATAACGAGATcaggaccaGTGACTGG
686		A GTTCAGACGTGT
90	index_8nt_	CAAGCAGAAGACGGCATAACGAGATgtcaaccaGTGACTGG
687		A GTTCAGACGTGT
91	index_8nt_	CAAGCAGAAGACGGCATAACGAGATagaagtaaGTGACTGG
691		A GTTCAGACGTGT
92	index_8nt_	CAAGCAGAAGACGGCATAACGAGATcatcctaaGTGACTGGA
693		G TTCAGACGTGT

93	index_8nt_	CAAGCAGAAGACGGCATAACGAGATgacgataaGTGACTGG
695		AGTTCAGACGTGT
94	index_8nt_	CAAGCAGAAGACGGCATAACGAGATtcatcgaaGTGACTGGA
700		GTTTCAGACGTGT
95	index_8nt_	CAAGCAGAAGACGGCATAACGAGATccgacgaaGTGACTGG
702		AGTTCAGACGTGT
96	index_8nt_	CAAGCAGAAGACGGCATAACGAGATtggatcaaGTGACTGGA
706		GTTTCAGACGTGT
97	index_8nt_	CAAGCAGAAGACGGCATAACGAGATgcgcgcaaGTGACTGG
708		AGTTCAGACGTGT
98	index_8nt_	CAAGCAGAAGACGGCATAACGAGATagctcaaGTGACTGG
709		AGTTCAGACGTGT
99	index_8nt_	CAAGCAGAAGACGGCATAACGAGATtctgcaaGTGACTGGA
710		GTTTCAGACGTGT
100	index_8nt_	CAAGCAGAAGACGGCATAACGAGATggagccaaGTGACTGG
711		AGTTCAGACGTGT

注：为了确保测序样品的 *Index* 碱基组成平衡，减少因 *index* 碱基不平衡而造成的测序错误。当样品数量小于 30 h，推荐使用下列任一 *index* 编号组合：

3.1 10 个样品

Set I: 27, 100, 208, 236, 336, 526, 527, 533, 685, 702

Set II: 1, 2, 68, 432, 436, 442, 554, 647, 687, 693

Set III: 10, 138, 254, 423, 435, 440, 527, 551, 639, 679

3.2 15 个样品

Set I: 1, 10, 89, 138, 254, 423, 435, 440, 527, 549, 551, 625, 639, 679, 711

Set II: 1, 10, 93, 233, 236, 285, 325, 517, 530, 533, 571, 654, 679, 687, 696

Set III: 2, 7, 9, 11, 288, 420, 435, 524, 526, 530, 559, 676, 684, 693, 695

3.3 20 个样品

Set I: 1, 2, 4, 124, 130, 255, 347, 420, 434, 435, 441, 522, 526, 527, 547, 647, 668, 686, 691, 693

Set II: 2, 3, 12, 93, 138, 237, 255, 309, 388, 419, 527, 533, 549, 558, 589, 625, 639, 695, 706, 711

Set III: 1, 2, 7, 9, 11, 288, 335, 420, 431, 435, 524, 526, 530, 547, 559, 647, 676, 684, 693, 695

3.4 25 个样品

Set I: 3, 7, 10, 13, 26, 200, 229, 235, 288, 325, 339, 381, 432, 487, 517, 522, 527, 530, 546, 609, 686,

687, 688, 703, 709

Set II: 1, 9, 20, 100, 129, 229, 332, 336, 383, 407, 421, 435, 440, 443, 513, 526, 527, 533, 546, 603,

628, 676, 684, 700, 706

Set III: 1, 8, 10, 11, 16, 129, 219, 254, 332, 334, 337, 362, 400, 517, 522, 526, 527, 565, 567, 654, 679,

683, 685, 691, 710

3.5 30 个样品

Set I: 1, 7, 10, 16, 50, 100, 129, 130, 255, 336, 339, 383, 388, 419, 431, 513, 517, 526, 527, 533, 547,

588, 617, 633, 639, 650, 688, 693, 695, 707

Set II: 2, 3, 4, 7, 27, 89, 134, 229, 255, 303, 313, 332, 388, 435, 436, 527, 530, 533, 546, 571, 582,

587, 604, 609, 620, 638, 672, 686, 693, 707

Set III: 1, 2, 4, 13, 27, 208, 219, 229, 237, 332, 340, 389, 407, 431, 435, 438, 482, 515, 526, 527, 530,

554, 567, 603, 633, 647, 678, 686, 687, 693

注：当样品数量超过 30 h，可选取任一 Index。

致谢

实验方案中文库构建和跨物种靶基因富集步骤根据作者 2013 年发表的方法修改 (Li *et al.*, 2013)，数据分析参考作者 2019 年发表的方法 (Yuan *et al.*, 2019)。

竞争性利益声明

作者声明没有利益冲突

参考文献

1. Meyer, M. and Kircher, M. (2010). [Illumina sequencing library preparation for highly multiplexed target capture and sequencing](#). *Cold Spring Harb Protoc* prot5448.
2. Li, C., Riethoven, J. J. M. and Naylor G. J. P. (2012). [EvolMarkers: a database for mining exon and intron markers for evolution, ecology and conservation studies](#). *Mol Ecol Resour* 12(5): 967-971.
3. Li, C., Hofreiter, M., Straube, N., Corrigan, S. and Naylor, G. J. P. (2013). [Capturing protein-coding genes across highly divergent species](#). *BioTechniques* 54: 321-326.
4. Yao, X., Li, C. and Dick, C. W. (2013). [Exon-primed intron-crossing \(EPIC\) markers for evolutionary studies of Ficus and other taxa in the fig family \(Moraceae\)](#). *Appl Plant Sci* 1(10): 1-4.
5. Stroeher, P. R., Li, C. and Pie, M. R. (2013). [Exon-primed intron-crossing \(EPIC\) markers as a tool for ant phylogeography](#). *Rev Bras Entomol* 57(4): 427-430.
6. Yuan, H., Jiang, J., Jimenez, F. A., Hoberg, E. P., Cook, J. A., Galbreath, K. E. and Li, C. H. (2016). [Target gene enrichment in the cyclophyllidean cestodes, the most diverse group of tapeworms](#). *Mol Ecol Resour* 16(5): 1095-1106.
7. Song, S., Zhao, J. and Li, C. (2017). [Species delimitation and phylogenetic reconstruction of the sinipercids \(Perciformes: Sinipercidae\) based on target enrichment of thousands of nuclear coding sequences](#). *Mol Phylogenet Evol* 111: 44-55.
8. Yin, G., Pan, Y., Sarker, A., Baki, M. A., Kim, J. K., Wu, H. and Li, C. (2019). [Molecular systematics of *Pampus* \(Perciformes: Stromateidae\) based on thousands of nuclear loci using target-gene enrichment](#). *Mol Phylogenet Evol* 140: 106595.
9. Yuan, H., Atta, C., Tornabene, L. and Li, C. (2019). [Assexon: assembling exon using gene capture data](#). *Evol Bioinform* 15: 1-13.