Protocol of Amplicon capture

JiaXuan Li, GuangCheng Han, Xiao Tian, Dan Liang*, Peng Zhang*

State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-Sen University

*Corresponding author:
Dan Liang. Email: liangd8@mail.sysu.edu.cn
Peng Zhang. Email: zhangp35@mail.sysu.edu.cn

Table of Contents

Materials and reagents

1. Centrifuge tube, 1.5 mL, 200 μL (BBI Life Sciences Corporation)
2. PCR 8-Tube and Cap Strip, 200 μL (BBI Life Sciences Corporation)
3. 96-Well Sample Plate (BBI Life Sciences Corporation)
4. Tip, 10 μL, 200 μL, 1 mL (BBI Life Sciences Corporation)
5. Agarose, Regular (Life Technologies)
6. Mineral oil (Sangon Biotech, Shanghai, China)
7. TransTaq-T DNA Polymerase (5U) (TransGen Biotech, Beijing, China)
8. TransTaq-T buffer (10×) (TransGen Biotech, Beijing, China)
9. dNTPs (10 mM) (Sangon Biotech, Shanghai, China)
10. AMPure XP beads (Beckman Coulter Inc.)
11. Ethanol absolute (Sangon Biotech, Shanghai, China)
12. ATP (100 mM) (Sangon Biotech, Shanghai, China)
13. T4 Polynucleotide Kinase (10 U/μL) (Fermentas)
14. T4 Polynucleotide Kinase buffer (10×) (Fermentas)
15. T4 DNA ligase (5 U/μL) (Fermentas)
16. T4 DNA ligase buffer (10×) (Fermentas)
17. PEG-4000 (50%) (Sangon Biotech, Shanghai, China)
18. NaCl (5 M) (Sangon Biotech, Shanghai, China)
19. Tris-HCl, pH 8.0 (1 M) (Sangon Biotech, Shanghai, China)
20. EDTA, pH 8.0 (0.5 M) (Sangon Biotech, Shanghai, China)
21. NEBNext Ultra DNA Library Prep Kit (New England BioLabs Inc.)
22. Human Cot1 (1 μg/μL) (Life Technologies)
23. SSPE (20×) (Sangon Biotech, Shanghai, China)
24. Denhardt's (50×) (Sangon Biotech, Shanghai, China)
25. SDS (10%) (Sangon Biotech, Shanghai, China)
26. SSC (20×) (Sangon Biotech, Shanghai, China)
27. Dynabeads MyOne Streptavidin C1 beads (Life Technologies)
28. Tween-20 (10%) (Sangon Biotech, Shanghai, China)
29. FastPure®gel DNA Extraction Mini Kit (Vazyme, Nanjing, China)
30. 1× TE Buffer (Sangon Biotech, Shanghai, China)
31. Bio-T linker (see Appendix)
32. Binding Buffer (see Appendix)
33. Washing Buffer 1 (see Appendix)
34. Washing Buffer 2 (see Appendix)

Instruments and equipment

1. 96-well Thermal circulator (VeriFlex™, Applied Biosystems)
2. Vortex-Genie2 (Scientific Industries)
3. Micro centrifuge (Pico 17, ThermoFisher scientific)
4. Scientz18-A ultrasonic processor (SCIENTZ, Zhejiang Province, China)
5. DynaMag™-2 (ThermoFisher scientific)
6. Thermo Shaker Incubator (MS-100, ALLSHENG, Zhejiang Province, China)
7. Low speed mini centrifuge (Mini-6KS, ALLSHENG, Zhejiang Province, China)
8. NanoDrop 2000 (ThermoFisher scientific)
9. Electrophoresis System (BAYGENE, Beijing, China)
10. Computing server (Linux operating system, ≥64G RAM, ≥500G disk space)

Software used in capture data analysis

1. Trimmomatic: http://www.usadellab.org/cms/?page=trimmomatic
2. MetaSPAdes: https://github.com/ablab/spades/
3. CD-HIT: https://github.com/weizhongli/cdhit/
4. BWA: https://github.com/lh3/bwa
5. SAMtools: https://github.com/samtools/samtools .
6. BLAST: https://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
7. Exonerate: https://github.com/nathanweeks/exonerate.git
8. MAFFT: https://mafft.cbrc.jp/alignment/software/
9. Gblocks: http://molevol.cmima.csic.es/castresana/Gblocks.html
10. Treeshrink: https://github.com/smirarab/TreeShrink
11. IQ-tree2: https://github.com/iqtree/iqtree2

Schematic overview of Amplicon capture

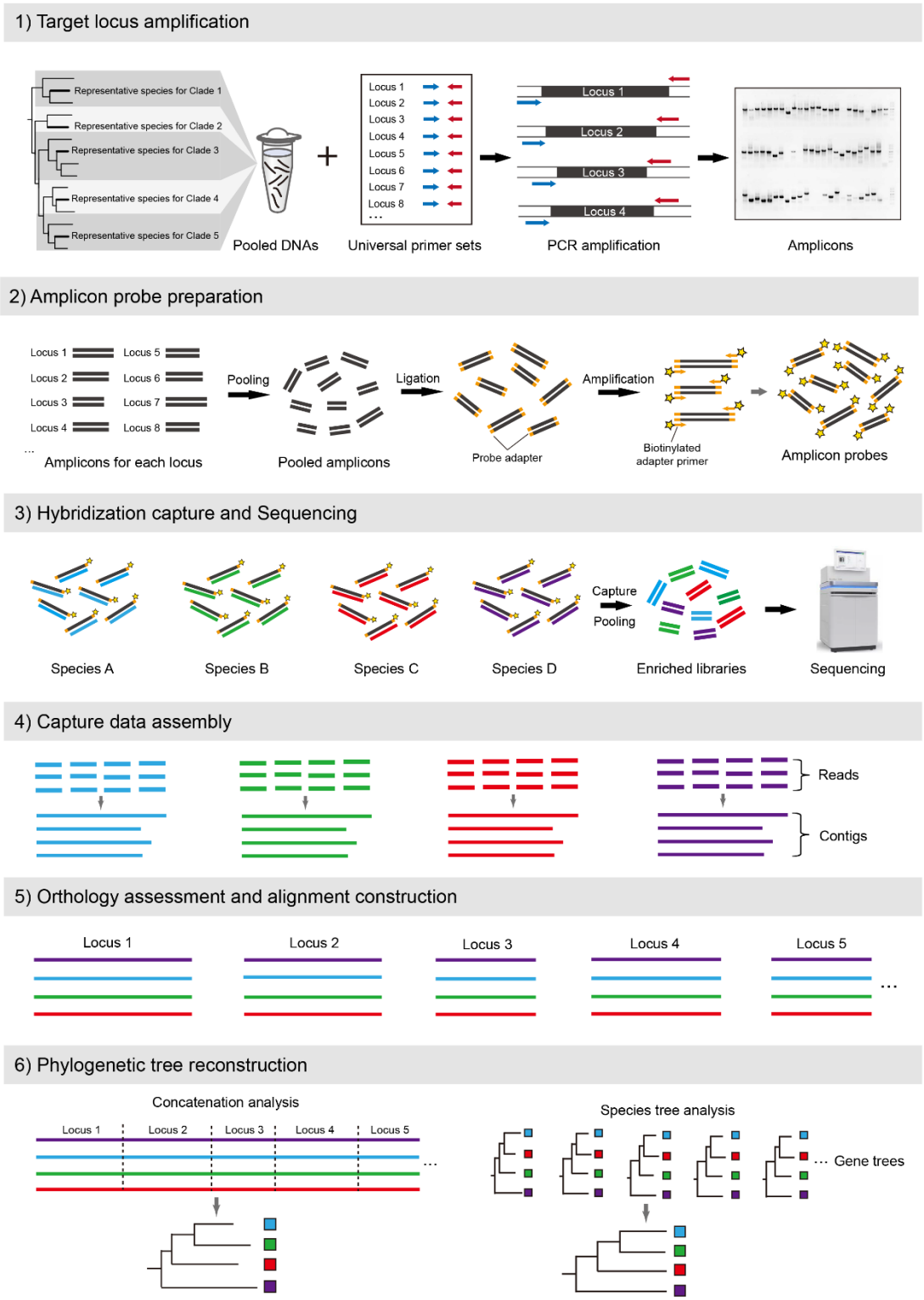

Figure 1. Flowchart of Amplicon capture.

Experimental procedure

1. Target locus amplification (Figure 1-1)

1.1 Development of NPCL primers by UPrimer for the target group

We present a UPrimer manual to guide researchers to develop NPCL primers for the target group (https://github.com/zhangpenglab/UPrimer). Once the primer design is completed, researchers need use the '***Highest-scoring_nested-PCR_primer_set_xxx.xls***' primer table to select target NPCLs. Next, researchers obtain the primer synthesis information (primer name and sequence) for selected NPCLs from the '***Synthesized_highest-scoring_nested-PCR_primer_xxx.xls***' table and provide this information to the company to synthesize primers.

1.2 Pooled DNA preparation

Prior to target NPCL amplification, it is necessary to prepare pooled DNAs comprising representative DNAs from different clades (Figure 1-1). To achieve this, researchers need to mix DNA solutions of all representative species in equal amounts and quantify the concentration of the resulting mixed DNA solution to ***10 ng/μL***.

1.3 Target NPCL amplification

Each target NPCL is amplified using two pairs of primers. The first pair, referred to as F1 and R1, is utilized in the initial PCR with pooled DNAs as a template, enriching the target region from the complex genomic background. The second pair, named F2 and R2, is used in the second PCR with the first-round PCR products as a template, specifically amplifying the target region.

First round PCR:

- Prepare a 25 μL PCR reaction as follows:

Reagent	Volume (μL)	Final concentration in 25 μL reaction
HiFi buffer (10×)	2.5	1×
dNTPs (10 mM each)	0.5	200 μM
HiFi Taq (5U/μL)	0.25	0.05 U/μL
Primer F1 (10 μM)	1.0	0.4 μM
Primer R1 (10 μM)	1.0	0.4 μM
Pooled DNAs (10 ng/μL)	1.0	
ddH_2O	18.75	
Total	25	

- Mix by pipetting followed by a quick spin to collect all liquid from the sides of the tube.

- Place in a thermocycler, with the heated lid on, and run the following program:

CYCLE STEP	TEMP	TIME	CYCLES
Initial denaturation	94°C	4 min	1
Denaturation Annealing Extension	94°C 50°C 72°C	45 seconds 40 seconds 2 minutes	35 cycles
Final extension	72°C	10 minutes	1
Hold	10°C	∞	

Second round PCR：

1) Prepare a 25 μL PCR reaction as follows:

Reagent	**Volume (μL)**	**Final concentration in 25 μL reaction**
HiFi buffer (10×)	2.5	1×
dNTPs (10 mM each)	0.5	200 μM
HiFi Taq (5U/μL)	0.25	0.05 U/μL
Primer F2 (10 μM)	1	0.4 μM
Primer R2 (10 μM)	1	0.4 μM
1st PCR products	1	
ddH_2O	18.75	
Total	25	

2) Mix by pipetting followed by a quick spin to collect all liquid from the sides of the tube.

3) Place in a thermocycler, with the heated lid on, and run the following program:

CYCLE STEP	**TEMP**	**TIME**	**CYCLES**
Initial denaturation	94°C	4 min	1
Denaturation Annealing Extension	94°C 50°C 72°C	45 seconds 40 seconds 2 minutes	35 cycles
Final extension	72°C	10 minutes	1
Hold	10°C	∞	

4) Use a 1% TAE agarose gel to check the second round of PCR products (Figure 2).

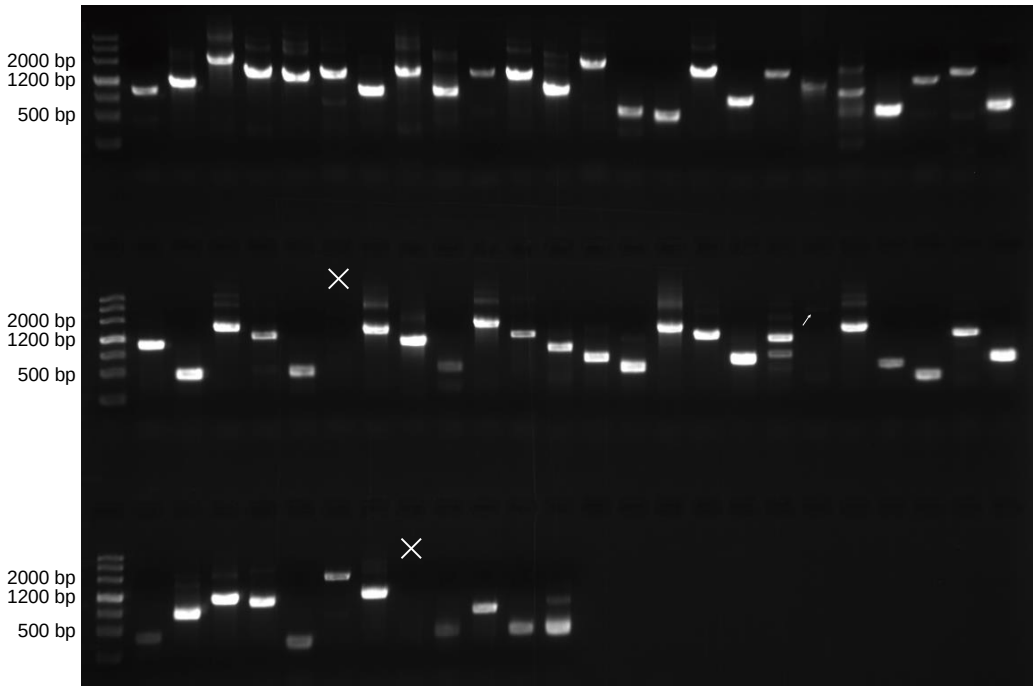

Figure 2. Schematic diagram of the PCR amplification result of NPCLs.

2. Amplicon probe preparation (Figure 1-2)

2.1 Pooled amplicons preparation

1) Mix in equal volumes all successfully amplified NPCLs from the second round of PCR (considering only those that were not amplified at all as failures. If the amplification result shows non-specific amplification outside the target region, consider these NPCLs as successfully amplified).

2) Cleanup pooled amplicons using ***FastPure®gel DNA Extraction Mini Kit*** (Vazyme, Nanjing, China).
#NOTE: It is also possible to use purification kits from other brands. The purpose of this step is to remove primer dimers and impurities from the PCR reaction system.

3) Use Nanodrop 2000 to measure the concentration of purified pooled amplicons. Keep the remaining purified pooled amplicons at -40°C.

2.2 Blunt-end repair and A-tailing

1) Take ***2 μg*** of mixed PCR product in a new centrifuge tube, ensuring that the total volume does not exceed 24 μL.

2) Prepare a 30 µL reaction as follows:

Reagent	Volume (µL)	Final concentration in 30 µL reaction
Tango Buffer (10×)	3	1×
dNTPs (10 mM)	0.3	0.1 mM
ATP (100 mM)	0.3	1 mM
T4 polynucleotide kinase (10 U/µL)	1.5	0.5 U/µL
T4 DNA polymerase (5 U/µL)	0.6	0.1 U/µL
Taq polymerase (5U/µL)	0.3	0.05 U/µL
Purified pooled amplicons + ddH_2O	24	
Total	30	

3) Mix by pipetting followed by a quick spin to collect all liquid from the sides of the tube.
! CAUTION: Don't vortex this reaction.

4) Place in a thermocycler, with the heated lid on, and run the following program:
30 minutes @ 20 °C
30 minutes @ 72 °C
! CAUTION: Proceed immediately to the ligation step.

2.3 BioT-linker ligation

1) Prepare a 30 µL reaction mixture as follows and add it to the 30 µL blunt-end repair and A-tailing reaction to obtain a final ligation reaction of 60 µL.

Reagent	Volume (µL)	Final concentration in 60 µL reaction
T4 DNA ligase buffer (10×)	3	0.5×
PEG-4000 (50 %)	3	2.5%
BioT-linker* (10 µM)	4	0.667 µM
T4 DNA ligase (5 U/µL)	1	0.083 U/µL
ddH_2O	19	
Total	30	

** The method for preparing the BioT-linker is described in the Appendix.*

2) Spin down the liquid by brief centrifugation, then incubate for 30 min at 25 °C in a thermocycler.
! CAUTION: Proceed immediately to the next step

2.4 Cleanup using AMPure XP beads

1) Vortex AMPure XP Beads to resuspend.

2) Add ***60 µL of resuspended AMPure XP beads*** to the the linker-ligated product. Mix thoroughly on a vortex mixer or by pipetting up and down at least 10 times

3) Incubate for 5 minutes at room temperature.

4) Quickly spin the tube and place the tube on an appropriate magnetic stand to separate the beads from the supernatant. After the solution is clear (about 5 minutes), carefully remove and discard the supernatant. Be careful not to disturb the beads that contain the DNA targets.

5) Add ***200 µL of 80% freshly prepared ethanol*** to the tube while in the magnetic stand. Incubate at room temperature for 30 seconds, and then carefully remove and discard the supernatant.

6) Repeat Step 5) once.

7) Air dry the beads for 5 minutes while the tube is on the magnetic stand with the lid open. **! CAUTION:** Do not overdry the beads. This may result in lower recovery of DNA target.

8) Remove the tube from the Magnetic Separation Rack. Elute the DNA target from the beads into ***30 µL of 0.1× TE***. Mix well on a vortex mixer or by pipetting up and down. Incubate for 2 minutes at room temperature.

9) Quickly spin the tube and place it on a magnetic stand. After the solution is clear (about 5 minutes), transfer 30 µL of the supernatant to a fresh, sterile 1.5 ml tube.

10) Measure the concentration using Nanodrop 2000. The final concentration of the purified linker-ligated amplicons should be 20 ng/µL.

2.5 Add biotin labels at both ends of amplicons through PCR amplification

1) Prepare a 25 µL PCR reaction as follows:

Reagent	**Volume (µL)**	**Final concentration in 25 µL reaction**
HiFi buffer (10×)	2.5	1×
dNTPs (10 mM each)	0.5	200 µM
HiFi Taq (5U/µL)	0.25	0.05 U/µL
Biotinylated linker primer (10 µM)	1.25	0.5 µM
Purified linker-ligated amplicons (20 ng/µL)	5.0	
ddH_2O	15.5	
Total	25	

2) Mix by pipetting followed by a quick spin to collect all liquid from the sides of the tube.
3) Place in a thermocycler, with the heated lid on, and run the following program:

CYCLE STEP	TEMP	TIME	CYCLES
Initial denaturation	98°C	30 seconds	1
Denaturation Annealing Extension	94°C 45°C 72°C	30 seconds 45 seconds 2 minutes	10 cycles
Final extension	72°C	10 minutes	1
Hold	10°C	∞	

4) Use a 1% TAE agarose gel to check amplicon probes (Figure 3).

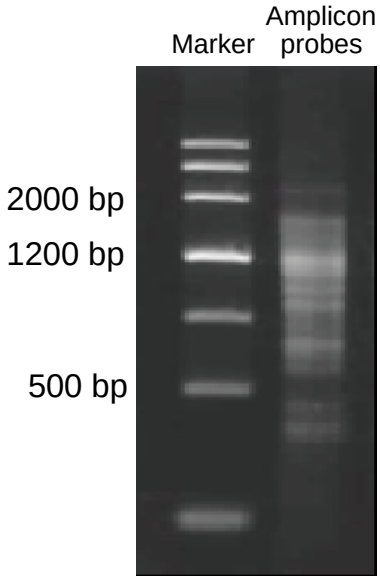

Figure 3. Schematic diagram of amplicon probes.

2.6 Cleanup using AMPure XP beads

1) Repeat step 2.4 1)-9)

2) Measure the concentration using Nanodrop 2000. The final concentration of purified amplicon probes should be 20 ng/μL.

3. Genomic DNA library preparation

3.1 DNA extraction

For each capture sample, genomic DNA was extracted from one to three legs using the TIANamp Genomic DNA Kit (TIANGEN Inc., Beijing, China). All DNA extracts were quantified using an ND-2000 spectrophotometer and diluted to a concentration of 20 ng/μL with 1× TE Buffer.

3.2 DNA fragmentation

1) Prepare a 20 μL PCR reaction as follows:

Reagent	Volume (μL)	Final concentration in 20 μL reaction
Genomic DNA (100 ng)	5	5 ng/μL
1× TE	17	
Total	22	

2) Use Scientz18-A ultrasonic processors (SCIENTZ, Zhejiang Province, China) to shear 100 ng genomic DNA of each sample to a size of 300-500 bp.

3) Use a 2% TAE agarose gel to check DNA fragmented products.

3.3 Illumina DNA library preparation

The purified fragmented DNA was used for Illumina library preparation with ***NEBNext Ultra DNA Library Prep Kit for Illumina*** (New England Biolabs). Detailed experimental procedures for Illumina library preparation can be seen the kit instructions (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). Finally, each sample was labeled with a unique 8-bp index sequence. ***Three or four libraries*** were mixed into a pooled library in equal concentrations for subsequent hybridization capture.

4. Hybridization capture (Figure 1-3)

4.1 Incubation

The final 50 μL hybridization reaction system:

Reagent	Volume (μL)
Library-probe mix	31.5
300 ng of pooled indexed libraries	
Human Cot1 and Blocking oligos	
Amplicon probes*	
Hybridization buffer	18.5
Total	50

** If 100 target NPCLs are captured, the recommended amount of probe is 20 ng.*

1) Prepare a hybridization buffer as follows:

Reagent	Volume (μL)	Final concentration in 50 μL reaction
SSPE (20×)	12.5	5×
EDTA (0.5 M)	0.5	5 mM
Denhardt's (50×)	5	5×
SDS (10%)	0.5	0.1%
Total	18.5	

2) Mix the hybridization buffer by pipetting gently and collect the liquid at the bottom of the tube by briefly centrifuging. Add ***12.5 μL mineral oil*** to the hybridization buffer (prevent volatilization). Incubate at 65°C in a hot-lid thermocycler.

3) Prepare a Library-probe mix as follows:

Reagent	Volume (μL)	Final concentration in 50 μL reaction
Human Cot1 (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.part2.F* (100 μM)	0.5	1 μM
Pooled indexed libraries (300 ng)	15	6 ng/μL
Amplicon probes + ddH_2O	11.5	
Total	31.5	

** Blocking oligo sequences can be found in the Appendix.*

4) Mix the library-probe mix by pipetting gently. Collect the liquid at bottom of the tube by briefly centrifuging. Add ***15 μL mineral oil*** to the library-probe mix (prevent volatilization). Incubate the library-probe mix at 95°C in a hot-lid thermocycler for 5 minutes. **Quickly transfer** all solution of the library-probe mix (including mineral oil) to the hybridization buffer, and thoroughly incorporate them by **quickly pipetting the reaction mixture up and down at 10 times.**

5) **Quickly close the lid of the tube**, and incubate the reaction using the following program:
   - ⇩ 65°C 6 hours
   - ⇩ 60°C 6 hours
   - ⇩ 55°C 6 hours
   - ⇩ 50°C 6 hours
   - ⇩ 45°C 6 hours

4.2 Binding hybridized libraries to beads and washing

1) Add n × ***3 µL*** (n is the number of hybridization reactions, n ≤ 40) of ***Dynabeads Myone streptavidin beads*** (Life Technologies) to a sterile 1.5 ml tube.

2) Spin the tube and place the tube on an appropriate magnetic stand to separate the beads from the supernatant. After the solution is clear (about 1 minutes), carefully remove and discard the supernatant.

3) Remove the tube from the Magnetic Separation Rack, add ***200 µL Binding Buffer*** (2 M NaCl, 10 mM Tris-HCL, 1 mM EDTA) and resuspend the beads by gently vortexing the tube or pipetting the reaction up and down at least 5 times. Place the tube back on the Magnetic Separation Rack. After the solution is clear (about 1 minutes), carefully remove and discard the supernatant.
#NOTE: Discard any loose beads that are not tightly attracted to the wall of the tube

4) Repeat step 3) twice for a total of three washes

5) Remove the tube from the Magnetic Separation Rack. Resuspend the tube in n × ***20 µL Binding Buffer*** and add n × ***0.1 µL 10% Tween-20***. Mix thoroughly on a vortex mixer or by pipetting up and down at least 10 times.

6) For each hybridization reaction, add ***180 µL Binding Buffer*** to a fresh 1.5 ml tube, and then add 20 µL resuspended beads to the tube. Thoroughly incorporate them by pipetting the reaction mixture up and down at least 10 times, and incubate the tube for 2 minutes at **last hybridization temperature** (45°C)

7) Transfer all hybridization solution (**! CAUTION:** Not including mineral oil) to the 200 µL of binding buffer/beads solution. Mix by pipetting followed by a quick spin to collect all liquid from the sides of the tube. Quickly place the tube on an appropriate **constant temperature oscillating metal bath**, and then incubate the tube at 1,500 rpm for 30 minutes at **last hybridization temperature** (45°C).

8) Remove the tube from the metal bath. Spin the tube to collect all liquid from the side of it, and then place the tube on the magnetic stand. After the solution is clear (about 3 minutes), **remove the supernatant completely**.

9) Remove the tube from the Magnetic Separation Rack, add ***200 µL Wash Buffer 1*** (1× SSC, 0.1% SDS) and resuspend the beads by gently vortexing or rocking the tube. Incubate for 10 minutes at room temperature, and then place the tube back on the Magnetic Separation Rack. After the solution is clear (about 3 minutes), remove and discard the supernatant.
#NOTE: After this step, preheat ***Wash Buffer 2*** (0.1× SSC, 0.1% SDS) to 45°C.

10) Repeat step 9) one more time for a total of ***two low stringency washes***.

11) Remove the tube from the Magnetic Separation Rack, add ***200 µL 45°C Wash Buffer 1*** (0.1× SSC, 0.1% SDS) and resuspend the beads by gently vortexing or rocking the tube. Incubate for 10 minutes at 45°C, and then place the tube back on the Magnetic Separation Rack. After the solution is clear (about 3 minutes), remove and discard the supernatant.

12) Repeat step 11) twice for a total of ***three high stringency washes***.
! CAUTION: After the last wash, make sure all additional supernatant is removed.

13) Remove the tube from the Magnetic Separation Rack, add ***30 µL of 0.1× TE*** and resuspend the beads by gently pipetting the reaction up and down at least 10 times. The resuspended beads were directly used for subsequent PCR amplification.

4.3 Pre-hybridization PCR (for the second hybridization capture)

#NOTE: We used off-beads amplification (Fisher et al. 2011). This avoids the need to denature and elute the captured target from the baits using sodium hydroxide. The procedure is less problematic and results in more captured products.

1) Prepare a 25 µL PCR reaction as follows:

Reagent	**Volume (µL)**	**Final concentration in 25 µL reaction**
HiFi buffer (10×)	2.5	1×
dNTPs (10 mM each)	0.5	200 µM
HiFi Taq (5U/µL)	0.25	0.05 U/µL
P5 primer (10 µM)*	0.5	0.2 µM
P7 primer (10 µM)*	0.5	0.2 µM
Resuspended beads (1st hybridization products)	3	
ddH_2O	15.75	
Total	25	

**P5 and P7 primers are included in the NEBNext Ultra DNA Library Prep Kit for Illumina.*

2) Mix by pipetting followed by a quick spin to collect all liquid from the sides of the tube.

3) Place in a thermocycler, with the heated lid on, and run the following program:

CYCLE STEP	TEMP	TIME	CYCLES
Initial denaturation	95°C	30 seconds	1
Denaturation Annealing & Extension	98°C 65°C	10 seconds 75 seconds	14 cycles
Final extension	65°C	5 minutes	1
Hold	10°C	∞	

4) Use a 2% TAE agarose gel to check the 1st hybridization PCR products.

4.4 Cleanup using AMPure XP beads

1) Repeat step 2.4 1)-7)

2) Remove the tube from the Magnetic Separation Rack. Elute the DNA target from the beads into ***20 μL of 0.1× TE***. Mix well on a vortex mixer or by pipetting up and down. Incubate for 2 minutes at room temperature.

3) Quickly spin the tube and place it on a magnetic stand. After the solution is clear (about 5 minutes), transfer 20 μL of the supernatant (purified 1st hybridized library) to a fresh, sterile 1.5 ml tube labeled as **"name + 1st_cap"**.

4) Use Nanodrop 2000 to measure the concentration of purified hybridized library. Keep the remaining purified hybridized library at -40°C.

5. The second round of hybridization capture and Sequencing

#NOTE: To improve capture efficiency, we recommend performing a second round of hybridization capture. The experimental protocol for the second round of hybridization capture is exactly the same as that for the first round, except that the library used for preparing the probe mixture is **the enriched library obtained from the first round of hybridization**.

5.1 Incubation

Repeat 4.1

5.2 Binding hybridized libraries to beads and washing

Repeat 4.2

5.3 Pre-hybridization PCR (for high-throughput sequencing)

Repeat 4.3

5.4 Cleanup using AMPure XP beads

1) Repeat step 2.4 1)-7)

2) Remove the tube from the Magnetic Separation Rack. Elute the DNA target from the beads into ***20 μL of 0.1× TE***. Mix well on a vortex mixer or by pipetting up and down. Incubate for

2 minutes at room temperature.

3) Quickly spin the tube and place it on a magnetic stand. After the solution is clear (about 5 minutes), transfer 20 μL of the supernatant (purified 2nd hybridized library) to a fresh, sterile 1.5 ml tube labeled as "**name + 2nd_cap**".

4) Use Nanodrop 2000 to measure the concentration of purified hybridized library. Keep the remaining purified hybridized library at -40°C.

5.5 Pooling multiple samples for sequencing

Combine all the 2nd hybridized libraries in equal amounts to form a pool. For the Illumina HiSeq X-ten sequencer, extract DNA fragments in the size range of 300-500 bp using the gel-cutting method. The final concentration of the pooled library should be approximately 10 ng/μL with a minimum volume of 25 μL.

Data analysis process

1. Capture data assembly (Figure 1-4)

1.1 Data quality control

1) Use ***Trimmomatic v0.32*** to filter out adapter sequences and low-quality nucleotides from the raw reads of each species.

```
$java -jar trimmomatic-0.39.jar PE sample1_R1.fq.gz sample1_R2.fq.gz
sample1_R1_paired.fq.gz sample1_R1_unpaired.fq.gz sample1_R2_paired.fq.gz
sample1_R2_unpaired.fq.gz ILLUMINACLIP:TruSeq3-PE.fa:2:30:10:2:keepBothReads
LEADING:3 TRAILING:3 MINLEN:36
```

1.2 Data assembly

2) Use ***metaSPAdes v3.13.0*** to assembly the resulting clean reads for each species in to contigs.

```
$metaspades.py -t 20 -1 sample1_R1_paired.fq.gz -2 sample1_R2_paired.fq.gz -o
sample1_contigs.fasta
```

1.3 Contig quality control

1) Use ***CD-HIT-EST v 4.8.1*** with a 95% similarity cutoff to remove redundant contigs.

```
$cd-hit-est -i sample1_contigs.fasta -o sample1_contigs_cd_hit.fasta -c 0.95
```

2) Use ***BWA 0.7.10-r789*** and ***SAMtools v 1.4*** to calculate the average sequencing depth for non-redundancy contigs.

```
$bwa index sample1_contigs_cd_hit.fasta
$bwa mem –t 20 sample1_R1_paired.fq.gz sample1_R2_paired.fq.gz > sample1_bwa.sam
$samtools view -bS -@ 20 sample1_bwa.sam > sample1_bwa.bam
$samtools sort -@ 20 sample1_bwa.bam > sample1_bwa_sorted.bam
$samtools depth sample1_bwa_sorted.bam > sample1_depth
```

3) Retained contigs with an average sequencing depth of ≥ 5×.

2. Orthology assessment and alignment construction (Figure 1-5)

2.1 Orthologous sequence group extraction

#NOTE: We developed a custom Python script called ***'Extract_orthologous_sequence_groups_from_assembled_contigs.py'*** to extract orthologous sequence groups (OGs) from filtered contigs. The script can be downloaded from the following link: https://github.com/zhangpenglab/UPrimer/tree/main/Accessory. The detailed usage information of this script can be referred to at the following link: https://github.com/zhangpenglab/UPrimer#extract-orthologous-sequence-groups-from-assembled-contigs.

1) Prior to running the script, researchers need to prepare the following input files:
   - a folder named "***contigs***" contains filtered contigs for each species
   - a folder named "***reference***" contains reference nucleotide and peptide sequences generated by UPrimer;
   - a folder named "***pep_for_exonerate***" contains individual reference peptide sequences.

2) Run this script using the following command:

```
$python Extract_orthologous_sequence_groups_from_assembled_contigs.p -c /path/to/contigs
```

3) Collect the file of the orthologous sequence groups (OGs) and proceed with the subsequent steps.

2.2 Multiple sequence alignment

- Use ***MAFFT v 7.0.1*** to align the nucleotide sequences of target NPCLs based on their deduced protein sequences with default settings.

```
$mafft --auto Target_NPC_1_OG.fasta > Target_NPC_1_MSA.fasta
```

2.3 Alignment refining

1) Use ***Gblocks v 0.91*** to refine alignments under codon mode (-t = c) with half gaps allowed.

```
$Gblocks Target_NPC_1_MSA.fasta -t=c -b5=h
```

2) Use ***IQ-Tree2 v. 2.1.0*** to reconstruct single gene tree for each trimmed alignment under the GTR+G model.

```
$iqtree2 –s Target_NPC_1_MSA_trimmed_MSA.fasta –b 100 –m GTR+G
```

3) Use ***TreeShrink v. 1.3.3*** to eliminate long branches:

```
$ python treeshrink.py -i Target_NPC_1_trimmed_MSA.trees
```

4) Use ***MAFFT v 7.0.1*** to realign all refined alignments.

```
$mafft --auto Target_NPC_1_trimmed_MSA.fasta > Target_NPC_1_refined_MSA.fasta
```

3. Phylogenetic tree reconstruction (Figure 1-6)

3.1 Concatenation analysis

1) Combine all refined alignments in to a concatenated supermatrix.

2) Use ***IQ-Tree2 v. 2.1.0*** to reconstruct a ML tree.

```
$iqtree2 –s Target_NPCL_supermatrix.fasta –p three_codon_partitions.txt –b 500 –m MFP
```

#NOTE: To ensure the inferred trees were stabilizing, repeat three independent IQ-TREE runs.

3.2 Species tree analysis

1) Use ***IQ-Tree2 v. 2.1.0*** to reconstruct single gene tree for each refined alignment.

```
$iqtree2 –s Target_NPC_1_refined_MSA.fasta –b 300 –m MFP+MERGE
```

2) Use ***ASTRAL-III*** to reconstruct a species tree based on all single gene best trees and their bootstrap files.

```
$ java -jar astral.5.7.8.jar –i All_NPCL_best_trees.tre –b All_NPCL_bootstrap_trees.tre
–o ASTRAL_result
```

Appendix

1. Solution recipe for ***10× oligonucleotide buffer*** (used for Bio-T linker preparation):

Reagent	**Volume (µL)**	**Final concentration in 1 mL reaction**
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
ddH_2O	888	
Total	1000	

2. ***10 µM Bio-T linker*** preparation method

- Preparing a reaction system as follows:

Reagent	**Volume (µL)**	**Final concentration in 100 µL reaction**
Bio-T (100 µM)*	10	10 µM
TR (100 µM)*	10	10 µM
10× oligonucleotide buffer	10	1×
ddH_2O	70	
Total	1000	

**Bio-T: 5'-Biotin-CAAGGACATCCGT-3'*
** TR: 5'-CGGATGTCCTTGC-3'*

- Mix by pipetting followed by a quick spin to collect all liquid from the sides of the tube.

- Place in a thermocycler, with the heated lid on, and run the following program:

CYCLE STEP	TEMP	TIME
Initial denaturation	95°C	5 min
Annealing	-0.1° C/s	decrease 0.1 degrees Celsius per second
Hold	12°C	∞

3. Solution recipe for ***Binding Buffer***:

Reagent	**Volume (µL)**	**Final concentration in 1 mL reaction**
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
ddH_2O	588	
Total	1000	

4. Solution recipe for ***Washing Buffer 1***:

Reagent	**Volume (µL)**	**Final concentration in 1 mL reaction**
SSC (20×)	50	1×
SDS (10%)	10	0.1%
ddH_2O	940	
Total	1000	

5. Solution recipe for ***Washing buffer 2***:

Reagent	**Volume (µL)**	**Final concentration in 1 mL reaction**
SSC (20×)	5	0.1×
SDS (10%)	10	0.1%
ddH_2O	985	
Total	1000	

6. Blocking oligo sequences:

Name	**Sequence (5'-3')**
BO1.P5.F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-Phosphate
BO5.P7.part1.F	AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC-Phosphate
BO7.P7.part2.F	ATCTCGTATGCCGTCTTCTGCTTG-Phosphate