FEB 02, 2023

🌐 High-throughput DNA barcoding library construction and sequencing protocol for BIOSCAN using unpurified non-destructively extracted DNA from arthropods
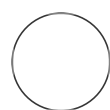
Naomi Park[1], Emma Betteridge[1], Scott Thurston[1], Abdulrahman Tuameh[1], Marco Mosca[1], Lyndall Pereira da Conceicoa[1], Ian Johnston[1], Mara Lawniczak[1]

[1]Wellcome Sanger Institute

Coronavirus Method Development Community

Emma Betteridge

DOI:

dx.doi.org/10.17504/protocols.io.8epv5jzxdl1b/v1

Protocol status: Working
We use this protocol and it's working

Created: Oct 07, 2022

Last Modified: Feb 02, 2023

ABSTRACT

This SOP describes the procedure for high-throughput generation of mitochondrial cytochrome c oxidase subunit I (COI) DNA barcode amplicons using very small quantities of crude DNA extracted non-destructively (i.e., without grinding or disruption to the organism) from arthropods LysisCextractionSOPV1.pdf - Google Drive. The use of an inhibitor-tolerant polymerase enables amplification of crude lysate without purification, which can add significant cost. The first PCR amplifies the target of choice using untailed primers. Here, we target the Cytochrome Oxidase I mitochondrial locus, but in principle, the locus could be any amplicon. In a second PCR step, long read compatible 16- mer combinatorial dual indexed amplicons are then made directly from the first PCR product. Although full length indexed amplicons can be made in a single PCR step, by incorporating the use of non-tailed COI primers first, the sensitivity to low template inputs is markedly improved. Insects alone can range across three orders of magnitude in size and can be as small as 0.2 mm, so increasing sensitivity to low quantity inputs without oversequencing individuals with much greater DNA quantities is desirable. After the two step PCR is complete, as many as 9216 PCRs are then equivolume pooled and quantitated, prior to long-read library construction. This single library is then sequenced on a single Pacbio 8M SMRT Cell.

This SOP is entitled BIOSCAN as it supports the current global endeavour of the International Barcode of Life (https://ibol.org/programs/bioscan/) to massively increase species discovery using barcoding. Additionally, this SOP is being used for the Sanger BIOSCAN project to study 1M insects across the UK (https://www.sanger.ac.uk/collaboration/bioscan/).

This 2-step indexing PCR approach is an adaptation of the COVID-19 ARTIC Illumina library construction - tailed method, which can be found here: COVID-19 ARTIC v4.1 Illumina library construction and sequencing protocol - tailed method (protocols.io)

GUIDELINES

It is vital PCR 1 setup is performed in a laboratory in which post PCR-COI amplicons are not present, to minimise any risk of sample contamination.

Note: Throughout the protocol we have indicated the liquid handling automation in use at the Wellcome Sanger Institute for specific parts of the process. However, these steps could be performed on alternative liquid handlers or manually.

COI amplification (PCR1)

1 Important! This step must be performed in a pre-PCR environment in which post PCR COI amplicons are not present, to minimise risk of sample contamination.

⚠️

Input into COI amplification is unpurified non-destructively extracted DNA from arthropods.

2   Generate the COI primer pool ([M] 2.5 micromolar (µM) each primer) by combining the following in a 2mL Eppendorf DNA LoBind tube and vortex to mix.

> **Note**
>
> Aliquot primer pool into useful sizes (125uL is sufficient for 1 x 384 plate including 20% overage). Aliquots are stable at 🌡 -20 °C or may be stored short term at 🌡 4 °C

	A	B	C	D
	Non-tailed COI primer	**Sequence**	**Concentration (µM)**	**Volume (µl)**
	LepF1	ATTCAACCA ATCATAAAG ATATTGG	100	40
	LepR1	TAAACTTCT GGATGTCCA AAAAATCA	100	40
	LCO1490	GGTCAACAA ATCATAAAG ATATTGG	100	40
	HC02198	TAAACTTCA GGGTGACCA AAAAATCA	100	40
	Qiagen EB			1440
	Total			1600

COI non-tailed primer mix. Order STD purification. Pool volumes may be scaled to required sample number throughput

3

Prepare the following COI PCR master mix and mix thoroughly by vortexing on full power. Keep on ice whilst preparing for subsequent steps.

⊗ RepliQa HiFi ToughMix® **VWR international Ltd Catalog #95200-500**

A	B	C
Weighted PCR Primer Pool 1 Master Mix	**Vol/PCR RXN (µl)**	**Vol/384 plate (µl) inc. 20% excess**
COI Primer mix (2.5µM each)	0.25	115
RepliQa HiFi ToughMix	2.5	1150

A	B	C
Nuclease-free water	2.15	989
Total	**4.9**	2254

4 Use the SPT Labtech Dragonfly Discovery to predispense 🜇 4.9 µL mastermix per well into 384 well plates.

> **Note**
>
> The SPT Labtech Dragonfly Discovery uses positive displacement syringes for non-contact reagent dispensing. This enables efficient and accurate, low volume dispensing with minimal syringe consumption. The Dragonfly is very flexible and easy to programme.

5 Select 4 x 96 well plates containing crude lysate and centrifuge at 2000rpm for 2 minutes and remove the seal

> **Note**
>
> Crude lysate plates should contain 100µL volume, and require centrifugation immediately prior to liquid transfer, concentrating inhibitors towards the well bottom. By careful sampling from the upper 50µL of the well, the amount of inhibitor is usually sufficiently low to enable amplification.

6 Use the SPT Labtech Mosquito LV to transfer 🜇 100 nL of crude lysate into the plate containing the COI PCR master mix maintaining the same well locations throughout. The Mosquito LV must be setup to fix the aspirate height to aspirate from the upper 50µL of the 100µL well contents. Immediately proceed to the next step.

> **Note**
>
> The SPT Labtech Mosquito LV is used for highly accurate, low volume liquid transfers. It utilises multi-channel positive displacement pipetting, with a range of 25nl to 1.2ul. It enables miniaturisation of methods which reduces costs.

7 Heat seal and mix the plate e.g. on a BioShake iQ for 1 minute at 2000rpm, and centrifuge briefly at 3000rpm.

Important! Heat seal to minimise evaporation during PCR.

Place the plates onto a thermocycler and run the following program:

> **Note**
>
> Amplification should ideally be performed in a different lab to minimise the risk of contamination.

A	B	C
Step	**Temperature**	**Time**
1	98°C	10 seconds
2	45°C	5 seconds
3	68°C	5 seconds
4	Repeat steps 1 - 3 for a total of 40 cycles	
5	10°C	∞

> **Note**
>
> **Optional QC step**: Dilute a small proportion of wells 1:10 with Elution Buffer and run directly on TapeStation High Sensitivity D5000. A single peak ~658bp is expected although the residual salts cause the sizing to run ~150bp smaller. Inhibition is indicated by complete absence of any product, in contrast to insufficent template which is indicated by a short product ~30bp.

8 **PAUSE POINT** Amplified DNA can be stored at 4°C (overnight) or -20°C (up to 6 months).

Indexing amplified DNA (PCR2)

9

> **Note**
>
> Long read compatible indexed DNA barcodes are generated from a small aliquot of the
> amplified template from PCR1 using KAPA HiFi HotStart ReadyMix, combinatorial dual
> indexed 16-mer barcoding primers and pools of tailed versions of the primers used for the
> DNA amplification.

The tailed primer pools used in this stage correspond to those used in the COI amplification stage, with the following modifications:

- The 5' end of the tailed COI primers contain a /5AmMC6/ modification, which is a 5' blocker so only full length indexed PCR 2 products can ligate to Pacbio / ONT adapters in case of incomplete conversion
- GCAGTCGAACATGTAGCTGACTCAGGTCAC appended to the 5' end of both forward primers
- TGGATCACTTGTGCAAGCATCACATCGTAG appended to the 5' end of both reverse primers

A	B
Tailed primer name	Tailed primer sequence
LepF1_tail	/5AmMC6/GCAGTCGAACATGTAGCTGACTCAGGTCACATTCAACCAATCATAAAGATATTGG
LepR1_tail	/5AmMC6/TGGATCACTTGTGCAAGCATCACATCGTAGTAAACTTCTGGATGTCCAAAAAATCA
LCO1490_tail	/5AmMC6/GCAGTCGAACATGTAGCTGACTCAGGTCACGGTCAACAAATCATAAAGATATTGG
HC02198_tail	/5AmMC6/TGGATCACTTGTGCAAGCATCACATCGTAGTAAACTTCAGGGTGACCAAAAAATCA

Due to the complexity of processing 24 x 384 dual indexing primer combinations, both the indexing primers and tailed primer pools are predispensed to plates and frozen down in advance for ease of processing.

The tailed primer is combined with EB (containing [M] 0.01 % volume Triton-X), forward and reverse indexes to create plates of ⏳ 6.15 µL per well, with indexing primers at [M] 2 micromolar (µM) each and tailed primers at [M] 4 nanomolar (nM) each. We use the SPT Labtech Dragonfly Discovery to first dispense ⏳ 6 µL of all components excluding the indexing primers, followed by the Beckman Coulter Echo 525 liquid handler to dispense 75nL of the appropriate forward and reverse primers (96 forward indexes x 96 reverse indexes = 9216 unique combinations and 24 differently indexed 384 plates).

📎 bioscan indexing primers.xlsx

⊗ 2x Kapa HiFi Hotstart Readymix **Kapa Biosystems Catalog #KK2602**

10 Defrost the COI indexing plates, being careful to record which index plate # is to be combined with which PCR 1 plate.

Note

Up to 24 indexing plates may be pooled for a sequencing run and it is vital to carefully track processing to ensure each version is only used once within a final pool.

11 Use the SPT Labtech Mosquito LV to transfer ⧗ 100 nL of COI PCR 1 product into the dual indexed plate containing the tailed primers, maintaining the same well locations throughout. Immediately proceed to the next step.

12 Use the SPT Labtech Dragonfly Discovery to dispense ⧗ 6.25 µL of Kapa HiFi 2X Mastermix into the dual indexed plate from step 11, and place 🌡 On ice immediately. The dispense is sufficient to mix all the reagents.

Note

The final PCR volume is ⧗ 12.5 µL
The final concentration of each tailing primer in the reaction will be [M] 2 nanomolar (nM)
The final concentration of each barcoding primer in the reaction will be
[M] 1 micromolar (µM)
The amplified COI template forms [M] 0.8 % (v/v) of the total PCR volume

13 Heat seal and place the plate onto a thermocycler and run the following program.
Important! Heat seal to minimise evaporation.

	A	B	C
	Step	Temperature	Time
	1	95°C	5 minutes
	2	98°C	30 seconds
	3	53°C	20 minutes
	4	72°C	2 minutes
	Repeat steps 2-4 once more		
	5	98°C	30 seconds
	6	62°C	30 seconds
	7	72°C	2 minutes
	Repeat steps 5-7 six more times		
	8	72°C	5 minutes
	9	10°C	∞

Note

The long annealing times of the first two cycles of PCR ensure efficient annealing of the tailed primers to their targets in the amplified COI template (and therefore incorporation of the tail sequences) in spite of their very low concentration in the PCR. In the following seven cycles of PCR the much shorter annealing time and increased annealing temperature make the annealing of the tailed primers inefficient, therefore only the indexing primers participate in the PCR. This ensures that the vast majority of products formed at the end of the PCR are of full length.

14   **PAUSE POINT** Amplified indexed products can be stored at 4°C (overnight) or -20°C (up to 6 months).

Construction of equivolume pool

15

In a post-PCR lab, use a VBLOK200 reservoir to collect the entire contents of a single post indexed COI plate by upside down centrifugation at 1000rpm for 1 minute.

> **Note**
>
> Do not exceed 1000rpm to ensure the integrity of the VBLOK200 reservoir is maintained.

16 Transfer the contents in the reservoir to a 5mL Eppendorf tube and vortex to mix. The same VBLOK200 reservoir may be used to collect the contents of multiple plates which will eventually be pooled together (up to a maximum of 24 plates)

> **Note**
>
> Subsequent pools processed with the same VBLOK200 reservoir will contain low-levels of the previous samples. Therefore, only use the same VBLOK200 for pooling samples which will be sequenced together.

17 **Optional QC step**: Dilute each pool 1:10 with Elution Buffer and run directly on TapeStation High Sensitivity D5000. A single peak ~890bp is expected although the residual salts cause the sizing to run ~150bp smaller.

18 **PAUSE POINT** Pools can be stored at 4°C (overnight) or -20°C (up to 6 months).

19 Manually combine ⏳ 30 µL of each of the 24 pools together, and mix by vortexing to form an equivolume pool of 9216 samples.

Equivolume pool SPRI bead cleanup

20 Allow AMPure XP beads to equilibrate to room temperature (~30 minutes). Ensure solution is homogenous prior to use.

21 Add **0.6X** volume (⏳ 300 µL) of AMPure XP beads per ⏳ 500 µL of pooled product, and mix well by vortexing.

22 Incubate for ⏱ `00:06:00` at 🌡 `Room temperature` . `6m`

23 Transfer the tube to a magnet, allow ⏱ `00:04:00` for the beads to form a pellet. `4m`

24 Carefully remove and discard the supernatant, taking care not to disturb the bead pellet.

25 Wash the beads with ⚗ `1000 µL` 75% ethanol for ⏱ `00:00:15` then carefully remove ethanol `15s`
and discard.
(First wash)

26 Wash the beads with ⚗ `1000 µL` 75% ethanol for ⏱ `00:00:15` then carefully remove ethanol `15s`
and discard.
(Second wash)

27 Pulse spin the tube and return to magnet to remove residual 75% ethanol. Leave ~1 minute to dry
(being careful not to overdry)

28 Remove tube from magnet and resuspend beads in ⚗ `100 µL` elution buffer, mix well by
vortexing.

29 Incubate for ⏱ `00:03:00` at 🌡 `Room temperature` `3m`

30 Transfer tube to magnet, allow 🕑 00:05:00 for the beads to form a pellet. `5m`

31 Carefully transfer supernatant into a new tube, taking care not to disturb the bead pellet.

32 The clean equivolume pool may be quantified using Qubit Fluorometer, and sizing checked on TapeStation D5000.

PacBio Library Preparation and Sequencing

33

We currently prepare our amplicon pool for PacBio sequencing using the protocol attached below, 'Preparing SMRTbell Libraries using PacBio Barcoded Universal Primers for Multiplexing Amplicons', starting with DNA Damage Repair.

The library, containing 9216 samples, is sequenced on a SMRT Cell 8M using the Sequel IIe system.

Sample setup recommendations for sequencing amplicon libraries <3 kb:
Sequencing Primer: Sequencing Primer v4
Binding Kit: Sequel II Binding Kit 2.1
Binding Time: 1 Hour
Sequencing Kit: Sequel II Sequencing Plate 2.0
On-Plate Loading Concentration: 100 pM

Recommended Run parameters:
Movie Time (hours): 10
Pre-Extension Time (hours): 0.5
Immobilization Time (hours): 2 (default)

📄 Procedure-Checklist-Preparing-SMRTbell-Libraries-using-PacBio-Barcoded-Universal-Primers-for-Multiplexing-Amplicons.pdf

> **Note**
>
> At Sanger, we plan to adopt SMRTbell Prep Kit 3.0 and Binding Kit 3.1 in Q1 2023.

Analysis using mBRAVE

34 PacBio sequence data de-multiplexing is performed using the rapid and highly configurable mBRAVE (Multiplex Barcode Research And Visualization Environment) online analysis platform http://www.mbrave.net/. mBRAVE builds on the BOLD platform, http://www.boldsystems.org/, to support species identification and discovery.

The index set currently in use at Sanger is registered on mBRAVE as 'Sanger_BIOSCAN_v1'.

For more information on how to use mBRAVE for data analysis, please follow the 'Contact' tab on the mBRAVE web page.

ONT Library Preparation and Sequencing

35 The amplicon pool generated in steps 1-32 is also compatible with Oxford Nanopore sequencing.

The amplicon pool can be prepared for Oxford Nanopore sequencing using the protocol attached below, 'Ligation sequencing amplicons V14 (SQK-LSK114)'.

The library is then sequenced on an R10.4.1 MinION flow cell (FLO-MIN114).

    📎 ligation-sequencing-amplicons-sqk-lsk114-ACDE_9163_v114_revJ_29Jun2022-gridion.pdf

36 **Custom demultiplexing for Oxford Nanopore sequence data**

Each sample was identified by a pair of index sequences: a front index f_i and a rear index r_j. Individual index sequences are not unique, i.e. a front index is paired with more than one rear index and vice versa (f_1-sample1-r_1, f_2-sample2-r_1, …). The pair f_i + r_j uniquely identifies a sample *s*.

Since the ONT deplexer (guppy_barcoder) cannot handle non-unique single indexes, the deplexing was customised. ONT advised us to use nanoplexer to perform custom deplexing. Nanoplexer (v0.1.2) takes as input a fastq/fastq.gz file and a configuration file describing a set of indexes. It outputs one file per index containing the classified reads. In order to deplex the

pooled samples, the software was run twice; firstly, for a rear index set R and secondly, for a front index set F. The following steps were used to deplex the sample pool:

1. Deplex by rear indexes $r_j \in R$

2. For each set of classified reads (by r_j)

    a. Deplex the set by front indexes $f_i \in F$