



Jan 21, 2025

An innovative method for multiplexed targeted sequencing of cDNA

DOI

<https://dx.doi.org/10.17504/protocols.io.eq2lywx8pvx9/v1>

Ange TCHAKOUNTE¹, CAMPBELL Nathan², DELORME Quentin¹, MARIN-MENENDEZ Alejandro¹, TALMAN Arthur¹

¹MIVEGEC, Université de Montpellier, IRD, CNRS, Montpellier, France;

²GT seek, Twin Falls, ID 83301, United States

Ange TCHAKOUNTE: Corresponding author : ange.tchakounte@yahoo.com;



Ange TCHAKOUNTE

IRD

Create & collaborate more with a free account

Edit and publish protocols, collaborate in communities, share insights through comments, and track progress with run records.

Create free account

OPEN  ACCESS



DOI: <https://dx.doi.org/10.17504/protocols.io.eq2lywx8pvx9/v1>

Protocol Citation: Ange TCHAKOUNTE, CAMPBELL Nathan, DELORME Quentin, MARIN-MENENDEZ Alejandro, TALMAN Arthur 2025. An innovative method for multiplexed targeted sequencing of cDNA . **protocols.io**

<https://dx.doi.org/10.17504/protocols.io.eq2lywx8pvx9/v1>

License: This is an open access protocol distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: Working

We use this protocol and it's working

Created: June 21, 2024

Last Modified: January 21, 2025

Protocol Integer ID: 102195

Keywords: Malaria, Targeted RNA sequencing, Intra-host dynamics, *Plasmodium falciparum*, parallel capture of target rna biomarker, multiplexed capture of rna transcript, multiplexed rna quantification from sample, multiplexed rna quantification, apicomplexan parasite plasmodium falciparum, targeted sequencing, target rna biomarker, few rna biomarker, incorporation of unique molecular identifier, rna transcript, unique molecular identifier, deep sequencing, sequencing index, parallelized assay, reverse transcription, rna, rna content, specific mrna, antisense primer, pool of antisense primer, rna content of cell, indexing via conventional pcr, multiplexed capture

Funders Acknowledgements:

IHU

RIVOC

IRD

Disclaimer

The information presented in this protocol is provided for informational purposes and does not replace the advice of a qualified professional. The content published on this platform has not been peer-reviewed. This protocol is intended for experienced users and must be followed under appropriate conditions and with suitable equipment to minimize the risks of errors or hazards. Any decision or action taken based on this information is solely your responsibility. By using this protocol, you acknowledge that neither the platform, the authors, the contributors, nor any associated individuals can be held liable for any consequences arising from its use.

Abstract

RNA sequencing allows for quantifying the RNA content of cells or samples, but it requires deep sequencing, which can be onerous and limits high-throughput analysis. Some applications only require the measurement of a few RNA biomarkers, and RNA-seq is not suitable for highly parallelized assays. This protocol presents an innovative method, named HAPPI, developed for multiplexed RNA quantification from samples of the apicomplexan parasite *Plasmodium falciparum*.

The HAPPI method relies on the multiplexed capture of RNA transcripts and the incorporation of unique molecular identifiers (UMIs) during reverse transcription, followed by amplification and indexing via conventional PCR. This approach enables precise and parallel capture of target RNA biomarkers while minimizing amplification biases. Specific mRNAs are captured using a pool of antisense primers containing a UMI linked to an adaptor. The synthesised primary cDNAs are amplified from the pool of specific sense primers and a portion of the adaptor sequence from the antisense primers, without prior purification. A second amplification step adds sequencing indexes and adaptors, enabling targeted and efficient analysis.

Guidelines

- All PCR amplification steps were performed using the QIAGEN Multiplex PCR Plus kit (Qiagen, 206152), specifically designed for simple and sensitive multiplex PCR without the need for optimization. This kit, based on exclusive multiplex PCR technology, includes a pre-optimized master mix containing HotStarTaq Plus DNA polymerase and an innovative PCR buffer system developed specifically for multiplex PCR. The stringent hot-start mechanism provided by HotStarTaq Plus DNA polymerase and the unique buffer composition enhance the specificity of the multiplex reaction by preventing extension of non-specifically hybridized primers and primer dimers.
- The kit is supplied with two optional reagents (Qiagen, 206152): Q-Solution and CoralLoad dye. Q-Solution promotes amplification of difficult targets, such as GC-rich regions or models with complex secondary structures. CoralLoad dye facilitates easy visualization of DNA migration during electrophoresis.
- A primer concentration of 0.1 μ M was found to be as effective as 0.2 μ M for PCR reactions

Materials

- *Plasmodium falciparum* RNA
- Trizol reagent (Invitrogen, 12034977)
- Chloroform (Fisher, 10488400)
- Isopropanol (Fisher, P-7507-17)
- Glycoblue (Invitrogen, 10301575)
- RNase-free water (InvitrogenTM 10977015)
- DNase kit (Ambion, AM1907)
- WarmStart Reverse Transcriptase 15U/ μ l (NEB, M0380L)
- Suprase Inhibitor 20 U/ μ l (InvitrogenTM, 10773267)
- dNTPs 10 mM (Fisher, 11853933)
- Qiagen Multiplex PCR Plus Kit (Qiagen, 206152)
- SPRIselect magnetic beads (Beckman Coulter, B23317)
- NucleoMag[®] NGS Clean up and Size Select beads (MACHEREY-NAGEL, 744970.5)
- Ampure XP (Beckman Coulter, A63881)
- PCR tube (InvitrogenTM, AM12450)
- QubitTM dsDNA High Sensitivity Assay Kit (InvitrogenTM, Q32851)
- InvitrogenTM ColibriTM Library Quantification Kit (InvitrogenTM, 15920523)
- high-sensitivity DNA analysis kit (Agilent, 5067-4626)
- RSB suspension buffer (Illumina, 20050023)
- PhiX V3 10% control (Illumina, FC-110-3001)
- Illumina iSeq 100 Kit (Illumina, 20008584)



Before start

- Ensure proper RNA preservation for optimal results.
- Verify that pipettes are correctly calibrated, as library preparations are highly sensitive to pipetting errors.
- Thoroughly clean the bench and pipettes before and between different steps of the manipulation to avoid contamination.
- Properly thaw reagents before use and strictly follow the manufacturer's instructions.
- Before each use, vortex and micro-centrifuge each reagent to ensure no reagent is stuck in the cap or on the side of the tube.
- Regularly change or disinfect gloves with RNA Away.
- Distinctly label your primers (sense, antisense) and perform dilutions separately to avoid contamination during mixing.
- Preferably perform reverse transcription under a hood pre-cleaned with DNA Away and exposed to UV light for about 10 minutes.

Overview of molecular of the protocol

- 1 Specific mRNAs are captured using a pool of antisense primers, each containing a unique molecular identifier (UMIs) linked to an adaptor sequence. The primary cDNA is then amplified using a pool of specific sense primers and a portion of the adaptor sequence that is incorporated during reverse transcription (RT), without requiring prior purification. This step enables both cDNA amplification and the incorporation of a second adaptor. Subsequently, a second amplification step uses the adaptor sequences to add sequencing barcodes (Figure 1).

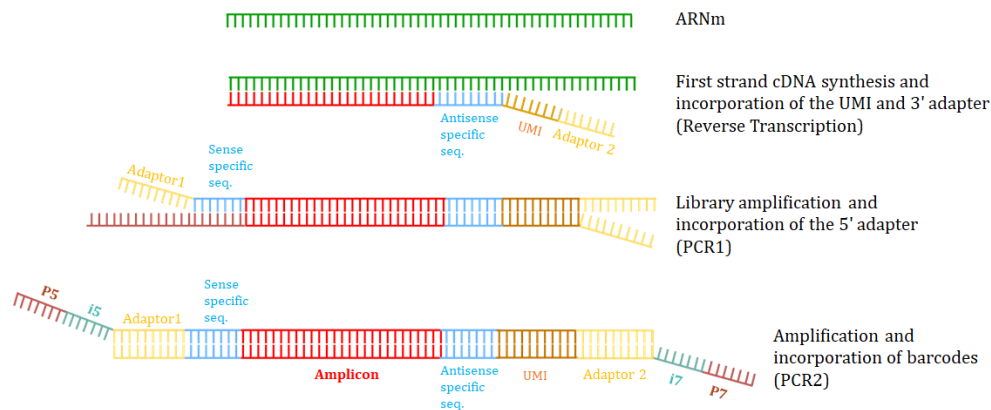


Figure 1: Molecular design for targeted amplification of RNA transcripts in multiplex.

Primer design

- 2 The primers were designed using the primer design services of GTseek LLC, according to the high-throughput genotyping by sequencing (GT-seq) method described by Campbell *et al.* (2015). For each transcript, an antisense primer was specifically designed for cDNA synthesis. This primer includes a gene-targeting sequence (20 to 26 bases), an 8-base UMI, and a 34-base R2 adaptor (5'-GTGACTGGAGTTCAGACGTGCTTCCGATCT-NNNNNNNN-[antisense sequence]-3'). The antisense primers have an annealing temperature between 57 °C and 63 °C, which allows for high-fidelity cDNA synthesis during reverse transcription.

For PCR amplification of the synthesized cDNA, sense primers were used to capture an amplicon approximately 300 bases in size, using the UMI-labeled cDNA as a template. These primers include a 33-base R1 adaptor sequence, followed by a gene-specific recognition site of 17 to 26 bases, with an annealing temperature between 57 °C and 63 °C (5'-ACCTCCACGACGCTTCCGATCT-[sense primer sequence]-3'). A

third universal primer, using a portion of the R2 sequence, served as the antisense primer. It is 20 bases long and has an annealing temperature of 59.6 °C.

Once amplification was completed, barcodes incorporation was performed using the adaptors from the R1 and R2 sites in a second PCR step. The primers used contain the R1 and R2 Illumina adaptor sequences, along with a 6-base index and 24-base P5 and P7 sequences that anneal to the flow cell.

Pooling primers

- 3 Specific antisense primers were resuspended at a concentration of 100 µM and pooled in a 1.5 mL tube, resulting in a concentration of 1.8 µM for each primer. The sense primers were pooled in the same manner. The mixtures were aliquoted and stored at -20°C until use.

Reverse transcription

- 4 Parasite RNA was extracted using Trizol reagent (Invitrogen, 12034977), following the manufacturer's protocol. Briefly, thawed samples were mixed with Trizol to lyse cells and dissolve proteins while preserving RNA. The addition of chloroform (Fisher, 10488400) allowed separation of aqueous and organic phases, with RNA recovered in the aqueous phase. RNA was then precipitated with isopropanol (Fisher, P-7507-17), using glycoblue (Invitrogen, 10301575) to facilitate pellet visualization. After centrifugation, the RNA pellet was washed with 70% ethanol to remove impurities and salts. Finally, the RNA was resuspended in sterile RNase-free water (Invitrogen™ 10977015). and incubated at 55°C to solubilize the pellet.
- 5 RNA was treated with DNase (Ambion, AM1907) to remove residual gDNA. Briefly, the extracted RNA was mixed with buffer and DNase I, followed by incubation at 37°C during 20-30 minutes. To terminate the reaction, the inactivation reagent was added, and the mixture was incubated at room temperature for 2 minutes, then centrifuged at maximum speed for 1.5 minutes. The purified RNA was collected from the liquid phase and stored at -80°C.
- 6 For reverse transcription, we used a concentration of 500 pM of antisense primers to minimize the presence of primer residues and avoid the need for a purification step before amplification of cDNA strands.
- 7 We used the highly specific WarmStart Reverse Transcriptase (NEB, M0380L) and assembled the following reaction:

A	B	C
Reagent	For 1 reaction (µl)	Final concentration
Buffer 10x	2	1x



A	B	C
SupraseIN 20 U/μl	0.5	0.5 U/μl
RTase Warmstart RTx 15 U/μl	0.25	0.19 U/μl
RNA	10.65	-
Antisense primer mix (1.8 nM each)	5.6	500 pM
dNTPs mix 10 mM	1	0.5 mM
Total reaction	20	-

- 8 The cDNA synthesis was carried out at 60°C (10 min), followed by heat inactivation at 80°C (10 min). A no-RT control was systematically included.

Library amplification_PCR1

- 9 The Qiagen Multiplex PCR Plus enzyme (Qiagen, 206152) was used to amplify the cDNA by assembling the following reaction:

A	B	C
Reagent	For 1 reaction (μl)	Final concentration
Master mix 2x	10	1x
Sense primer mix (1.8 μM each)	1.1	0.1 μM
Universal Primer 10 μM	0.2	0.1 μM
cDNA	8.7	-
Total reaction	20	-

- 10 Cycling conditions outlined below:

A	B	C	D
Initial activation step	5 min	95°C	25 cycles
Denaturation	30 s	95°C	
Annealing	30 s	65°C	
Extension	30 s	72°C	



	A	B	C	D
	Final extension	10 min	68°C	

Bead cleanup

- 11 For this step, SPRIselect magnetic beads (Beckman Coulter, B23317) were used at a concentration of 0.7X to clean-up reagent residues and remove nonspecific products below 200 bp (the smallest expected correct amplicons measured 259 bp, including adaptors). Alternatively, NucleoMag[®] NGS Clean-up and Size-Select beads (MACHEREY-NAGEL, 744970.5) can be used at an equivalent concentration with similar efficacy.
- 12 The following steps provide a detailed description of the cleaning process and the selection of specific products.
 - 12.1 Resuspend the SPRI beads (equilibrated at RT) by gentle shaking.
 - 12.2 Add the required volume of beads according to the PCR volume (for a bead concentration of 0.7X in a PCR volume of 20µl, use $0.7 \times 20 = 14$ µl of beads).
 - 12.3 Mix by pipetting several times and incubate for 5 minutes at room temperature.
 - 12.4 Place the strip on the magnetic support for 2 minutes until the beads have formed a pellet and the supernatant is completely clear.
 - 12.5 Gently remove the supernatant and rinse twice with 200µl of 70% ethanol.
 - 12.6 Remove the strip from the magnetic holder and centrifuge briefly. Replace on the magnetic holder for 2 minutes and remove excess alcohol with a 10 µl pipette tip.
 - 12.7 Remove the strip from the holder and leave to dry for 1 minute or until the pellet loses its lustre without closing the lid. Do not allow the pellet to dry completely to avoid it cracking and being difficult to resuspend.
 - 12.8 Elute by adding 30 µl of EB elution buffer. Mix well and leave for approximately 5 minutes at room temperature.



- 12.9 Gently remove the amplified cDNA-containing supernatant and transfer to new low adhesion tubes for storage.
- 12.10 Place the strip on the magnetic stand for approximately 2 minutes until the beads have pelleted and the supernatant is completely clear.

Qubit quantification of libraries (Optional)

- 13 After PCR1, you can quantify your libraries to normalize them before incorporating the barcodes for better yield. The libraries were quantified using the Qubit™ high-sensitivity double-stranded DNA assay kit (2 ng) (Invitrogen™, Q32851), according to the manufacturer's protocol, to ensure consistent input for the subsequent steps.

Incorporation of barcodes and sequencing adaptors_PCR2

- 14 This step involves indexing the libraries for parallel sequencing.
- 15 The Qiagen Multiplex PCR Plus enzyme was used to amplify the cDNA by assembling the following reaction:

A	B	C
Reagent	For 1 reaction (μl)	Final concentration
Master mix 2x	10	1x
Sense tagging primer 10μM	0.2	0.1 μM
Antisense tagging primer 10μM	0.2	0.1 μM
Product PCR 1	9.6	-
Total reaction	20	-

- 16 Cycling conditions outlined below:

A	B	C	D
Initial activation step	5 min	95°C	
Denaturation	30 s	95°C	20 cycles
Annealing	30 s	65°C	



	A	B	C	D
	Extension	30 s	72°C	
	Final extension	10 min	68°C	

Bead cleanup

- 17 This involves removing reagent residues without size selection. You can use AMPure XP (Beckman Coulter, A63881), SPRIselect or NucleoMag® NGS Clean-up and Size Select beads.
- 18 We used a concentration of 1.5X. The procedure follows the same steps as described in step 12.

Qubit quantification of libraries

- 19 The libraries were quantified with Qubit for normalisation before being pooled. We used the Qubit™ high-sensitivity (2 ng) double-stranded DNA assay kit (Invitrogen™, Q32851) according to the manufacturer's protocol.

Normalization and pooling of libraries

- 20 We converted the values reported by the Qubit (ng/μl) to nM using the following formula:
Concentration in nM = [(Concentration in ng/μl) / (660 g/mol × average size of library in bp)] × 10⁶
- 21 Subsequently, the libraries were diluted with RNase-free water.
- 22 10 μl of each library normalised to 10 nM was introduced into a RNase-free low-adhesion PCR tube (Invitrogen™, AM12450) and the suspension was pipetted and then microcentrifuged.

Qualitative analysis of the library pool

- 23 The sizes of the pooled libraries were analyzed on a BioAnalyzer, with the high-sensitivity DNA analysis kit (Agilent, 5067-4626), according to the manufacturer's protocol.
- 24 An example of a correctly amplified mix of amplicons (Figure 2).

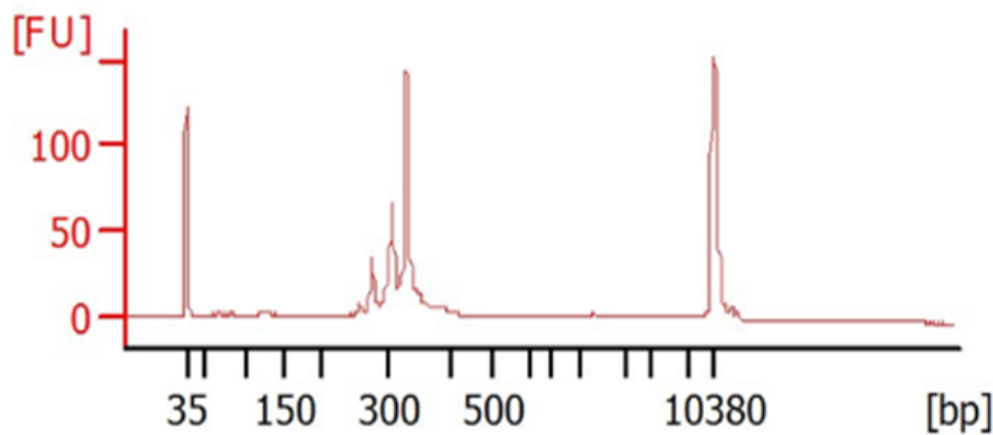


Figure 2: Size distribution of the DNA sample after purification with SPRISelect beads at 0.7x, analyzed on a BioAnalyzer (Agilent 2100), fluorescence units (FU) versus size (in bp).

qPCR quantification of pooled libraries

- 25 The final library was quantified by qPCR using the LightCycler® 96 instrument (Roche, 13072) with the Invitrogen™ Collibri™ Library Quantification Kit (Invitrogen™, 15920523).
- 26 It was diluted 1:1000 and 1:10 000 in Collibri™ dilution buffer inside RNase-free non-stick microfuge tubes according to the manufacturer's protocol.
- 27 Note: The dilution concentration depends on the starting concentration of the libraries. Tests should be performed to determine the appropriate dilution concentration. The manufacturer recommends starting from a concentration of 1:10 000 to 1:100 000. For this study, 1:1000 and 1:10 000 dilutions proved suitable for our libraries diluted to 10 nM. We chose to use a 1:10 000 dilution.
- 28 Preparation of the reaction
16 µl of master mix were distributed to the relevant wells of the qPCR plates. 4 µl of the diluted library was added in triplicate to the designated wells, along with 4 µl of each DNA standard and 4 µl of dilution buffer for the negative controls .

The assembly reaction is presented in the following table:

A	B	C
Reagent	For 1 reaction (µl)	Final concentration

A	B	C
Master mix 1.25x	16	1x
Diluted library 1:10 000 (1pM) Standart NTC	4	0.2 pM
Total reaction	20	-

29 The qPCR cycling conditions have been as follows:

A	B	C	D
Initial denaturation	2 min	95°C	1 cycle
Denaturation	30 s	95°C	35 cycles
Annealing/Extension	45 s	65°C	

Dilution of the final library and addition of the PhiX control

- 30 After quantification of the libraries to be sequenced by qPCR, the final library was diluted to 200pM with illumina RSB suspension buffer (Illumina, 20050023). The suspension was mixed and microcentrifuged.
- 31 Note: The input concentration of the library ready for sequencing depends on the kit used to prepare the library and the sequencer. For custom libraries, tests should be carried out to determine the optimum input concentration.
- 32 10% PhiX V3 control (FC-110-3001) diluted to 200 pM was added to the final library prior to sequencing.
- 33 Note: The PhiX control must be diluted to the same concentration as the final library. It is optional and its concentration depends on the level of nucleotide representation.

Illumina sequencing

- 34 Dual indexing sequencing was performed on the iSeq 100 paired read flow cell (Illumina, 20008584) with 151 cycles for sequencing reads and 6 cycles for index reads. Note: Dual indexing sequencing on the iSeq 100 requires the reverse complement of the i5 index primer sequence for reading. The primer design must include the corresponding adaptor sequence for reading the second i5 index, which differs from that used by other sequencers (MiSeq, HiSeq 2000/2500, NovaSeq).

- 35 20 µl of the final library was introduced into the bottom of the cartridge reservoir to be sequenced according to the manufacturer's protocol.

Data analysis

- 36 The demultiplexed reads (FASTQ), were analyzed using the reference sequences of the target genes. These sequences were extracted from the reference genome in FASTA format and then used to build an index with the BWA tool (version 0.7.17-r1188) via the "bwa index" command (Li, 2013). This step optimized the alignment of the reads by facilitating the rapid localization of corresponding regions on the reference sequence.
- 37 Next, the UMIs were extracted from the reads using the UMI-tools software (version 1.1.4) (Smith *et al.*, 2017). Once extracted, these UMIs were added to the header of each read, allowing each sequence to be associated with its unique molecular identifier and enabling the tracking of the original reads throughout the analysis process.
- 38 Subsequently, the reads were aligned to the reference sequence using the bwa mem command, generating SAM files. These files were filtered using Samtools (version 1.10) to retain only uniquely aligned reads, ensuring the specificity of the alignments (Danecek *et al.*, 2021).
- 39 Finally, a custom script was used to quantify the reads and their associated UMIs for each reference sequence, producing a summary table.

Command lines

- 40 Index Construction:
bwa index/path/to/amplicon_all_target.fasta.
- 41 UMI extraction:
umi_tools extract -l /path/to/reads.R2.fastq.gz --read2-in='/path/to/reads.R1.fastq.gz'
--bc-pattern=NNNNNNNN -S sample.output.R2.fastq.gz--read2-
out=sample.output.R1.fastq.gz
- 42 Read alignment:
bwa mem '/path/to/reference/sequences.fasta' 'output.R1.fastq.gz'
'sample.output.R2.fastq.gz' >
sample.aln.sam.
- 43 Filtering non-aligned reads:
samtools view -F 4 -h sample.aln.sam > mapped_sample.aln.sam

```
44   Excluding secondary and multiple alignments:
      samtools view mapped_sample.aln.sam -h | grep -v -e 'XA:Z' -e 'SA:Z' >
      unique_mapped_sample.aln.sam

45   Custom script:
      #Count UMIs and reads per target sequences in alignment file (sam)
      import argparse
      import sys

      #ARGUMENTS
      parser = argparse.ArgumentParser(description='')
      parser.add_argument("-A", "--alignment", dest="aln_file" , type=str, help="QC paired end
      alignment file")
      parser.add_argument("-O", "--output", dest="output_path" , type=str, help="output
      path")
      parser.add_argument("-S", "--sample", dest="sample_name" , type=str, help="sample
      name")
      args = parser.parse_args()

      aln_file = args.aln_file
      output_path = args.output_path
      sample_name = args.sample_name

      ref2umi = {} #keys : target sequences ; values : list of UMIs matching sequences
      ref2reads = {} #keys : target sequences ; values : list of reads matching sequences
      with open(aln_file, 'r') as leSam:
          for line in leSam:
              if line[0] != "@":
                  line = line.rstrip()
                  read_mapped = line.split('\t')[0]
                  corresponding_umi = read_mapped.split('_')[1]
                  reference_mapped = line.split('\t')[2]
                  if reference_mapped not in ref2umi:
                      ref2umi[reference_mapped] = [corresponding_umi]
                  else:
                      ref2umi[reference_mapped].append(corresponding_umi)
                  if reference_mapped not in ref2reads:
                      ref2reads[reference_mapped] = [read_mapped]
                  else:
                      ref2reads[reference_mapped].append(read_mapped)

      #Output : tab separated file
      #First column : target sequence
```

```
#Second column : number of aligned UMIs
```

```
#Third column : number of aligned reads
```

```
with open(output_path + "/" + sample_name + "_output.csv", 'w') as leOut:
```

```
    for element in ref2umi:
```

```
        leOut.write(">" + element + '\t' + str(len(set(ref2umi[element]))) + '\t' +  
str(len(ref2reads[element])) + '\n')
```

Protocol references

Campbell, N. R., Harmon, S. A., & Narum, S. R. (2015). Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. *Molecular ecology resources*, 15(4), 855-867.

Danecek, P., Bonfield J. K., Liddle J., Marshall J., Ohan V., *et al.* (2021). Twelve years of SAMtools and BCFtools, *GigaScience* , Volume 10, Numéro 2, giab008.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.

Smith, T., Heger, A., & Sudbery, I. (2017). UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome research*, 27(3), 491-499.

Acknowledgements

We would like to thank the Institut Méditerranée Infection in Marseille for the scholarship that enabled the completion of this work. We also express our gratitude to the MIVEGEC-IRD laboratories in Montpellier for their warm welcome and technical support. This work also benefited from partial funding by the Risques Infectieux et Vecteurs en Occitanie (RIVOC) program, and we are grateful for its contribution.