**Supplementary File 1: Primer scheme design and optimization and amplicon read depth analysis**

NOTE: This Supplementary File contains further advice on how to design, optimize and assess the performance of a primer scheme. This is for guidance only. It assumes a level of familiarity with bioinformatic tools and/or programming experience. If users are not comfortable with these steps, input from an experienced user is highly recommended.

1. **Address incomplete or non-representative reference data for the input to Primal Scheme.**

**NOTE:** If the reference data which are available are incomplete or not representative of the study area, there may be gaps in the sequencing coverage due to ‘amplicon drop-out’ (failure of the primers to bind and amplify the genome in this region). This is more likely to occur if there is limited existing knowledge of the diversity of circulating lineages or if users are sequencing viruses sampled across a wide geographic range.

* 1. Choose publicly available data
     1. In the absence of publicly available data, ideally, some preliminary metagenomic sequencing should be performed to obtain 1 or more reference sequences.
     2. If initial metagenomics is not possible, always try to choose the most complete genomes available from the area.
     3. One can choose genomes from a different geographical area as long as it is identified as closely related lineages to the ones circulating in the study area.
     4. Incorporate partial genomes obtained from GenBank. Gap-fill sequencing can be concatenated and spliced into the reference genome(s). Take care to ensure each part of the genome is correctly aligned and not to overrepresent any particular partial sequence.

NOTE:The majority of publicly available RABV sequences are partial genome data, particularly for nucleoprotein and glycoprotein genes. Therefore, there is potentially a lot to gain by including existing partial genome information in primer design. This is most useful to fill in missing ends of otherwise representative genomes, since the ends are generally highly conserved, or if a particular region is failing to sequence for certain lineage(s). Similarly, targeted gap-fill sequencing (across problematic regions of amplicon drop-out) can improve reference sequence data when redesigning primers in problematic genome regions (at relatively low cost) to allow primer redesign in these areas. (Refer to part 2).

* + 1. Use Primal Scheme to generate a new primer scheme.

NOTE:If you are unsure about these steps, you may instead modify individual primer sequences according to section 3 below.

1. **Sequence across large internal gaps using targeted PCRs.**

Targeted singleplex PCRs attempt to sequence across the regions of amplicon drop-out by generating amplicons larger than 400 bp to span across the gap. The partial genomes (large amplicons) generated can be used in Step 1.1.4. above.

NOTE:Before beginning this step, one may seek bioinformatics help to check if there is already any low-level representation of larger ‘gap-filling’ amplicons in the existing sequence data, where primers flanking the problematic region have naturally paired and amplified across it.

* 1. Identify the closest forward and reverse primers flanking the problematic region that generated amplicons with good read depth in previous rounds of sequencing.

NOTE***:*** If the gap is at the end of the sequence, try using a universal rabies primer (since sequence ends are highly conserved).

* 1. Select a PCR polymerase that is optimized for amplifying the size of the amplicon being attempted (a standard polymerase effectively amplifies fragments up to approximately 5 kb).

NOTE: The generation of larger amplicons relies on greater sample integrity.

* 1. Carry out singleplex PCR according to the manufacturer's instructions and annealing temperature of the selected primers.
  2. Check for the presence of the gap-fill amplicon product on an agarose gel.
  3. If present, process the gap-fill amplicon for sequencing following the Workflow section 5 onwards.

NOTE: Molarity calculations will vary for different amplicon lengths, adjust accordingly. Multiple gap-fill amplicons should be normalized in equimolar quantities to each other and any other samples being processed in parallel.

1. **Individual primer modification and amplicon depth equalization.**

Consistent amplicon dropout or low average read depth (refer to section 5 below) may result from *(a)* poor annealing of one or both of the primers (from the pair) to the genome variant, or *(b)* interaction between primers (in the scheme) resulting in primer sequestration as mentioned by Itokawa et al.43. The sequence and/or final concentration of individual primers may be optimized to equalize any large differences in average read depth per amplicon. Equalizing average read depth between amplicons helps to achieve full genome coverage earlier in the run, promoting re-use of the flow cell and cost-saving.

NOTE:If there are a large number of amplicons with low average read depth, or if new lineage diversity has emerged during the initial sequencing, it is recommended to generate a new primer scheme in Primal Scheme using an updated input fasta which includes the new sequences or partial genomes (refer to section 1).

* 1. Optimizing individual primer sequences.

NOTE***:*** Primer sequnences may be modified with nucleotide substitutions either within the original primer if appropriate, or by creating an alternative primer to represent a subset of the circulating diversity. Or alternatively, the position of the whole primer can be shifted to an area which is better conserved among the diversity of circulating lineages.

* + 1. Create an alignment of available sequences to the ‘index reference’ genome of the whole primer scheme.

NOTE:These may include whole and/or partial genomes from GenBank, your own sequencing or gap-fill sequences (refer to section 2).

* + 1. For each amplicon with low depth of coverage, visually inspect the annealing sites of both forward and reverse primers and compare the mismatching of the primers to the various genomes in the alignment.

NOTE: An appropriate genome alignment, manipulation or visualization software may be used in place of visual inspection if comfortable with this computationally.

* + 1. Determine whether one or both of the primers should be redesigned and how (e.g., nucleotide substitution, alternative primer, shift position of primer).

NOTE:If both primers appear not to have significant mismatches, low read depth could be due to hybridization and sequestration by another primer in the scheme. If this is the case, consider increasing the concentration of both primers in the pair. Primer-primer interactions can be checked with a primers analyzer tool. If changing the position of a primer, make sure the new position is known with respect to the index reference genome.

* + 1. Try to avoid introducing new primer-primer interactions to the scheme with manually generated primers. Primers should be checked for secondary structure formation and hybridization to other primers in the scheme using a primer analyzer tool.
    2. Once primer modifications are finalized, create new version configuration directories of the RAMPART protocol and ARTIC bioinformatics pipeline following the instructions in **Supplemental File 2**.
  1. Optimizing the final concentration of individual primers.

NOTE:Perform this step only after several initial rounds of sequencing.

* + 1. Refer to the “Achieving more even genome coverage” section of COVID-19 ARTIC v3 Illumina library construction and sequencing protocol V5. (<https://www.protocols.io/view/covid-19-artic-v3-illumina-library-construction-an-j8nlke665l5r/v5> ) for guidance before attempting.

NOTE: The protocol refers to Illumina sequencing but is fully transferable to amplicon sequencing on Nanopore platforms.

* + 1. If amplicons are consistently under-represented or over-represented, increase or decrease the final concentration of the primer pair respectively.

1. **Assess the expected performance of a primer scheme against set(s) of reference sequences in silico**

After generating a primer scheme, one may wish to see how well it is expected to perform against sets of sequences representing different diversity before starting sequencing.

* 1. Use a DNA manipulation and visualization program such as ApE by Davis et. Al.44 or nucleotide alignment software such as BLASTn suite-2Sequences to align primer oligo sequences to input sequences, or use a specific multiplex primer validation tool such as the R package openPrimeR by Doring et. Al.45.
  2. Assess mismatches between the input sequence(s) and each primer. One may consider summing and plotting the proportion of primers in the scheme which anneals with a given number of mismatches (e.g., 0, 1, 2, 3, unbound).
  3. Primers annealing with more than 3 mismatches indicate sequence variation that could affect amplification. It may be possible to redesign individual primers to account for extra diversity (see above, section 3), or change the reference sequences to better capture the intended diversity. If representative reference sequences are not available, refer to Part 1 above.

1. **Analyze average read depth per amplicon**

Analyze the depth of coverage per nucleotide of the reference genome to generate coverage profiles, as shown in **Figure 6** of the main text.

NOTE:Coverage profiles inform how well the primer scheme is performing across the genome. Excessive peaks or troughs appearing consistently between several samples likely indicate a bias or lack of annealing of primer pair(s) during the amplification step. These areas should be prioritized for optimization following the guidance provided above in section 3 of this Supplemental File.

* 1. Use the script **mappingSummary.sh** located in the directory **‘artic-rabv/other-scripts/’** to generate: a) summary statistics of mapped reads and b) depth of coverage per base. To use the script:
     1. Ensure the script is made executable in your local environment using the following command: **chmod u+x /path/mappingSummary.sh**
     2. Run the script from the directory containing the output from artic-rabv bioinformatics scripts:

**cd /path/analysis/run\_name**

**/path/artic-rabv/other-scripts/mappingSummary.sh**

* + 1. This script automatically finds bam files in the current directory and extract depth and coverage summary statistics that will be output in files ‘\_mappingStats.txt’ and ‘\_depth.txt’ for each sample.
  1. Use the **depth\_plots.R** script located in the directory **‘artic-rabv/DEPTHS/’** to import the mappingSummary.sh outputs to R and plot.