Frequently asked questions

1. **I'm losing a lot of DNA during SPRI bead clean-ups. Why would this be?**
    - Inexperience
    - Pellet overdries
    - % Ethanol not correct or not used at room temp
    - SPRI beads not at room temp or not properly mixed
    - Incorrect SPRI ratio
    - Not using low bind tubes
    - Inaccurate quantification
    - Not as much input material as you thought
2. **What is normalization and how do I normalize my DNA?**
   The purpose of normalization is to balance the representation of samples on a multiplex sequencing run- to achieve the same levels of coverage in the same amount of time. Inherent biases in sample sets mean that there is variation between (e.g. sample quality) and within samples (e.g. PCR amplicon bias). In order to do normalization you first need to quantify your DNA, fluorometer-based quantification methods are required to get accurate readings. You also need to know the size of your DNA, which can be obtained using prior knowledge (e.g. we amplified 400bp products so we know the expected size) or methods to size DNA (e.g. agarose gels, bioanalyzer, tapestation).
   For sequencing we want to normalize the *molarity* so we first need to calculate that using the formula below:

M=w/MW Where: *M:* DNA molarity, in mol *w:* DNA weight, in g *MW:* DNA molecular weight, in g/mol If only DNA length is known, the molecular weight is calculated as: *MW = DNA Length (bp) × DNA/RNA base weight* The approx. double strand DNA base weight is 660 Dalton, for single strand is 330, for RNA is 340. *M=w/MW* will give a value in moles, to convert to femtomolar (fmol) multiply by 1e+15

   Or, you could make things easier for yourself by using an online molarity calculator (e.g. https://nebiocalculator.neb.com) or an excel sheet formula (see the Nanopore_library_prep_EXCEL_protocol.xlsx sheet in 3_nanopore_sequencing, associated_resources folder).

3. **How do I design primers?**

If you are designing primers it means you already know something about your pathogen of interest, for example you have some reference sequences. This means you have something to base your primer design on! For the multiplex PCR strategy, for viruses, there is a tool called Primal Scheme (https://primalscheme.com/) that will design a scheme for you if you provide reference sequences as input, plus state your desired amplicon length and overlap. Note that this tool cannot deal with high levels of diversity, in which case you might want to consider other approaches like bait capture. Look at the primer_design tutorial in the extra materials folder for practice on primer design for rabies virus and also read the original reference for Primal Scheme (Quick et al, 2016 in 2_intro_to_workshop>masterclass_resources>associated_resources) plus our Jove paper (Bautista et al, 2023 in 1_opening_day>associated_resources) that has an appendix with helpful tips on optimizing primer sets.

4. How many times can I reuse a flowcell?

There isn't a straightforward answer to this question- it depends. The maximum lifespan of a flowcell is 72hours (but it's never actually that long) so in theory you should be able to reuse the flowcell until you have collectively used it for this length of time. However, in my experience it's pretty variable in terms of how many times it can be used. Sometimes once, sometimes five times! The ability to keep washing and reusing can be affected by contaminants in the library, which might degrade pores faster than normal, washes themselves may damage the pores if not done well. Overall though- pores will naturally die off over time and can't be recovered. So there is a limit to reuse.

5. How long do I sequence for?

Again, it depends. You want to sequences for as long as it takes to get the amount of data you want. And that varies depending on your pathogen/region of interest (e.g. genome size), your desired coverage threshold, number of samples being multiplexed, variation in the sample set etc etc. If you want to get a rough idea, you could use the following formula

C=LN/G
C stands for coverage
G is the haploid genome length L is the read length
N is the number of reads

Example
I want to get 100X coverage of rabies genome, which is ~12000bp in length. I am using an amplicon based approach so my reads are 400bp long. Therefore:
C=LN/G
100= 400xN / 12000
N=3000 reads
However, there is a lot of variability in my library due to PCR bias and sample variation so in reality I probably need more. Based on previous experience I aim for about 100,000 reads/sample.

6. How much does a sequencing run cost?

This very much depends on your method of sample and library preparation, the amount of samples per run (multiplexing), the number of times you can wash and resuse flowcells and how much you might be able to bulk buy to save costs. Generally, you should be able to

multiplex a lot of virus samples (because their genomes are small) if you are using a target enrichment based approach, like PCR amplification, because then you have a lot of input material. So the multiplex PCR approach is quite cost-effective compared to other methods like metagenomic sequencing. We have estimated a cost of about £38/sample (including costs from sample extraction right up to sequence) using this method for rabies but this is quite conservative and only considers 24 samples per run with only 1 wash so it could be cheaper. Note that sometimes the costs/sample given by companies or other researchers might only be considering library costs (i.e. not all the other prep like sample extraction, PCR etc), be careful to interpret accordingly.

7. What is the total yield from a nanopore flowcell?

With the R10.4.1 flowcells the maximum theoretical yield is 50 GB (gigabases) but it is highly unlikely that you will achieve this. Nanopore users suggest that about 5-15GB per run is a more realistic expectation.

8. What is RAMPART and how can I use it for my own sequencing run?

RAMPART is a software tool developed by the ARTIC network (https://artic.network/) and is open-source. It's available, with installation and running instructions, here: https://artic.network/rampart . RAMPART runs concurrently with MinKNOW and shows you demuxing / mapping results in real time. We use it to get an idea of coverage per sample at the amplicon level (e.g. we can see which amplicons are under or over performing) and when we have achieved our desired level of coverage for each sample (we aim for at least 100x coverage at each position in the genome. Note that RAMPART is a few years old now and hasn't been maintained recently so it won't work on newer computers like macs with M chips.