# Mapping putative spacer sequences

## Parse raw seq files

1. In the Terminal, change to the directory with the seq files
2. Run collate\_seq\_files.py (may need to adapt the source code if file headers contain unusual characters that will mess with the BLASTN search)
3. The fasta file will be saved in that directory

## BLASTN

1. Go to the BLASTN page: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>
2. Upload your newly created fasta file
3. Enter a helpful job title
4. In “Organism” choose the phage genome to search against
5. In Program Selection, make sure “Highly similar sequences (megablast)” is selected
6. Hit the BLAST button and wait
7. On the results page, go to “Download” and select “Multiple-file XML2”. This will save a zip archive of the results for each query. Unzip it once the download is finished.

## Geneious

1. Open Geneious
2. Make sure you have a local copy of the phage genome
3. Go to File > Import… > From Multiple Files…
4. In the dialogue, select “BLAST XML summaries” in the File Type. Navigate to the unzipped directory of the BLAST results, select all the file and hit Import
5. You might get an error like “Unable to open file”. You can safely ignore this – it’s only because there are search results that didn’t map to the phage genome
6. Select all the imported alignment documents and the phage genome file, and go to Align/Assemble > Map To Reference. Ensure that the phage genome is set as the Reference Sequence. The default setting should be adequate
7. Hit OK and wait
8. Go to the new assembly document. In General options, under Show, select “Query” and sort by “Read direction”. This will give the contigs ordered by forward and reverse reads, with the original name of the sequence (i.e. the clone they came from)