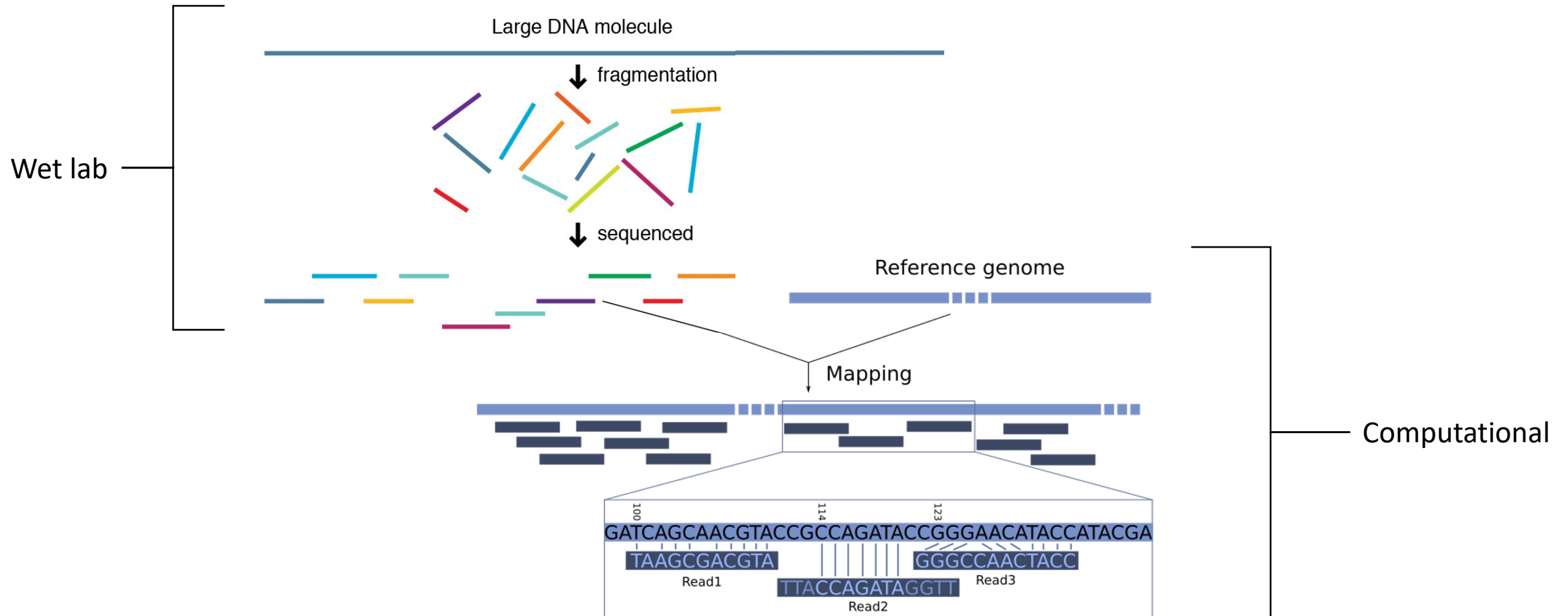
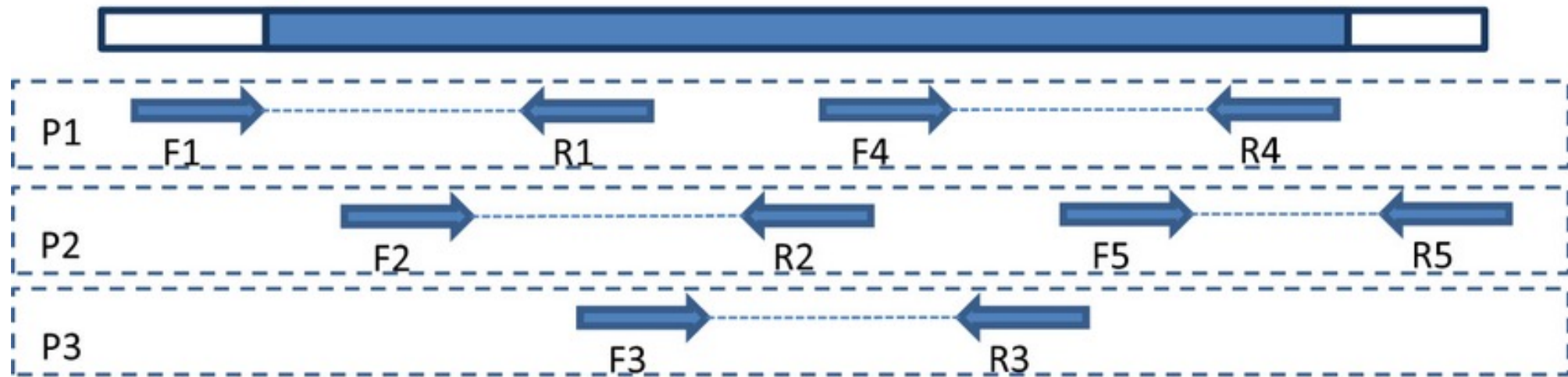


Shotgun Sequencing Pipeline

What is shotgun sequencing?



ARTIC sequencing



Preparing tools and conda environment

```
conda env create --file shotgun.yml  
conda activate shotgun
```

```
wget https://snpeff.blob.core.windows.net/versions/snpEff_latest_core.zip
```

```
unzip snpEff_latest_core.zip  
cd snpEff  
pwd  
*save the pwd path
```

- Conda environment:
 - seqkit
 - fastp
 - seqtk
 - bwa
 - samtools
 - lofreq
 - pandas
- snpEff

Download necessary scripts

```
git clone https://github.com/shandley/viral_genome_sop.git
```

```
mv viral_genome_sop/* ./
```

```
vi covid_snpotate.sh  
i  
*replace /home/shandley/install_files/ with your path  
*press esc  
:wq!
```

Clean raw reads

```
./isolate_cleaning.sh
```

```
cat cleaned_seqs/input_stats.txt
```

- seqkit:
 - Statistics of raw reads → input_stats.txt
- fastp:
 - Quality control → cleaned_seq directory
- num_seqs:
 - Good to keep under 10,000,000 reads

Subsampling (optional)

```
for i in *.fastq.gz; do seqtk sample -s100 $i 100000000 > sub_$i ; done
```

```
mkdir subsample  
mv sub_* subsample/  
mv *.sh subsample  
mv NC_045512.2.fasta subsample/  
cd subsample/
```

```
./isolate_cleaning.sh
```

- Seqtk
 - Subsamples raw reads

Alignment, sort, variant calling

```
./isolate_variant_caller.sh -r NC_044512.fasta -t 12
```

```
mv covid_snp* variants  
cd variants
```

- Bwa:
 - Maps raw reads against reference genome → .sam
- Samtools:
 - sort → sorted .bam
 - Removes duplicates
- LoFreq:
 - Corrects mapping errors
 - Inserts indel qualities
 - Inserts alignment qualities
 - Variant calling → _vars.vcf

Filter and annotate variants

```
./covid_snp_filter.sh
```

- Lofreq:
 - Filters variants based on coverage (> 75)

```
./covid_snpotate.sh
```

- snpEff:
 - Annotates variants

Filtering annotated vcf files

```
git clone https://github.com/anajung/viral_genomes.git
```

```
mv viral_genomes/* ./  
pwd  
*copy path  
vi run_info_splitter.sh  
i  
*replace '/Users/anajung/Documents/HandleyLab_Code'  
with own path  
bash run_info_splitter.sh
```