Supplementary Worksheet 1 - BioHansel

Geneviève Labbé et al

2019-11-26

Detailed parameter settings for all programs used for validation

Concordance Assessment - *S. typhi*

Command-lines for BioHansel and Snippy found in Nextflow workflow nf-biohansel-snippy-comparison

BioHansel command-line:

```
mkdir -p reads
ln -s `realpath *.fastq.gz` reads/
hansel \\
  -v \\
  -t 1 \\
  --min-kmer-freq $mincov \\
  -s $scheme_fasta \\
  -D reads/ \\
  -o $summary_report \\
  -O $detailed_report
```

Parameters:

- `-v` verbose logging
- `-t 1` run BioHansel with one CPU thread
- `-D reads/` run BioHansel on reads symlinked to `reads/` directory
- `$scheme_fasta` path to *S. typhi* BioHansel scheme
- `$mincov` minimum kmer frequency value of 3, 6, 8
- `$summary_report` BioHansel summary report output file
- `$detailed_report` BioHansel detailed results output file

Snippy command-line:

```
snippy --prefix $accession \\
    --outdir $accession \\
```

```
--cpus 16 \\
--ram 16 \\
--mincov $mincov \\
--R1 $reads1 \\
--R2 $reads2 \\
--ref $ref_genbank \\
--tmpdir ./
```

- `$accession` SRA run accession (sample name)
- `$mincov` minimum coverage depth of 3, 6, 8
- `$reads1`/`$reads2` forward and reverse paired-end Illumina reads paths
- `$ref_genbank` GenBank file for reference genome NC_003198.1 (*Salmonella enterica* subsp. enterica serovar Typhi str. CT18, complete genome.)