

Output description for IRMA service for Influenza

Pipeline overview

- [FastQC](#) v0.11.8 - read quality control.
- [Trimmomatic](#) v0.33 - adapter and low quality trimming.
- [IRMA](#) v0.9.3 - Influenza assembly, variant calling and phasing.

Preprocessing

FastQC

[FastQC](#) gives general quality metrics about your reads. It provides information about the quality score distribution across your reads, the per base sequence content (%T/A/G/C). You get information about adapter contamination and other overrepresented sequences.

For further reading and documentation see the [FastQC help](#).

Output directory: `01-fastqc`

- `{sample_id}/{sample_id}_R[12]_fastqc.html`
 - html report. This file can be opened in your favourite web browser (Firefox/chrome preferable) and it contains the different graphs that fastqc calculates for QC.
- `{sample_id}/{sample_id}_R[12]_fastqc`
 - folder with fastqc output in plain text.
- `{sample_id}/{sample_id}_R[12]_fastqc.zip`
 - zip file containing the FastQC report, tab-delimited data file and plot images

Trimming

[Trimmomatic](#) is used for removal of adapter contamination and trimming of low quality regions. Parameters included for trimming are:

- Nucleotides with phred quality < 10 in 3'end.
- Mean phred quality < 20 in a 4 nucleotide window.
- Read length < 50

Results directory: `02-preprocessing`

- Files:
 - `{sample_id}/{sample_id}_R[12]_filtered.fastq.gz`: contains high quality reads with both forward and reverse tags surviving.
 - `{sample_id}/{sample_id}_R[12]_unpaired.fastq.gz`: contains high quality reads with only forward or reverse tags surviving.

Note: To see how your reads look after trimming, look at the FastQC reports in the ANALYSIS/{ANALYSIS_ID}/03-preprocQC directory

Note: From now on, all the steps will be host specific.

IRMA

IRMA was designed for the robust assembly, variant calling, and phasing of highlyfrag variable RNA viruses. Currently IRMA is deployed with modules for influenza and ebolavirus. IRMA is free to use and parallelizes computations for both cluster computing and single computer multi-core setups.

Results directory: `04-irma/{sample_id}`

- Files:
 - `amended_consensus`: Assembled consensus per gene segment w/mixed base calls
 - `figures`
 - `{fragment_id}-coverageDiagram.pdf`: Shows coverage and variant calls.
 - `{fragment_id}-heuristics.pdf`: Heuristic graphs for the fragment.
 - `{fragment_id}-EXPENRD.pdf`: Fragment's variant phasing using 'experimental enrichment' distances.
 - `{fragment_id}-JACCARD.pdf`: Fragment's variant phasing using 'modified Jaccard' distances.
 - `{fragment_id}-MUTUALD.pdf`: Fragment's variant phasing using 'mutual association' distances.
 - `{fragment_id}-NJJOINTP.pdf`: Fragment's variant phasing using 'normalized joint pobability' distances.
 - `READ_PERCENTAGED.pdf`: Break down for reads assembled.
 - `intermediate`: intermediate data for each step.
 - `logs`:

- `ASSEMBLY_log.txt`: SSW scores per all rounds tried in the iterative refinement.
- `NR_COUNTS_log.txt`: Read pattern counts at various stages.
- `QC_log.txt`: Quality control output.
- `READ_log.txt`: Counts of assembled reads from BAM files.
- `FLU-Mixture_Example.sh`: Configuration file corresponding to this IRMA run.
- `run_info.txt`: Table of parameters used by the IRMA run.
- `matrices`: Phasing matrices used to generate heat maps.
- `secondary`:
- `unmatched_read_patterns.tar.gz`: Archive of left over read patterns that did not match FLU.
- `tables`:
- `{fragment_id}-pairingStats.txt`: Summary of paired-end merging statistics, if applicable.
- `{fragment_id}-coverage.txt`: Summary coverage statistics for the assembly.
- `{fragment_id}-allAlleles.txt`: Statistics for every position & allele in the assembly.
- `{fragment_id}-insertions.txt`: Called insertion variants.
- `{fragment_id}-deletions.txt`: Called deletion variants.
- `{fragment_id}-variants.txt`: Called single nucleotide variants.
- `READ_COUNTS.txt`: Read counts for various points in the assembly process.
- `{fragment_id}.fasta`: Final assembled plurality consensus (no mixed calls) for the fragment.
- `{fragment_id}.vcf`: Custom variants call file for called IRMA variants for that fragment.
- `{fragment_id}.bam`: Sorted BAM file for the final fragment's assembly (merged if applicable).
- `{fragment_id}.bam.bai`: BAM file index.