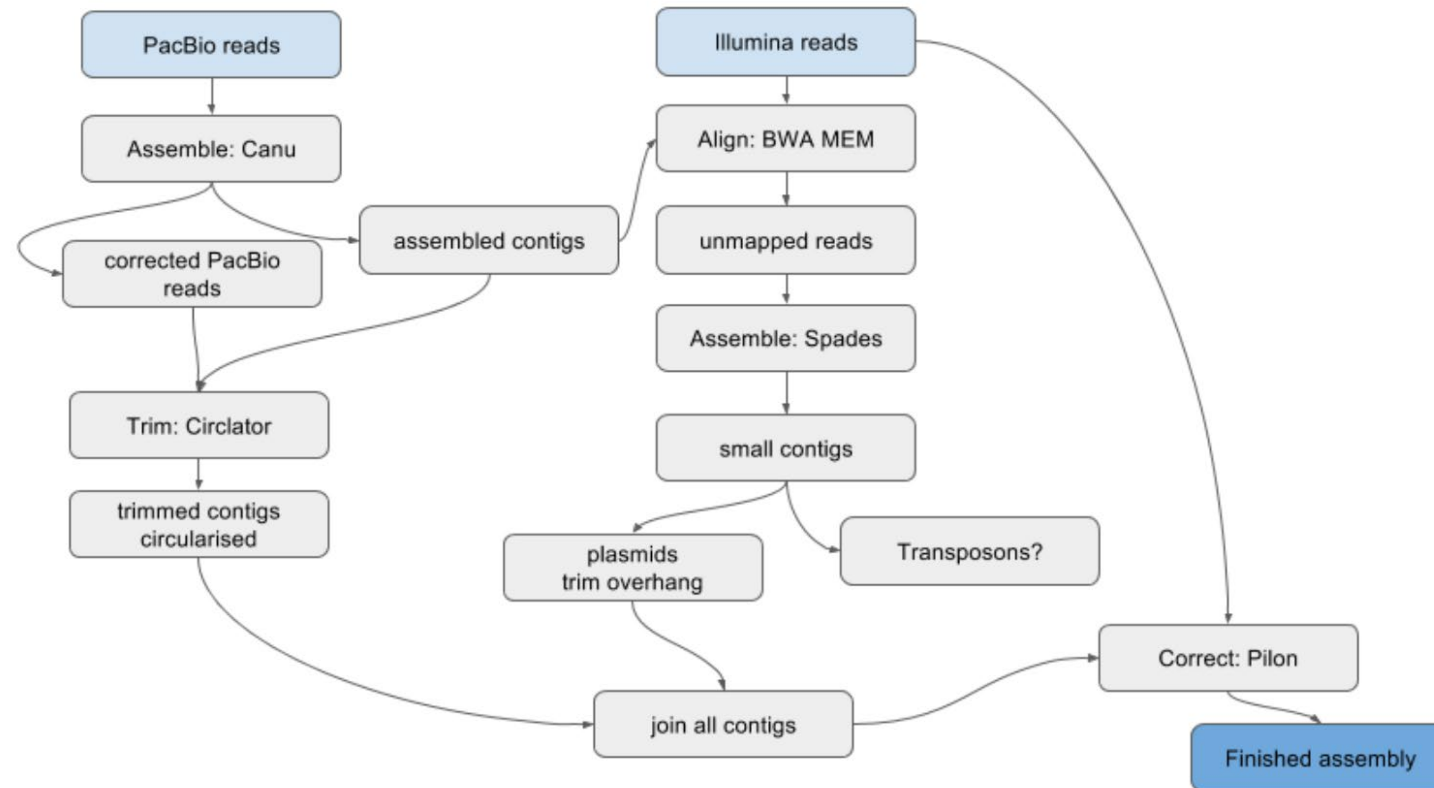
PacBio

Typical PacBio workflow



Running Canu in Galaxy

Galaxy Europe

Analyze Data Workflow Visualize Shared Data Help User

Using 2%

Tools

canu

Show Sections

Canu assembler Assembler optimized for long error-prone reads such as PacBio, Oxford Nanopore

GLASSgo sRNA homolog finder

Interproscan functional predictions of ORFs Interproscan functional predictions of ORFs

Interproscan functional predictions of ORFs Interproscan functional predictions of ORFs

WORKFLOWS

All workflows

Canu assembler Assembler optimized for long error-prone reads such as PacBio, Oxford Nanopore (Galaxy Version 1.8)

Favorite Versions Options

Input reads

17: R1.fq
16: pacbio.fq
6: S288C_reference_sequence_R64-chromosome_names.fasta
5: H3_TTAGGC_L002_R2_001.fastq.gz
4: H3_TTAGGC_L002_R1_001.fastq.gz
3: GSM461180_1.fastqsanger
2: GSM461177_2.fastqsanger

Mode

PacBio raw

To restrict canu to only a specific stage, use

all

Estimated genome size (e.g. 80m, 15k, 2g)

30000

(genomeSize)

Maximum raw overlap mismatch

The defaults are 0.300 for PacBio reads and 0.500 for Nanopore reads. (rawErrorRate)

Maximum corrected overlap mismatch

The allowed difference in an overlap between two corrected reads. Assemblies of low coverage or data with biological differences will benefit from a slight increase in this. Defaults are 0.045 for PacBio reads and 0.144 for Nanopore reads. (correctedErrorRate)

Minimum read length

History

search datasets

RNAseq

16 shown, 7 deleted

4.84 GB

23: Canu assembler on data 16 (trimmed reads)
22: Canu assembler on data 16 (corrected reads)
21: Canu assembler on data 16 (unitigs)
20: Canu assembler on data 16 (unassembled)
19: Canu assembler on data 16 (contigs)
18: R2.fq
17: R1.fq
16: pacbio.fq
8: 189_ featureCounts on collection 116_ Counts _H1.tgz

Canu commandline

```
usage:   canu [-version] [-citation] \
           [-correct | -trim | -assemble | -trim-assemble] \
           [-s <assembly-specifications-file>] \
            -p <assembly-prefix> \
            -d <assembly-directory> \
            genomeSize=<number>[g|m|k] \
           [other-options] \
           [-pacbio-raw |
            -pacbio-corrected |
            -nanopore-raw |
            -nanopore-corrected] file1 file2 ...

example: canu -d run1 -p godzilla genomeSize=1g -nanopore-raw reads/*.fasta.gz
```

```
To restrict Canu to only a specific stage, use:
   -correct       - generate corrected reads
   -trim          - generate trimmed reads
   -assemble      - generate an assembly
   -trim-assemble - generate trimmed reads and then assemble them
```

Canu options

The assembly is computed in the -d <assembly-directory>, with output files named using the -p <assembly-prefix>. This directory is created if needed. It is not possible to run multiple assemblies in the same directory.

The genome size should be your best guess of the haploid genome size of what is being assembled. It is used primarily to estimate coverage in reads, NOT as the desired assembly size. Fractional values are allowed: '4.7m' equals '4700k' equals '4700000'

A full list of options can be printed with '-options'.

Reads can be either FASTA or FASTQ format, uncompressed, or compressed with gz, bz2 or xz.

Reads are specified by the technology they were generated with, and any processing performed:

```
-pacbio-raw          <files>     Reads are straight off the machine.
-pacbio-corrected    <files>     Reads have been corrected.
-nanopore-raw        <files>
-nanopore-corrected  <files>
```