Sep 02, 2022

De-novo assembly of Xanthomonas genomes from Illumina NovaSeq reads

David J Studholme[1], Jamie Harrison[1]

[1]University of Exeter

1 Works for me Share

This protocol is published without a DOI.

David J Studholme

DISCLAIMER

DISCLAIMER – FOR INFORMATIONAL PURPOSES ONLY; USE AT YOUR OWN RISK

The protocol content here is for informational purposes only and does not constitute legal, medical, clinical, or safety advice, or otherwise; content added to protocols.io is not peer reviewed and may not have undergone a formal approval of any kind. Information presented in this protocol should not substitute for independent professional judgment, advice, diagnosis, or treatment. Any action you take or refrain from taking using or relying upon the information presented here is strictly at your own risk. You agree that neither the Company nor any of the authors, contributors, administrators, or anyone else associated with protocols.io, can be held responsible for your use of the information contained in or linked to this protocol or any of our Sites/Apps and Services.

ABSTRACT

This protocol describes the *de-novo* assembly of *Xanthomonas* genome sequences from short-read genomic shotgun sequencing data. It includes quality control of the raw sequence reads, assembly and finally polishing of the assembly based on alignment of reads against the preliminary assembly.

PROTOCOL CITATION

David J Studholme, Jamie Harrison 2022. De-novo assembly of Xanthomonas genomes from Illumina NovaSeq reads. **protocols.io** https://protocols.io/view/de-novo-assembly-of-xanthomonas-genomes-from-illum-cfiktkcw

FUNDERS ACKNOWLEDGEMENT

UKRI

Grant ID: BB/T010916/1

LICENSE

CREATED

Aug 19, 2022

LAST MODIFIED

Sep 02, 2022

PROTOCOL INTEGER ID

68908

DISCLAIMER:

> DISCLAIMER – FOR INFORMATIONAL PURPOSES ONLY; USE AT YOUR OWN RISK
>
> The protocol content here is for informational purposes only and does not constitute legal, medical, clinical, or safety advice, or otherwise; content added to protocols.io is not peer reviewed and may not have undergone a formal approval of any kind. Information presented in this protocol should not substitute for independent professional judgment, advice, diagnosis, or treatment. Any action you take or refrain from taking using or relying upon the information presented here is strictly at your own risk. You agree that neither the Company nor any of the authors, contributors, administrators, or anyone else associated with protocols.io, can be held responsible for your use of the information contained in or linked to this protocol or any of our Sites/Apps and Services.

1 Software pre-requisites.

> This protocol assumes that you have already installed fastp, SPAdes, SAMtools, BowTie2 and Pilon. I also assumes that the paired Illumina sequence data comprises two gzipped FASTQ files called *name_r1.fq.gz* and *name_r1.fq.gz*.

2 Perform quality-based filtering and adapter trimming using fastp.

```
mkdir name_fastp_out
```

Creates a directory for the fastp QC report files.

```
fastp -i name_r1.fq.gz -I name_r2.fq.gz -o name_trimmed_r1.fq.gz -O name_trimmed_r2.fq.gz --unpaired1 name_trimmed_unp.fq.gz --unpaired2 name_trimmed_unp.fq.gz -r --cut_right_window_size 5 --cut_right_mean_quality 20 -c -l 50 -j name_fastp_out/name_fastp_report.json -h name_fastp_out/name_fastp_report.html
```

Generates trimmed and filtered sequence files and QC reports on the Illumina NovaSeq FASTQ sequence files.

3 Perform *de-novo* assembly using SPAdes.

```
spades.py -1 name_trimmed_r1.fq.gz -2 name_trimmed_r2.fq.gz -s name_trimmed_unp.fq.gz --careful --cov-cutoff auto -o name_spades_out
```

Performs the de-novo assembly.

4 Polishing with Pilon

This step assumes that the SPAdes assembly is contained in a file in the current working directory called *name.fasta*. It is assumed that the two trimmed-and-filtered gzipped FASTQ files are also in the current working directory. If these files are located elsewhere, then you can make symbolic links to them in the current working directory.

```
bowtie2-build name.fasta name
```

Creates BowTie2 index files with 'name' as the prefix for their filenames.

```
bowtie2 -x name -1 name_trimmed_r1.fq.gz -2 name_trimmed_r2.fq.gz -S name_vs_name.sam
```

Performs alignment of the trimmed-and-filtered reads against the genome assembly to generate an alignment in SAM format.

```
samtools view -b -T name.fasta name_vs_name.sam -o name_vs_name.sam.bam
```

Converts the SAM-formatted file into BAM format.

```
samtools sort --reference name.fasta name_vs_name.sam.bam -o name_vs_name.sam.bam.sorted.bam
```

Sorts the BAM file.

```
samtools index name_vs_name.sam.bam.sorted.bam
```

Indexes the sorted BAM file.

```
rm name_vs_name.sam.bam $name_vs_$name.sam
```

Removes the intermediate files to save disk space.

```
pilon --genome name.fasta --frags name_vs_name.sam.bam.sorted.bam
--output name.pilon --outdir name_pilon_out
```

Generates a modified genome assembly based on reconciling discrepancies between assembly and aligned reads.

The polished genome assembly in FASTA format can be found in the Pilon output directory: ./*name_pilon_out/name.pilon.fasta*

This file can now be subjected to further quality control and/or submitted to public repositories.

5 Bibliography

Chen S, Zhou Y, Chen Y, Gu J (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor.. Bioinformatics (Oxford, England). https://doi.org/10.1093/bioinformatics/bty560

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing.. Journal of computational biology : a journal of computational molecular cell biology.
https://doi.org/10.1089/cmb.2012.0021

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H (2021). Twelve years of SAMtools and BCFtools.. GigaScience.
https://doi.org/pii:giab008.10.1093/gigascience/giab008

Langmead B, Salzberg SL (2012). Fast gapped-read alignment with Bowtie 2.. Nature methods.
https://doi.org/10.1038/nmeth.1923

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement.. PloS one.
https://doi.org/10.1371/journal.pone.0112963