

COMPARISON OF PAIRWISE LONG READ ALIGNERS

Type of project*: Bachelor End

Student background:** Bachelor CS/NB/LST

Background

In the past few years, with the improvements in sequencing technologies and introduction of long reads, many new aligners have been developed which can work with long read sequences. As pairwise alignment is input of de novo genome assembly which is an important bioinformatics problem, a fair comparison between aligners is essential.

The goal of this project is to provide a fair comparison between long read pairwise aligners.

Aims

The aim of this project is to find out

Which aligner produces the best alignments for de novo assembly?

What is the impact of sequencing technology on the performance of aligners?

Approach:

- 1- Simulate Pacbio SMRT or ONT Nanopore long reads from *S.cerevisiae* and *S.toberosum* chromosome 11.
- 2- Perform pairwise alignment of simulated reads and do *de novo* assembly based on the overlaps.
- 3- Assess and compare genome assembly performance with respect to a golden standard reference and measure intrinsic characteristics (Contiguity, N50, size, ...).

List of Aligners (select few): Minimap2, Daligner, Cosine, MHAP, Bwa-sw, Blast, Star, Mummer, BBmap

List of Assemblers (select few): Canu, Miniasm, SMARTdenovo, PBcR

Suggested literature

<https://academic.oup.com/bioinformatics/article-abstract/34/18/3094/4994778?redirectedFrom=fulltext>

https://link.springer.com/chapter/10.1007/978-3-662-44753-6_5

<https://academic.oup.com/nar/article/45/14/e132/3861609>

<https://www.biorxiv.org/content/early/2014/08/14/008003>

<https://academic.oup.com/bioinformatics/article/26/5/589/211735>

<https://www.nature.com/scitable/topicpage/basic-local-alignment-search-tool-blast-29096>

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3530905/>

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5802927/>

https://jgi.doe.gov/wp-content/uploads/2013/11/BB_User-Meeting-2014-poster-FINAL.pdf

<https://www.ncbi.nlm.nih.gov/pubmed/28298431>

<https://academic.oup.com/bioinformatics/article-abstract/34/18/3094/4994778?redirectedFrom=fulltext>

<https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbx147/4590140>

<https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbx147/4590140>

<https://github.com/rrwick/Basecalling-comparison>

Responsible supervisor: Thomas Abeel – t.abeel@tudelft.nl - EEMCS Pattern Recognition and Bioinformatics

*Types of project: Bachelor seminar (TI3706): 5 ECTS literature review course // Research Assignment (IN5010): 15 ECTS bioinformatics literature review // BEP: 10-15 ECTS Bachelor End Project // MEP: 30-60 ECTS Master End (thesis) Project // internship: 3 month no credit project. The type of project you are completing will impact the scope and depth that you will be expected to accomplish.

** Student background: The Delft Bioinformatics Lab serves a broad student community with a variety of projects. The background mentioned here is a suggestion and not a restriction. We will adapt the scope and focus of the project to connect well with your expertise and program requirements.