

The Edena v3 assembler, development version 110705

Reference manual

License

Edena is free only for non-profit academic uses. Using this program implies the acceptance by the user of the terms of the license "LICENSE.TXT" included in the package.

Development version

You downloaded a development version of Edena. This means that some features are incomplete and that the program has not been thoroughly tested. If you experience weird results or behaviors, please report the case to david.hernandez@genomic.ch.

Overview

A quick description of the program options is obtained by typing: `edena -h`

Edena is an overlaps graph based short reads assembler. The maximum supported read length is 128 nucleotides. However, reads that are too long may impair the assembly performance. In such a case, you should consider the "3' end reads truncation" procedure described below.

This program requires the reads to be all the same length, as Illumina GA reads are. This is due to historical reasons and because it greatly simplifies several computational steps. 454 or Sanger reads are therefore not suited to Edena. If you provide multiple files with different reads lengths, Edena will trim the 3' end of the reads so that the reads are all the same length as the shortest reads in the files.

The program was developed in a framework of whole genome bacterial assemblies. It is therefore more suited for this kind of task though we also successfully used it for other types of projects.

An assembly with Edena is a two step process: *overlapping* and *assembling*.

Overlapping mode: The reads files are provided to the program which computes the transitively reduced overlaps graph. This structure is then stored together with the sequence reads in a binary file suffixed with ".ovl".

Assembling mode: The ".ovl" file is provided to the program, as well as some assembly parameters. A set of contigs in FASTA format is outputted. The purpose of having a two step process is that the .ovl file is computed only once and can then be used to produce assemblies with different parameters.

Usage: Overlapping mode

Edena can accept both unpaired and paired files, fastq and fasta format. Note that for technical reasons, all reads are required to be of the same length. You can however provide the program with different files containing different reads length. In such case, Edena will trim the 3' ends of the longer reads so that they fit the shorter length. It is however required that reads within each individual file are the same length (as Illumina GA reads are). (See also the 3'end reads truncation option).

-r <files...> **unpaired files**

Unpaired files are provided using the flag -r

```
>edena -r file1 file2 ...
```

-paired <files...> **paired-end files**

Paired files are specified by pairs, multiple pairs can be specified. **!At the moment, Edena considers all the paired files to be issued from the same library size. Mixing paired files with different sizes may seriously reduce the assembly performance!**

```
>edena -paired file1_pair1 file1_pair2 file2_pair1 file2_pair2 ...
```

-t <value> **3' end reads truncation**

Use this option to truncate the 3'end of the reads such that the resulting length is <value>. You may consider reads truncation since it can significantly improve the assembly. Since Edena computes exact overlaps, only error free reads can take part to the assembly. Since errors are likely to occur at the 3' ends, shortening the reads by some nucleotides may increase the number of errors-free reads in the dataset, and thus increase the assembly performance.

-M <value> **minimum overlap size to compute**

If not specified, this value is set to half of the reads length. When the sequencing coverage is sufficient, you can increase this value which will reduce the computational time. Edena will compute the overlaps whose sizes range from this value to the reads length.

-p <name> **prefix name for output files**

Usage: Assembling mode

-e <.ovl file> **Edena “.ovl” file**

Specify here the Edena “.ovl” file obtained from the overlapping step

Suggestions, comment, bugs

david.hernandez@genomic.ch.