

SOLiD™ System Application Documentation:

solid_denovo_postprocessor.pl

1.0	10/22/08	Vrunda Sheth	First draft.
-----	----------	--------------	--------------

Program

Program Name

solid_denovo_postprocessor.pl

Program Version

solid_denovo_postprocessor.pl v1

Description

Application

De novo assembly of SOLiD™ colorspace reads.

Description

The solid_denovo_postprocessor.pl script is a part of the Velvet de novo assembly pipeline.

Release Notes

First release

Algorithm/Script Description

This script parses the read tracking information from Velvet and combines it with the individual colorspace read sequences to create an input file for the de novo adaptor program.

Installation

Installation Instructions

Download the script and make it executable. You may need to change the first line of the script if your Perl is in a non-standard location (i.e., not /usr/bin/perl). You may need to install the Getopt::Long, IO::Handle, FileHandle, Tie::File, English, and File::Basename Perl modules from CPAN. Requires Perl version 5.8.5.

Usage

Usage Parameters - required

--afgfile <file> afg file from Velvet run.

--csfasta <file> colorspace_input.csfasta (output from solid_denovo_preprocessor.pl)

Multiple input csfasta files may be specified using additional --csfasta file parameter, but order must match the order in which they were submitted to Velvet.

--output <file> Name of your output file. Give full path name.

Usage Parameters - optional

--run fragment | mates

Specifies whether input files come from a mate-pair or a fragment run (default is fragment)

-h | --help

Prints a usage summary

-v | --version

Prints the program version

Usage Parameters: is argument order important?

If multiple csfasta input parameters are specified, order must match the order in which they were submitted to Velvet.

Usage Example

Example1

```
./solid_denovo_postprocessor.pl --afgfile Velvet_asm.afg --csfasta  
colorspace_input.csfasta --run fragment
```

List of programs/scripts included

This script belongs to the Velvet de novo assembly pipeline. Refer to Solid_denovo_HowTo for more information on how to do de novo assembly of SOLiD™ reads.

Other scripts called by this program

n/a.

Path constraints

Specify full path names of input and output files.

System Input Files

- One or more csfasta files from the solid_denovo_preprocessor
- Read tracking file from Velvet in afg format

Input File Versions supported

n/a

Additional Input Files

None

Input File Comments

The csfasta input file should be the output of `solid_denovo_preprocessor.pl`, and must correspond directly to the double encoded input file that was used as input to Velvet. The read-tracking file produced by Velvet has the '.afg' extension.

Output File

The output file contains the reads and their coordinates within a contig. Contigs are denoted using '=NODE_N' to indicate the start of a contig and '-end' to indicate the end of the contig. Reads are denoted in Fasta format with the read ID followed by a comma and number indicating the position of the read in the contig. For example,

```
=NODE_1
>6_8_474_R3,669
G2331011230330332332123322
...
>19_967_247_R3,611
G0321233311013221032100033
-end
=NODE_2
>6_813_306_F3,347
T1123330313031130313231031
...
>6_1264_270_R3,676
G1303303323321233221133313
-end
```

See the documentation of the `denovoadp` for further information.

Output File Comments

The output file can be submitted to the `denovoadp` to obtain base space contigs.

Sample Input Files

- Velvet_asm.afg
- colorspace_input.csfasta

Sample Output File

read_tracking_file

Sample output was generated using the following options:

```
$ solid_denovo_postprocessor --afgfile Velvet_asm.afg
--csfasta colorspace_input.csfasta --output read_tracking_file
--run mates
```

Supports AB kit or protocol or sample prep method

n/a

Other

Development language

Perl

Compiled for:

n/a

PBS is Required?

No.

Comments

This script is a part of the Velvet de novo assembly pipeline for SOLiD™ reads.

Date

10/06/08