Atlas-Link Manual

1 Introduction:

The Atlas-Link software implements a greedy algorithm and uses graph theory to link and orient assembled existing contigs quickly and accurately using mate pair information.

2 Version:

This is 0.01 release of the software. Development is still ongoing.

3 Author:

Jixin Deng
jdeng@bcm.edu
10/18/2010

4 Prerequisites:

A. System:
Unix & Linux (tested on Redhat)
B. Perl 5.8.0
C. Perl Module:
Graph
(http://search.cpan.org/ jhi/Graph-0.94/lib/Graph.pod)
ClusterPoints
(http://search.cpan.org/ salva/Algorithm-ClusterPoints-0.08/lib/Algorithm/ClusterPoints.pm)
XML::DOM
(http://search.cpan.org/ enno/libxml-enno-1.02/lib/XML/DOM.pm)
Statistics/Descriptive/Weighted
(http://search.cpan.org/ dhard/Statistics-Descriptive-Weighted-0.5/lib/Statistics/Descriptive/Weighted.pm)

5 Install:

In Atlas-link download folder, do
perl Makefile.PL PREFIX=(current working folder)
make
make install

6 Run Atlas-link:

It requires at least three input files and a optional AGP file:
1. example.lib: file for library information
2. example.contig file for mate pair information, format is TIGR Assembler's adaptation of the GDE alignment format (usl: http://www.cbcb.umd.edu/research/contig_representation.shtml).
3. example.configure file for configure information in xml format
4 A optional AGP file can be given for upgrading mode.

7 Command line:

Your_working_dir>perl do-A-Link.pl -l example.lib -t example.contig -f example.configure (<option>-a example.agp)

8 Important parameters in configure file:

level: the step count. range from 1 to n
type: 1 (de novo scaffolding or gap-filling for upgrading mode) 2 (superscaffolding for upgrading mode)
lib_size_limit: define the lower limit (min) and upper limit (max) of libraries to be used in each step
min_links: minimum mate pairs to initiate a link between two contigs (default 2)
deviate_factor: the starndard deviation of library size differ allowed in each step of linking(default 3)
treat_minus_gap_size_as: In agp file, if a minus gap was estimated, put this number in output AGP (default 50).

9 A typical configure file

The following example configure file present a typical way to set up Atlas-link run. It consist of two steps, the first step is de-novo scaffolding mode and second step is to update the scaffold from first step using same set of mate-pair data but with superscaffolding mode:

```
<?xml version="1.0"?>
<procedure>
<step level="1" type="1">      set up the first step
<mate_pair>
<lib_size_limit min="0" max="6000"></lib_size_limit>
</mate_pair>
<min_links>2</min_links>
<deviate_factor>5</deviate_factor>
</step>

<step level="2" type="2">      set up the second step
<mate_pair>
<lib_size_limit min="0" max="6000"></lib_size_limit>
</mate_pair>
<min_links>2</min_links>
<deviate_factor>5</deviate_factor>
</step>
</procedure>
```

10 License: