

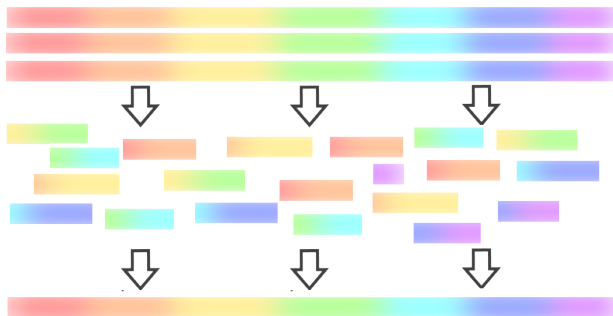
Scaffold graph visualization and RNA-Seq scaffolding

Chernikova Olga
Supervisor: Andrey Prjibelski

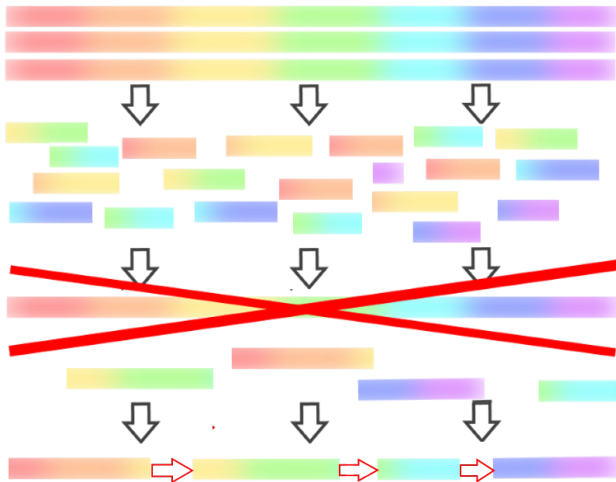
Center for Algorithmic Biotechnology

07/27/2017

Genome assembly



Genome assembly

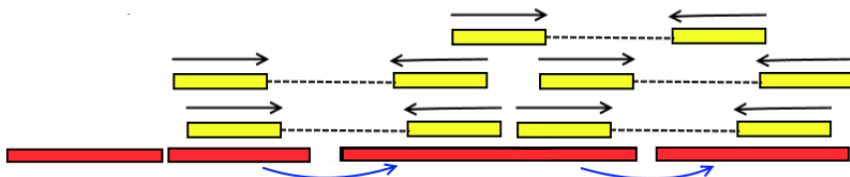


Connection by pair DNA reads

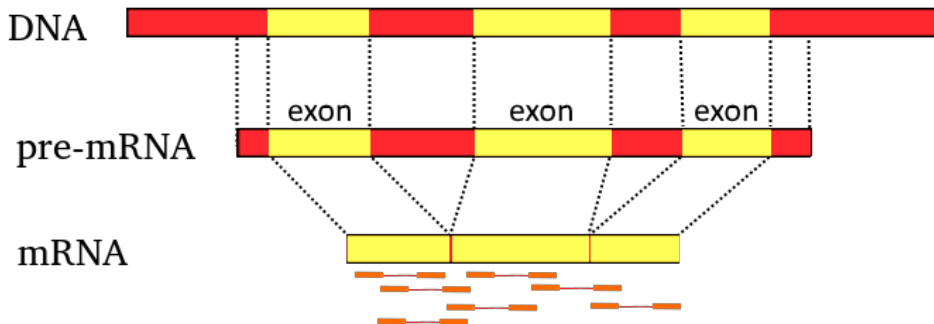
■ Pair reads:



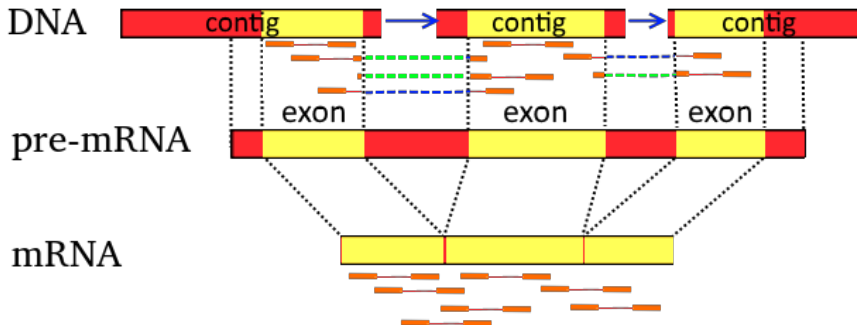
■ Finding connections by pair reads:



By RNA-seq reads



By RNA-seq reads



Goal and tasks

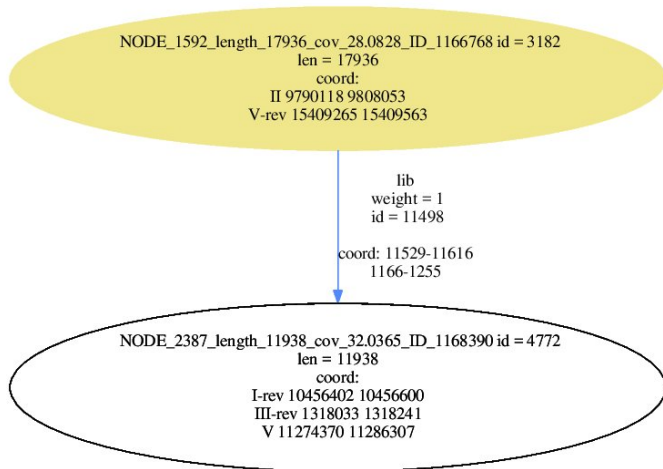
Goal

Building scaffolds by RNA-seq reads

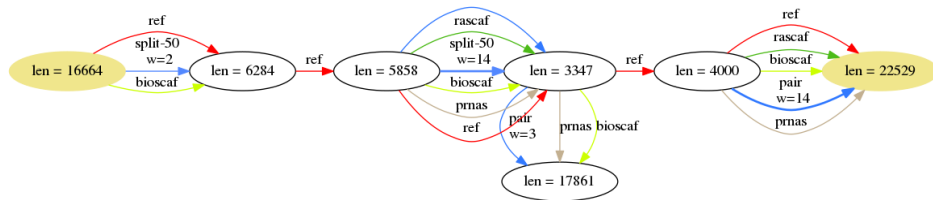
Tasks

- Building the contig graph
- Building scaffolds by obtained connections
- Creating tool for visualizing a contig graph
- Comparing results with other tools for building scaffolds by RNA-seq

Visualization



Visualization



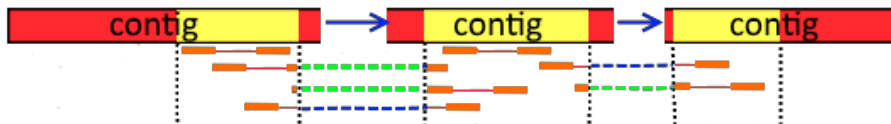
Visualization

Options for filtering the graph:

- by edge weight and contig size
- showing only parts with difference in two libs
- showing only parts with one lib but without the second one
- showing only parts with errors
- etc.

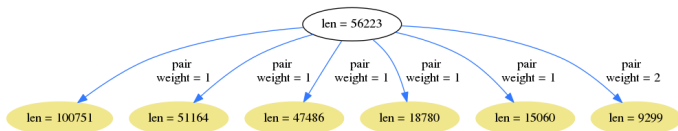
Building a contig graph

- Alignment of pair RNA-seq reads
- Building a contig graph on the basis of this alignment
- Splitting reads into two parts
- Alignment of these parts
- Building graph by these parts of the reads
- Saving graph



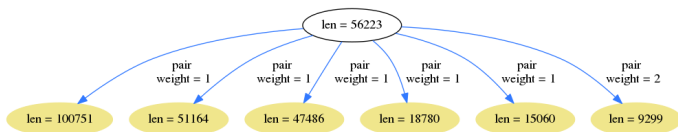
Simplification of the graph

■ Deleting small weight edges



Simplification of the graph

■ Deleting small weight edges

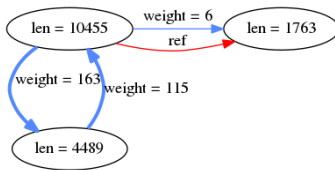


■ Edge projection



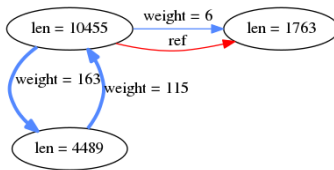
Simplification of the graph

■ Deleting cycles

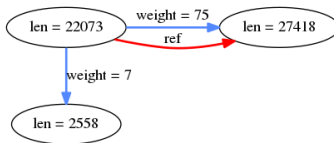


Simplification of the graph

■ Deleting cycles

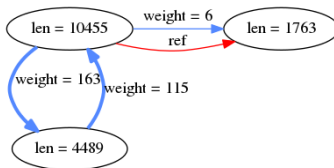


■ Fork with big difference in weight

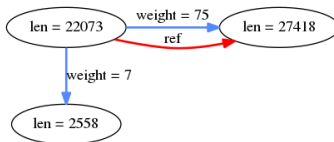


Simplification of the graph

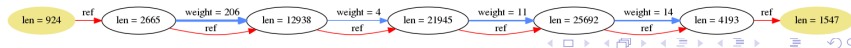
■ Deleting cycles



■ Fork with big difference in weight



■ Connecting simple paths into scaffolds



Comparison

C.elegans, SRR1560107

	bio_scaffolder	P_RNA_scaffolder	rascaf
NG50	36855	36075	32879
NG75	17299	17188	18395
NGA50	30383	28828	27116
NGA75	12735	12489	11667
LGA50	918	955	995
misassemblies	529	621	521

Results and plans

Results

- Creating the program for RNA-seq scaffolding
- Creating the tool for visualization a contig graph

Plans

- Testing and comparing with other tools given bigger diversity of data
- Making program faster
- Implementing new idea for making scaffolds
- Writing manual and interface
- Writing an article

Thank you!

https://github.com/olga24912/bio_scaffolder