Scaffold graph visualization and RNA-Seq scaffolding
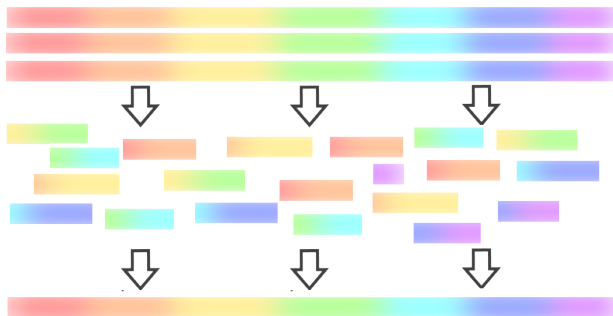
Chernikova Olga
Supervisor: Andrey Prjibelski
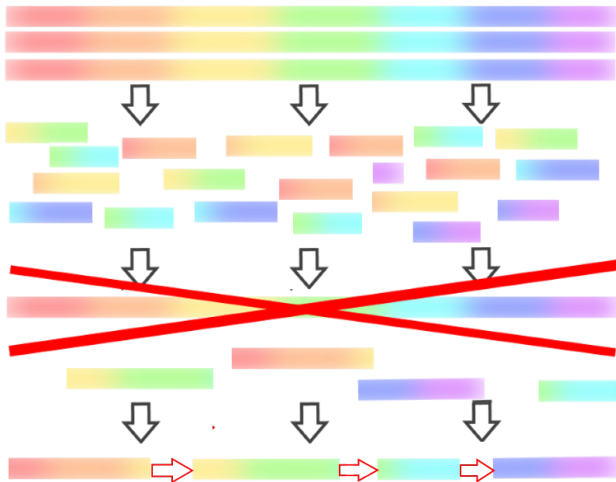
Center for Algorithmic Biotechnology

07/27/2017
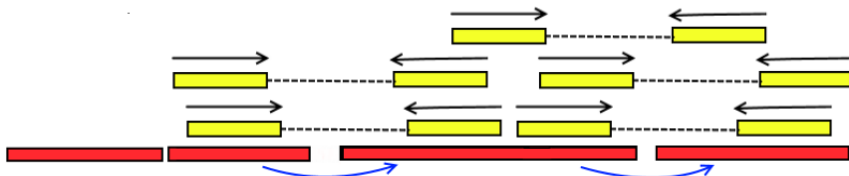
Genome assembly

Genome assembly
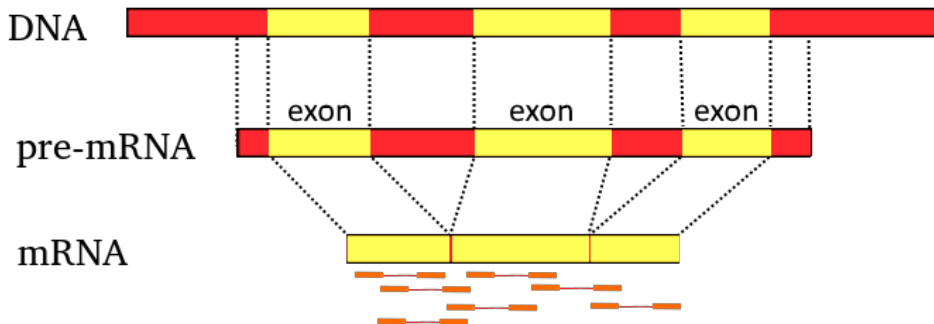
Connection by pair DNA reads
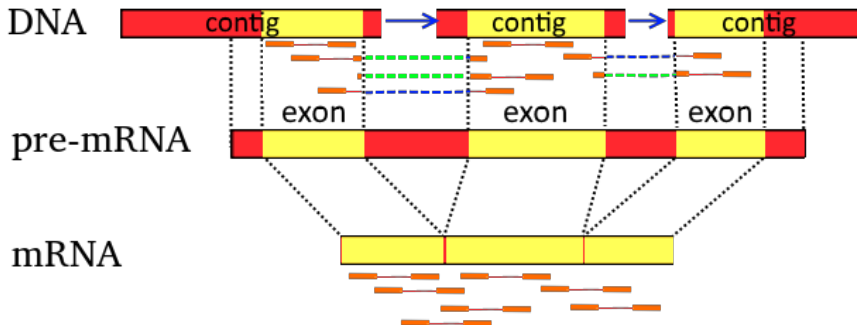
- Pair reads:

- Finding connection by pair reads:

By RNA-seq reads

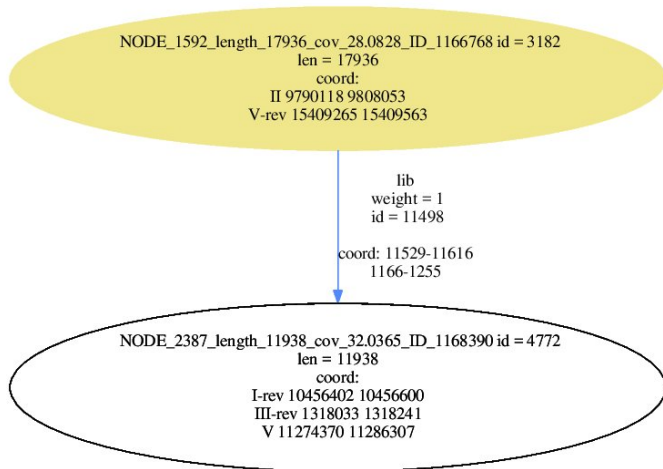# By RNA-seq reads

Goal and tasks

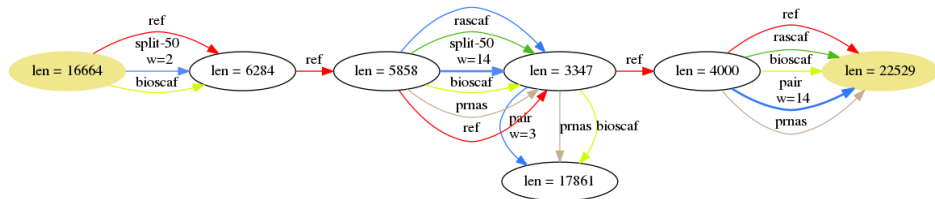## Goal

Build scaffolds by RNA-seq reads

Tasks

- Build the contig graph
- Build scaffolds by finding connection.
- Create tool for visualizing contig graph
- Compare results with other tools for building scaffolds by RNA-seq.

Visualization

Visualization

Visualization

Options for filtering the graph:

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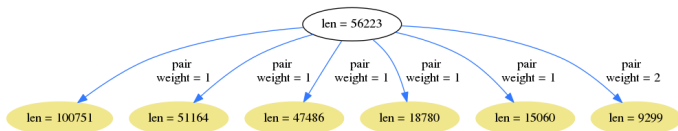- show only parts with errors
- etc.

Build a contig graph

- Alignment pair RNA-seq read
- Build a contig graph by this alignment
- Split reads on two parts
- Alignment this parts
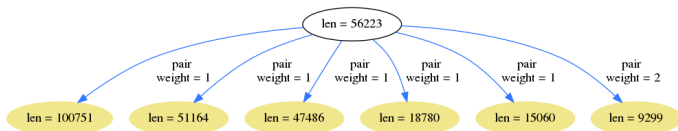- Build graph by this parts of the reads
- Save graph
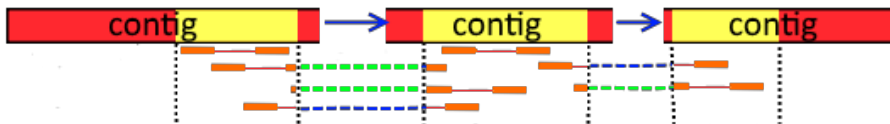
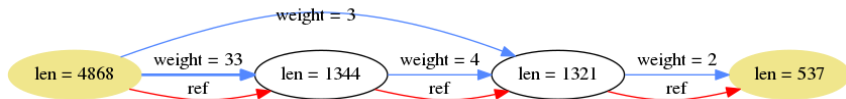# Simplification of the graph

- Delete small weight edge

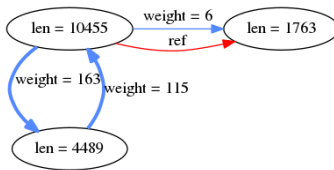# Simplification of the graph

■ Delete small weight edge
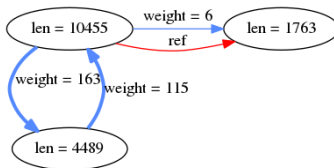

■ Edge projection

Simplification of the graph
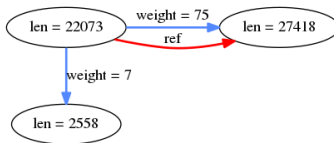
- Delete small cycle

Simplification of the graph
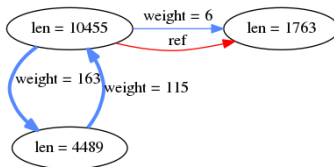
- Delete small cycle

- Big difference in weight

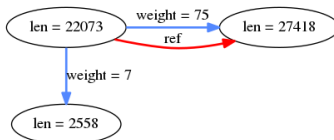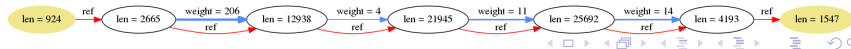# Simplification of the graph

- Delete small cycle

- Big difference in weight

- Build 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

- Create program for RNA-seq scaffolding
- Create tool for visualization a contig graph

Plans

- Testing and comparing with other tools on bigger diversity of data.
- Make program faster
- Use new idea for making scaffolds
- Write manual and interface
- Write an article

-

Thank you!

https://github.com/olga24912/bio_scaffolder