

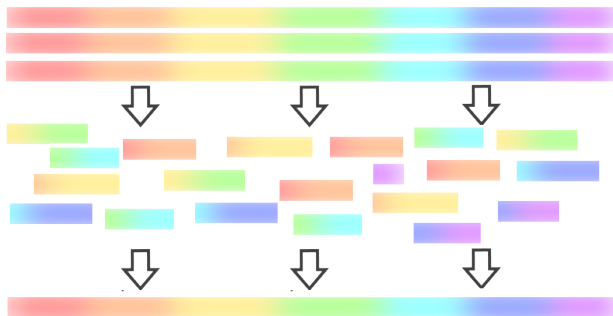
Scaffold graph visualization and RNA-Seq scaffolding

Kunyavskaya Olga
Supervisor: Andrey Prjibelski

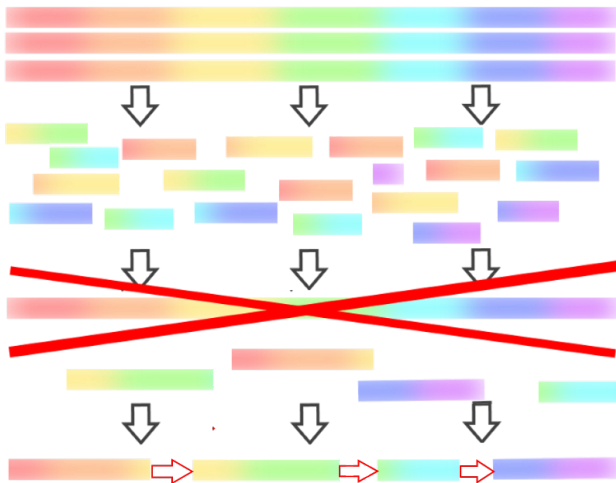
Center for Algorithmic Biotechnology

09/08/2017

Genome assembly



Genome assembly

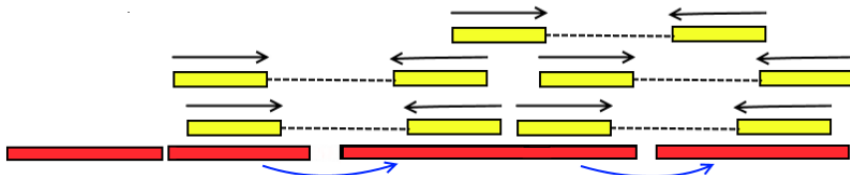


Connection using DNA paired-end reads

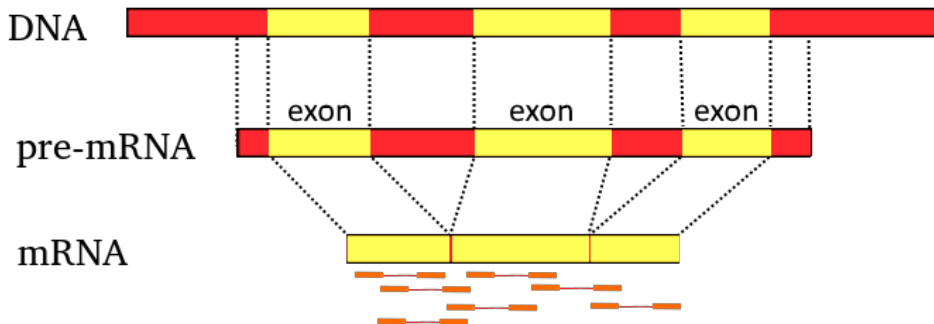
■ Paired reads:



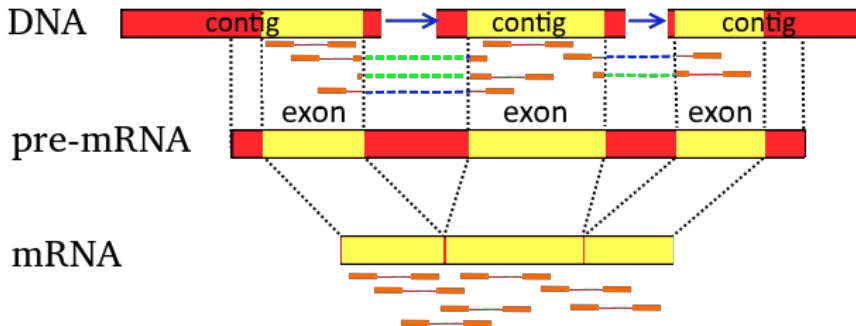
■ Finding connections using paired reads:



Using RNA-seq reads



Using RNA-seq reads



Goal and tasks

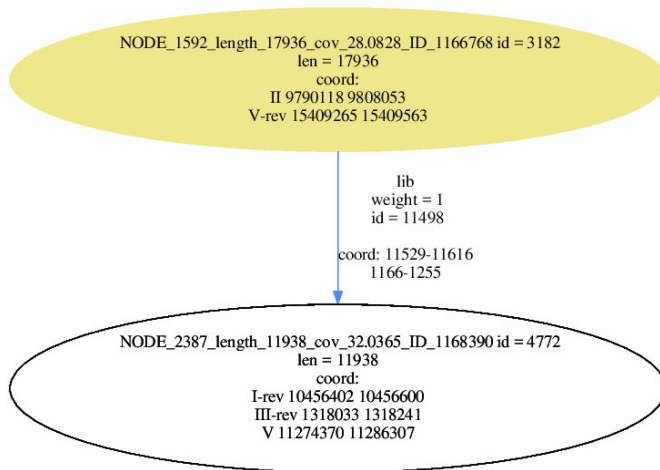
Goal

Build scaffolds using RNA-seq reads

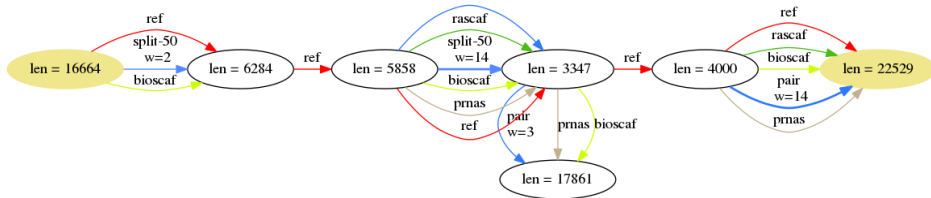
Tasks

- Build the scaffold graph
- Build scaffolds using obtained connections
- Create tool for visualizing a scaffold graph
- Compare results with other tools

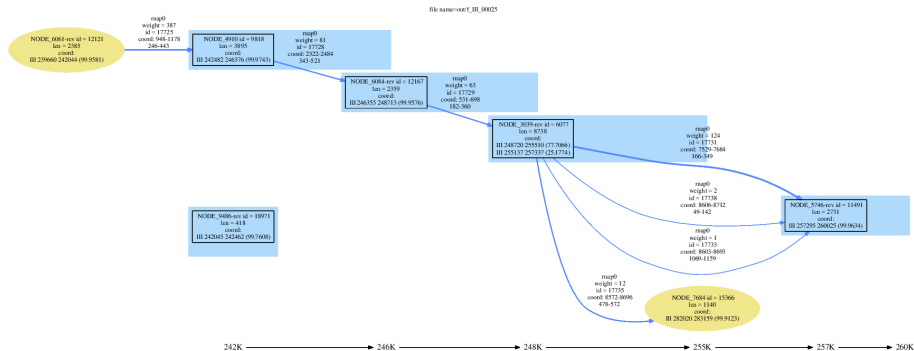
Visualization



Visualization



Visualization



Visualization

Drawing interesting parts:

- difference between two libs
- one lib is present and another is not
- difference with reference (possible missassembly)
- etc

Existing RNA-seq scaffolders

- **L_RNA_scaffolder** (2013) uses long transcriptome reads.
- **AGOUTI** (2015) uses paired-end RNA-seq reads and gene predictions.
- **rascaf** (2016) uses paired-end RNA-seq reads.
- **P_RNA_scaffolder** uses paired-end RNA-seq reads, unpublished, details unknown.



Dhruv S. Shankar Li Song and Liliana Florea. "Rascaf: Improving Genome Assembly with RNA Sequencing Data." In: *Plant Genome*. doi: 10.3835/plantgenome2016.03.0027 (2016).

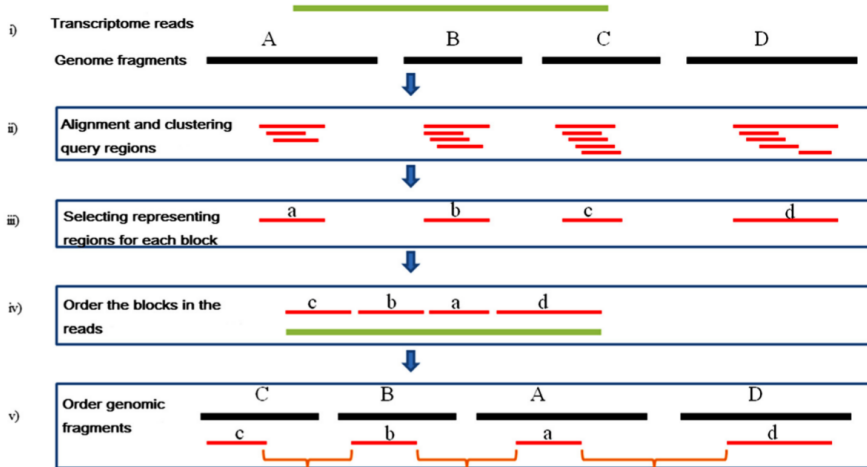


J.T. Li Y.P. Zhu G.Y. Hou X.F. Kong Y.Y. Kuang Xue W. and X.W. Sun. "L_RNA_scaffolder: Scaffolding genomes with transcripts". In: *BMC Genomics* 14:604. doi:10.1186/1471-2164-14-604 (2013).

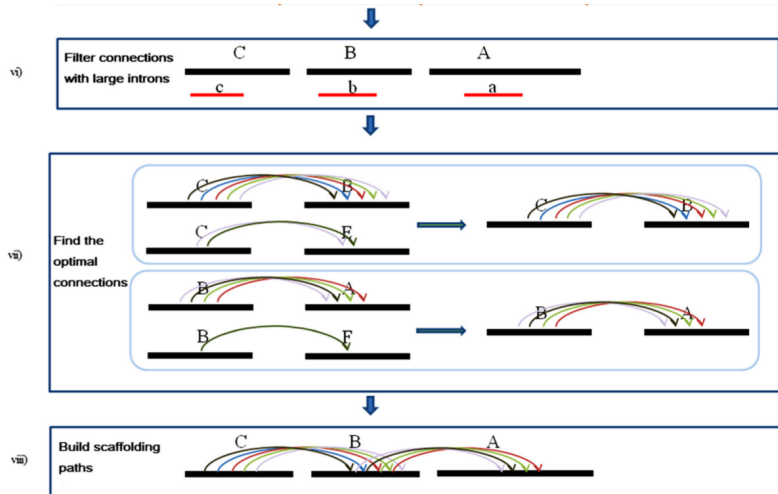


L. Zhuo Zhang S.V. and M.W. Hahn. "AGOUTI: Improving genome assembly and annotation using transcriptome data." In: *Bioinformatics* doi:10.1101/033019 (in press). (2015).

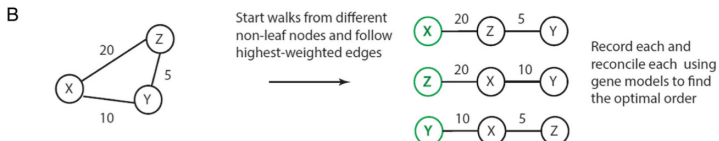
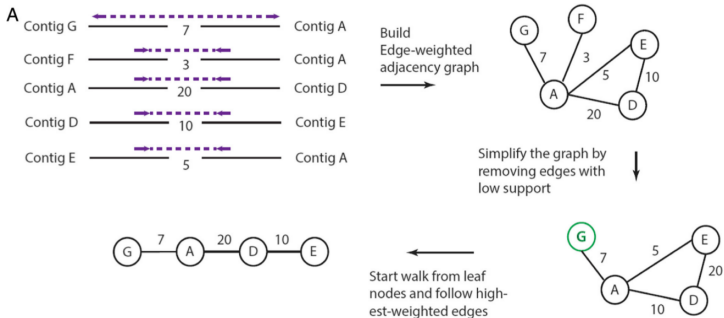
L_RNA_scaffolder



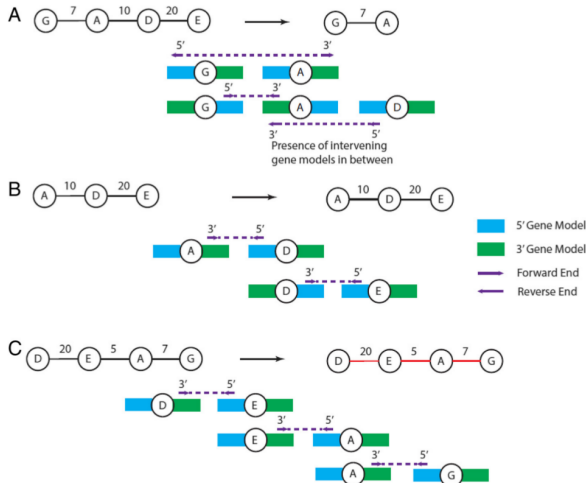
L_RNA_scaffolder



AGOUTI

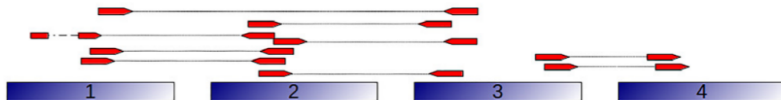


AGOUTI



Rascaf

1. Contigs and alignments of reads



2. Build exon blocks by coverage



3. Build gene blocks from exon blocks and splice junctions



4. Build a gene block graph



Rascaf

5. Find contig connections



6. Build a contig graph



7. Join the contigs

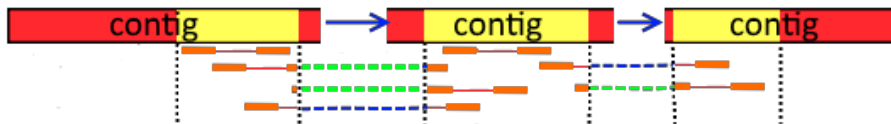


Pipeline

- Align reads
- Build scaffold graph
- Filter graph
- Output scaffolds

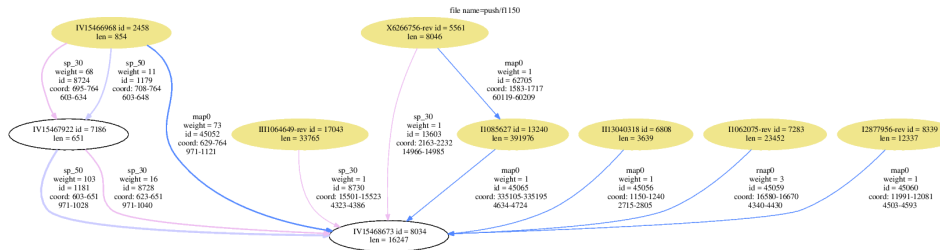
Building a scaffold graph

- Align RNA-seq paired-end reads
- Build a scaffold graph using these alignments
- Split every read into two parts in the middle
- Align these parts independently
- Build graph using these parts of the reads
- Save graph

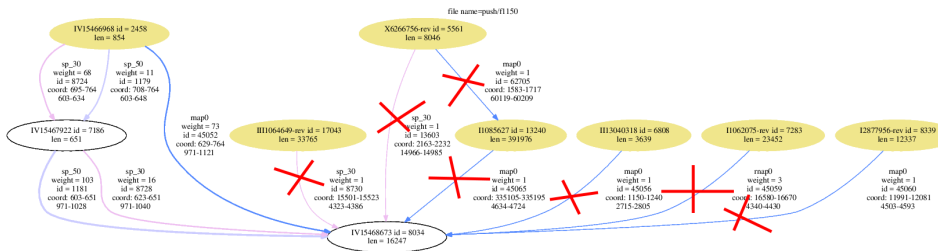


Graph simplification

Delete low-weight edges

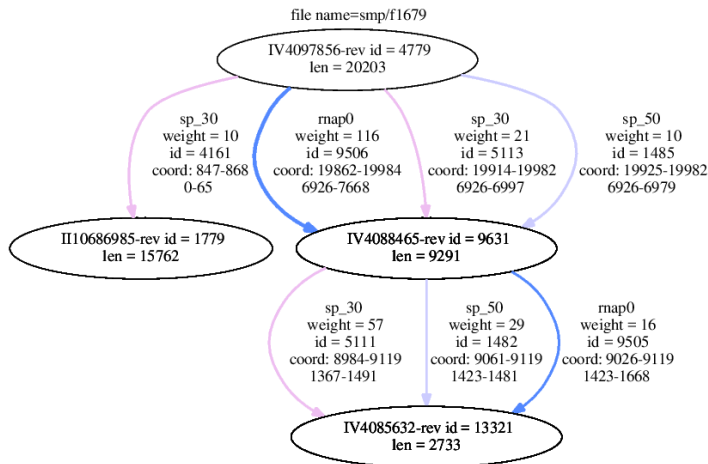


Delete low-weight edges



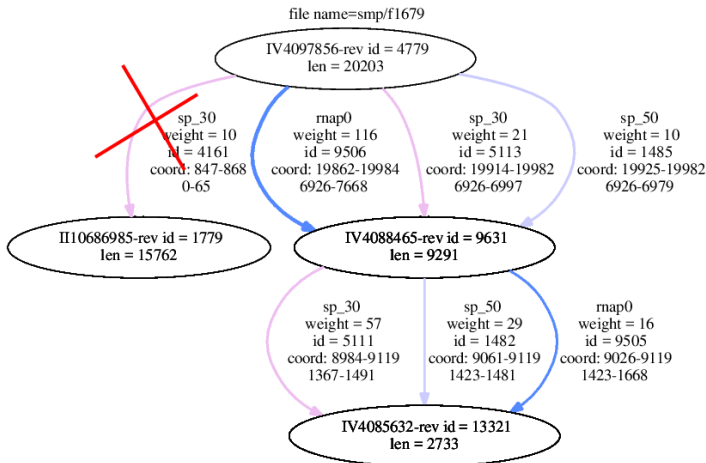
Graph simplification

Fork with big difference in weight



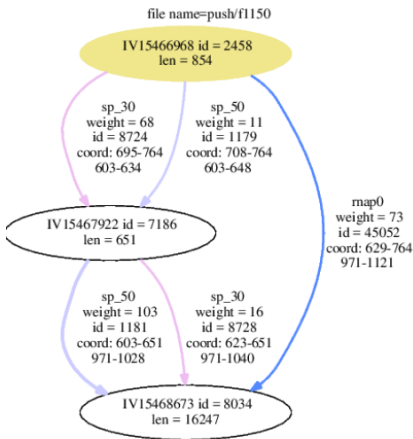
Graph simplification

Fork with big difference in weight



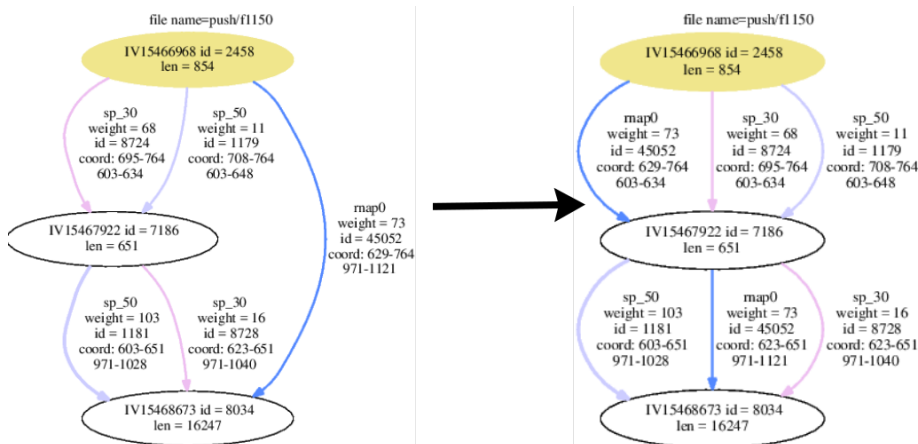
Graph simplification

Edge projection



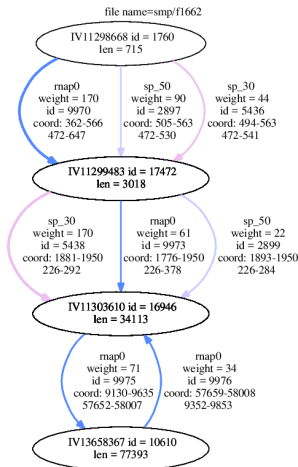
Graph simplification

Edge projection



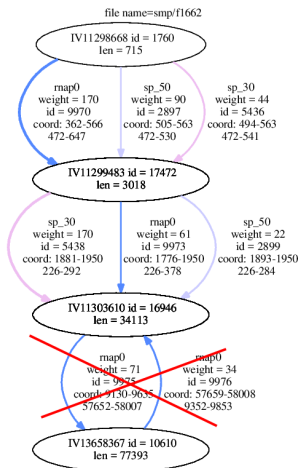
Graph simplification

Delete cycles



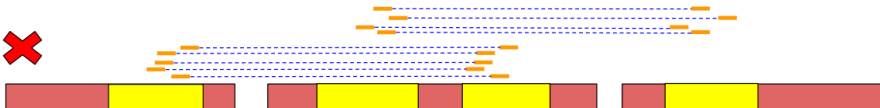
Graph simplification

Delete cycles



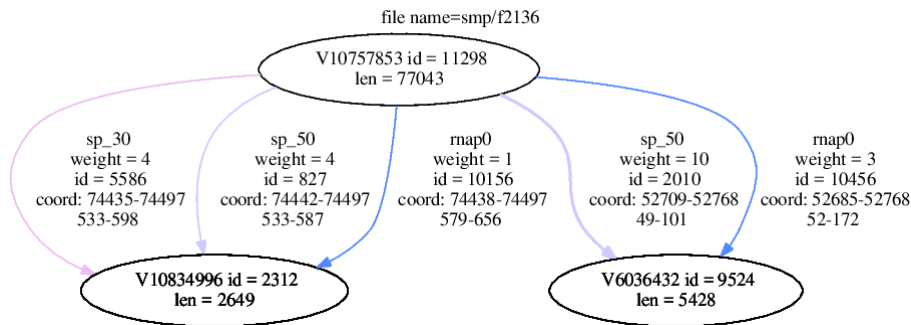
Graph simplification

Check coordinates order



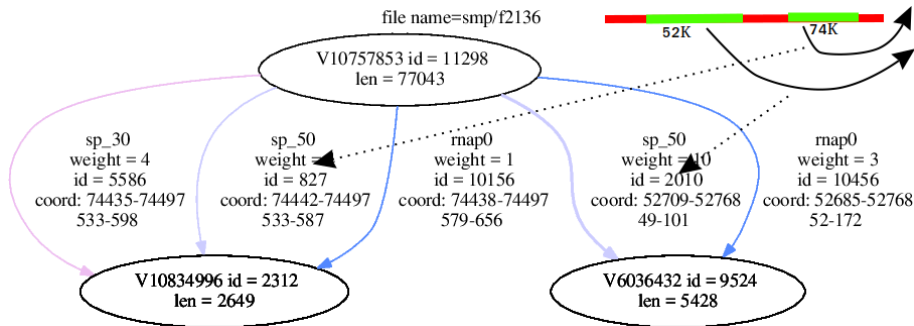
Graph simplification

Check coordinates order



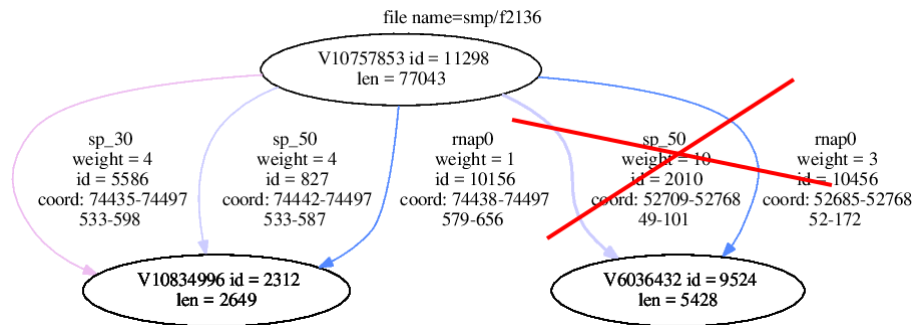
Graph simplification

Check coordinates order



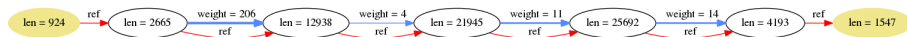
Graph simplification

Check coordinates order



Graph simplification

Connect simple paths into scaffolds



Plans

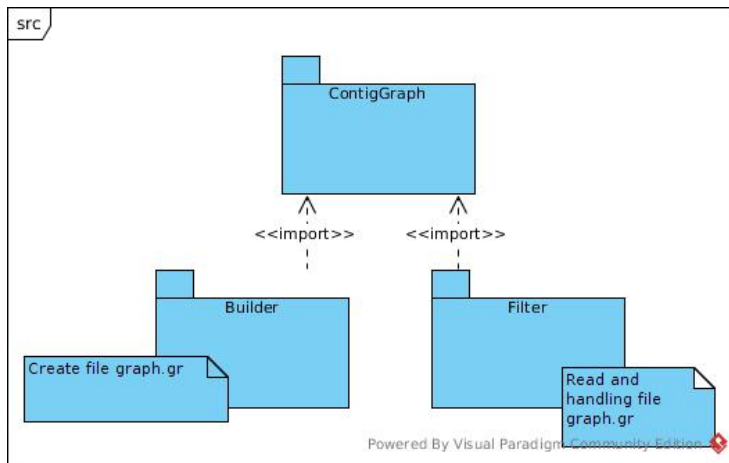
- Coverage depth
- Detect exon blocks
- Additional validation of split links

Comparison

C.elegans

		SRR1560107	SRR4181037	SRR5067781
bio_scaf	NGA50	31168	30974	27650
	mis	550	554	509
P_RNA_scaffolder	NGA50	28828	33645	29786
	mis	621	750	571
rascaf	NGA50	27116	30403	27266
	mis	521	561	507
AGOUTI	NGA50	25384	27047	25275
	mis	545	613	510
contigs	NGA50	21938	21938	21938
	mis	352	352	352

Architecture



Results and plans

Results

- Designed an algorithm for RNA-seq scaffolding
- Developed a tool for visualization a scaffold graph

Plans

- Compare and test against other tools on more datasets
- Improve performance
- Implement new ideas for scaffolding
- Write manual and interface
- Write a paper

Thank you!

https://github.com/olga24912/bio_scaffolder