Automating riboSeed assembly assessment

Lab Meeting

Nick Waters

June 12, 2017

Department of Microbiology
School of Natural Sciences
National University of Ireland Galway

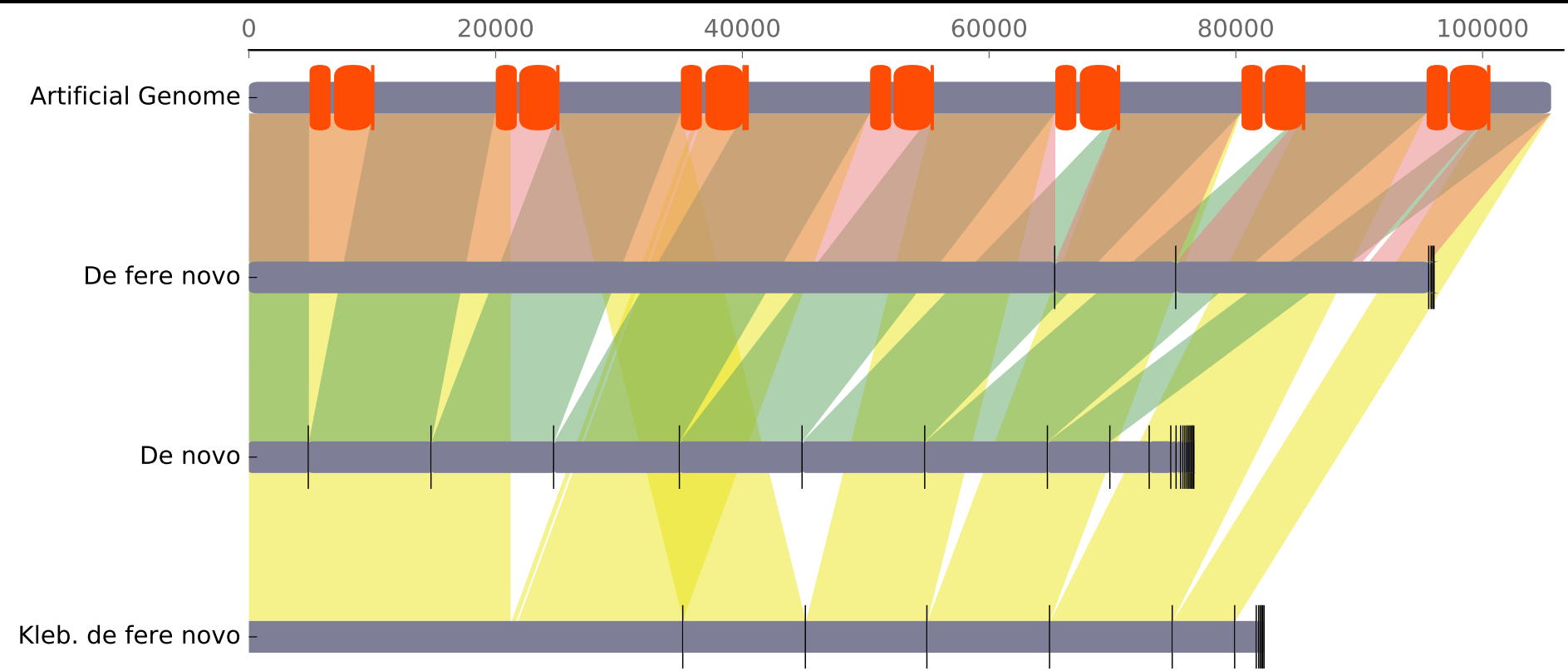
Problem

Problem

Table: Runs of riboSeed by manuscript section

	runs	repetitions
background	8	8
simulated genome	6	1
simulated reads	12	4
GAGE-B	16	
other	10	

Problem: Sample Result Visualization



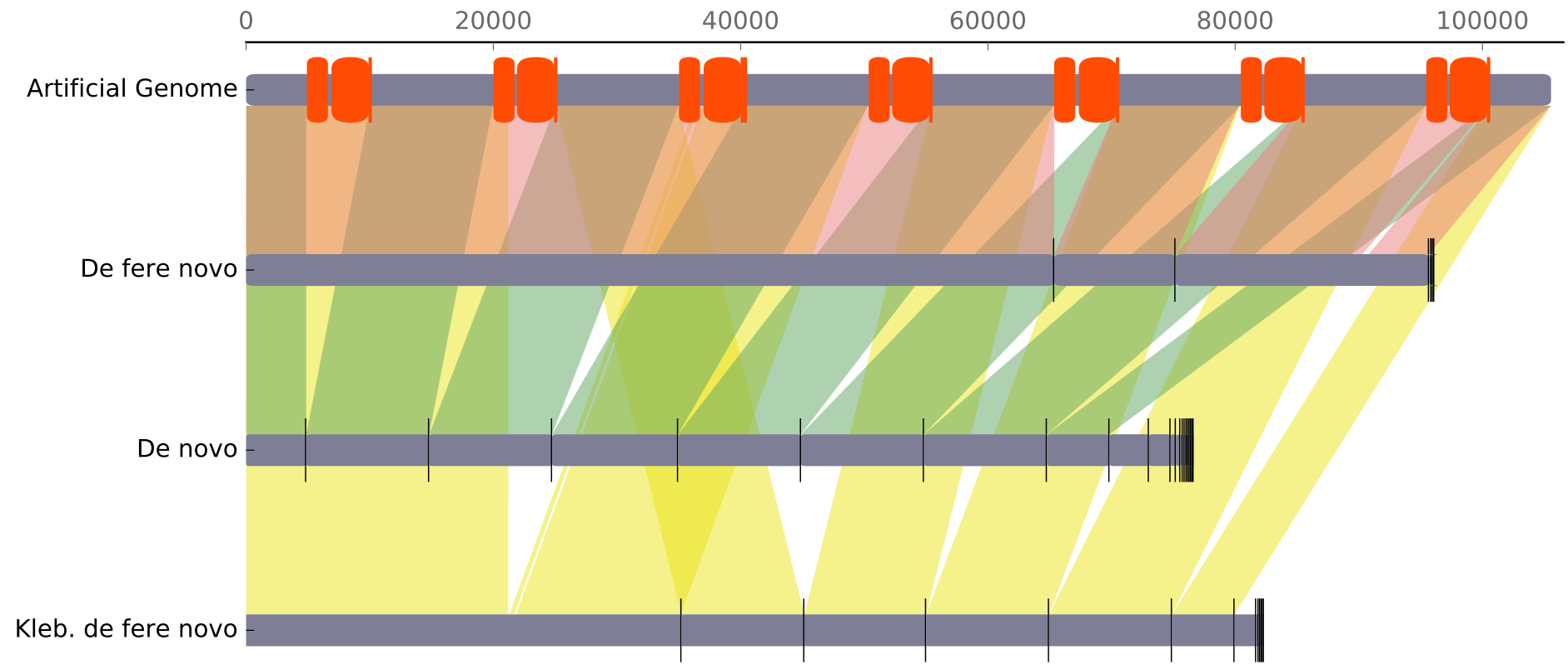
Solution 1: Full length Matches

Solution 1

- Annotate and extract rDNAs and flanking regions from reference and assemblies
- Use BLAST or Mummer to find best reciprocal matches

Solution 1: Results

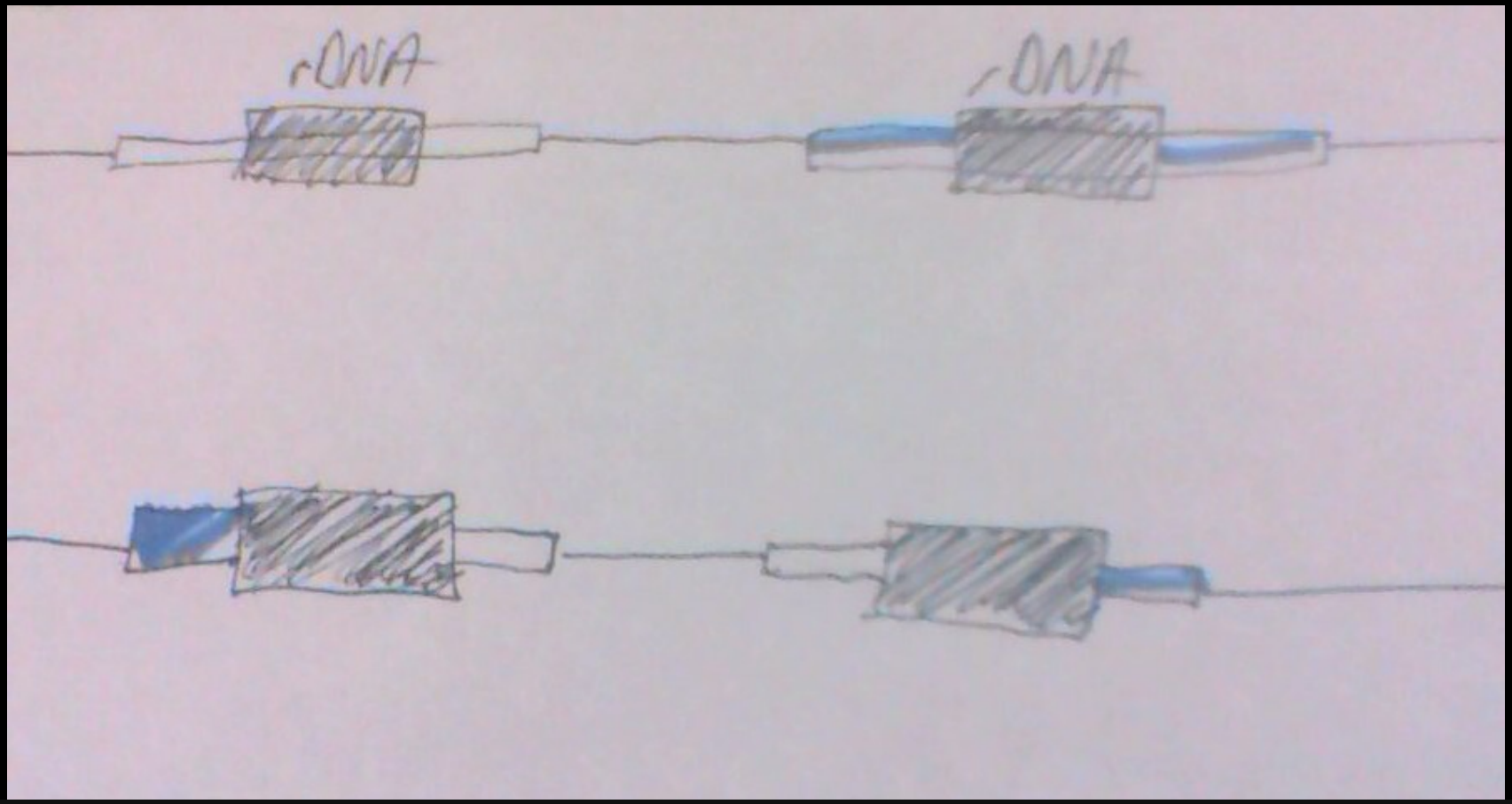
Accurately characterized true positives



Failed to detect false positives

Solution 2

- Annotate rDNAs and extract just flanking regions from reference and assemblies
- Use BLAST to find best reciprocal matches
- Parse fragment pairs



Solution 2: Results

Accurately characterized true positives Accurately detected false positives

Work for the Week

- Apply riboScore to all the results
- Determine effectiveness on real data
- Make a pipeline-specific version for automated error correction