

# RAGOUT ASSEMBLY QUALITY ANALYSIS

---

Student: Geller Mark

Advisor: Mikhail Kolmogorov

December 12, 2014

Bioinformatics Institute

Ragout (Reference-Assisted Genome Ordering UTility) is a tool for assisted assembly using multiple references. To assess the quality of it's work we need to compare two genome assemblies

Given two synteny block permutations find their rearrangement distance

# WORK PROCESS

---

- Get Ragout Assembly
- Find markers
- ???
- profit Score

input: two sets of contigs represented as synteny blocks permutations

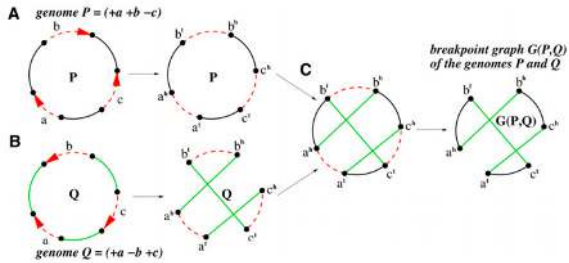
Desired result: Find minimum number of rearrangements between assemblies

first idea: create two-layered graph, minimize number of crossing edges.

# BREAKPOINT GRAPH

possible operations:

- Inversion
- Fission
- Fussion





second idea: block ordering problem<sup>1</sup> – algorithm to maximize number of cycles in breakpoint graph induced by those permutations, using fragmented breakpoint graph and edge matching graph

output result as percent of maximum possible number of cycles

Ordering Partially Assembled Genomes Using Gene Arrangements, Eric Gaul and Mathieu Blanchette, 2006

# RESULTS

---

**Table:** Ragout Score on Drosophila

Block size	Raw contigs	1 reference	2 references	3 references
1000	0.955	0.853	0.875	0.888
5000	0.925	0.685	0.685	0.765

This results are in agreement with expectations: Ragout assembly quality improves with increase of phylogenetic tree

Future possible improvements:

- Correctly process repeats
- Enhance edge-matching blocks scoring

QUESTIONS?