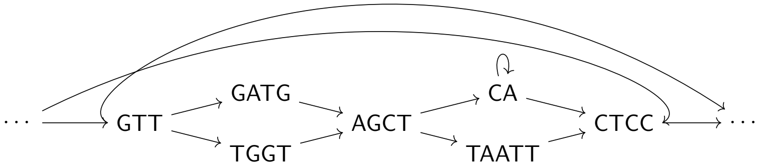
CARP - Quantifying Structural Complexity of Pangenomes

Leonard Bohnenkämper

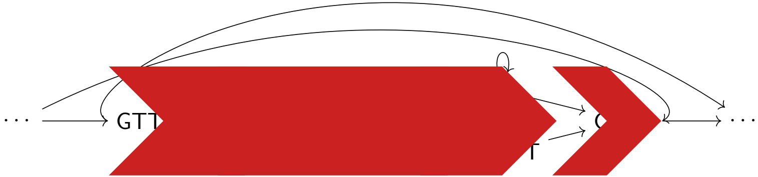
March 5th, 2025

Pangenomes, Rearrangements and Recombinations

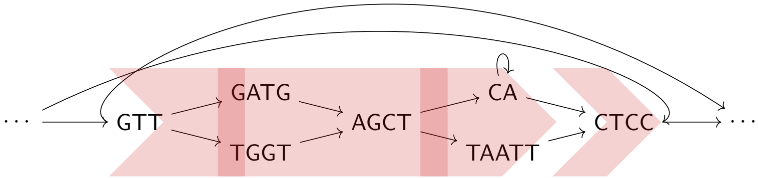
Abstracting from Local Variations



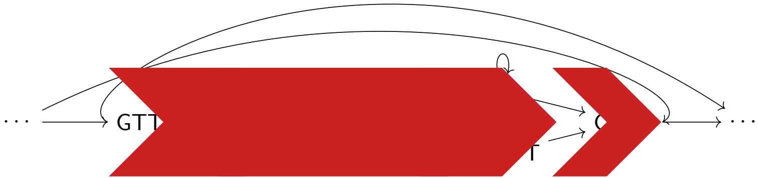
Abstracting from Local Variations



Abstracting from Local Variations



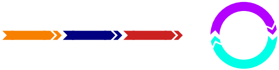
Abstracting from Local Variations



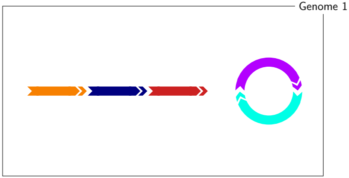
→ Collinear Block Detection.

Pangenome

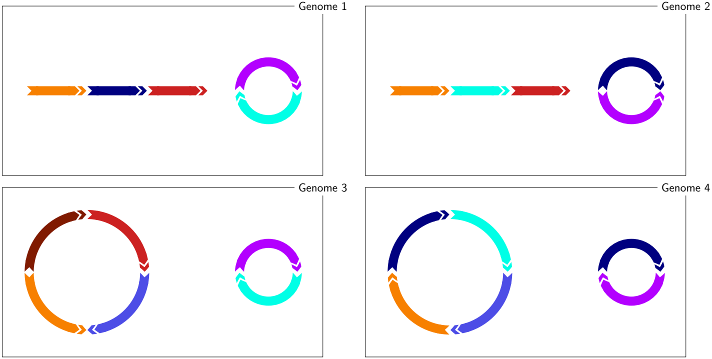
Pangenome



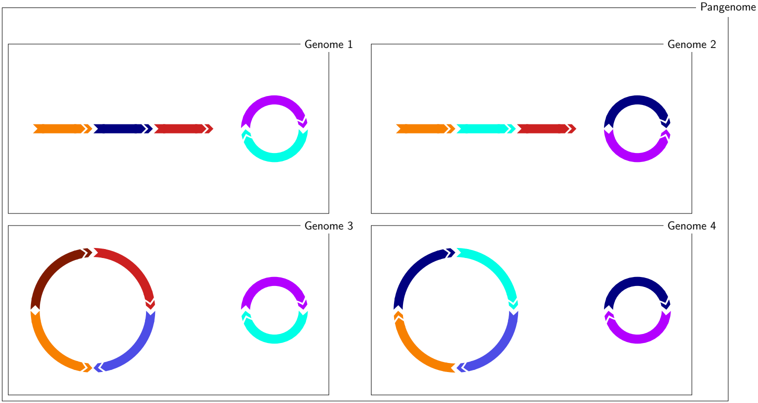
Pangenome



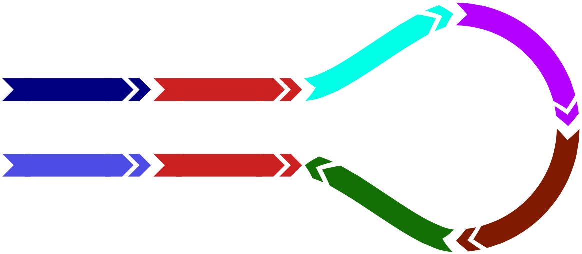
Pangenome



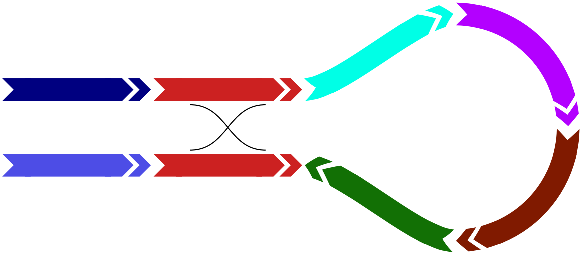
Pangenome



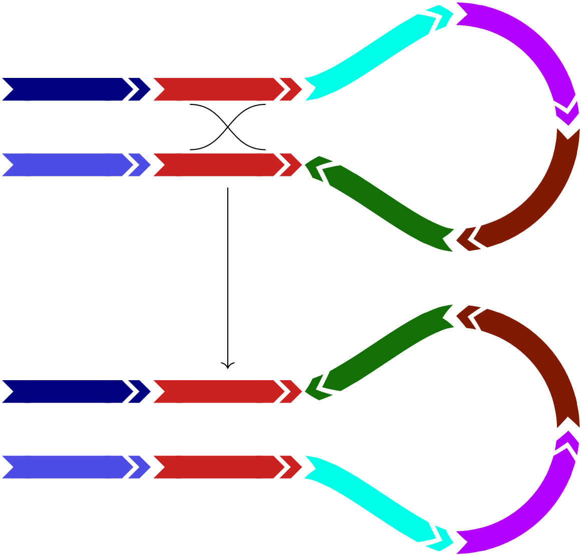
Rearrangements and Homologous Recombinations



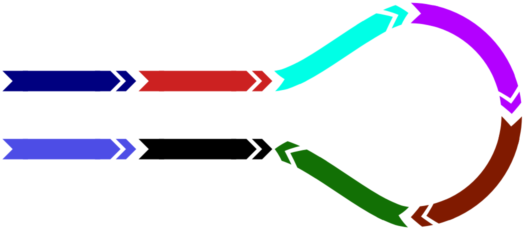
Rearrangements and Homologous Recombinations



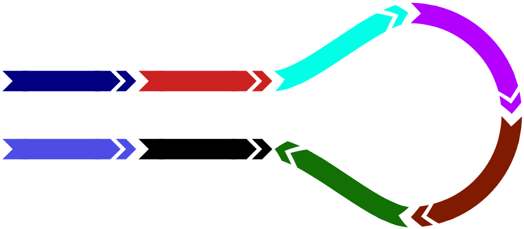
Rearrangements and Homologous Recombinations



Rearrangements and Homologous Recombinations

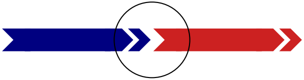


Rearrangements and Homologous Recombinations

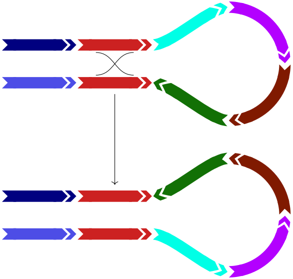


- No rearrangements via Homologous Recombination.

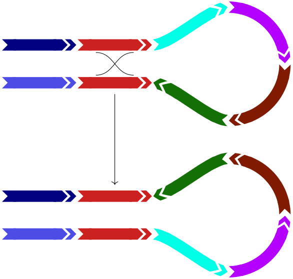
Flash Forward: Adjacencies



(Non)-Adjacency Modifying Operations

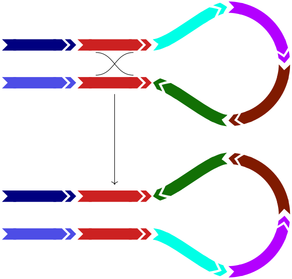


(Non)-Adjacency Modifying Operations



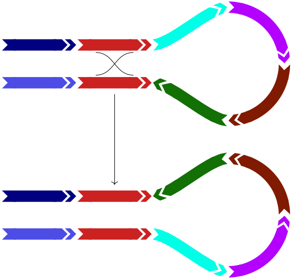
- No change in adjacencies!

(Non)-Adjacency Modifying Operations



- No change in adjacencies!
  → Non-Adjacency Modifying Operations (Namos)

(Non)-Adjacency Modifying Operations

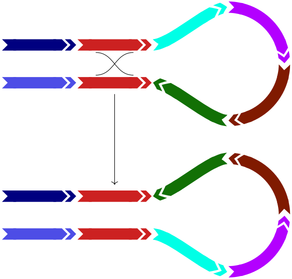


- No change in adjacencies!

  → Non-Adjacency Modifying Operations (Namos)

  ↔ Adjacency Modifying Operations (Amos)

(Non)-Adjacency Modifying Operations



- No change in adjacencies!

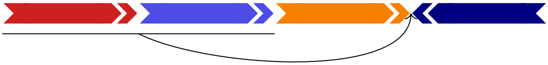
  → Non-Adjacency Modifying Operations (Namos)

  ↔ Adjacency Modifying Operations (Amos)

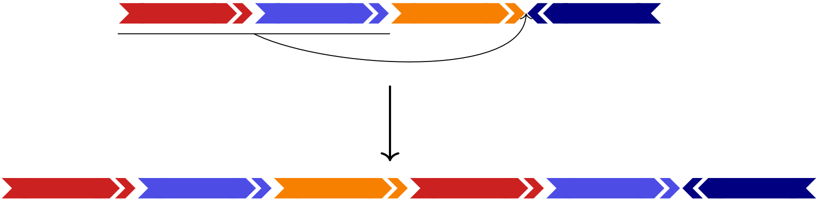
Amos can introduce structural complexity!

Amos - Example: Segmental Duplication

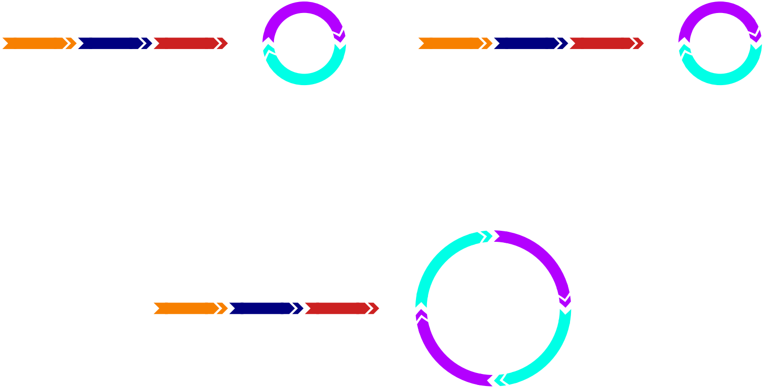
Amos - Example: Segmental Duplication



Amos - Example: Segmental Duplication



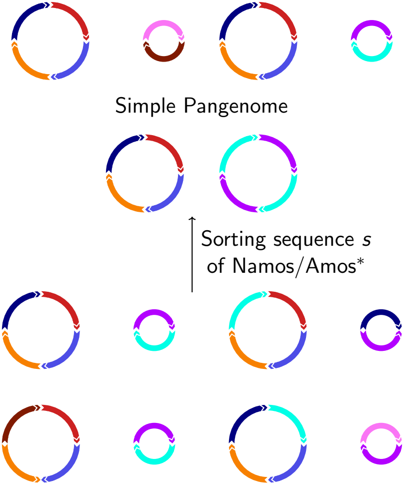
Pangenomes with Minimal Structural Complexity



~~Boring~~ **Simple Pangenome**

General Ancestral Reconstruction Problem (GARP)

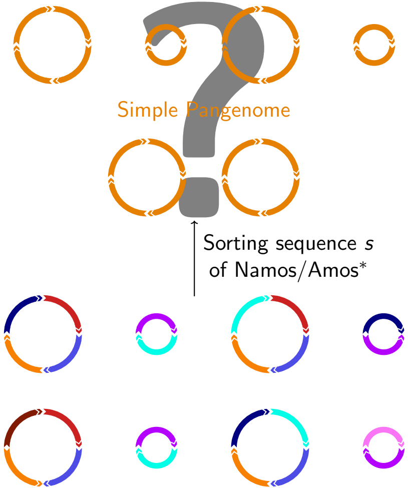
General Ancestral Reconstruction Problem (GARP)



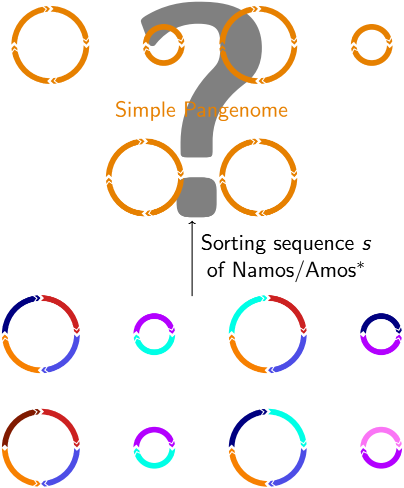
Simple Pangenome

Sorting sequence *s* of Namos/Amos*

General Ancestral Reconstruction Problem (GARP)

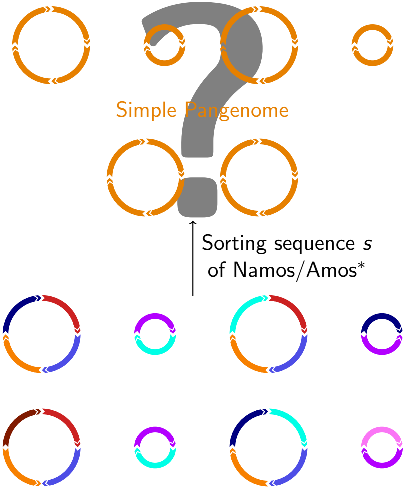


General Ancestral Reconstruction Problem (GARP)



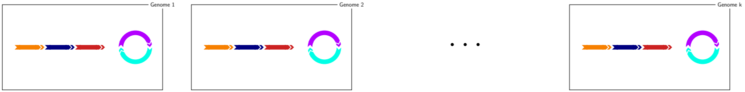
- #Amos in *s*: measure of the Structural Complexity.

General Ancestral Reconstruction Problem (GARP)

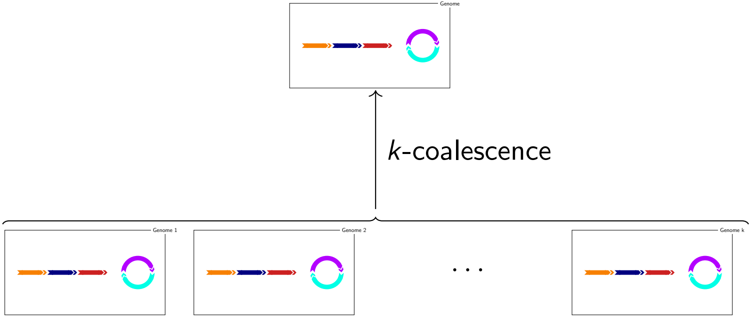


- #Amos in *s*: measure of the Structural Complexity.
- * with minimal #Amos

k-Coalescence – an important Namo



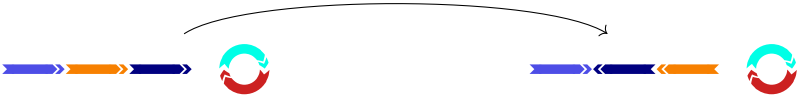
k-Coalescence – an important Namo



k-coalescence

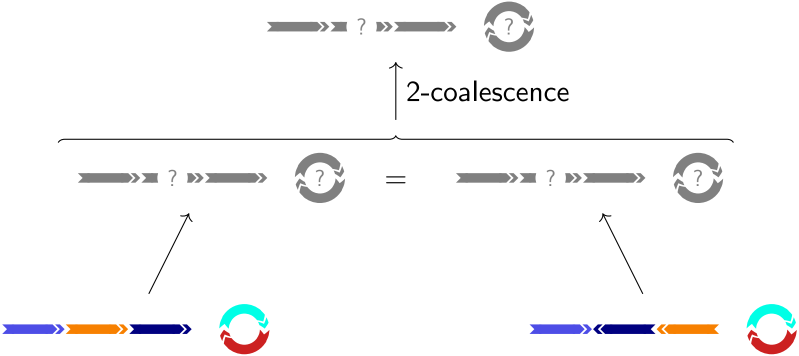
Classical Problems and GARP

Distance Problem and GARP

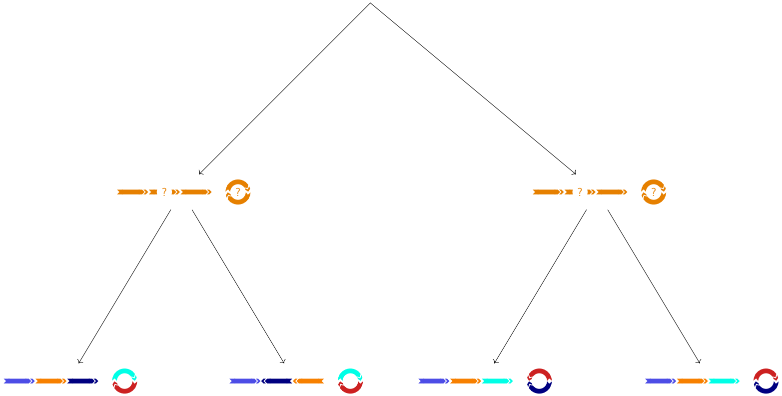


Distance Problem and GARP

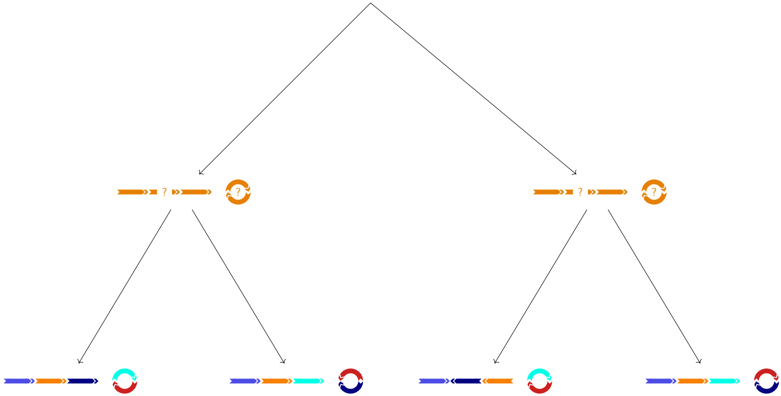
Distance Problem and GARP



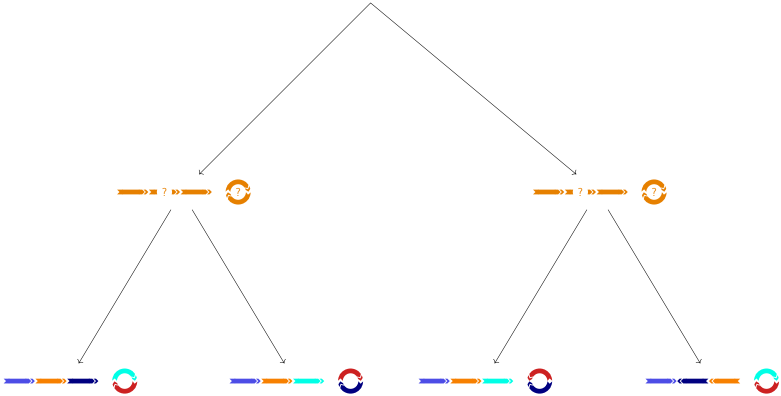
Large Parsimony Problem and GARP



Large Parsimony Problem and GARP



Large Parsimony Problem and GARP

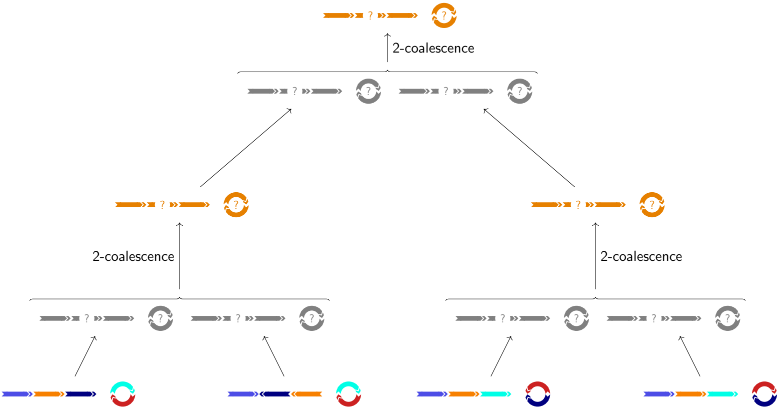


Large Parsimony Problem and GARP

2-coalescence

2-coalescence

Large Parsimony Problem and GARP



Complete Ancestral Reconstruction for Pangenomes (CARP)

GARP and Classical Problem Formulations

- Very limited interaction between genomes (only Coalescence)

GARP and Classical Problem Formulations

- Very limited interaction between genomes (only Coalescence)
- Pangenomes

GARP and Classical Problem Formulations

- Very limited interaction between genomes (only Coalescence)
- Pangenomes$\rightarrow$ phylogenetic closeness

GARP and Classical Problem Formulations

- Very limited interaction between genomes (only Coalescence)
- Pangenomes→ phylogenetic closeness→ more interaction/horizontal effects

GARP and Classical Problem Formulations

- Very limited interaction between genomes (only Coalescence)
- Pangenomes→ phylogenetic closeness→ more interaction/horizontal effects

→ Classical Problem Formulations would overestimate structural complexity for pangenomes (missing Namos substituted with Amos)

GARP and Classical Problem Formulations

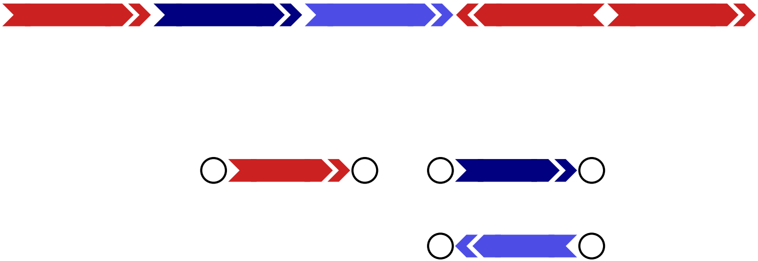
- Very limited interaction between genomes (only Coalescence)
- Pangenomes→ phylogenetic closeness→ more interaction/horizontal effects

→ Classical Problem Formulations would overestimate structural complexity for pangenomes (missing Namos substituted with Amos)

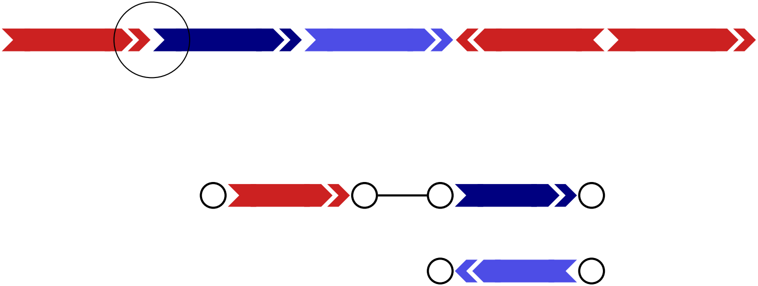
→ For a lower bound, we need a "maximally powerful" set of Namos

Multi-Breakpoint Graph (MBPG)

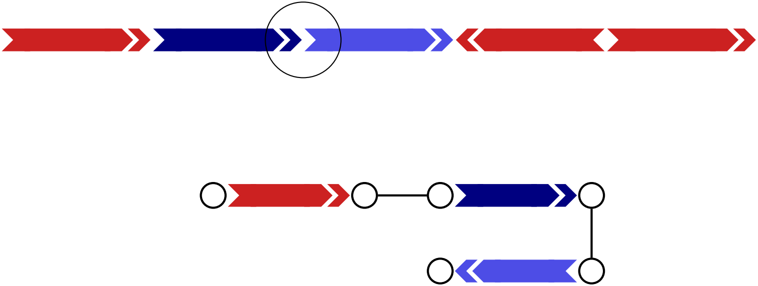
Multi-Breakpoint Graph (MBPG)



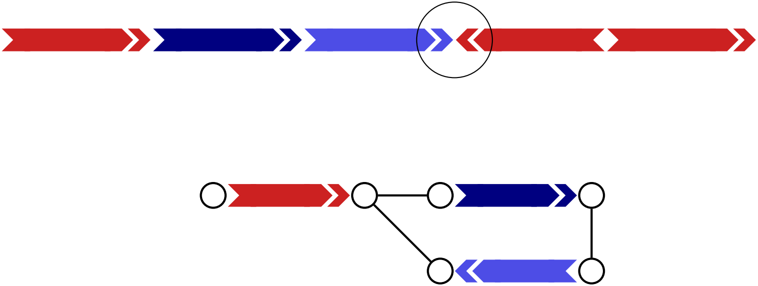
Multi-Breakpoint Graph (MBPG)



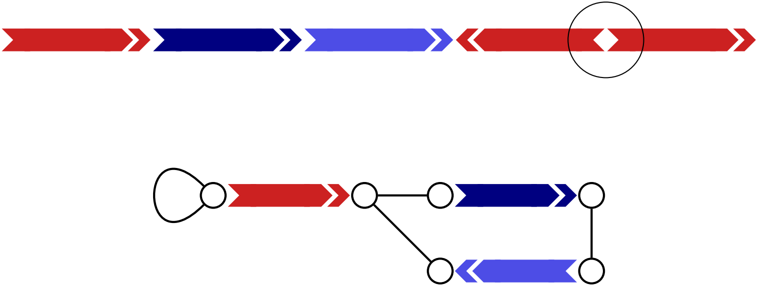
Multi-Breakpoint Graph (MBPG)



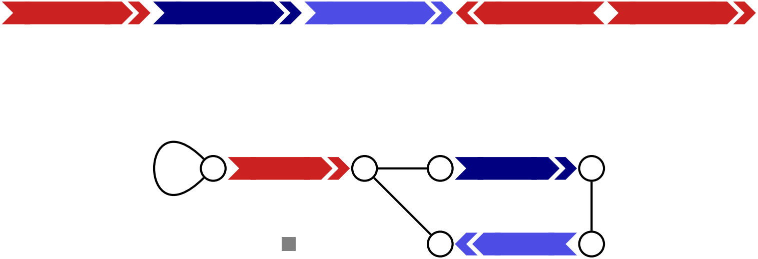
Multi-Breakpoint Graph (MBPG)



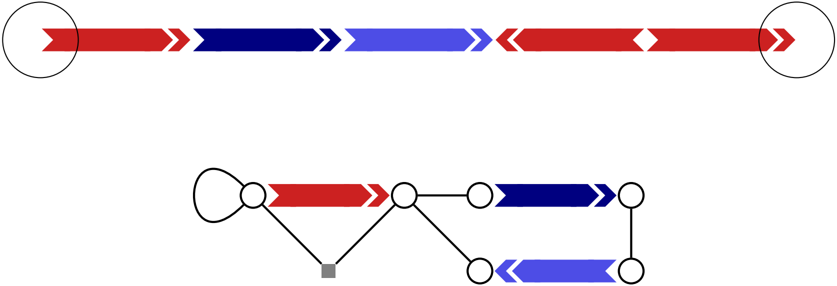
Multi-Breakpoint Graph (MBPG)



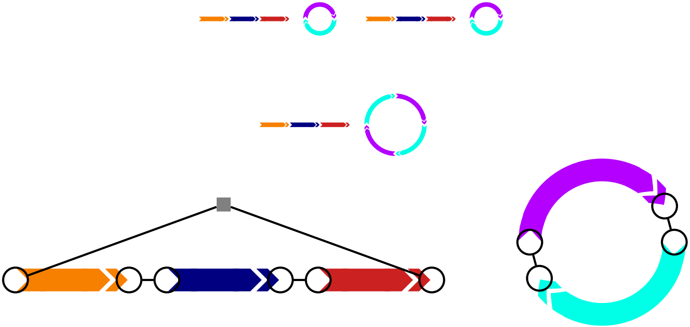
Multi-Breakpoint Graph (MBPG)



Multi-Breakpoint Graph (MBPG)

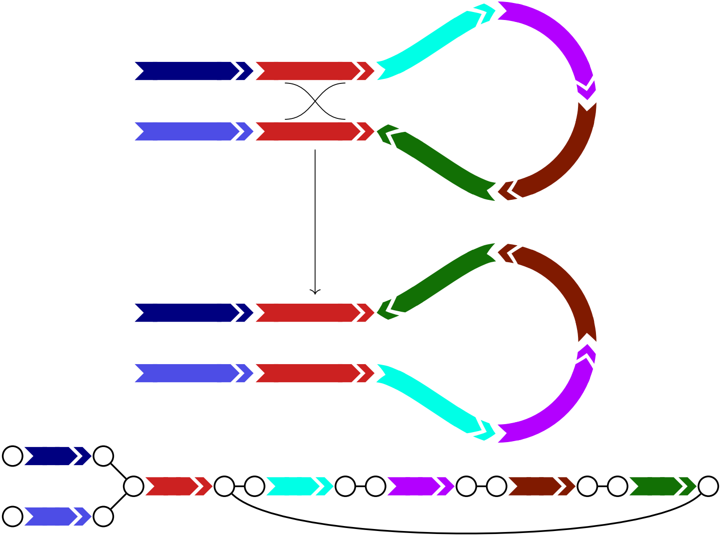


MBPG of a Simple Pangenome



- Simple pangenome $\iff$ adjacencies are a perfect matching

MBPG and Namos



- Namos don't change the graph!

Conclusions from the MBPG

Observation

Given a Namo o and pangenomes $\mathbb{P}, \mathbb{P}'$, *where* $\mathbb{P} \xrightarrow{o} \mathbb{P}'$, *the MBPGs of* $\mathbb{P}$ *and* $\mathbb{P}'$ *are identical.*

Conclusions from the MBPG

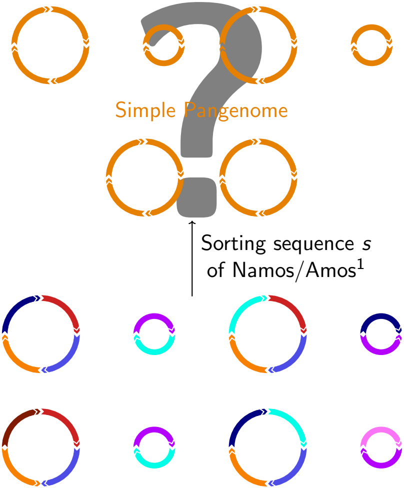
Observation

Given a Namo o and pangenomes $\mathbb{P}, \mathbb{P}'$*, where* $\mathbb{P} \xrightarrow{o} \mathbb{P}'$*, the MBPGs of* $\mathbb{P}$ *and* $\mathbb{P}'$ *are identical.*

Definition

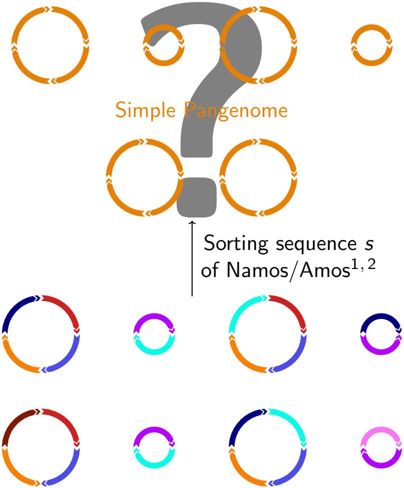
A set of Namos N is called MBPG-complete, if for all pairs of pangenomes $\mathbb{P}, \mathbb{P}'$ with the same MBPG there is a sequence of Namos $o_1 o_2 \dots o_k \in N^*$, such that $\mathbb{P} \xrightarrow{o_1}\xrightarrow{o_2} \cdots \xrightarrow{o_k} \mathbb{P}'$.

Complete Ancestral Reconstruction for Pangenomes (CARP)



- #Amos in s: measure of the Structural Complexity.
- [1] with minimal #Amos

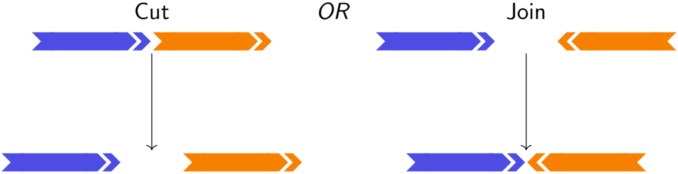
Complete Ancestral Reconstruction for Pangenomes (CARP)



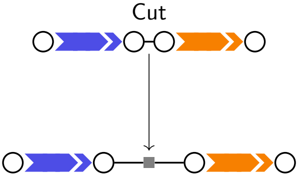
- #Amos in *s*: measure of the Structural Complexity.
- [1] with minimal #Amos
- [2] MBPG-complete set of Namos

SCJ-CARP

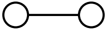
Amos: SCJ-operations



Amos: SCJ-operations

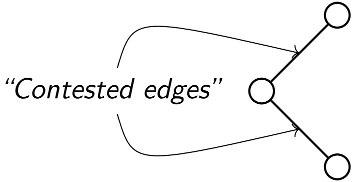


Modifying the MBPG with SCJs



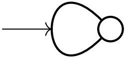
Nothing to be done for adjacencies at non-branching nodes.

Modifying the MBPG with SCJs



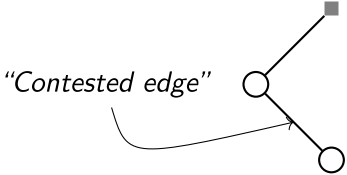
“Contested edges”

Modifying the MBPG with SCJs



"Contested edge"

Modifying the MBPG with SCJs



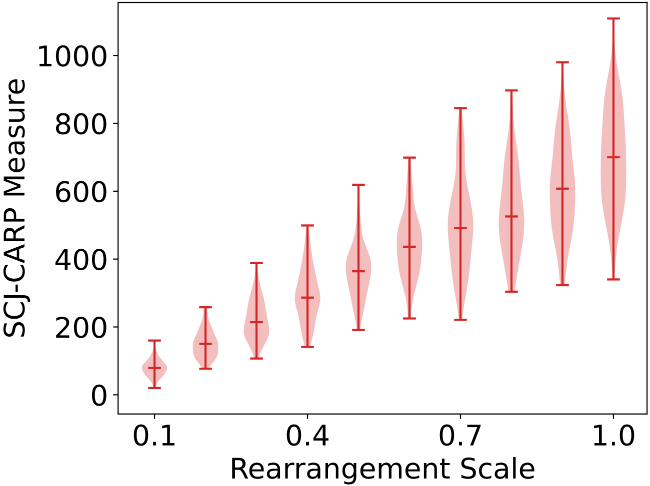
“Contested edge”

Modifying the MBPG with SCJs

Lemma

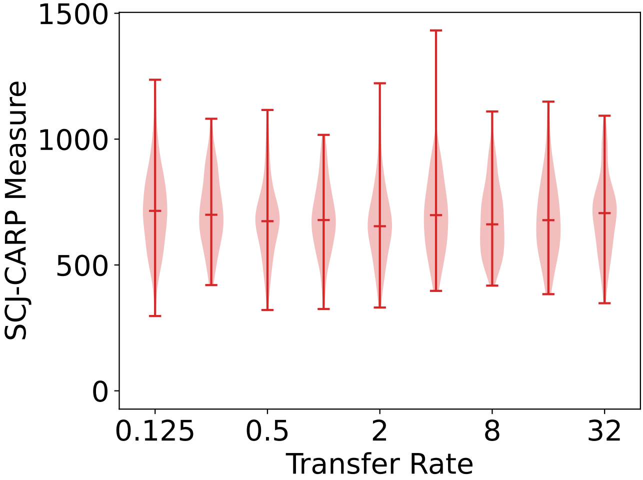
Need exactly $|E_C|$ SCJs to transform the MBPG into a simple MBPG where E_C is the set of contested adjacencies. We call $|E_C|$ the SCJ-CARP measure.

SCJ-CARP Tracks Structural Complexity



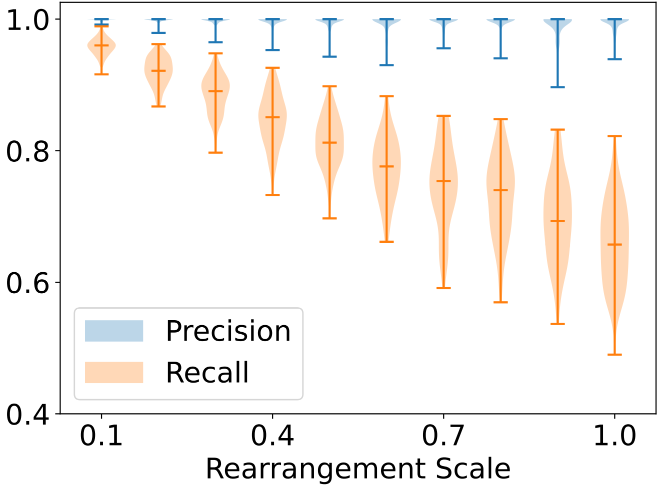
Pearson Coefficient: 0.88

SCJ-CARP Does Not Track Horizontal Effects



Pearson Coefficient: -0.01

SCJ-CARP Reconstructs Ancestral Adjacencies (to a Limited Extent)



Conclusion

Open Questions

- CARP for more complex rearrangement models (DCJ/HP/BI...)

Open Questions

- CARP for more complex rearrangement models (DCJ/HP/BI...)
  - → Characterize Pangenome by vector of CARP measures $(m_{SCJ}, m_{DCJ}, m_{HP}, m_{BI}, \ldots)$.
  - → Compare vectors to compare pangenomes

Open Questions

- CARP for more complex rearrangement models (DCJ/HP/BI...)
  - → Characterize Pangenome by vector of CARP measures $(m_{SCJ}, m_{DCJ}, m_{HP}, m_{BI}, \ldots)$.
  - → Compare vectors to compare pangenomes
- Comparing pangenomes $\mathbb{P}_a, \mathbb{P}_b$ on the same marker set

Open Questions

- CARP for more complex rearrangement models (DCJ/HP/BI...)
  - → Characterize Pangenome by vector of CARP measures $(m_{SCJ}, m_{DCJ}, m_{HP}, m_{BI}, \ldots)$.
  - → Compare vectors to compare pangenomes
- Comparing pangenomes $\mathbb{P}_a, \mathbb{P}_b$ on the same marker set
  - Simple:
    - Calculate CARP ancestors $\mathbb{A}_a, \mathbb{A}_b$
    - $d(\mathbb{P}_a, \mathbb{P}_b) = d(\mathbb{A}_a, \mathbb{A}_b)$

Open Questions

- CARP for more complex rearrangement models (DCJ/HP/BI...)
  - → Characterize Pangenome by vector of CARP measures $(m_{SCJ}, m_{DCJ}, m_{HP}, m_{BI}, \ldots)$.
  - → Compare vectors to compare pangenomes
- Comparing pangenomes $\mathbb{P}_a, \mathbb{P}_b$ on the same marker set
  - Simple:
    - Calculate CARP ancestors $\mathbb{A}_a, \mathbb{A}_b$
    - $d(\mathbb{P}_a, \mathbb{P}_b) = d(\mathbb{A}_a, \mathbb{A}_b)$
  - Advanced:
    - Among all possible CARP ancestors A_a, A_b
    - $d(\mathbb{P}_a, \mathbb{P}_b) = \min_{\mathbb{A}_a \in A_a, \mathbb{A}_b \in A_b} d(\mathbb{A}_a, \mathbb{A}_b)$

Open Questions

- CARP for more complex rearrangement models (DCJ/HP/BI...)
  - → Characterize Pangenome by vector of CARP measures $(m_{SCJ}, m_{DCJ}, m_{HP}, m_{BI}, \ldots)$.
  - → Compare vectors to compare pangenomes
- Comparing pangenomes $\mathbb{P}_a, \mathbb{P}_b$ on the same marker set
  - Simple:
    - Calculate CARP ancestors $\mathbb{A}_a, \mathbb{A}_b$
    - $d(\mathbb{P}_a, \mathbb{P}_b) = d(\mathbb{A}_a, \mathbb{A}_b)$
  - Advanced:
    - Among all possible CARP ancestors A_a, A_b
    - $d(\mathbb{P}_a, \mathbb{P}_b) = \min_{\mathbb{A}_a \in A_a, \mathbb{A}_b \in A_b} d(\mathbb{A}_a, \mathbb{A}_b)$
- (Formal definition for) Marker segmentation on pangenomes

Thank you!

slides

github