Topics of today:

Singular DCJ-indel distance and sorting:

1. Review
2. Capped relational graph of canonical genomes
3. Capped relational graph of singular genomes
4. Indel-potential of cycles via transitions

Components of a relational graph

Cycle with k extremity-edges: k-cycle or c_k

Path with k extremity-edges: k-path or p_k

if $k = 0$ the component is a **singleton**

Disjoint sets of components:

$$\begin{cases} \mathcal{C} = \{c_k : k \geq 2\} : \text{set of cycles } (k \text{ is even}) \\ \mathcal{S} = \{c_k : k = 0\} : \text{set of circular singletons} \\ \mathcal{P}_{\mathbb{A}\mathbb{A}} = \{p_k : \text{starts and ends in } \mathbb{A}\} : \\ \qquad \text{set of } \mathbb{A}\mathbb{A}\text{-paths } (k \geq 0 \text{ is even}) \\ \mathcal{P}_{\mathbb{B}\mathbb{B}} = \{p_k : \text{starts and ends in } \mathbb{B}\} : \\ \qquad \text{set of } \mathbb{B}\mathbb{B}\text{-paths } (k \geq 0 \text{ is even}) \\ \mathcal{P}_{\mathbb{A}\mathbb{B}} = \{p_k : \text{starts in } \mathbb{A} \text{ and ends in } \mathbb{B}\} : \\ \qquad \text{set of } \mathbb{A}\mathbb{B}\text{-paths } (k \geq 1 \text{ is odd}) \end{cases}$$

DCJ-sorted (or short) components: 2-cycles and 1-paths (and 0-cycles and 0-paths)

Long components: k-cycles (with $k \geq 4$) and k-paths (with $k \geq 2$)

DCJ-sorting a long component C: transforming C into a set of DCJ-sorted components with DCJ-operations

Types of DCJ operation

With respect to the position of the cuts:

Internal: either a single-cut operation or two cuts applied in the same component

Recombination: each cut is applied in a distinct component

With respect to the effect on the relational graph:

Gaining: creates one cycle or two $\mathbb{AB}$-paths
$\Delta_{\text{DCJ}} = 0$

Neutral: preserves the number of cycles and of $\mathbb{AB}$-paths
$\Delta_{\text{DCJ}} = 1$

Losing: destroys one cycle or two $\mathbb{AB}$-paths
$\Delta_{\text{DCJ}} = 2$

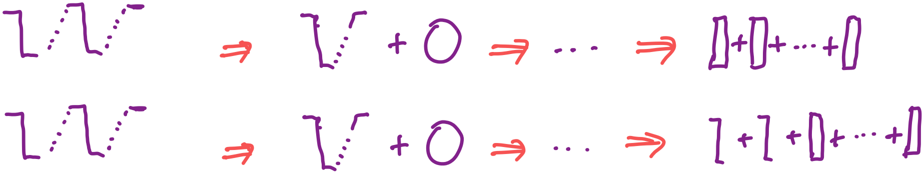
Each component can be sorted separately...

...with an internal gaining DCJ at each step:

Cycle: creates a new cycle at each step

$\mathbb{AB}$-path: creates a new cycle at each step

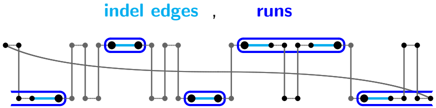
$\mathbb{AA}$-path: creates a new cycle at each step, eventually one step is a single cut (on $\mathbb{B}$) that creates two $\mathbb{AB}$-paths



$\mathbb{BB}$-path: analogous to $\mathbb{AA}$-path

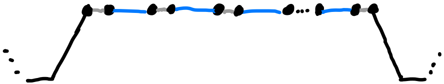
Accummulating runs

One indel-enclosing cycle:



$\Lambda = 4$

Accummulated run:



Each **run** can be **accumulated** with internal gaining DCJ operations and then inserted/deleted at once

⇒ Second upper bound:

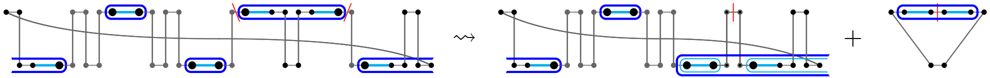
$$\mathbf{d}_{\mathrm{DCJ}}^{\mathrm{ID}}(\mathbb{A},\mathbb{B}) \leq n - |\mathcal{C}| - \frac{|\mathcal{P}_{\mathbb{A}\mathbb{B}}|}{2} + \sum_{C \in RG(\mathbb{A},\mathbb{B})} \Lambda(C)$$

Merging runs with internal gaining DCJ operations

DCJ operations can modify the number of runs by at most two:

$$\text{A DCJ operation can have} \begin{cases} \Delta_\Lambda = -2 & \text{(merges two pairs of runs)} \\ \Delta_\Lambda = -1 & \text{(merges one pair of runs)} \\ \Delta_\Lambda = 0 & \text{(preserves the runs)} \\ \Delta_\Lambda = 1 & \text{(splits one run)} \\ \Delta_\Lambda = 2 & \text{(splits two runs)} \end{cases}$$

A **gaining DCJ operation** applied to two adjacency-edges belonging to the same indel-enclosing component can **decrease** the number of runs:



$\Lambda = 4 \quad \rightsquigarrow \quad 2 \quad + \quad 1 \; = \; 3 \;\; (\Delta_\Lambda{=}{-1})$

Indel-potential $\lambda(C)$ of a component C:

minimum number of runs that we can obtain by DCJ-sorting C with internal gaining DCJ operations

Indel-potential of a cycle C - with $\Lambda(C) = 0, 1, 2, 4, 6, 8, \ldots$

We will show that $\lambda(C)$ depends only on the value $\Lambda(C)$: denote $\lambda(C) = \lambda(\Lambda(C))$

$$\Lambda(C) = 0 \Rightarrow \lambda(0) = 0$$
$$\Lambda(C) = 1 \Rightarrow \lambda(1) = 1$$
$$\Lambda(C) = 2 \Rightarrow \lambda(2) = 2$$
$$\Lambda(C) = 4 \Rightarrow \lambda(4) = 3 \text{ (can be verified by listing all cases)}$$

$\Lambda(C) \geq 6$: extract 3 runs from C into a new cycle $\rightarrow$ garantees that $\Delta_\Lambda = -2$

$$\text{two resulting cycles: } \begin{cases} \text{one with 2 runs} \\ \text{one with } \Lambda(C) - 4 \text{ runs} \end{cases}$$

$$\Rightarrow \lambda(6) = \lambda(2) + \lambda(2) = 2 + 2 = 4$$
$$\Rightarrow \lambda(8) = \lambda(2) + \lambda(4) = 2 + 3 = 5$$
$$\Rightarrow \lambda(10) = \lambda(2) + \lambda(6) = 2 + 4 = 6$$

Λ	λ'
0	0
1	1
2	2
4	3
6	4
8	5
$\vdots$	$\vdots$

$$\text{Induction: } \begin{cases} \text{hypothesis: } \lambda(\Lambda(C)) = \frac{\Lambda(C)}{2} + 1 \\ \text{base cases: } \lambda(1) = 1,\ \lambda(2) = 2 \text{ and } \lambda(4) = 3 \end{cases}$$

Induction step: in general, for $\Lambda(C) \geq 6$, we can state

$$\lambda(\Lambda(C)) = \lambda(2) + \lambda(\Lambda(C) - 4)$$
$$= 2 + \left(\frac{\Lambda(C) - 4}{2} + 1\right)$$
$$= \frac{\Lambda(C)}{2} + 1$$

Indel-potential λ of a path P - with $\Lambda(P) = 0, 1, 2, 3, 4, 5, 6, 7, 8, \ldots$

Since $\lambda(P)$ depends only on the value $\Lambda(P)$, we can denote $\lambda(P) = \lambda(\Lambda(P))$

$$\Lambda(P) = 0 \Rightarrow \lambda(0) = 0$$

$$\Lambda(P) = 1 \Rightarrow \lambda(1) = 1$$

$$\Lambda(P) = 2 \Rightarrow \lambda(2) = 2$$

$$\Lambda(P) \geq 3 : \begin{cases} \text{if } \Lambda(P) \text{ is even, then } \lambda(\Lambda(P)) = \frac{\Lambda(P)}{2} + 1 \\ \text{else } \lambda(\Lambda(P)) = \lambda(\Lambda(P) - 1) \end{cases}$$

Λ	λ''
0	0
1	1
2 3	2 2
4 5	3 3
6 7	4 4
$\vdots$	$\vdots$

In general, for $\Lambda(P) \geq 2$, we have

$$\lambda(\Lambda(P)) = \left\lceil \frac{\Lambda(P) + 1}{2} \right\rceil$$

Indel-potential λ of a component C

If C is a singleton: $\lambda(C) = 1$

If C is a cycle:

$$\lambda(C) = \begin{cases} 0 & \text{if } \Lambda(C) = 0 \ (C \text{ is indel-free}) \\ 1 & \text{if } \Lambda(C) = 1 \\ \frac{\Lambda(C)}{2} + 1 & \text{if } \Lambda(C) \geq 2 \end{cases}$$

If C is a path:

$$\lambda(C) = \begin{cases} 0 & \text{if } \Lambda(C) = 0 \ (C \text{ is indel-free}) \\ \left\lceil \frac{\Lambda(C)+1}{2} \right\rceil & \text{if } \Lambda(C) \geq 1 \end{cases}$$

In general, for any component C:

$$\lambda(C) = \begin{cases} 0 & \text{if } \Lambda(C) = 0 \ (C \text{ is indel-free}) \\ \left\lceil \frac{\Lambda(C)+1}{2} \right\rceil & \text{if } \Lambda(C) \geq 1 \end{cases}$$

Λ	λ	
0	0	paths and cycles
1	1	paths, cycles and singletons
2	2	paths and cycles
3	2	paths
4	3	paths and cycles
5	3	paths
6	4	paths and cycles
7	4	paths
⋮	⋮	

Third upper bound: $$d_{\mathrm{DCJ}}^{\mathrm{ID}}(\mathbb{A}, \mathbb{B}) \leq n - |\mathcal{C}| - \frac{|\mathcal{P}_{\mathbb{AB}}|}{2} + \sum_{C \in RG} \lambda(C)$$

(gaining DCJ operations + indels sorting components separately)

Effect of a DCJ operation on the third upper bound:

DCJ-types of DCJ operation $\begin{cases} \Delta_{\text{DCJ}} = 0 \text{ (gaining): creates one cycle or two } \mathbb{AB}\text{-paths} \\ \Delta_{\text{DCJ}} = 1 \text{ (neutral): preserves the numbers of cycles and of } \mathbb{AB}\text{-paths} \\ \Delta_{\text{DCJ}} = 2 \text{ (losing): destroys one cycle or two } \mathbb{AB}\text{-paths} \end{cases}$

Indel-types of DCJ operation $\begin{cases} \Delta_\lambda = -2 & : \text{decreases the overall indel-potential by two} \\ \Delta_\lambda = -1 & : \text{decreases the overall indel-potential by one} \\ \Delta_\lambda = 0 & : \text{does not change the overall indel-potential} \\ \Delta_\lambda = 1 & : \text{increases the overall indel-potential by one} \\ \Delta_\lambda = 2 & : \text{increases the overall indel-potential by two} \end{cases}$

Effect of a DCJ operation ρ on the third upper bound: $\Delta^\lambda_{\text{DCJ}}(\rho) = \Delta_{\text{DCJ}}(\rho) + \Delta_\lambda(\rho)$

DCJ Operations that can decrease the third upper bound: $\begin{cases} \Delta_{\text{DCJ}} = 0 \text{ (gaining) and } \Delta_\lambda = -2 & : \Delta^\lambda_{\text{DCJ}} = -2 \\ \Delta_{\text{DCJ}} = 0 \text{ (gaining) and } \Delta_\lambda = -1 & : \Delta^\lambda_{\text{DCJ}} = -1 \\ \Delta_{\text{DCJ}} = 1 \text{ (neutral) and } \Delta_\lambda = -2 & : \Delta^\lambda_{\text{DCJ}} = -1 \end{cases}$

- By definition: any internal gaining DCJ operation ρ (applied to a single component) has $\Delta_\lambda(\rho) \geq 0$ and, consequently, $\Delta^\lambda_{\text{DCJ}}(\rho) \geq 0$
- Any losing DCJ operation ρ has $\Delta^\lambda_{\text{DCJ}}(\rho) \geq 0$

DCJ operations involving cycles

Λ	λ
0	0
1	1
2	2
4	3
6	4
8	5
⋮	⋮

- Any recombination involving two cycles is losing and has $\Delta^{\lambda}_{\text{DCJ}} \geq 0$ (cannot decrease the DCJ-indel distance)

- An internal DCJ operation ρ applied to a cycle C can be:
  - Gaining, with $\Delta^{\lambda}_{\text{DCJ}}(\rho) \geq 0$ (cannot decrease the DCJ-indel distance)
  - Neutral ($\Delta_{\text{DCJ}}(\rho) = 1$):

    If $\Lambda(C) \geq 4$, the DCJ ρ can merge at most two pairs of runs: $\Delta_{\Lambda}(\rho) \geq -2$ and $\Delta_{\lambda}(\rho) \geq -1$

    $\Rightarrow$ Any internal neutral DCJ operation applied to a cycle has $\Delta^{\lambda}_{\text{DCJ}} \geq 0$ (cannot decrease the DCJ-indel distance)

If singular genomes $\mathbb{A}$ and $\mathbb{B}$ are circular, the graph $RG(\mathbb{A}, \mathbb{B})$ has only cycles (and eventually singletons).

In this case:

$$d^{\text{ID}}_{\text{DCJ}}(\mathbb{A}, \mathbb{B}) = n - |\mathcal{C}| + \sum_{C \in RG} \lambda(C)$$

DCJ operations involving paths

Λ	λ
0	0
1	1
2 3	2 2
4 5	3 3
6 7	4 4
⋮	⋮

- Any recombination involving a path and a cycle is losing and has $\Delta^{\lambda}_{\mathrm{DCJ}} \geq 0$ (cannot decrease the DCJ-indel distance)

- An internal DCJ operation ρ applied to a path P can be:
  - Gaining, with $\Delta^{\lambda}_{\mathrm{DCJ}}(\rho) \geq 0$ (cannot decrease the DCJ-indel distance)
  - Neutral ($\Delta_{\mathrm{DCJ}}(\rho) = 1$):

    If $\Lambda(P) \geq 4$, the DCJ ρ can merge at most two pairs of runs: $\Delta_{\Lambda}(\rho) \geq -2$ and $\Delta_{\lambda}(\rho) \geq -1$

    $\Rightarrow$ Any internal neutral DCJ operation applied to a path has $\Delta^{\lambda}_{\mathrm{DCJ}} \geq 0$ (cannot decrease the DCJ-indel distance)

Path recombinations can have $\Delta^{\lambda}_{\mathrm{DCJ}} \leq -1$

A gaining (**deducting**) path recombination with $\Delta^{\lambda}_{\mathrm{DCJ}} = -2$:

Sources

$(\sum \lambda = 2 + 2 = 4)$

$\mathbb{AA}$ + $\mathbb{BB}$

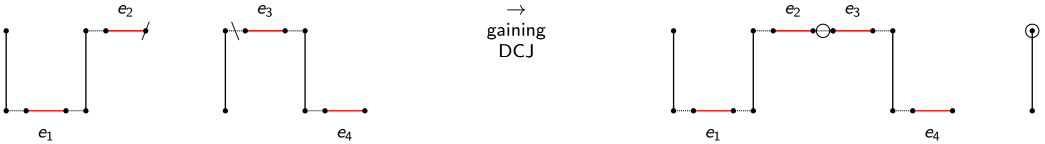
2 runs + 2 runs

Resultants

$(\sum \lambda = 2 + 0 = 2)$

$\mathbb{AB}$ + $\mathbb{AB}$

3 runs + no run



$$\mathbb{AA}_{\mathcal{BA}} + \mathbb{BB}_{\mathcal{AB}} = \begin{cases} \mathbb{AB}_{\mathcal{BAB}} + \mathbb{AB}_{\varepsilon} \\ (\mathbb{AB}_{\mathcal{ABA}} + \mathbb{AB}_{\varepsilon}) \\ (\mathbb{AB}_{\mathcal{A}} \quad + \mathbb{AB}_{\mathcal{B}}) \end{cases}$$

(all variants have $\Delta^{\lambda}_{\mathrm{DCJ}} = -2$)

Deducting path recombinations

have $\Delta^{\lambda}_{\mathrm{DCJ}} \leq -1$

General DCJ-indel distance formula:

$$d^{\mathrm{ID}}_{\mathrm{DCJ}}(\mathbb{A}, \mathbb{B}) = n - |\mathcal{C}| - \frac{|\mathcal{P}_{\mathbb{AB}}|}{2} + \sum_{C \in RG} \lambda(C) - \delta,$$

where δ is the value obtained by optimizing deducting path recombinations

Optimizing deducting path recombinations (for computing δ)

Run-type of a path $\begin{cases} \varepsilon & \equiv & \varepsilon & \text{(empty)} \\ \mathcal{ABAB}\dots\mathcal{A} & \equiv & \mathcal{A} & \text{(odd)} \\ \mathcal{BABA}\dots\mathcal{B} & \equiv & \mathcal{B} & \text{(odd)} \\ \mathcal{ABAB}\dots\mathcal{AB} & \equiv & \mathcal{AB} & \text{(even)} \\ \mathcal{BABA}\dots\mathcal{BA} & \equiv & \mathcal{BA} & \text{(even)} \end{cases}$

Path types $\begin{cases} \mathbb{AA}_\varepsilon, \mathbb{AA}_\mathcal{A}, \mathbb{AA}_\mathcal{B}, \mathbb{AA}_\mathcal{AB} (\equiv \mathbb{AA}_\mathcal{BA}) \\ \mathbb{BB}_\varepsilon, \mathbb{BB}_\mathcal{A}, \mathbb{BB}_\mathcal{B}, \mathbb{BB}_\mathcal{AB} (\equiv \mathbb{BB}_\mathcal{BA}) \\ \mathbb{AB}_\varepsilon, \mathbb{AB}_\mathcal{A}, \mathbb{AB}_\mathcal{B}, \mathbb{AB}_\mathcal{AB}, \mathbb{AB}_\mathcal{BA} \\ \quad \Rightarrow \text{an } \mathbb{AB}\text{-path is always read from } \mathbb{A} \text{ to } \mathbb{B} \end{cases}$

Deducting path recombinations that allow the best reuse of the resultants:

sources	resultants	Δ_λ	Δ_{DCJ}	$\Delta^\lambda_{\text{DCJ}}$
$\mathbb{AA}_\mathcal{AB} + \mathbb{BB}_\mathcal{AB}$	$\bullet + \bullet$	-2	0	-2
$\mathbb{AA}_\mathcal{AB} + \mathbb{BB}_\mathcal{A}$	$\bullet + \mathbb{AB}_\mathcal{BA}$	-1	0	-1
$\mathbb{AA}_\mathcal{AB} + \mathbb{BB}_\mathcal{B}$	$\bullet + \mathbb{AB}_\mathcal{AB}$	-1	0	-1
$\mathbb{AA}_\mathcal{A} + \mathbb{BB}_\mathcal{AB}$	$\bullet + \mathbb{AB}_\mathcal{AB}$	-1	0	-1
$\mathbb{AA}_\mathcal{B} + \mathbb{BB}_\mathcal{AB}$	$\bullet + \mathbb{AB}_\mathcal{BA}$	-1	0	-1
$\mathbb{AA}_\mathcal{A} + \mathbb{BB}_\mathcal{A}$	$\bullet + \bullet$	-1	0	-1
$\mathbb{AA}_\mathcal{B} + \mathbb{BB}_\mathcal{B}$	$\bullet + \bullet$	-1	0	-1

sources	resultants	Δ_λ	Δ_{DCJ}	$\Delta^\lambda_{\text{DCJ}}$
$\mathbb{AA}_\mathcal{AB} + \mathbb{AA}_\mathcal{AB}$	$\mathbb{AA}_\mathcal{A} + \mathbb{AA}_\mathcal{B}$	-2	$+1$	-1
$\mathbb{BB}_\mathcal{AB} + \mathbb{BB}_\mathcal{AB}$	$\mathbb{BB}_\mathcal{A} + \mathbb{BB}_\mathcal{B}$	-2	$+1$	-1
$\mathbb{AA}_\mathcal{AB} + \mathbb{AB}_\mathcal{AB}$	$\bullet + \mathbb{AA}_\mathcal{A}$	-2	$+1$	-1
$\mathbb{AA}_\mathcal{AB} + \mathbb{AB}_\mathcal{BA}$	$\bullet + \mathbb{AA}_\mathcal{B}$	-2	$+1$	-1
$\mathbb{BB}_\mathcal{AB} + \mathbb{AB}_\mathcal{AB}$	$\bullet + \mathbb{BB}_\mathcal{B}$	-2	$+1$	-1
$\mathbb{BB}_\mathcal{AB} + \mathbb{AB}_\mathcal{BA}$	$\bullet + \mathbb{BB}_\mathcal{A}$	-2	$+1$	-1
$\mathbb{AB}_\mathcal{AB} + \mathbb{AB}_\mathcal{BA}$	$\bullet + \bullet$	-2	$+1$	-1

$\bullet \begin{cases} \mathbb{AB}_\varepsilon \\ \mathbb{AB}_\mathcal{A} \\ \mathbb{AB}_\mathcal{B} \end{cases}$

Path recombinations with $\Delta^\lambda_{\text{DCJ}} = 0$ creating resultants that can be used in deducting recombinations:

sources	resultants	Δ_λ	Δ_{DCJ}	$\Delta^\lambda_{\text{DCJ}}$
$\mathbb{AA}_\mathcal{A} + \mathbb{AB}_\mathcal{BA}$	$\bullet + \mathbb{AA}_\mathcal{AB}$	-1	$+1$	0
$\mathbb{AA}_\mathcal{B} + \mathbb{AB}_\mathcal{AB}$	$\bullet + \mathbb{AA}_\mathcal{AB}$	-1	$+1$	0
$\mathbb{BB}_\mathcal{A} + \mathbb{AB}_\mathcal{AB}$	$\bullet + \mathbb{BB}_\mathcal{AB}$	-1	$+1$	0
$\mathbb{BB}_\mathcal{B} + \mathbb{AB}_\mathcal{BA}$	$\bullet + \mathbb{BB}_\mathcal{AB}$	-1	$+1$	0

sources	resultants	Δ_λ	Δ_{DCJ}	$\Delta^\lambda_{\text{DCJ}}$
$\mathbb{AA}_\mathcal{A} + \mathbb{BB}_\mathcal{B}$	$\bullet + \mathbb{AB}_\mathcal{AB}$	0	0	0
$\mathbb{AA}_\mathcal{B} + \mathbb{BB}_\mathcal{A}$	$\bullet + \mathbb{AB}_\mathcal{BA}$	0	0	0
$\mathbb{AB}_\mathcal{AB} + \mathbb{AB}_\mathcal{AB}$	$\mathbb{AA}_\mathcal{A} + \mathbb{BB}_\mathcal{B}$	-2	$+2$	0
$\mathbb{AB}_\mathcal{BA} + \mathbb{AB}_\mathcal{BA}$	$\mathbb{AA}_\mathcal{B} + \mathbb{BB}_\mathcal{A}$	-2	$+2$	0

Sources:

$\mathbb{AA}_\mathcal{AB}$: $\mathtt{W}$

$\mathbb{AA}_\mathcal{A}$: $\overline{\mathtt{W}}$

$\mathbb{AA}_\mathcal{B}$: $\underline{\mathtt{W}}$

$\mathbb{BB}_\mathcal{AB}$: $\mathtt{M}$

$\mathbb{BB}_\mathcal{A}$: $\overline{\mathtt{M}}$

$\mathbb{BB}_\mathcal{B}$: $\underline{\mathtt{M}}$

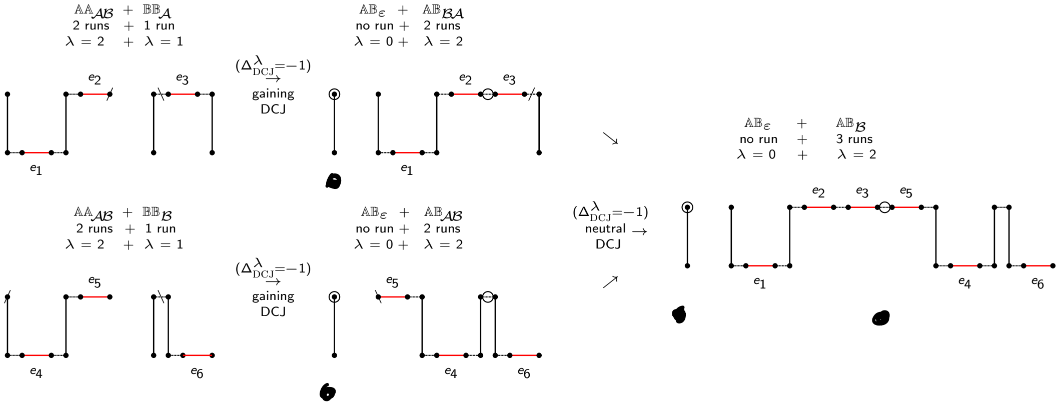
$\mathbb{AB}_\mathcal{AB}$: $\mathtt{Z}$

$\mathbb{AB}_\mathcal{BA}$: $\mathtt{N}$

Optimizing deducting path recombinations (for computing δ)

Deducting chain of path recombinations

transforming	$2 \times \mathbb{AA}_{\mathcal{AB}} + \mathbb{BB}_{\mathcal{A}} + \mathbb{BB}_{\mathcal{B}}$
into	$3 \times \mathbb{AB}_{\varepsilon} + \mathbb{AB}_{\mathcal{B}}$
with	overall $\Delta^{\lambda}_{\mathrm{DCJ}} = -3$



	id	sources			resultants				$\mathbf{\Delta}^{\boldsymbol{\lambda}}_{\mathrm{DCJ}}$	scr
$\mathcal{P}$	$\mathtt{WM}$	$\mathbb{AA}_{\mathcal{AB}}$	$\mathbb{BB}_{\mathcal{AB}}$	—	—	—	—	$2 \times \bullet$	-2	-1
$\mathcal{Q}$	$\mathtt{WW\overline{M}\underline{M}}$	$2 \times \mathbb{AA}_{\mathcal{AB}}$	$\mathbb{BB}_{\mathcal{A}} + \mathbb{BB}_{\mathcal{B}}$	—	—	—	—	$4 \times \bullet$	-3	$-3/4$
	$\mathtt{MM\overline{W}\underline{W}}$	$\mathbb{AA}_{\mathcal{A}} + \mathbb{AA}_{\mathcal{B}}$	$2 \times \mathbb{BB}_{\mathcal{AB}}$	—	—	—	—	$4 \times \bullet$	-3	$-3/4$
$\mathcal{T}$	$\mathtt{WZ\overline{M}}$	$\mathbb{AA}_{\mathcal{AB}}$	$\mathbb{BB}_{\mathcal{A}}$	$\mathbb{AB}_{\mathcal{AB}}$	—	—	—	$3 \times \bullet$	-2	$-2/3$
	$\mathtt{WW\overline{M}}$	$2 \times \mathbb{AA}_{\mathcal{AB}}$	$\mathbb{BB}_{\mathcal{A}}$	—	$\mathbb{AA}_{\mathcal{B}}$	—	—	$2 \times \bullet$	-2	$-2/3$
	$\mathtt{WN\underline{M}}$	$\mathbb{AA}_{\mathcal{AB}}$	$\mathbb{BB}_{\mathcal{B}}$	$\mathbb{AB}_{\mathcal{BA}}$	—	—	—	$3 \times \bullet$	-2	$-2/3$
	$\mathtt{WW\underline{M}}$	$2 \times \mathbb{AA}_{\mathcal{AB}}$	$\mathbb{BB}_{\mathcal{B}}$	—	$\mathbb{AA}_{\mathcal{A}}$	—	—	$2 \times \bullet$	-2	$-2/3$
	$\mathtt{MN\overline{W}}$	$\mathbb{AA}_{\mathcal{A}}$	$\mathbb{BB}_{\mathcal{AB}}$	$\mathbb{AB}_{\mathcal{BA}}$	—	—	—	$3 \times \bullet$	-2	$-2/3$
	$\mathtt{MM\overline{W}}$	$\mathbb{AA}_{\mathcal{A}}$	$2 \times \mathbb{BB}_{\mathcal{AB}}$	—	—	$\mathbb{BB}_{\mathcal{B}}$	—	$2 \times \bullet$	-2	$-2/3$
	$\mathtt{MZ\underline{W}}$	$\mathbb{AA}_{\mathcal{B}}$	$\mathbb{BB}_{\mathcal{AB}}$	$\mathbb{AB}_{\mathcal{AB}}$	—	—	—	$3 \times \bullet$	-2	$-2/3$
	$\mathtt{MM\underline{W}}$	$\mathbb{AA}_{\mathcal{B}}$	$2 \times \mathbb{BB}_{\mathcal{AB}}$	—	—	$\mathbb{BB}_{\mathcal{A}}$	—	$2 \times \bullet$	-2	$-2/3$
$\mathcal{S}$	$\mathtt{ZN}$	—	—	$\mathbb{AB}_{\mathcal{AB}} + \mathbb{AB}_{\mathcal{BA}}$	—	—	—	$2 \times \bullet$	-1	$-1/2$
	$\mathtt{\overline{WM}}$	$\mathbb{AA}_{\mathcal{A}}$	$\mathbb{BB}_{\mathcal{A}}$	—	—	—	—	$2 \times \bullet$	-1	$-1/2$
	$\mathtt{\underline{WM}}$	$\mathbb{AA}_{\mathcal{B}}$	$\mathbb{BB}_{\mathcal{B}}$	—	—	—	—	$2 \times \bullet$	-1	$-1/2$
	$\mathtt{W\overline{M}}$	$\mathbb{AA}_{\mathcal{AB}}$	$\mathbb{BB}_{\mathcal{A}}$	—	—	—	$\mathbb{AB}_{\mathcal{BA}}$	$\bullet$	-1	$-1/2$
	$\mathtt{W\underline{M}}$	$\mathbb{AA}_{\mathcal{AB}}$	$\mathbb{BB}_{\mathcal{B}}$	—	—	—	$\mathbb{AB}_{\mathcal{AB}}$	$\bullet$	-1	$-1/2$
	$\mathtt{WZ}$	$\mathbb{AA}_{\mathcal{AB}}$	—	$\mathbb{AB}_{\mathcal{AB}}$	$\mathbb{AA}_{\mathcal{A}}$	—	—	$\bullet$	-1	$-1/2$
	$\mathtt{WN}$	$\mathbb{AA}_{\mathcal{AB}}$	—	$\mathbb{AB}_{\mathcal{BA}}$	$\mathbb{AA}_{\mathcal{B}}$	—	—	$\bullet$	-1	$-1/2$
	$\mathtt{WW}$	$2 \times \mathbb{AA}_{\mathcal{AB}}$	—	—	$\mathbb{AA}_{\mathcal{A}} + \mathbb{AA}_{\mathcal{B}}$	—	—	—	-1	$-1/2$
	$\mathtt{M\overline{W}}$	$\mathbb{AA}_{\mathcal{A}}$	$\mathbb{BB}_{\mathcal{AB}}$	—	—	—	$\mathbb{AB}_{\mathcal{AB}}$	$\bullet$	-1	$-1/2$
	$\mathtt{M\underline{W}}$	$\mathbb{AA}_{\mathcal{B}}$	$\mathbb{BB}_{\mathcal{AB}}$	—	—	—	$\mathbb{AB}_{\mathcal{BA}}$	$\bullet$	-1	$-1/2$
	$\mathtt{MZ}$	—	$\mathbb{BB}_{\mathcal{AB}}$	$\mathbb{AB}_{\mathcal{AB}}$	—	$\mathbb{BB}_{\mathcal{B}}$	—	$\bullet$	-1	$-1/2$
	$\mathtt{MN}$	—	$\mathbb{BB}_{\mathcal{AB}}$	$\mathbb{AB}_{\mathcal{BA}}$	—	$\mathbb{BB}_{\mathcal{A}}$	—	$\bullet$	-1	$-1/2$
	$\mathtt{MM}$	—	$2 \times \mathbb{BB}_{\mathcal{AB}}$	—	—	$\mathbb{BB}_{\mathcal{A}} + \mathbb{BB}_{\mathcal{B}}$	—	—	-1	$-1/2$

	id	sources			resultants				$\Delta^{\lambda}_{\text{DCJ}}$	scr
$\mathcal{M}$	$\mathtt{ZZ\underline{W}\overline{M}}$	$\mathbb{AA}_{\mathcal{B}}$	$\mathbb{BB}_{\mathcal{A}}$	$2 \times \mathbb{AB}_{\mathcal{AB}}$	—	—	—	$4 \times \bullet$	-2	$-1/2$
	$\mathtt{NN\overline{W}\underline{M}}$	$\mathbb{AA}_{\mathcal{A}}$	$\mathbb{BB}_{\mathcal{B}}$	$2 \times \mathbb{AB}_{\mathcal{BA}}$	—	—	—	$4 \times \bullet$	-2	$-1/2$
$\mathcal{N}$	$\mathtt{Z\underline{W}\overline{M}}$	$\mathbb{AA}_{\mathcal{B}}$	$\mathbb{BB}_{\mathcal{A}}$	$\mathbb{AB}_{\mathcal{AB}}$	—	—	$\mathbb{AB}_{\mathcal{BA}}$	$2 \times \bullet$	-1	$-1/3$
	$\mathtt{ZZ\underline{W}}$	$\mathbb{AA}_{\mathcal{B}}$	—	$2 \times \mathbb{AB}_{\mathcal{AB}}$	$\mathbb{AA}_{\mathcal{A}}$	—	—	$2 \times \bullet$	-1	$-1/3$
	$\mathtt{ZZ\overline{M}}$	—	$\mathbb{BB}_{\mathcal{A}}$	$2 \times \mathbb{AB}_{\mathcal{AB}}$	—	$\mathbb{BB}_{\mathcal{B}}$	—	$2 \times \bullet$	-1	$-1/3$
	$\mathtt{N\overline{W}\underline{M}}$	$\mathbb{AA}_{\mathcal{A}}$	$\mathbb{BB}_{\mathcal{B}}$	$\mathbb{AB}_{\mathcal{BA}}$	—	—	$\mathbb{AB}_{\mathcal{AB}}$	$2 \times \bullet$	-1	$-1/3$
	$\mathtt{NN\overline{W}}$	$\mathbb{AA}_{\mathcal{A}}$	—	$2 \times \mathbb{AB}_{\mathcal{BA}}$	$\mathbb{AA}_{\mathcal{B}}$	—	—	$2 \times \bullet$	-1	$-1/3$
	$\mathtt{NN\underline{M}}$	—	$\mathbb{BB}_{\mathcal{B}}$	$2 \times \mathbb{AB}_{\mathcal{BA}}$	—	$\mathbb{BB}_{\mathcal{A}}$	—	$2 \times \bullet$	-1	$-1/3$

Sources:

$\mathtt{W} : \mathbb{AA}_{\mathcal{AB}}$

$\mathtt{\overline{W}} : \mathbb{AA}_{\mathcal{A}}$

$\mathtt{\underline{W}} : \mathbb{AA}_{\mathcal{B}}$

$\mathtt{M} : \mathbb{BB}_{\mathcal{AB}}$

$\mathtt{\overline{M}} : \mathbb{BB}_{\mathcal{A}}$

$\mathtt{\underline{M}} : \mathbb{BB}_{\mathcal{B}}$

$\mathtt{Z} : \mathbb{AB}_{\mathcal{AB}}$

$\mathtt{N} : \mathbb{AB}_{\mathcal{BA}}$

DCJ-indel distance formula:

$$d^{\text{ID}}_{\text{DCJ}}(\mathbb{A}, \mathbb{B}) = n - |\mathcal{C}| - \frac{|\mathcal{P}_{\mathbb{AB}}|}{2} + \sum_{C \in RG} \lambda(C) - \delta,$$

where δ is the value obtained by optimizing deducting path recombinations:

$$\delta = 2\mathcal{P} + 3\mathcal{Q} + 2\mathcal{T} + \mathcal{S} + 2\mathcal{M} + \mathcal{N}$$

the values $\mathcal{P}$, $\mathcal{Q}$, $\mathcal{T}$, $\mathcal{S}$, $\mathcal{M}$ and $\mathcal{N}$ refer to the corresponding number of chains of deducting path recombinations of each type and can be obtained by a greedy approach (simple top-down screening of the table)

Singular DCJ-indel model - summary

DCJ-indel distance: $d_{\mathrm{DCJ}}^{\mathrm{ID}}(\mathbb{A}, \mathbb{B}) = n - |\mathcal{C}| - \frac{|\mathcal{P}_{\mathbb{AB}}|}{2} + \sum_{C \in RG} \lambda(C) - \delta,$ where δ is the value obtained by optimizing deducting path recombinations

$\mathbb{A}$ and $\mathbb{B}$ are circular: $d_{\mathrm{DCJ}}^{\mathrm{ID}}(\mathbb{A}, \mathbb{B}) = n - |\mathcal{C}| + \sum_{C \in RG} \lambda(C)$

Computing the distance and sorting can be done in **linear time.**

Quiz 1

1 Which of the following statements is correct?

A Any DCJ operation has $\Delta^{\lambda}_{\mathrm{DCJ}} \geq 0$.

B Any gaining DCJ operation has $\Delta^{\lambda}_{\mathrm{DCJ}} \geq 0$.

C Any internal gaining DCJ operation has $\Delta^{\lambda}_{\mathrm{DCJ}} \geq 0$.

2 Which of the following statements about the DCJ-indel model are true?

$\Delta_\lambda < 0$. ✓

A Any DCJ that decreases the number of runs has $\Delta_{\lambda} < 0$.

B If the input genomes are circular, sorting each component of the relational graph separately is an optimal approach.

C An optimal sequence of DCJ operations and indels sorting one singular genome into another can have gaining, neutral and losing DCJs.

D The triangular inequality holds for the DCJ-indel distance.

E The DCJ-indel distance can be distinct from the restricted DCJ-indel distance.

Capped relational graph

Capping is a procedure that circularizes all paths of a relational graph by adding **caps** (**artificial genes**):

- if the capping is optimal, the genomic distance is preserved
- from the capped relational diagram we can derive genomes composed only of circular chromosomes

A capping may require adjacencies between caps:

$\Gamma_{\mathbb{A}}$: represents an adjacency between caps in genome $\mathbb{A}$

$\Gamma_{\mathbb{B}}$: represents an adjacency between caps in genome $\mathbb{B}$.

Capped relational graph of canonical genomes

Optimally linking paths from $RG(\mathbb{A}, \mathbb{B})$ of canonical genomes $\mathbb{A}$ and $\mathbb{B}$ into cycles can be done as follows:

id	**paths**	**linking cycle**		**Δn**	**Δc**	**Δ(2$\mathbb{AB}$)**	**Δ_{DCJ}**
1	$\mathbb{AB}$	$(\mathbb{AB})$		$+0.5$	$+1$	-0.5	0
2	$\mathbb{AA} + \mathbb{BB}$	$(\mathbb{AA}, \mathbb{BB})$		$+1$	$+1$	0	0
3	$\mathbb{AA}$	$(\mathbb{AA}, \Gamma_{\mathbb{B}})$	$\cup$	$+1$	$+1$	0	0
4	$\mathbb{BB}$	$(\mathbb{BB}, \Gamma_{\mathbb{A}})$	$\cap$	$+1$	$+1$	0	0

{ Closing an $\mathbb{AA}$-path (over-represented in genome $\mathbb{A}$ and marked with a $\cup$) requires an adjacency $\Gamma_{\mathbb{B}}$.
Closing a $\mathbb{BB}$-path (over-represented in genome $\mathbb{B}$ and marked with a $\cap$) requires an adjacency $\Gamma_{\mathbb{A}}$.

Any capping producing linking cycles as indicated on the table above is optimal:

- The value $\Delta_{\mathrm{DCJ}} = \Delta n - \Delta c - \Delta(2\mathbb{AB})$ is the DCJ-effect produced by each type of linking cycle.
- All given linking cycles have $\Delta_{\mathrm{DCJ}} = 0$, therefore they preserve the DCJ distance.

Let { $\kappa_{\mathbb{A}}$: number of linear chromosomes in $\mathbb{A}$
$\kappa_{\mathbb{B}}$: number of linear chromosomes in $\mathbb{B}$

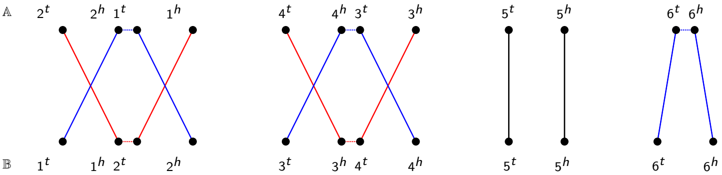
The difference between the number of $\mathbb{AA}$- and of $\mathbb{BB}$-paths is equal to the difference between $\kappa_{\mathbb{A}}$ and $\kappa_{\mathbb{B}}$.

An optimal capping that maximizes the number of linking cycles of type **2** minimizes the number of caps:

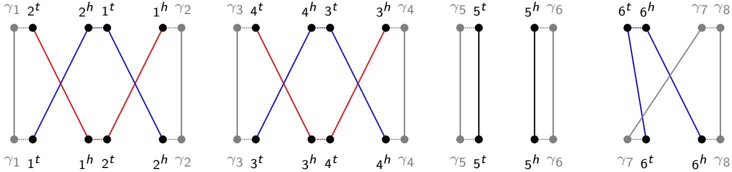
{ The number of caps to be added is exactly $p_* = \max\{\kappa_{\mathbb{A}}, \kappa_{\mathbb{B}}\}$.
The number of adjacencies between caps is exactly $a_* = |\kappa_{\mathbb{A}} - \kappa_{\mathbb{B}}|$.

Capped relational graph of canonical genomes - example

$\mathbb{A} = [2\ 1]\ [4\ 3]\ [5]\ (6)$ and $\mathbb{B} = [1\ 2]\ [3\ 4]\ [5]\ [6]$; $p_* = 4$ and $a_* = 1$



$$
\begin{aligned}
d_{\mathrm{DCJ}} &= n - |\mathcal{C}| - \frac{|\mathcal{P}_{\mathbb{A}\mathbb{B}}|}{2} \\
&= 6 - 0 - 1 \\
&= 5
\end{aligned}
$$



$$
\begin{aligned}
d_{\mathrm{DCJ}} &= n + p_* - |\mathcal{C}| \\
&= 6 + 4 - 5 \\
&= 5
\end{aligned}
$$

Any way of pairing the cap extremities $\gamma_1, \gamma_2, \ldots, \gamma_8$ is valid; possible derived circular genomes are:

$\mathbb{A}_\circ = (2\ 1\ \mathtt{W})\ (4\ 3\ \mathtt{X})\ (5\ \mathtt{Y})\ (6)\ (\mathtt{Z})$ and $\mathbb{B}_\circ = (1\ 2\ \mathtt{W})\ (3\ 4\ \mathtt{X})\ (5\ \mathtt{Y})\ (6\ \mathtt{Z})$

$(\mathtt{W}^h = \gamma_1,\ \mathtt{W}^t = \gamma_2,\ \mathtt{X}^h = \gamma_3,\ \mathtt{X}^t = \gamma_4,\ \mathtt{Y}^h = \gamma_5,\ \mathtt{Y}^t = \gamma_6,\ \mathtt{Z}^h = \gamma_7,\ \mathtt{Z}^t = \gamma_8)$

or

$\mathbb{A}_\circ = (2\ 1\ \mathtt{W}\ 4\ 3\ \mathtt{X}\ 5\ \ \mathtt{Y}\ \mathtt{Z})\ (6)$ and $\mathbb{B}_\circ = (1\ 2\ \mathtt{W}\ 3\ 4\ \mathtt{X}\ 5\ \mathtt{Y}\ 6\ \mathtt{Z})$

$(\mathtt{W}^h = \gamma_3,\ \mathtt{W}^t = \gamma_2,\ \mathtt{X}^h = \gamma_5,\ \mathtt{X}^t = \gamma_4,\ \mathtt{Y}^h = \gamma_7,\ \mathtt{Y}^t = \gamma_6,\ \mathtt{Z}^h = \gamma_1,\ \mathtt{Z}^t = \gamma_8)$

Capping the relational graph - singular genomes

The sources of each chain of deducting recombinations must be properly linked together into a single cycle.

$$\begin{cases} \text{Unbalanced chains over-represented in genome } \mathbb{A} \text{ are marked with a } \cup \\ \quad \mathbb{BB}_\varepsilon \prec \Gamma_\mathbb{B}\text{: a path } \mathbb{BB}_\varepsilon \text{ is preferred to close a } \cup\text{-unbalanced chain; if it does not exist, an adjacency } \Gamma_\mathbb{B} \text{ is used} \\ \text{Unbalanced chains over-represented in genome } \mathbb{B} \text{ are marked with a } \cap \\ \quad \mathbb{AA}_\varepsilon \prec \Gamma_\mathbb{A}\text{: a path } \mathbb{AA}_\varepsilon \text{ is preferred to close a } \cap\text{-unbalanced chain; if it does not exist, an adjacency } \Gamma_\mathbb{A} \text{ is used} \end{cases}$$

In order to give the correct order of linking $\begin{cases} \text{a path } \mathbb{AB}_{\mathcal{AB}} \text{ can be represented by } \mathbb{BA}_{\mathcal{BA}} \\ \text{a path } \mathbb{AB}_{\mathcal{BA}} \text{ can be represented by } \mathbb{BA}_{\mathcal{AB}} \end{cases}$

	id	sources	linking cycle		Δn	Δc	$\Delta(2\mathbb{AB})$	$\Delta\lambda$	$\Delta^{\lambda}_{\text{DCJ}}$
$\mathcal{P}$	$\mathtt{WM}$	$\mathbb{AA}_{\mathcal{AB}} + \mathbb{BB}_{\mathcal{AB}}$	$(\mathbb{AA}_{\mathcal{AB}}, \mathbb{BB}_{\mathcal{BA}})$		$+1$	$+1$	0	-2	-2
$\mathcal{Q}$	$\mathtt{WW\overline{M}\underline{M}}$	$2 \times \mathbb{AA}_{\mathcal{AB}} + \mathbb{BB}_{\mathcal{A}} + \mathbb{BB}_{\mathcal{B}}$	$(\mathbb{AA}_{\mathcal{AB}}, \mathbb{BB}_{\mathcal{B}}, \mathbb{AA}_{\mathcal{BA}}, \mathbb{BB}_{\mathcal{A}})$		$+2$	$+1$	0	-4	-3
	$\mathtt{MM\overline{W}\underline{W}}$	$2 \times \mathbb{BB}_{\mathcal{AB}} + \mathbb{AA}_{\mathcal{A}} + \mathbb{AA}_{\mathcal{B}}$	$(\mathbb{BB}_{\mathcal{AB}}, \mathbb{AA}_{\mathcal{B}}, \mathbb{BB}_{\mathcal{BA}}, \mathbb{AA}_{\mathcal{A}})$		$+2$	$+1$	0	-4	-3
$\mathcal{T}$	$\mathtt{WZ\overline{M}}$	$\mathbb{AA}_{\mathcal{AB}} + \mathbb{BB}_{\mathcal{A}} + \mathbb{AB}_{\mathcal{AB}}$	$(\mathbb{AB}_{\mathcal{AB}}, \mathbb{AA}_{\mathcal{BA}}, \mathbb{BB}_{\mathcal{A}})$		$+1.5$	$+1$	-0.5	-3	-2
	$\mathtt{WW\overline{M}}$	$2 \times \mathbb{AA}_{\mathcal{AB}} + \mathbb{BB}_{\mathcal{A}}$	$(\mathbb{AA}_{\mathcal{BA}}, \mathbb{BB}_{\mathcal{A}}, \mathbb{AA}_{\mathcal{AB}}, \mathbb{BB}_\varepsilon \prec \Gamma_\mathbb{B})$	$\cup$	$+2$	$+1$	0	-3	-2
	$\mathtt{WN\underline{M}}$	$\mathbb{AA}_{\mathcal{AB}} + \mathbb{BB}_{\mathcal{B}} + \mathbb{AB}_{\mathcal{BA}}$	$(\mathbb{AB}_{\mathcal{BA}}, \mathbb{AA}_{\mathcal{AB}}, \mathbb{BB}_{\mathcal{B}})$		$+1.5$	$+1$	-0.5	-3	-2
	$\mathtt{WW\underline{M}}$	$2 \times \mathbb{AA}_{\mathcal{AB}} + \mathbb{BB}_{\mathcal{B}}$	$(\mathbb{AA}_{\mathcal{AB}}, \mathbb{BB}_{\mathcal{A}}, \mathbb{AA}_{\mathcal{AB}}, \mathbb{BB}_\varepsilon \prec \Gamma_\mathbb{B})$	$\cup$	$+2$	$+1$	0	-3	-2
	$\mathtt{MN\overline{W}}$	$\mathbb{BB}_{\mathcal{AB}} + \mathbb{AA}_{\mathcal{A}} + \mathbb{AB}_{\mathcal{BA}}$	$(\mathbb{AB}_{\mathcal{BA}}, \mathbb{AA}_{\mathcal{A}}, \mathbb{BB}_{\mathcal{AB}})$		$+1.5$	$+1$	-0.5	-3	-2
	$\mathtt{MM\overline{W}}$	$2 \times \mathbb{BB}_{\mathcal{AB}} + \mathbb{AA}_{\mathcal{A}}$	$(\mathbb{BB}_{\mathcal{BA}}, \mathbb{AA}_{\mathcal{A}}, \mathbb{BB}_{\mathcal{AB}}, \mathbb{AA}_\varepsilon \prec \Gamma_\mathbb{A})$	$\cap$	$+2$	$+1$	0	-3	-2
	$\mathtt{MZ\underline{W}}$	$\mathbb{BB}_{\mathcal{AB}} + \mathbb{AA}_{\mathcal{B}} + \mathbb{AB}_{\mathcal{AB}}$	$(\mathbb{AB}_{\mathcal{AB}}, \mathbb{AA}_{\mathcal{B}}, \mathbb{BB}_{\mathcal{BA}})$		$+1.5$	$+1$	-0.5	-3	-2
	$\mathtt{MM\underline{W}}$	$2 \times \mathbb{BB}_{\mathcal{AB}} + \mathbb{AA}_{\mathcal{B}}$	$(\mathbb{BB}_{\mathcal{AB}}, \mathbb{AA}_{\mathcal{B}}, \mathbb{BB}_{\mathcal{BA}}, \mathbb{AA}_\varepsilon \prec \Gamma_\mathbb{A})$	$\cap$	$+2$	$+1$	0	-3	-2

	id	sources	linking cycle		Δn	Δc	$\Delta(2\mathbb{AB})$	$\Delta\lambda$	$\Delta^{\lambda}_{\text{DCJ}}$
$\mathcal{S}$	ZN	$\mathbb{AB}_{\mathcal{AB}} + \mathbb{AB}_{\mathcal{BA}}$	$(\mathbb{AB}_{\mathcal{AB}}, \mathbb{AB}_{\mathcal{BA}})$		$+1$	$+1$	-1	-2	-1
	$\overline{\texttt{WM}}$	$\mathbb{AA}_{\mathcal{A}} + \mathbb{BB}_{\mathcal{A}}$	$(\mathbb{AA}_{\mathcal{A}}, \mathbb{BB}_{\mathcal{A}})$		$+1$	$+1$	0	-1	-1
	$\underline{\texttt{WM}}$	$\mathbb{AA}_{\mathcal{B}} + \mathbb{BB}_{\mathcal{B}}$	$(\mathbb{AA}_{\mathcal{B}}, \mathbb{BB}_{\mathcal{B}})$		$+1$	$+1$	0	-1	-1
	$\texttt{W}\overline{\texttt{M}}$	$\mathbb{AA}_{\mathcal{AB}} + \mathbb{BB}_{\mathcal{A}}$	$(\mathbb{AA}_{\mathcal{BA}}, \mathbb{BB}_{\mathcal{A}})$		$+1$	$+1$	0	-1	-1
	$\texttt{W}\underline{\texttt{M}}$	$\mathbb{AA}_{\mathcal{AB}} + \mathbb{BB}_{\mathcal{B}}$	$(\mathbb{AA}_{\mathcal{AB}}, \mathbb{BB}_{\mathcal{B}})$		$+1$	$+1$	0	-1	-1
	WZ	$\mathbb{AA}_{\mathcal{AB}} + \mathbb{AB}_{\mathcal{AB}}$	$(\mathbb{AA}_{\mathcal{BA}}, \mathbb{BB}_{\varepsilon} \prec \Gamma_{\mathbb{B}}, \mathbb{AB}_{\mathcal{AB}})$	$\cup$	$+1.5$	$+1$	-0.5	-2	-1
	WN	$\mathbb{AA}_{\mathcal{AB}} + \mathbb{AB}_{\mathcal{BA}}$	$(\mathbb{AA}_{\mathcal{AB}}, \mathbb{BB}_{\varepsilon} \prec \Gamma_{\mathbb{B}}, \mathbb{AB}_{\mathcal{BA}})$	$\cup$	$+1.5$	$+1$	-0.5	-2	-1
	WW	$\mathbb{AA}_{\mathcal{AB}} + \mathbb{AA}_{\mathcal{AB}}$	$(\mathbb{AA}_{\mathcal{AB}}, \mathbb{BB}_{\varepsilon} \prec \Gamma_{\mathbb{B}}, \mathbb{AA}_{\mathcal{BA}}, \mathbb{BB}_{\varepsilon} \prec \Gamma_{\mathbb{B}})$	$\cup$	$+2$	$+1$	0	-2	-1
	$\texttt{M}\overline{\texttt{W}}$	$\mathbb{BB}_{\mathcal{AB}} + \mathbb{AA}_{\mathcal{A}}$	$(\mathbb{AA}_{\mathcal{A}}, \mathbb{BB}_{\mathcal{AB}})$		$+1$	$+1$	0	-1	-1
	$\texttt{M}\underline{\texttt{W}}$	$\mathbb{BB}_{\mathcal{AB}} + \mathbb{AA}_{\mathcal{B}}$	$(\mathbb{AA}_{\mathcal{B}}, \mathbb{BB}_{\mathcal{BA}})$		$+1$	$+1$	0	-1	-1
	MZ	$\mathbb{BB}_{\mathcal{AB}} + \mathbb{AB}_{\mathcal{AB}}$	$(\mathbb{BB}_{\mathcal{BA}}, \mathbb{AB}_{\mathcal{AB}}, \mathbb{AA}_{\varepsilon} \prec \Gamma_{\mathbb{A}})$	$\cap$	$+1.5$	$+1$	-0.5	-2	-1
	MN	$\mathbb{BB}_{\mathcal{AB}} + \mathbb{AB}_{\mathcal{BA}}$	$(\mathbb{BB}_{\mathcal{AB}}, \mathbb{AB}_{\mathcal{BA}}, \mathbb{AA}_{\varepsilon} \prec \Gamma_{\mathbb{A}})$	$\cap$	$+1.5$	$+1$	-0.5	-2	-1
	MM	$\mathbb{BB}_{\mathcal{AB}} + \mathbb{BB}_{\mathcal{AB}}$	$(\mathbb{BB}_{\mathcal{AB}}, \mathbb{AA}_{\varepsilon} \prec \Gamma_{\mathbb{A}}, \mathbb{BB}_{\mathcal{BA}}, \mathbb{AA}_{\varepsilon} \prec \Gamma_{\mathbb{A}})$	$\cap$	$+2$	$+1$	0	-2	-1
$\mathcal{M}$	$\texttt{ZZ}\underline{\texttt{W}}\overline{\texttt{M}}$	$2 \times \mathbb{AB}_{\mathcal{AB}} + \mathbb{AA}_{\mathcal{B}} + \mathbb{BB}_{\mathcal{A}}$	$(\mathbb{AB}_{\mathcal{AB}}, \mathbb{AA}_{\mathcal{B}}, \mathbb{BA}_{\mathcal{BA}}, \mathbb{BB}_{\mathcal{A}})$		$+2$	$+1$	-1	-4	-2
	$\texttt{NN}\overline{\texttt{W}}\underline{\texttt{M}}$	$2 \times \mathbb{AB}_{\mathcal{BA}} + \mathbb{AA}_{\mathcal{A}} + \mathbb{BB}_{\mathcal{B}}$	$(\mathbb{AB}_{\mathcal{BA}}, \mathbb{AA}_{\mathcal{A}}, \mathbb{BA}_{\mathcal{AB}}, \mathbb{BB}_{\mathcal{B}})$		$+2$	$+1$	-1	-4	-2
$\mathcal{N}$	$\texttt{Z}\underline{\texttt{W}}\overline{\texttt{M}}$	$\mathbb{AB}_{\mathcal{AB}} + \mathbb{AA}_{\mathcal{B}} + \mathbb{BB}_{\mathcal{A}}$	$(\mathbb{AB}_{\mathcal{AB}}, \mathbb{AA}_{\mathcal{B}}, \mathbb{BB}_{\mathcal{A}})$		$+1.5$	$+1$	-0.5	-2	-1
	$\texttt{ZZ}\underline{\texttt{W}}$	$2 \times \mathbb{AB}_{\mathcal{AB}} + \mathbb{AA}_{\mathcal{B}}$	$(\mathbb{AB}_{\mathcal{AB}}, \mathbb{AA}_{\mathcal{B}}, \mathbb{BA}_{\mathcal{BA}}, \mathbb{BB}_{\varepsilon} \prec \Gamma_{\mathbb{B}})$	$\cup$	$+2$	$+1$	-1	-3	-1
	$\texttt{ZZ}\overline{\texttt{M}}$	$2 \times \mathbb{AB}_{\mathcal{AB}} + \mathbb{BB}_{\mathcal{A}}$	$(\mathbb{BA}_{\mathcal{BA}}, \mathbb{BB}_{\mathcal{A}}, \mathbb{AB}_{\mathcal{AB}}, \mathbb{AA}_{\varepsilon} \prec \Gamma_{\mathbb{A}})$	$\cap$	$+2$	$+1$	-1	-3	-1
	$\texttt{N}\overline{\texttt{W}}\underline{\texttt{M}}$	$\mathbb{AB}_{\mathcal{BA}} + \mathbb{AA}_{\mathcal{A}} + \mathbb{BB}_{\mathcal{B}}$	$(\mathbb{AB}_{\mathcal{BA}}, \mathbb{AA}_{\mathcal{A}}, \mathbb{BB}_{\mathcal{B}})$		$+1.5$	$+1$	-0.5	-2	-1
	$\texttt{NN}\overline{\texttt{W}}$	$2 \times \mathbb{AB}_{\mathcal{BA}} + \mathbb{AA}_{\mathcal{A}}$	$(\mathbb{AB}_{\mathcal{BA}}, \mathbb{AA}_{\mathcal{A}}, \mathbb{BA}_{\mathcal{AB}}, \mathbb{BB}_{\varepsilon} \prec \Gamma_{\mathbb{B}})$	$\cup$	$+2$	$+1$	-1	-3	-1
	$\texttt{NN}\underline{\texttt{M}}$	$2 \times \mathbb{AB}_{\mathcal{BA}} + \mathbb{BB}_{\mathcal{B}}$	$(\mathbb{BA}_{\mathcal{AB}}, \mathbb{BB}_{\mathcal{B}}, \mathbb{AB}_{\mathcal{BA}}, \mathbb{AA}_{\varepsilon} \prec \Gamma_{\mathbb{A}})$	$\cap$	$+2$	$+1$	-1	-3	-1

	remaining paths	linking cycle		Δn	Δc	$\Delta(2\mathbb{AB})$	$\Delta\lambda$	$\Delta^{\lambda}_{\text{DCJ}}$
1	$\mathbb{AB}_*$	$(\mathbb{AB}_*)$		$+0.5$	$+1$	-0.5	0	0
2	$\mathbb{AA}_* + \mathbb{BB}_*$	$(\mathbb{AA}_*, \mathbb{BB}_*)$		$+1$	$+1$	0	0	0
3	$\mathbb{AA}_*$	$(\mathbb{AA}_*, \Gamma_{\mathbb{B}})$	$\cup$	$+1$	$+1$	0	0	0
4	$\mathbb{BB}_*$	$(\mathbb{BB}_*, \Gamma_{\mathbb{A}})$	$\cap$	$+1$	$+1$	0	0	0

Any capping producing linking cycles following a top-down screening of the table above is optimal:

- $\Delta^{\lambda}_{\text{DCJ}} = \Delta n - \Delta c - \Delta(2\mathbb{AB}) + \Delta\lambda$ gives the DCJ-indel-effect produced by each type of linking cycle.
- All given linking cycles have $\Delta^{\lambda}_{\text{DCJ}}$ equivalent to the respective chain of deducting recombinations, therefore they achieve the optimal DCJ-indel distance.

P1: After identifying chains of recombinations
- either there are no unbalanced chains
- or there are only $\cup$-unbalanced chains (over-repr. in $\mathbb{A}$)
- or there are only $\cap$-unbalanced chains (over-repr. in $\mathbb{B}$)

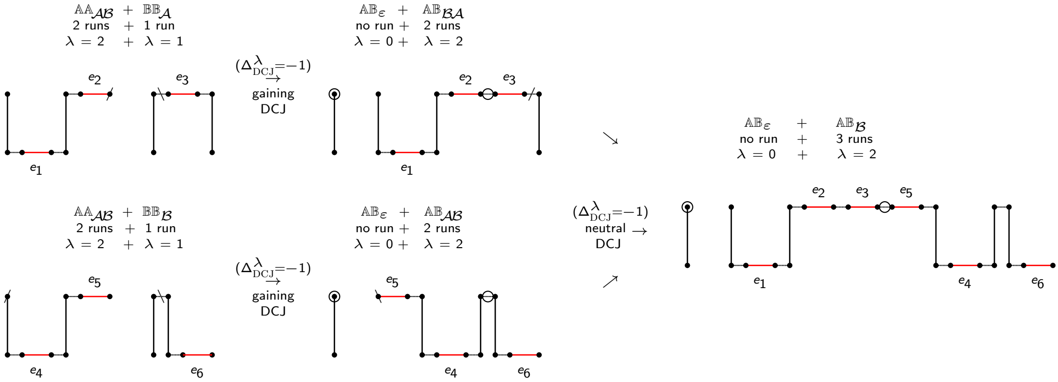
P2: When an unbalanced chain is being linked
- if there is a remaining indel-free $\mathbb{AA}_\varepsilon/\mathbb{BB}_\varepsilon$ (of the under-repr. genome), it is used to link the chain
- otherwise there is no remaining $\mathbb{AA}_*/\mathbb{BB}_*$ (of the under-repr. genome) and an adjacency $\Gamma_{\mathbb{A}/\mathbb{B}}$ links the chain

Any optimal capping that links all possible chains of deducting recombinations as described above and, for the remaining paths, maximizes the number of linking cycles of type **2** minimizes the number of caps:
- The number of caps to be added is exactly $p_* = \max\{\kappa_{\mathbb{A}}, \kappa_{\mathbb{B}}\}$.
- The number of adjacencies between caps is exactly $a_* = |\kappa_{\mathbb{A}} - \kappa_{\mathbb{B}}|$.

Capped relational graph of singular genomes - example

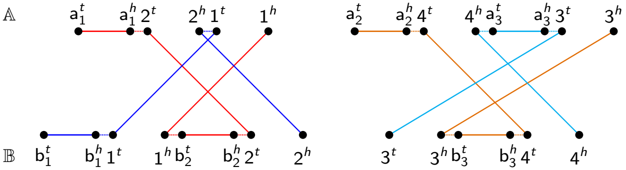
Deducting chain of path recombinations:

- transforming $2 \times \mathbb{AA}_{\mathcal{AB}} + \mathbb{BB}_{\mathcal{A}} + \mathbb{BB}_{\mathcal{B}}$
- into $3 \times \mathbb{AB}_{\varepsilon} + \mathbb{AB}_{\mathcal{B}}$
- with overall $\Delta^{\lambda}_{\mathrm{DCJ}} = -3$



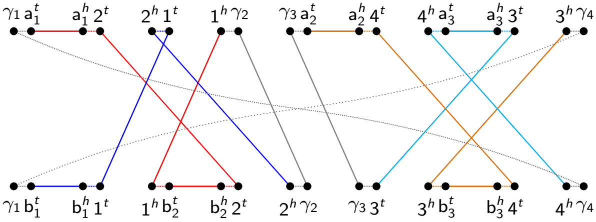
Capped relational graph of singular genomes - example

$\mathbb{A} = [\,\mathrm{a}_1 \; 2 \; 1\,] \; [\,\mathrm{a}_2 \; 4 \; \mathrm{a}_3 \; 3\,]$ and $\mathbb{B} = [\,\mathrm{b}_1 \; 1 \; \mathrm{b}_2 \; 2\,] \; [\,3 \; \mathrm{b}_3 \; 4\,]$; $p_* = 2$ and $a_* = 0$



Components: $2 \times \mathbb{AA}_{\mathcal{AB}}$, $\mathbb{BB}_{\mathcal{A}}$, $\mathbb{BB}_{\mathcal{B}}$

$$\mathrm{d}_{\mathrm{DCJ}}^{\mathrm{ID}} = n - |\mathcal{C}| - \frac{|\mathcal{P}_{\mathbb{AB}}|}{2} + \sum \lambda(C) - \delta$$
$$= 4 - 0 - 0 + 6 - 3$$
$$= 7$$



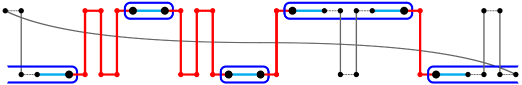
Linking cycle: $(\mathbb{AA}_{\mathcal{AB}}, \mathbb{BB}_{\mathcal{B}}, \mathbb{AA}_{\mathcal{BA}}, \mathbb{BB}_{\mathcal{A}})$

$$\mathrm{d}_{\mathrm{DCJ}}^{\mathrm{ID}} = n + p_* - |\mathcal{C}| + \sum \lambda(C)$$
$$= 4 + 2 - 1 + 2$$
$$= 7$$

The four sources of a chain of deducting recombinations are optimally linked into a single cycle.

Indel-potential via transitions

One indel-enclosing cycle:



$\aleph = \Lambda = 4$

$\Lambda(C)$ is the number of **runs** in cycle C

$\aleph(C)$ is the number of **transitions** in cycle C

Λ	$\aleph$	r	λ	
0	0	0	0	cycles
1	0	1	1	cycles and singletons
2	2	1	2	cycles
4	4	1	3	cycles
6	6	1	4	cycles
⋮	⋮	⋮	⋮	⋮

Indel-potential of a component C:

$$\lambda(C) = \begin{cases} 0 & \text{if } \Lambda(C) = 0 \ (C \text{ is indel-free}) \\ 1 & \text{if } \Lambda(C) = 1 \\ \frac{\Lambda(C)}{2} + 1 & \text{if } \Lambda(C) \geq 2 \end{cases}$$

$$\lambda(C) = \frac{\aleph(C)}{2} + r(C)$$

$$r(C) = \begin{cases} 1\,, & \text{component } C \text{ is indel-enclosing} \\ 0\,, & \text{component } C \text{ is indel-free} \end{cases}$$

Quiz 2

1 Which of the following statements about the capped relational graph are true?

A In an optimal capping, the distance computed based on the capped relational diagram must be equivalent to the distance computed based on the original relational diagram.

B Let $RG(\mathbb{A}, \mathbb{B})$ be a relational graph of **canonical** genomes.
An optimal capping of $RG(\mathbb{A}, \mathbb{B})$ that maximizes the number of cycles linking a pair $\mathbb{A}\mathbb{A} + \mathbb{B}\mathbb{B}$ has a minimum number of caps ($= \max\{\kappa_{\mathbb{A}}, \kappa_{\mathbb{B}}\}$).

C Let $\max\{\kappa_{\mathbb{A}_s}, \kappa_{\mathbb{B}_s}\} = \max\{\kappa_{\mathbb{A}_c}, \kappa_{\mathbb{B}_c}\}$.
An optimal capping of the relational graph of **singular** genomes $\mathbb{A}_s$ and $\mathbb{B}_s$ requires more caps than an optimal capping of the relational graph of **canonical** genomes $\mathbb{A}_c$ and $\mathbb{B}_c$.

D The indel-potential can be equivalently computed based on the number of runs or based on the number of transitions.

References

Double Cut and Join with Insertions and Deletions

(Marília D.V. Braga, Eyla Willing and Jens Stoye)

JCB, Vol. 18, No. 9 (2011)

Computing the Rearrangement Distance of Natural Genomes

(Leonard Bohnenkämper, Marília D. V. Braga, Daniel Doerr and Jens Stoye)

LNCS, 12074, pp 3-18 (2020)