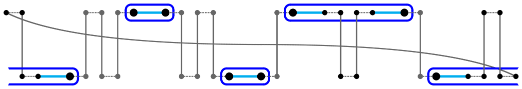
Topics of today:

Singular DCJ-indel distance and sorting:

1. Indel-potential
2. Deducting path recombinations
3. Restricted DCJ-indel model
4. The diameter of the DCJ-indel distance
5. Establishing the triangular inequality

Runs of indel-edges

One indel-enclosing cycle:



$\Lambda = 4$

indel edges , **runs**

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

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

Runs of indel-edges

Types of DCJ operation $\begin{cases} \Delta_{\mathrm{DCJ}} = 0 \textbf{ (gaining):} \text{ creates one cycle or two } \mathbb{AB}\text{-paths} \\ \Delta_{\mathrm{DCJ}} = 1 \textbf{ (neutral):} \text{ does not change the number of cycles nor of } \mathbb{AB}\text{-paths} \\ \Delta_{\mathrm{DCJ}} = 2 \textbf{ (losing):} \text{ destroys one cycle or two } \mathbb{AB}\text{-paths} \end{cases}$

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

$\Rightarrow$ Second upper bound:

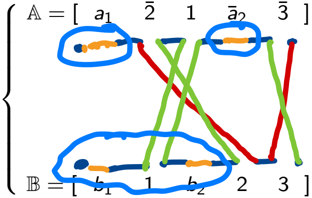
$$\mathrm{d}_{\mathrm{DCJ}}^{\mathrm{ID}}(\mathbb{A}, \mathbb{B}) \leq \underbrace{n - |\mathcal{C}| - \frac{|\mathcal{P}_{\mathbb{AB}}|}{2}}_{\text{DCJ part}} + \underbrace{\sum_{C \in RG} \Lambda(C)}_{\text{indel part}}$$

DCJ operations can modify the number of runs:

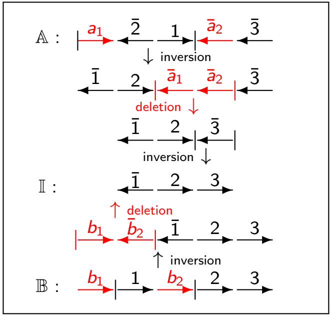
A DCJ operation can have $\begin{cases} \Delta_\Lambda = -2 & \text{(merges two pairs of runs)} \\ \Delta_\Lambda = -1 & \text{(merges one pair of runs)} \\ \Delta_\Lambda = 0 & \text{(preserves the runs)} \\ \cancel{\Delta_\Lambda = 1} & \cancel{\text{(splits one run)}} \\ \cancel{\Delta_\Lambda = 2}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

Runs can be merged and accumulated in both genomes

Singular genomes



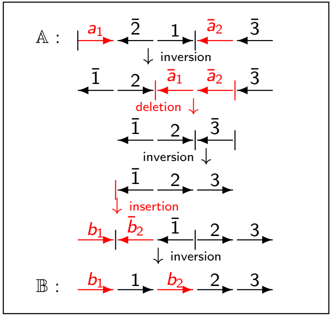
A sequence of 3 operations
sorting $\mathbb{A}$ into $\mathbb{I} = [\bar{1}\ 2\ 3]$



A sequence of 2 operations
sorting $\mathbb{B}$ into $\mathbb{I} = [\bar{1}\ 2\ 3]$

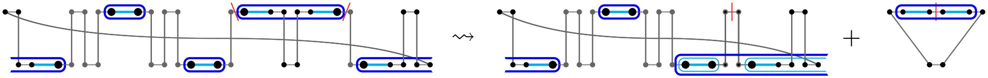
$\Rightarrow$

A sequence of 5 operations
sorting $\mathbb{A}$ into $\mathbb{B}$



Merging runs with "internal" gaining DCJ operations

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



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

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

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

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

Indel-potential λ' of a cycle C

$\Lambda(C) = 0, 1, 2, 4, 6, 8, \ldots$

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

$$\Lambda(C) = 1 \Rightarrow \lambda'(1) = 1$$

$$\Lambda(C) = 2 \Rightarrow \lambda'(2) = 2$$

$\Lambda(C) \geq 4 : \Lambda(C) = o_1 + o_2$ such that o_1 and o_2 are odd, and assume $o_1 \geq o_2$

$$\text{two resulting cycles: } \begin{cases} \text{one with } o_1 - 1 \text{ runs} \\ \text{one with either 1 run (if } o_2 = 1) \text{ or with } o_2 - 1 \text{ runs (if } o_2 \geq 3) \end{cases}$$

$$\Rightarrow \lambda'(4) = \lambda'(2) + \lambda'(1) = 2 + 1 = 3$$

$$\Rightarrow \lambda'(6) = \begin{cases} \lambda'(2) + \lambda'(2) = 2 + 2 = 4 \\ \lambda'(4) + \lambda'(1) = 3 + 1 = 4 \end{cases}$$

$$\Rightarrow \lambda'(8) = \begin{cases} \lambda'(4) + \lambda'(2) = 3 + 2 = 5 \\ \lambda'(6) + \lambda'(1) = 4 + 1 = 5 \end{cases}$$

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

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

Induction step: in general, for $\Lambda(C) \geq 4$, we can state

$$\begin{aligned} \lambda'(\Lambda(C)) &= \lambda'(\Lambda(C) - 2) + \lambda'(1) \\ &= \left(\frac{\Lambda(C) - 2}{2} + 1\right) + 1 \\ &= \frac{\Lambda(C)}{2} + 1 \end{aligned}$$

Indel-potential λ'' of a path P

$\Lambda(P) = 0, 1, 2, 3, 4, 5, 6, 7, 8, \ldots$

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

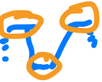
$\Lambda(P) = 1 \Rightarrow \lambda''(1) = 1$

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

$\Lambda(P) \geq 3 : \Lambda(P) = o_1 + o_2$ such that $o_1 \geq 1$ and o_2 is odd

two resulting components: $\begin{cases} \text{one path with either 1 run (if } o_1 = 1\text{) or with } o_1 - 1 \text{ runs (if } o_1 \geq 2\text{)} \\ \text{one cycle with either 1 run (if } o_2 = 1\text{) or with } o_2 - 1 \text{ runs (if } o_2 \in \{3, 5, \ldots\}\text{)} \end{cases}$

but we can get the same indel-potential if we extract all runs into a cycle:



$$\lambda''(3) = \begin{cases} \lambda''(1) + \lambda'(1) = 1+1 = 2 \\ \lambda'(2) = 2 \end{cases}$$

$$\lambda''(4) = \begin{cases} \lambda''(2) + \lambda'(1) = 2+1 = 3 \\ \lambda''(1) + \lambda'(2) = 1+2 = 3 \\ \lambda'(4) = 3 \end{cases}$$

$$\lambda''(5) = \begin{cases} \lambda''(3) + \lambda'(1) = 2+1 = 3 \\ \lambda''(1) + \lambda'(2) = 1+2 = 3 \\ \lambda'(4) = 3 \end{cases}$$

$$\lambda''(6) = \begin{cases} \ldots \\ \lambda'(6) = 4 \end{cases}$$

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 can state $\lambda''(\Lambda(P)) = \begin{cases} \lambda'(\Lambda(P)) & \text{if } \Lambda(P) \text{ is even} \\ \lambda'(\Lambda(P) - 1) & \text{if } \Lambda(P) \text{ is odd} \end{cases}$

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

Indel-potential λ of a component C

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

If C is a cycle:

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

If C is a path:

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

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
$\vdots$	$\vdots$	

In general, for any component C:

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

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

(gaining DCJ operations + indels sorting components separately)

Types of DCJ operation

DCJ-types of DCJ operation
$$\begin{cases} \Delta_{\text{DCJ}} = 0 \ \textbf{(gaining):} & \text{creates one cycle or two } \mathbb{AB}\text{-paths} \\ \Delta_{\text{DCJ}} = 1 \ \textbf{(neutral):} & \text{does not change the number of cycles nor of } \mathbb{AB}\text{-paths} \\ \Delta_{\text{DCJ}} = 2 \ \textbf{(losing):} & \text{destroys one cycle or two } \mathbb{AB}\text{-paths} \end{cases}$$

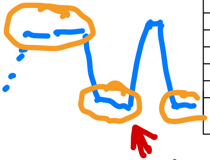
Indel-types of DCJ operation
$$\begin{cases} \Delta_\lambda = -2 & : \text{decreases the overall indel-potential by two} \\ \Delta_\lambda = -1 & : \text{decreases the overall indel-potential by one} \\ \Delta_\lambda = 0 & : \text{does not change the overall indel-potential} \\ \Delta_\lambda = \cancel{1}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

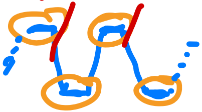
Effect of a DCJ operation ρ on the third upper bound: $\Delta^\lambda_{\text{DCJ}}(\rho) = \Delta_{\text{DCJ}}(\rho) + \Delta_\lambda(\rho)$

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
$\vdots$	$\vdots$

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

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

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

  $\Rightarrow$ Any neutral DCJ operation applied to a single cycle has $\Delta^{\lambda}_{\mathrm{DCJ}} \geq 0$
  (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:

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

Quiz 1

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

- A Any gaining DCJ operation applied to a single component has $\Delta^{\lambda}_{\mathrm{DCJ}} \geq 0$.
- B Any gaining DCJ operation has $\Delta^{\lambda}_{\mathrm{DCJ}} \geq 0$.
- C Any DCJ operation has $\Delta^{\lambda}_{\mathrm{DCJ}} \geq 0$.
- D Any DCJ that decreases the number of runs has $\Delta_{\lambda} < 0$.
- E If the input genomes are circular, we can obtain an optimal sequence of DCJ operations and indels that sort each component of the relational graph separately.

DCJ operations involving paths

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

- Any DCJ operation involving a path and a cycle is losing and has $\Delta^{\lambda}_{\mathrm{DCJ}} \geq 0$ (cannot decrease the DCJ-indel distance)
- A DCJ operation ρ applied to a single path P can be:
  - Gaining, with $\Delta^{\lambda}_{\mathrm{DCJ}}(\rho) \geq 0$ (cannot decrease the DCJ-indel distance)
  - Neutral ($\Delta_{\mathrm{DCJ}}(\rho) = 1$):

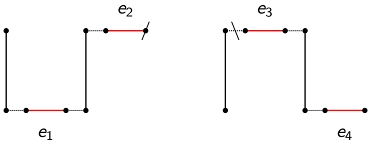
    If $\Lambda(P) \geq 4$, the DCJ ρ can merge at most two pairs of runs: $\Delta_{\Lambda}(\rho) \geq -2$ and $\Delta_{\lambda}(\rho) \geq -1$

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

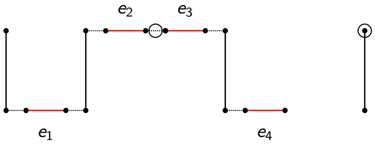
Path recombinations can have $\Delta^{\lambda}_{\text{DCJ}} \leq -1$

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

Sources	**Resultants**
$(\sum \lambda = 2 + 2 = 4)$	$(\sum \lambda = 2 + 0 = 2)$
$\mathbb{AA}$ + $\mathbb{BB}$	$\mathbb{AB}$ + $\mathbb{AB}$
2 runs + 2 runs	3 runs + no run



$\rightarrow$
gaining
DCJ



$$\mathbb{AA}_{\mathcal{BA}} + \mathbb{BB}_{\mathcal{AB}} = \begin{cases} \mathbb{AB}_{\mathcal{BAB}} + \mathbb{AB}_{\varepsilon} \\ (\mathbb{AB}_{\mathcal{ABA}} + \mathbb{AB}_{\varepsilon}) \\ (\mathbb{AB}_{\mathcal{A}} \quad + \mathbb{AB}_{\mathcal{B}}) \end{cases} \qquad \text{(all variants have } \Delta^{\lambda}_{\text{DCJ}} = -2\text{)}$$

Deducting path recombinations

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

General 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

Optimizing deducting path recombinations (for computing δ)

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

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}	$\Delta_{\text{DCJ}}^{\lambda}$
$\mathbb{AA}_{\mathcal{AB}} + \mathbb{BB}_{\mathcal{AB}}$	$\bullet + \bullet$	-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}	$\Delta_{\text{DCJ}}^{\lambda}$
$\mathbb{AA}_{\mathcal{AB}} + \mathbb{AA}_{\mathcal{AB}}$	$\mathbb{AA}_{\mathcal{A}} + \mathbb{AA}_{\mathcal{B}}$	-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

Sources:

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

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

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

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

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

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

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

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

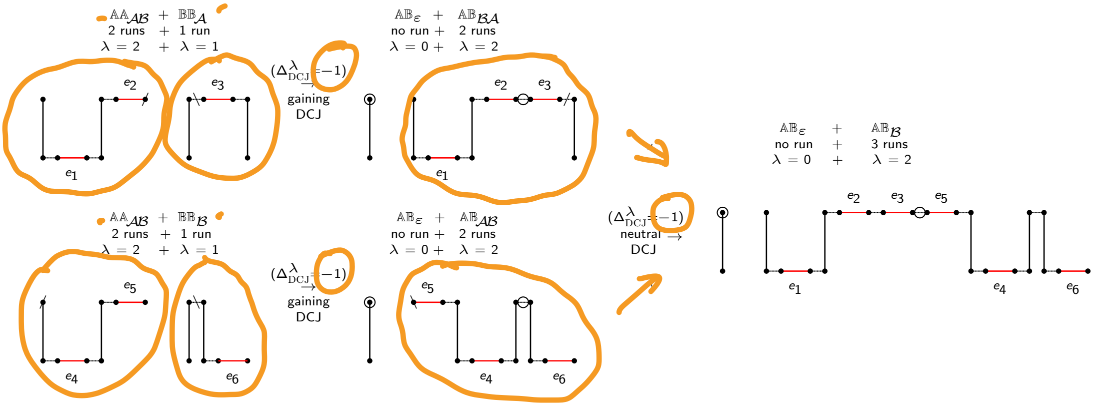
Path recombinations with $\Delta_{\text{DCJ}}^{\lambda} = 0$ creating resultants that can be used in deducting recombinations:

sources	resultants	Δ_λ	Δ_{DCJ}	$\Delta_{\text{DCJ}}^{\lambda}$
$\mathbb{AA}_{\mathcal{A}} + \mathbb{AB}_{\mathcal{BA}}$	$\bullet + \mathbb{AA}_{\mathcal{AB}}$	-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}	$\Delta_{\text{DCJ}}^{\lambda}$
$\mathbb{AA}_{\mathcal{A}} + \mathbb{BB}_{\mathcal{B}}$	$\bullet + \mathbb{AB}_{\mathcal{AB}}$	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

Optimizing deducting path recombinations (for computing δ)

Deducting chain of path recombinations $\begin{cases} \text{transforming} & 2 \times \mathbb{AA}_{\mathcal{AB}} + \mathbb{BB}_{\mathcal{A}} + \mathbb{BB}_{\mathcal{B}} \\ \text{into} & 3 \times \mathbb{AB}_{\varepsilon} + \mathbb{AB}_{\mathcal{B}} \\ \text{with} & \text{overall } \Delta^{\lambda}_{\text{DCJ}} = -3 \end{cases}$



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

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

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

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

Sorting genome $\mathbb{A}$ into genome $\mathbb{B}$ (with a minimum number of DCJs):

1. Apply all $\mathcal{P}$, $\mathcal{Q}$, $\mathcal{T}$, $\mathcal{S}$, $\mathcal{M}$ and $\mathcal{N}$ chains of deducting path recombinations, in this order.
2. For each component $C \in RG(\mathbb{A},\mathbb{B})$:
   - 2.1 Split C with **gaining** DCJs (that have $\boldsymbol{\Delta_\lambda = 0}$) until only components with at most two runs are obtained and the total number of runs in all new components is equal to $\lambda(C)$.
   - 2.2 Accumulate all runs in the smaller components derived from C with **gaining** DCJ operations (that have $\boldsymbol{\Delta_\lambda = 0}$).
   - 2.3 Apply **gaining** DCJ operations (that have $\boldsymbol{\Delta_\lambda = 0}$) in the smaller components derived from C until only DCJ-sorted components exist.
   - 2.4 **Delete** all runs in the DCJ-sorted components derived from C.

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

Singular DCJ-indel sorting: trade-off between DCJ and indels

and maximizes

The presented sorting algorithm maximizes gaining DCJs with $\Delta_\lambda = 0$ (~~minimizing~~ indels).

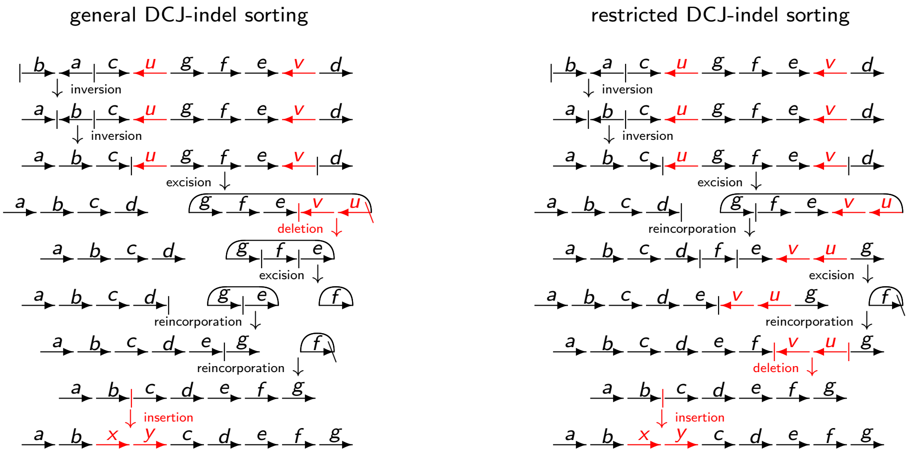
However, these gaining DCJs can often be replaced by $\begin{cases} \text{neutral DCJs with } \Delta_\lambda = -1 \\ \text{losing DCJs with } \Delta_\lambda = -2 \end{cases}$

$$\Downarrow$$

There is a big range of possibilities between the presented sorting algorithm and a sorting algorithm that minimizes gaining DCJs with $\Delta_\lambda = 0$ (~~maximizing~~ indels)

and minimizes

Restricted DCJ-indel-distance (singular linear genomes)



In any sorting sequence, it is always possible to $\begin{cases} \text{move } \textbf{deletions down} \\ \text{move } \textbf{insertions up} \end{cases}$

S : general sequence of DCJ and indel operations sorting linear $\mathbb{A}$ into linear $\mathbb{B}$

$$S \rightsquigarrow S' = S_{\text{INS}} \oplus S_{\text{DCJ}} \oplus S_{\text{DEL}} \rightsquigarrow R = S_{\text{INS}} \oplus R_{\text{DCJ}} \oplus S_{\text{DEL}} \quad \text{and} \quad |S| = |S'| = |R|$$

The diameter D^{ID}_{DCJ} of the DCJ-indel-distance

maximal subpath of C without extremity edges

For a given component C in a relational graph, let a **segment** of C be

$$\begin{cases} C \text{ itself} \quad (\text{if } C \text{ is a 0-cycle or a 0-path}) \\ \text{a minimal path flanked by two extremity-edges} \\ \text{a minimal path at the extremity of a path and connected to an extremity edge} \end{cases}$$

$s(C)$: number of segments in component C

$s(C)$	$d_{DCJ}(C)$	$\Lambda_{MAX}(C)$	$\lambda_{MAX}(C)$
1	0	1	1
2	0	2	2
3	1	3	2
4	1	4	3
5	2	5	3
6	2	6	4
7	3	7	4
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$s(C)$	$\left\lfloor \frac{s(C)-1}{2} \right\rfloor$	$s(C)$	$\left\lceil \frac{s(C)+1}{2} \right\rceil$

singletons

AB-paths / cycles

AA or BB paths

if $s(C)$ is odd:

$$d_{DCJ}(C) + \lambda_{MAX}(C) = \frac{s(C)-1}{2} + \frac{s(C)+1}{2} = s(C)$$

if $s(C)$ is even:

$$d_{DCJ}(C) + \lambda_{MAX}(C) = \frac{s(C)-2}{2} + \frac{s(C)+2}{2} = s(C)$$

Let $\begin{cases} \kappa(\mathbb{A}) : & \# \text{ linear chromosomes in } \mathbb{A} \\ \mathcal{S}(\mathbb{A}) : & \# \text{ (circular) singletons in } \mathbb{A} \\ \kappa(\mathbb{B}) : & \# \text{ linear chromosomes in } \mathbb{B} \\ \mathcal{S}(\mathbb{B}) : & \# \text{ (circular) singletons in } \mathbb{B} \end{cases}$

The number of segments in $RG(\mathbb{A}, \mathbb{B})$ is

$$s(RG(\mathbb{A}, \mathbb{B})) = 2n + \kappa(\mathbb{A}) + \mathcal{S}(\mathbb{A}) + \kappa(\mathbb{B}) + \mathcal{S}(\mathbb{B})$$

$$\begin{aligned} D^{ID}_{DCJ}(\mathbb{A}, \mathbb{B}) &= \sum_{C \in RG(\mathbb{A},\mathbb{B})} (d_{DCJ}(C) + \lambda_{MAX}(C)) \\ &= \sum_{C \in RG(\mathbb{A},\mathbb{B})} s(C) \\ &= s(RG(\mathbb{A}, \mathbb{B})) \end{aligned}$$

$$\boxed{D^{ID}_{DCJ}(\mathbb{A}, \mathbb{B}) = 2n + \kappa(\mathbb{A}) + \mathcal{S}(\mathbb{A}) + \kappa(\mathbb{B}) + \mathcal{S}(\mathbb{B})}$$

The triangular inequality does not hold for the DCJ-indel distance

Three singular genomes $\begin{cases} \mathbb{A} = [\,1\ 2\ 3\ 4\ 5\,] \\ \mathbb{B} = [\,1\ 3\ \bar{4}\ 2\ 5\,] \\ \mathbb{C} = [\,1\ 5\,] \end{cases}$.

The triangular inequality

$$d^{\mathrm{ID}}_{\mathrm{DCJ}}(\mathbb{A},\mathbb{B}) \leq d^{\mathrm{ID}}_{\mathrm{DCJ}}(\mathbb{A},\mathbb{C}) + d^{\mathrm{ID}}_{\mathrm{DCJ}}(\mathbb{B},\mathbb{C})$$

does not hold

$$\begin{cases} d^{\mathrm{ID}}_{\mathrm{DCJ}}(\mathbb{A},\mathbb{B}) = 3 \\ d^{\mathrm{ID}}_{\mathrm{DCJ}}(\mathbb{A},\mathbb{C}) = 1 \\ d^{\mathrm{ID}}_{\mathrm{DCJ}}(\mathbb{B},\mathbb{C}) = 1 \end{cases}$$

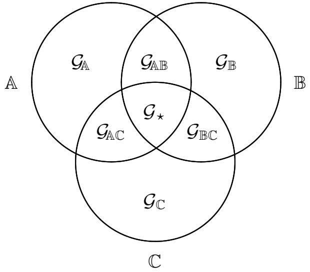
"Free lunch":
while sorting $\mathbb{A}$ into $\mathbb{C}$ and then $\mathbb{C}$ into $\mathbb{B}$,
a set of common genes of $\mathbb{A}$ and $\mathbb{B}$
are deleted and then reinserted

In the comparison of two genomes, our model prevents this problem:
common genes cannot be deleted or inserted

However, the triangular inequality is essential in other problems involving the DCJ-indel distance
for the comparison of three or more genomes (e.g. median)

Establishing the triangular inequality

Disjoint sets of genes $\mathcal{G}_{\mathbb{A}}$, $\mathcal{G}_{\mathbb{B}}$, $\mathcal{G}_{\mathbb{C}}$, $\mathcal{G}_{\mathbb{AB}}$, $\mathcal{G}_{\mathbb{BC}}$, $\mathcal{G}_{\mathbb{AC}}$ and $\mathcal{G}_{\star}$
for three genomes $\mathbb{A}, \mathbb{B}$ and $\mathbb{C}$



For each pair of genomes, we define the **corrected distance** $\mathsf{dk}_{\mathrm{DCJ}}^{\mathrm{ID}}$:

$$\mathsf{dk}_{\mathrm{DCJ}}^{\mathrm{ID}}(\mathbb{A},\mathbb{B}) = \mathsf{d}_{\mathrm{DCJ}}^{\mathrm{ID}}(\mathbb{A},\mathbb{B}) + \mathsf{k}(|\mathcal{G}_{\mathbb{A}}| + |\mathcal{G}_{\mathbb{AC}}| + |\mathcal{G}_{\mathbb{B}}| + |\mathcal{G}_{\mathbb{BC}}|)$$

$$\mathsf{dk}_{\mathrm{DCJ}}^{\mathrm{ID}}(\mathbb{A},\mathbb{C}) = \mathsf{d}_{\mathrm{DCJ}}^{\mathrm{ID}}(\mathbb{A},\mathbb{C}) + \mathsf{k}(|\mathcal{G}_{\mathbb{A}}| + |\mathcal{G}_{\mathbb{AB}}| + |\mathcal{G}_{\mathbb{C}}| + |\mathcal{G}_{\mathbb{BC}}|)$$

$$\mathsf{dk}_{\mathrm{DCJ}}^{\mathrm{ID}}(\mathbb{B},\mathbb{C}) = \mathsf{d}_{\mathrm{DCJ}}^{\mathrm{ID}}(\mathbb{B},\mathbb{C}) + \mathsf{k}(|\mathcal{G}_{\mathbb{B}}| + |\mathcal{G}_{\mathbb{AB}}| + |\mathcal{G}_{\mathbb{C}}| + |\mathcal{G}_{\mathbb{AC}}|)$$

The triangular inequality must hold for $\mathsf{dk}_{\mathrm{DCJ}}^{\mathrm{ID}}$:

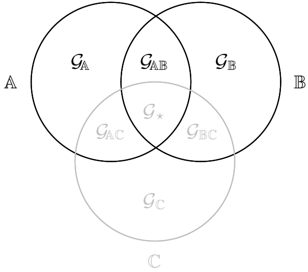
$$\mathsf{dk}_{\mathrm{DCJ}}^{\mathrm{ID}}(\mathbb{A},\mathbb{B}) \leq \mathsf{dk}_{\mathrm{DCJ}}^{\mathrm{ID}}(\mathbb{A},\mathbb{C}) + \mathsf{dk}_{\mathrm{DCJ}}^{\mathrm{ID}}(\mathbb{B},\mathbb{C})$$

$$\begin{aligned}
\mathsf{d}_{\mathrm{DCJ}}^{\mathrm{ID}}(\mathbb{A},\mathbb{B}) + \mathsf{k}(|\mathcal{G}_{\mathbb{A}}| + |\mathcal{G}_{\mathbb{AC}}| + |\mathcal{G}_{\mathbb{B}}| + |\mathcal{G}_{\mathbb{BC}}|) &\leq \mathsf{d}_{\mathrm{DCJ}}^{\mathrm{ID}}(\mathbb{A},\mathbb{C}) + \mathsf{k}(|\mathcal{G}_{\mathbb{A}}| + |\mathcal{G}_{\mathbb{AB}}| + |\mathcal{G}_{\mathbb{C}}| + |\mathcal{G}_{\mathbb{BC}}|) + \\
&\quad\ \mathsf{d}_{\mathrm{DCJ}}^{\mathrm{ID}}(\mathbb{B},\mathbb{C}) + \mathsf{k}(|\mathcal{G}_{\mathbb{B}}| + |\mathcal{G}_{\mathbb{AB}}| + |\mathcal{G}_{\mathbb{C}}| + |\mathcal{G}_{\mathbb{AC}}|)
\end{aligned}$$

$$\mathsf{d}_{\mathrm{DCJ}}^{\mathrm{ID}}(\mathbb{A},\mathbb{B}) \leq \mathsf{d}_{\mathrm{DCJ}}^{\mathrm{ID}}(\mathbb{A},\mathbb{C}) + \mathsf{k}(|\mathcal{G}_{\mathbb{AB}}| + |\mathcal{G}_{\mathbb{C}}|) + \mathsf{d}_{\mathrm{DCJ}}^{\mathrm{ID}}(\mathbb{B},\mathbb{C}) + \mathsf{k}(|\mathcal{G}_{\mathbb{AB}}| + |\mathcal{G}_{\mathbb{C}}|)$$

$$\mathsf{d}_{\mathrm{DCJ}}^{\mathrm{ID}}(\mathbb{A},\mathbb{B}) \leq \mathsf{d}_{\mathrm{DCJ}}^{\mathrm{ID}}(\mathbb{A},\mathbb{C}) + \mathsf{d}_{\mathrm{DCJ}}^{\mathrm{ID}}(\mathbb{B},\mathbb{C}) + 2\mathsf{k}(|\mathcal{G}_{\mathbb{AB}}| + |\mathcal{G}_{\mathbb{C}}|)$$

Establishing the triangular inequality



$$\begin{cases} \mathrm{d}^{\mathrm{ID}}_{\mathrm{DCJ}}(\mathbb{A},\mathbb{B}) \;\leq\; \mathrm{d}^{\mathrm{ID}}_{\mathrm{DCJ}}(\mathbb{A},\mathbb{C}) + \mathrm{d}^{\mathrm{ID}}_{\mathrm{DCJ}}(\mathbb{B},\mathbb{C}) + 2\mathrm{k}(|\mathcal{G}_{\mathbb{AB}}| + |\mathcal{G}_{\mathbb{C}}|) \\ \mathrm{d}^{\mathrm{ID}}_{\mathrm{DCJ}}(\mathbb{A},\mathbb{C}) \;\leq\; \mathrm{d}^{\mathrm{ID}}_{\mathrm{DCJ}}(\mathbb{A},\mathbb{B}) + \mathrm{d}^{\mathrm{ID}}_{\mathrm{DCJ}}(\mathbb{B},\mathbb{C}) + 2\mathrm{k}(|\mathcal{G}_{\mathbb{AC}}| + |\mathcal{G}_{\mathbb{B}}|) \\ \mathrm{d}^{\mathrm{ID}}_{\mathrm{DCJ}}(\mathbb{B},\mathbb{C}) \;\leq\; \mathrm{d}^{\mathrm{ID}}_{\mathrm{DCJ}}(\mathbb{A},\mathbb{B}) + \mathrm{d}^{\mathrm{ID}}_{\mathrm{DCJ}}(\mathbb{A},\mathbb{C}) + 2\mathrm{k}(|\mathcal{G}_{\mathbb{BC}}| + |\mathcal{G}_{\mathbb{A}}|) \end{cases}$$

Assume $\begin{cases} \mathrm{d}^{\mathrm{ID}}_{\mathrm{DCJ}}(\mathbb{A},\mathbb{B}) \geq \mathrm{d}^{\mathrm{ID}}_{\mathrm{DCJ}}(\mathbb{A},\mathbb{C}) \\ \mathrm{d}^{\mathrm{ID}}_{\mathrm{DCJ}}(\mathbb{A},\mathbb{B}) \geq \mathrm{d}^{\mathrm{ID}}_{\mathrm{DCJ}}(\mathbb{B},\mathbb{C}) \end{cases}$

Let $\begin{cases} \xi(\mathbb{A}): & \text{\# chromosomes in } \mathbb{A} \\ \kappa(\mathbb{A}): & \text{\# linear chromosomes in } \mathbb{A} \\ \mathcal{S}(\mathbb{A}): & \text{\# (circular) singletons in } \mathbb{A} \\ \xi(\mathbb{B}): & \text{\# chromosomes in } \mathbb{B} \\ \kappa(\mathbb{B}): & \text{\# linear chromosomes in } \mathbb{B} \\ \mathcal{S}(\mathbb{B}): & \text{\# (circular) singletons in } \mathbb{B} \end{cases}$

$\kappa(\mathbb{A}) + \mathcal{S}(\mathbb{A}) \leq \xi(\mathbb{A})$

and

$\kappa(\mathbb{B}) + \mathcal{S}(\mathbb{B}) \leq \xi(\mathbb{B})$

We need to find a value k that guarantees:
$\mathrm{d}^{\mathrm{ID}}_{\mathrm{DCJ}}(\mathbb{A},\mathbb{B}) \;\leq\; \mathrm{d}^{\mathrm{ID}}_{\mathrm{DCJ}}(\mathbb{A},\mathbb{C}) + \mathrm{d}^{\mathrm{ID}}_{\mathrm{DCJ}}(\mathbb{B},\mathbb{C}) + 2\mathrm{k}(|\mathcal{G}_{\mathbb{AB}}| + |\mathcal{G}_{\mathbb{C}}|)$

In the worst case genome $\mathbb{C}$ is empty:

$$\mathrm{d}^{\mathrm{ID}}_{\mathrm{DCJ}}(\mathbb{A},\mathbb{C}) = \xi(\mathbb{A}) \quad \text{and} \quad \mathrm{d}^{\mathrm{ID}}_{\mathrm{DCJ}}(\mathbb{B},\mathbb{C}) = \xi(\mathbb{B})$$

$$\mathrm{D}^{\mathrm{ID}}_{\mathrm{DCJ}}(\mathbb{A},\mathbb{B}) = 2|\mathcal{G}_{\mathbb{AB}}| + \kappa(\mathbb{A}) + \mathcal{S}(\mathbb{A}) + \kappa(\mathbb{B}) + \mathcal{S}(\mathbb{B})$$

$$\mathrm{D}^{\mathrm{ID}}_{\mathrm{DCJ}}(\mathbb{A},\mathbb{B}) \leq \xi(\mathbb{A}) + \xi(\mathbb{B}) + 2\mathrm{k}|\mathcal{G}_{\mathbb{AB}}|$$

$$\vdots$$

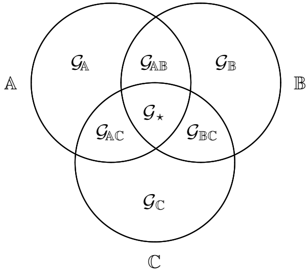
$$2|\mathcal{G}_{\mathbb{AB}}| \;\leq\; 2\mathrm{k}|\mathcal{G}_{\mathbb{AB}}| \;\Rightarrow\; \boxed{\mathrm{k} \geq 1}$$

Establishing the triangular inequality

$$\mathsf{dk}^{\mathrm{ID}}_{\mathrm{DCJ}}(\mathbb{A},\mathbb{B}) = \mathsf{d}^{\mathrm{ID}}_{\mathrm{DCJ}}(\mathbb{A},\mathbb{B}) + \mathsf{k}(|\mathcal{G}_{\mathbb{A}}| + |\mathcal{G}_{\mathbb{A}\mathbb{C}}| + |\mathcal{G}_{\mathbb{B}}| + |\mathcal{G}_{\mathbb{B}\mathbb{C}}|)$$

$$\mathsf{dk}^{\mathrm{ID}}_{\mathrm{DCJ}}(\mathbb{A},\mathbb{C}) = \mathsf{d}^{\mathrm{ID}}_{\mathrm{DCJ}}(\mathbb{A},\mathbb{C}) + \mathsf{k}(|\mathcal{G}_{\mathbb{A}}| + |\mathcal{G}_{\mathbb{A}\mathbb{B}}| + |\mathcal{G}_{\mathbb{C}}| + |\mathcal{G}_{\mathbb{B}\mathbb{C}}|)$$

$$\mathsf{dk}^{\mathrm{ID}}_{\mathrm{DCJ}}(\mathbb{B},\mathbb{C}) = \mathsf{d}^{\mathrm{ID}}_{\mathrm{DCJ}}(\mathbb{B},\mathbb{C}) + \mathsf{k}(|\mathcal{G}_{\mathbb{B}}| + |\mathcal{G}_{\mathbb{A}\mathbb{B}}| + |\mathcal{G}_{\mathbb{C}}| + |\mathcal{G}_{\mathbb{A}\mathbb{C}}|)$$



$\boxed{\text{The triangular inequality holds for the corrected distance } \mathsf{dk}^{\mathrm{ID}}_{\mathrm{DCJ}} \text{ for any } \mathsf{k} \geq 1}$

Quiz 2

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

- A A sequence of DCJ operations and indels that sort each component of the relational graph separately is always optimal.
- B 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.
- C The triangular inequality does not hold for the DCJ-indel distance, but a simple correction can be done.
- E The DCJ-indel distance can be distinct from the restricted DCJ-indel distance.

2 The best known algorithm for the restricted DCJ-indel sorting runs in...

- A $O(n)$ time.
- B $O(n \log n)$ time.
- C $O(n^2)$ time.

References

Double Cut and Join with Insertions and Deletions

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

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

Sorting Linear Genomes with Rearrangements and Indels

(Marília D. V. Braga and Jens Stoye)

TCBB, vol 12, issue 3, pp. 500-506 (2015)

On the weight of indels in genomic distances

(Marília D. V. Braga, Raphael Machado, Leonardo C. Ribeiro and Jens Stoye)

BMC Bioinformatics, vol. 12, S13 (2011)