

Topics of today:

- More about framed conserved intervals and inversion distance

- Relations $\begin{cases} \text{inversion} \times \text{DCJ distance} \\ \text{DCJ} \times \text{SCJ distance} \\ \text{inversion} \times \text{SCJ distance} \end{cases}$

- Singular DCJ-indel distance and sorting:

1. Indels: insertions and deletions
2. Relational graph of singular genomes
3. Runs and indel-potential
4. Deducing path recombinations
5. Restricted DCJ-indel model

Components are framed conserved intervals

Assuming that $\mathbb{B} = (1 \ 2 \ 3 \ \dots \ 16)$, let us identify its **framed conserved intervals** with respect to

$$\mathbb{A} = (1 \ \bar{4} \ 2 \ 3 \ 5 \ 7 \ 6 \ 8 \ \bar{16} \ \bar{14} \ \bar{15} \ \bar{13} \ \bar{11} \ \bar{12} \ \bar{10} \ 9)$$

For given $i \geq 1$ and $j \geq 1$ such that $i+j \leq n+1$:

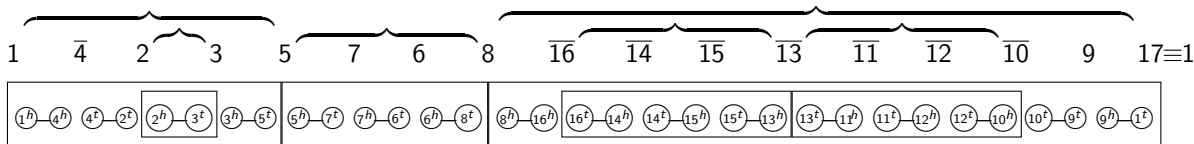
Conserved interval: interval of \mathbb{A} composed of values $i, i+1, \dots, i+j$ (assuming $n+1 \equiv 1$)

Framed conserved interval $\begin{cases} \text{direct: first element is } i \text{ and last element is } i+j; \text{ or} \\ \text{reverse: first element is } \bar{i+j} \text{ and last element is } \bar{i} \end{cases}$

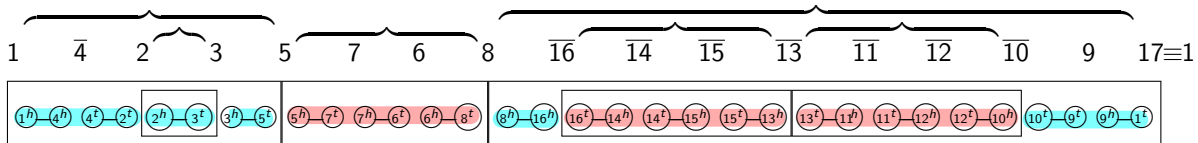
Direct: $[1..5]; [2..3]; [5..8]; [8..17]$ Reverse: $[\bar{16}..\bar{13}]; [\bar{13}..\bar{10}], [\bar{16}..\bar{10}]$

Component: framed conserved interval that is not a union of framed conserved intervals

Direct: $[1..5]; [2..3]; [5..8]; [8..17]$ Reverse: $[\bar{16}..\bar{13}]; [\bar{13}..\bar{10}]$



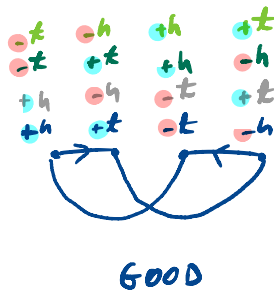
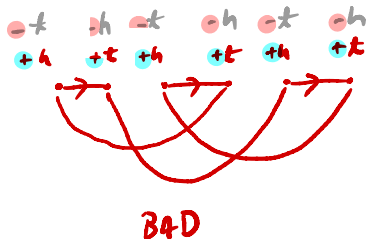
Identifying good and bad framed conserved intervals



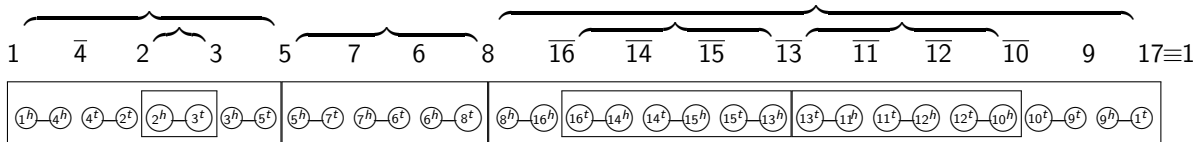
Component: framed conserved interval that is not a union of framed conserved intervals

Direct: [1..5]; [2..3]; [5..8]; [8..17]
good *trivial* *bad* *good*

Reverse: [$\overline{16}..\overline{13}$]; [$\overline{13}..\overline{10}$]
bad *bad*



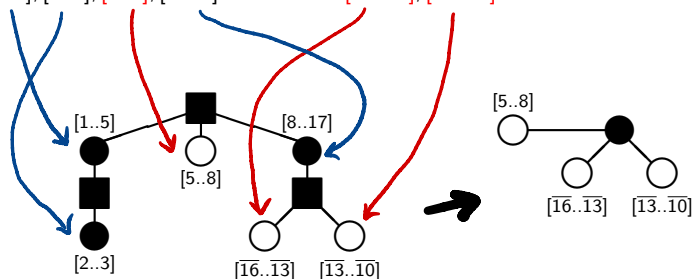
Component tree based on framed conserved intervals



Component: framed conserved interval that is not a union of framed conserved intervals

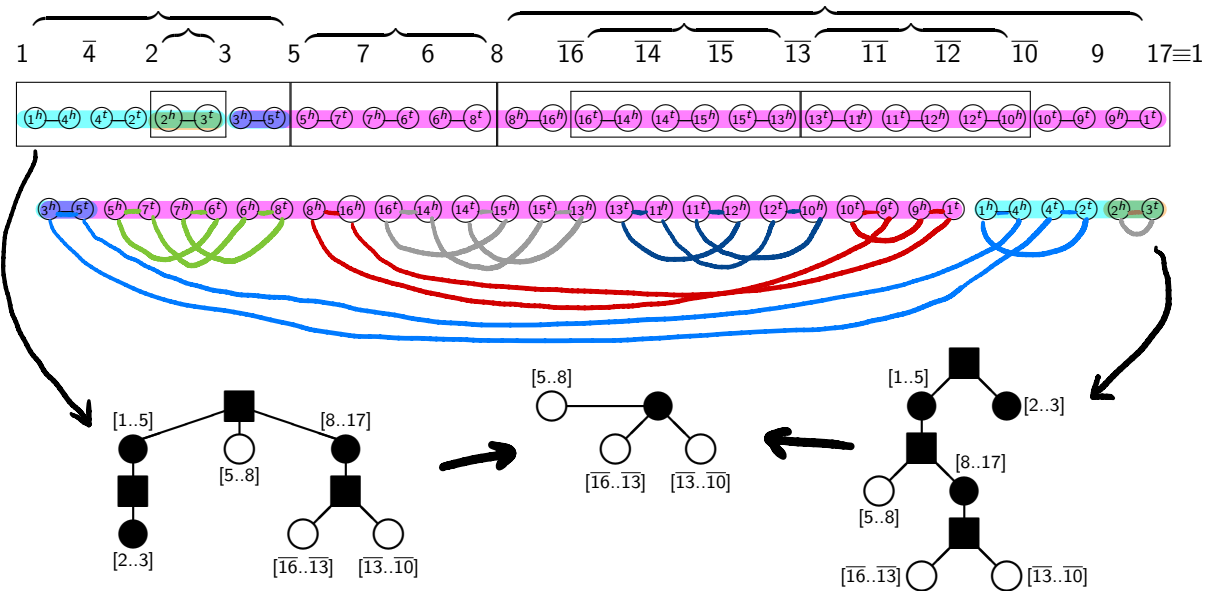
Direct: [1..5]; [2..3]; [5..8]; [8..17]

Reverse: [$\overline{16}..13$]; [$\overline{13}..10$]



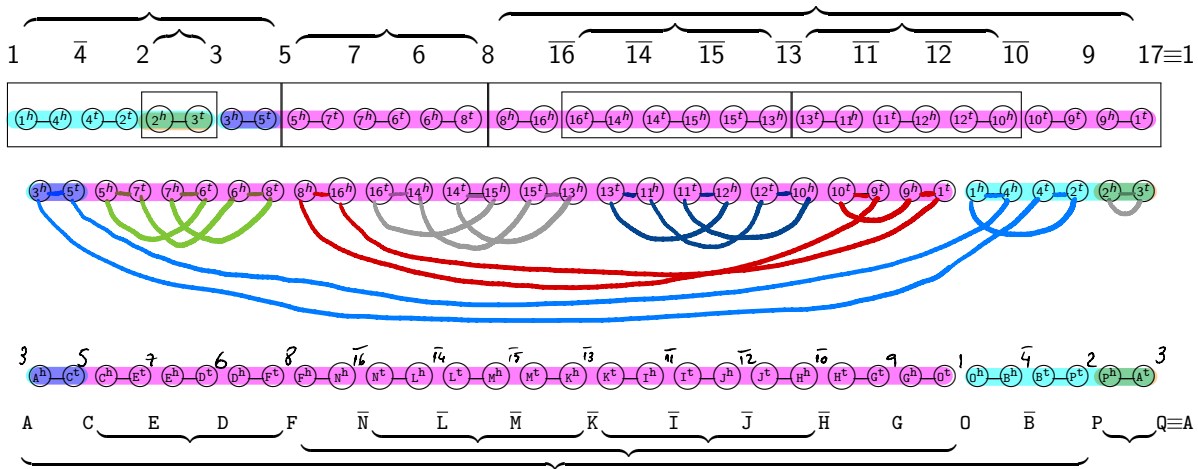
Any rotation of the diagram gives the same component tree

Components: Direct: [1..5]; [2..3]; [5..8]; [8..17] Reverse: [$\overline{16}..\overline{13}$]; [$\overline{13}..\overline{10}$]



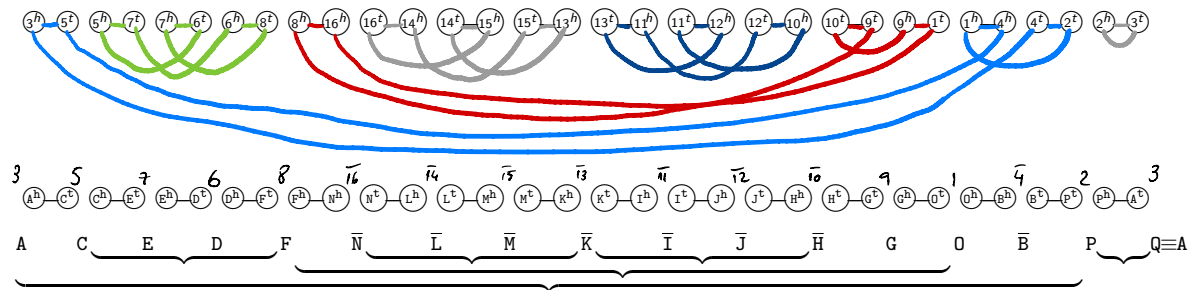
Any rotation of the diagram gives the same component tree

Components: Direct: [1..5]; [2..3]; [5..8]; [8..17] Reverse: [$\overline{16}..\overline{13}$]; [$\overline{13}..\overline{10}$]

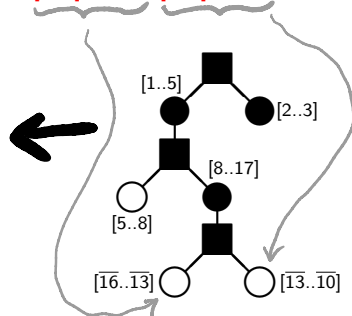
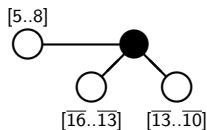


Components: Direct: $[A..P] \equiv [1..5]$; $[P..A] \equiv [2..3]$; $[C..F] \equiv [5..8]$; $[F..O] \equiv [8..17]$ Reverse: $[\bar{N}..\bar{K}] \equiv [\bar{16}..\bar{13}]$; $[\bar{K}..\bar{H}] \equiv [\bar{13}..\bar{10}]$

Any rotation of the diagram gives the same component tree



Components: Direct: $[A..P] \equiv [1..5]$; $[P..A] \equiv [2..3]$; $[C..F] \equiv [5..8]$; $[F..O] \equiv [8..17]$ Reverse: $[\bar{N}..\bar{K}] \equiv [\bar{16}..\bar{13}]$; $[\bar{K}..\bar{H}] \equiv [\bar{13}..\bar{10}]$

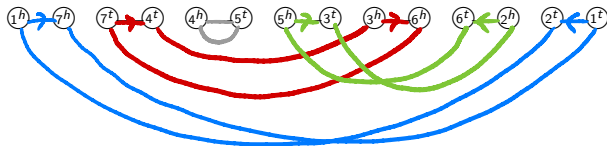


Inversion \times DCJ distance

For unichromosomal circular canonical genomes

$$\begin{cases} d_{\text{DCJ}}(\mathbb{A}, \mathbb{B}) = n - |\mathcal{C}| \\ d_{\text{INV}}(\mathbb{A}, \mathbb{B}) = d_{\text{DCJ}}(\mathbb{A}, \mathbb{B}) + \tau(\Upsilon_o) \end{cases} \quad \begin{aligned} &= n - |\mathcal{C}| + \tau(\Upsilon_o) \\ &\text{and } \tau(\Upsilon_o) = h + f \end{aligned}$$

$$\mathbb{A} = (1 \bar{7} 4 5 3 \bar{6} \bar{2})$$



$$n = |\mathcal{G}_*| = 7$$

$$|\mathcal{C}| = 4$$

$$h = 0 \text{ and } f = 0$$

$$d_{\text{INV}}(\mathbb{A}, \mathbb{B}) = d_{\text{DCJ}}(\mathbb{A}, \mathbb{B}) = n - |\mathcal{C}| = 7 - 4 = 3$$

$$\mathbb{B} = (1 2 3 4 5 6 7)$$

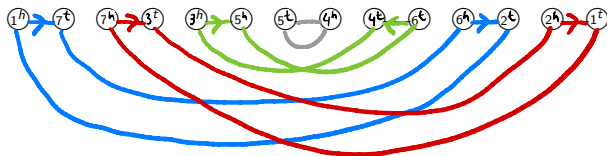
DCJ and Inversion optimal scenario

$$\left\{ \begin{array}{l} (1 \bar{7} 4 5 3 \bar{6} \bar{2}) \\ (1 \bar{7} 4 5 6 \bar{3} \bar{2}) \\ (1 \bar{7} \bar{6} \bar{5} 4 \bar{3} \bar{2}) \\ (1 \bar{2} \bar{3} \bar{4} \bar{5} \bar{6} \bar{7}) \end{array} \right\}$$

$$\left\{ \begin{array}{l} (1 \bar{7} 4 5 3 \bar{6} \bar{2}) \\ (1 \bar{7} \bar{6} \bar{2}) (4 5 3) \\ (1 \bar{7} \bar{6} \bar{5} 4 \bar{3} \bar{2}) \\ (1 \bar{2} \bar{3} \bar{4} \bar{5} \bar{6} \bar{7}) \end{array} \right\} \text{DCJ optimal scenario}$$

Inversion \times DCJ distance

$$\mathbb{A} = (1 \ 7 \ 3 \ \bar{5} \ \bar{4} \ 6 \ 2)$$



$$\mathbb{B} = (1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7)$$

$$n = |\mathcal{G}_*| = 7$$

$$|\mathcal{C}| = 4$$

$$d_{\text{DCJ}}(\mathbb{A}, \mathbb{B}) = n - |\mathcal{C}| = 7 - 4 = 3$$

$$h = 1 \text{ and } f = 0$$

$$d_{\text{INV}}(\mathbb{A}, \mathbb{B}) = n - |\mathcal{C}| + h = 7 - 4 + 1 = 4$$

inversion
optimal
scenario

- $(1 \ 7 \ 3 \ \bar{5} \ \bar{4} \ 6 \ 2)$
- $(1 \ \bar{6} \ 4 \ 5 \ \bar{3} \ \bar{7} \ 2)$
- $(1 \ \bar{6} \ 4 \ 5 \ \bar{3} \ \bar{2} \ 7)$
- $(1 \ 2 \ 3 \ \bar{5} \ \bar{4} \ 6 \ 7)$
- $(1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7)$

DCJ
optimal
scenario

- $(1 \ 7 \ 3 \ \bar{5} \ \bar{4} \ 6 \ 2)$
- $(1 \ 2) \ (7 \ 3 \ \bar{5} \ \bar{4} \ 6)$
- $(1 \ 2 \ 3 \ \bar{5} \ \bar{4} \ 6 \ 7)$
- $(1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7)$

SCJ \times DCJ distance

A DCJ can correspond to

- a single SCJ: operation creating or “destroying” a single adjacency
(fusion / fission / circularization / linearization)
- two SCJs: operations rearranging one adjacency and one telomere
- four SCJs: operations rearranging two adjacencies

$$d_{\text{DCJ}}(\mathbb{A}^f, \mathbb{B}^f) \leq d_{\text{SCJ}}(\mathbb{A}^f, \mathbb{B}^f) \leq 4 d_{\text{DCJ}}(\mathbb{A}^f, \mathbb{B}^f)$$

SCJ \times inversion distance

An inversion can correspond to $\begin{cases} \text{two SCJs: inversion at the end of a linear chromosome} \\ \text{four SCJs: inversion in the "middle" of a chromosome} \end{cases}$

$$2 d_{\text{INV}}(\mathbb{A}^f, \mathbb{B}^f) \leq d_{\text{SCJ}}(\mathbb{A}^f, \mathbb{B}^f) \leq 4 d_{\text{INV}}(\mathbb{A}^f, \mathbb{B}^f)$$

Quiz 1 - quick review

1 Which of the following statements about the inversion model are true?

☒ The inversion distance depends only on the number of cycles in the breakpoint diagram.

☒ A good component can always be sorted with (safe) split inversions.

☒ Every bad component in the diagram is a hurdle.

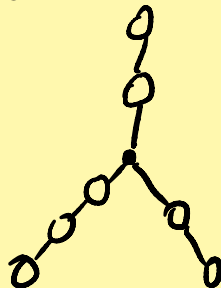
☒ If there is a bad component there is a hurdle in the diagram.

☒ If there are exactly two bad components there are two hurdles in the diagram.

☒ If there are exactly three bad components there are three hurdles in the diagram.

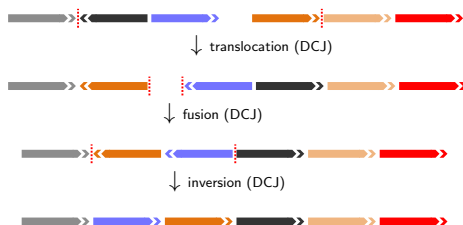
☒ A diagram needs at least six bad components to be a fortress.

☒ A diagram with an odd number of bad components can be a fortress.

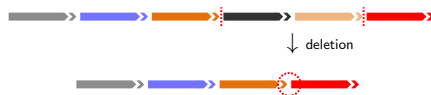


DCJ and indels

▶ DCJ: structural rearrangements



▶ Modifying the content: insertions and deletions (**indels**)



Singular DCJ-indel model

Recall that $\mathcal{G}_\star = \mathcal{G}(\mathbb{A}) \cap \mathcal{G}(\mathbb{B})$

Let $\begin{cases} \mathcal{A} = \mathcal{G}(\mathbb{A}) \setminus \mathcal{G}_\star & (\text{set of genes exclusive to genome } \mathbb{A}) \\ \mathcal{B} = \mathcal{G}(\mathbb{B}) \setminus \mathcal{G}_\star & (\text{set of genes exclusive to genome } \mathbb{B}) \end{cases}$

Restrictions for indel operations:

- ▶ At most one chromosome can be deleted or inserted at once
- ▶ Only genes of set \mathcal{A} can be deleted
- ▶ Only genes of set \mathcal{B} can be inserted

Singular DCJ-indel model

Given two singular genomes \mathbb{A} and \mathbb{B} ,...

Singular DCJ-indel Distance Problem: Compute the minimum number of DCJ and indel operations required to transform \mathbb{A} into \mathbb{B} .

Denote by $d_{\text{DCJ}}^{\text{ID}}(\mathbb{A}, \mathbb{B})$ the DCJ-indel distance of \mathbb{A} and \mathbb{B} .

Singular DCJ-indel Sorting Problem: Find a sequence of $d_{\text{DCJ}}^{\text{ID}}(\mathbb{A}, \mathbb{B})$ DCJ and indel operations that transform \mathbb{A} into \mathbb{B} .

Relational graph of singular genomes

Given two singular genomes \mathbb{A} and \mathbb{B} , their **relational graph** $RG(\mathbb{A}, \mathbb{B}) = (V, E)$ is described as follows:

1. $V = V(\xi(\mathbb{A})) \cup V(\xi(\mathbb{B}))$: there is a vertex for each extremity of each gene in \mathbb{A}
and a vertex for each extremity of each gene in \mathbb{B}

Each vertex v has a label $\ell(v)$, that corresponds to the extremity it represents.

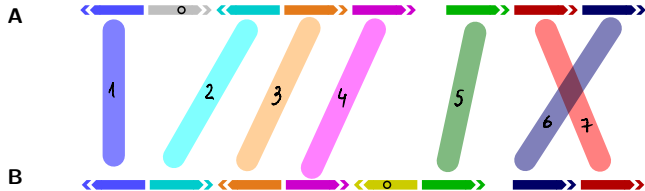
2. $E = E_{\Gamma}(\mathbb{A}) \cup E_{\Gamma}(\mathbb{B}) \cup E_{\xi} \cup E_{\text{ID}}(\mathbb{A}) \cup E_{\text{ID}}(\mathbb{B})$, where:

- ▶ **Adjacency edges:**
$$\begin{cases} E_{\Gamma}(\mathbb{A}) = \{uv : u, v \in V(\xi(\mathbb{A})) \text{ and } \ell(u)\ell(v) \in \Gamma(\mathbb{A})\} \\ E_{\Gamma}(\mathbb{B}) = \{uv : u, v \in V(\xi(\mathbb{B})) \text{ and } \ell(u)\ell(v) \in \Gamma(\mathbb{B})\} \end{cases}$$
- ▶ **Extremity edges:** $E_{\xi} = \{uv : u \in V(\xi(\mathbb{A})) \text{ and } v \in V(\xi(\mathbb{B})) \text{ and } \ell(u) = \ell(v)\}$
- ▶ **Indel edges:**
$$\begin{cases} E_{\text{ID}}(\mathbb{A}) = \{uv : \ell(u) = g^t \text{ and } \ell(v) = g^h \text{ and } g \in \mathcal{A}\} \\ E_{\text{ID}}(\mathbb{B}) = \{uv : \ell(u) = g^t \text{ and } \ell(v) = g^h \text{ and } g \in \mathcal{B}\} \end{cases}$$

Graph model for the DCJ-indel distance of *singular* genomes

Relational graph

$n = \#$ common families ($n = 7$)



Graph model for the DCJ-indel distance of *singular* genomes

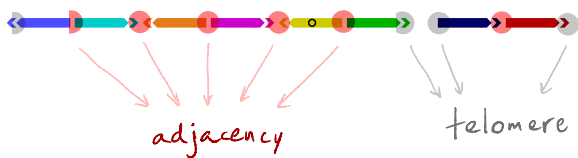
Relational graph

$n = \#$ common families ($n = 7$)

A



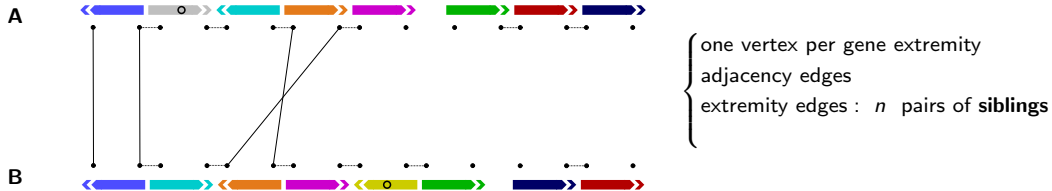
B



Graph model for the DCJ-indel distance of *singular* genomes

Relational graph

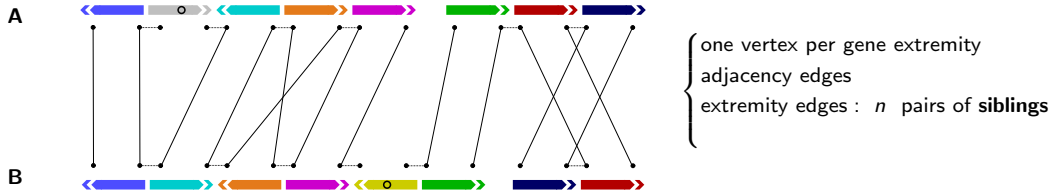
$n = \#$ common families ($n = 7$)



Graph model for the DCJ-indel distance of *singular* genomes

Relational graph

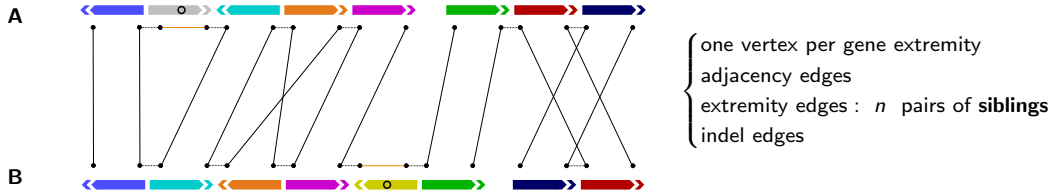
$n = \#$ common families ($n = 7$)



Graph model for the DCJ-indel distance of *singular* genomes

Relational graph

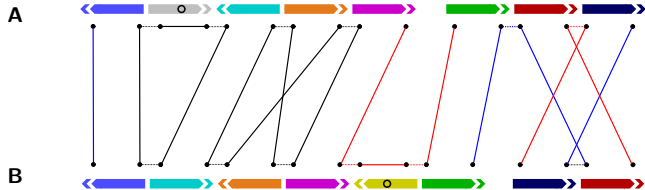
$n = \#$ common families ($n = 7$)



Graph model for the DCJ-indel distance of *singular* genomes

Relational graph

$n = \#$ common families ($n = 7$)



each vertex has degree 1 or 2:

connected components:

cycles, AB-paths, AA- and BB-paths

$\mathcal{S} \cup \mathcal{C}$ \mathcal{P}_{AB} \mathcal{P}_{AA} \mathcal{P}_{BB}

0-cycles (≥ 2)-cycles

$\left\{ \begin{array}{l} \text{indel-enclosing} \\ \text{indel-free} \end{array} \right.$

$|\mathcal{P}_{AB}|$ is even

length of a component: $\#$ of extremity edges

$\left\{ \begin{array}{l} \text{cycles, AA- and BB-paths have even length} \\ \text{AB-paths have odd length} \end{array} \right.$



\emptyset -cycle

\emptyset -AA-path

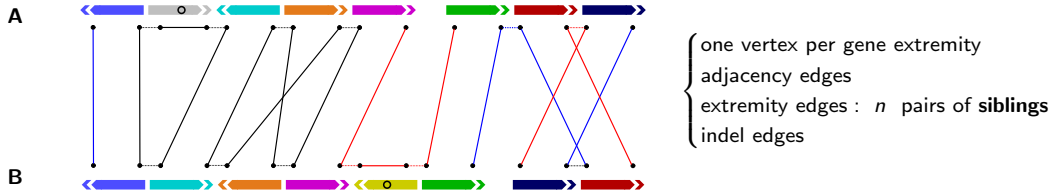
$\left\{ \begin{array}{l} \text{one vertex per gene extremity} \\ \text{adjacency edges} \\ \text{extremity edges : } n \text{ pairs of siblings} \\ \text{indel edges} \end{array} \right.$

Graph model for the DCJ-indel distance of *singular* genomes

Relational graph

$n = \#$ common families ($n = 7$)

$$A \stackrel{\text{dcj}}{\neq} B : n > |\mathcal{C}| + \frac{|\mathcal{P}_{AB}|}{2}$$



each vertex has degree 1 or 2:

connected components:

cycles, **AB-paths**, **AA-** and **BB-paths**

$\mathcal{S} \cup \mathcal{C}$ \mathcal{P}_{AB} \mathcal{P}_{AA} \mathcal{P}_{BB}

0-cycles (≥ 2) -cycles

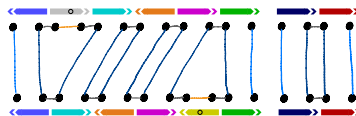
{ indel-enclosing
 indel-free

$|\mathcal{P}_{AB}|$ is even

length of a component: $\#$ of extremity edges

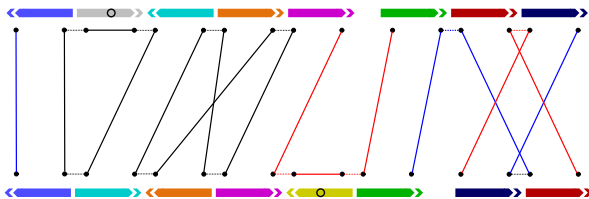
{ cycles, **AA-** and **BB-paths** have even length
AB-paths have odd length

$A \stackrel{\text{dcj}}{=} B$: 2-cycles, 1-paths and { 0-cycles
 0-paths



$$2n = 2|\mathcal{C}| + |\mathcal{P}_{AB}| \Rightarrow n = |\mathcal{C}| + \frac{|\mathcal{P}_{AB}|}{2}$$

Relational graph of singular genomes



components can be **indel-inclosing** or **indel-free**

Every vertex has degree one or two:

$RG(\mathbb{A}, \mathbb{B})$ is a collection of paths and cycles

cycle with k edges in E_ξ : k -cycle or c_k

path with k edges in E_ξ : k -path or p_k

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

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

$|\mathcal{P}_{\mathbb{A}\mathbb{B}}|$ is even (E_ξ has $2n$ edges)

$|\mathcal{P}_{\mathbb{A}\mathbb{A}}| + |\mathcal{P}_{\mathbb{B}\mathbb{B}}| + |\mathcal{P}_{\mathbb{A}\mathbb{B}}| = \kappa(\mathbb{A}) + \kappa(\mathbb{B})$

If $\mathbb{A}_c^f = \mathbb{B}_c^f$,

$RG(\mathbb{A}, \mathbb{B})$ has only 2-cycles and 1-paths:

$$2n = 2|\mathcal{C}| + |\mathcal{P}_{\mathbb{A}\mathbb{B}}| \Rightarrow n = |\mathcal{C}| + \frac{|\mathcal{P}_{\mathbb{A}\mathbb{B}}|}{2}$$

Otherwise, if $\mathbb{A}_c^f \neq \mathbb{B}_c^f$:

$$n > |\mathcal{C}| + \frac{|\mathcal{P}_{\mathbb{A}\mathbb{B}}|}{2}$$

Singular DCJ-indel model

First upper bound:

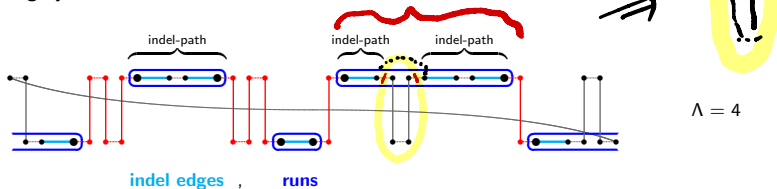
$$d_{\text{DCJ}}^{\text{ID}}(\mathbb{A}, \mathbb{B}) \leq d_{\text{DCJ}}(\mathbb{A}_c^f, \mathbb{B}_c^f) + |\mathcal{A}| + |\mathcal{B}|$$

where $\begin{cases} \mathbb{A}_c^f \text{ is the genome obtained from } \mathbb{A} \text{ by simply removing the genes of } \mathcal{A} \\ \mathbb{B}_c^f \text{ is the genome obtained from } \mathbb{B} \text{ by simply removing the genes of } \mathcal{B} \end{cases}$

Types of DCJ operation $\left\{ \begin{array}{l} \Delta_{\text{DCJ}} = 0 \text{ (**gaining**): creates one cycle or two AB-paths} \\ \Delta_{\text{DCJ}} = 1 \text{ (**neutral**): does not change the number of cycles nor of AB-paths} \\ \Delta_{\text{DCJ}} = 2 \text{ (**losing**): destroys one cycle or two AB-paths} \end{array} \right.$

Runs of indel-paths

One indel-enclosing cycle:



Each **run** can be inserted/deleted at once

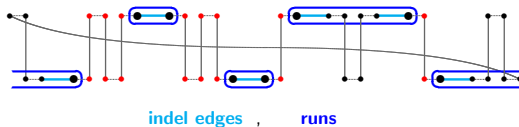
Each **indel-path** can be inserted/deleted at once

Each **run** can be accumulated into one **indel-path** with *gaining* DCJs

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

Runs of indel-paths

One indel-enclosing cycle:



$$\Lambda = 4$$

$\Lambda(C)$ 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
\vdots	\vdots
\vdots	\vdots

Each **run** can be inserted/deleted at once

\Rightarrow Second upper bound:

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

Runs of indel-paths

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

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

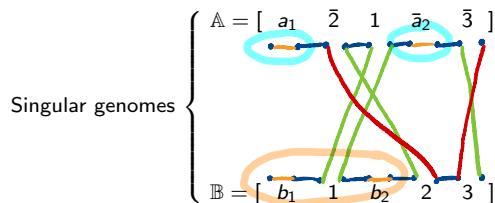
\Rightarrow Second upper bound:

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

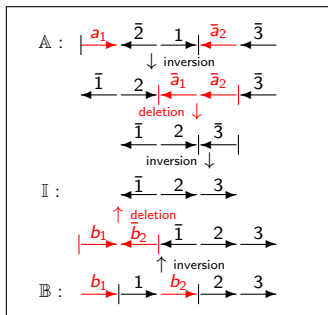
DCJ operations can modify the number of runs:

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

Runs can be merged and accumulated in both genomes

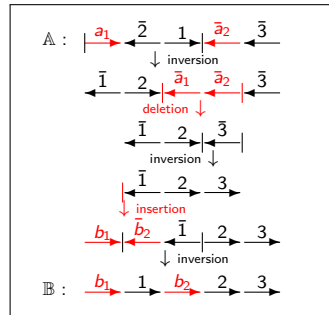


A sequence of 3 operations
sorting \mathbb{A} into $\mathbb{I} = [\bar{1} \ 2 \ 3]$



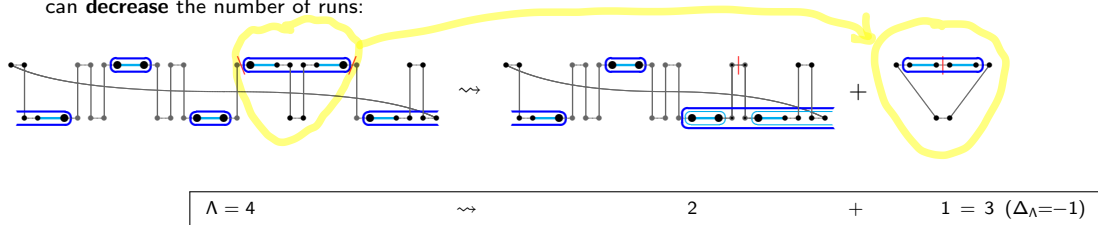
A sequence of 2 operations
sorting \mathbb{B} into $\mathbb{I} = [\bar{1} \ 2 \ 3]$

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:



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

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

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

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

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

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

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
...	...
...	...

Handwritten red annotations: A bracket on the left groups the rows for $\Lambda=4, 6, 8$ with a '+2' next to it. A bracket on the right groups the rows for $\Lambda=4, 6, 8$ with a '+1' next to it.

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 $\lambda(\Lambda(C)) = \lambda(\Lambda(C) - 2) + \lambda(1)$

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

Indel-potential λ of a path P

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

$$\text{If } \Lambda(P) = 1 \text{ is odd} \Rightarrow \lambda(P) = 1$$

$$\text{If } \Lambda(P) \geq 2 \text{ is even} \Rightarrow \lambda(P) = \frac{\Lambda(P)}{2} + 1$$

$$\text{If } \Lambda(P) \geq 3 \text{ is odd} \Rightarrow \lambda(P) = \lambda(\Lambda(P) - 1)$$

$$\text{In general, for } \Lambda(P) \geq 1, \text{ we can state } \lambda(\Lambda(P)) = \left\lceil \frac{\Lambda(P) + 1}{2} \right\rceil$$

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

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:

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

In general, for any component C :

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

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

paths and cycles

paths, cycles and singletons

paths and cycles

paths

paths and cycles

paths

paths and cycles

paths

Third upper bound:

$$d_{\text{DCJ}}^{\text{ID}}(\mathbb{A}, \mathbb{B}) \leq n - |\mathcal{C}| - \frac{|\mathcal{P}_{\mathbb{A}\mathbb{B}}|}{2} + \sum_{C \in \mathcal{R}\mathcal{G}} \lambda(C)$$

(gaining DCJ operations + indels sorting components separately)

Types of DCJ operation

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

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

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

DCJ Operations that can decrease the third upper bound: $\left\{ \begin{array}{l} \Delta_{\text{DCJ}} = 0 \text{ (gaining) and } \Delta_{\lambda} = -2 : \Delta_{\text{DCJ}}^{\lambda} = -2 \\ \Delta_{\text{DCJ}} = 0 \text{ (gaining) and } \Delta_{\lambda} = -1 : \Delta_{\text{DCJ}}^{\lambda} = -1 \\ \Delta_{\text{DCJ}} = 1 \text{ (neutral) and } \Delta_{\lambda} = -2 : \Delta_{\text{DCJ}}^{\lambda} = -1 \end{array} \right.$

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

DCJ operations involving cycles

Λ	λ
0	0
1	1
2	2
4	3
6	4
8	5
.	.
.	.
.	.

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

- ▶ A DCJ operation ρ applied to a single cycle C can be:

- ▶ Gaining, with $\Delta_{\text{DCJ}}^{\lambda}(\rho) \geq 0$ (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 neutral DCJ operation applied to a single cycle has $\Delta_{\text{DCJ}}^{\lambda} \geq 0$ (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:

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

Quiz 2

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

- ☒ A Any gaining DCJ operation applied to a single component has $\Delta_{\text{DCJ}}^{\lambda} \geq 0$.
- ☐ B Any gaining DCJ operation has $\Delta_{\text{DCJ}}^{\lambda} \geq 0$.
- ☐ C Any DCJ operation has $\Delta_{\text{DCJ}}^{\lambda} \geq 0$.
- ☐ 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.

$$\Delta_{\lambda} < 0$$

DCJ operations involving paths

- ▶ Any DCJ operation involving a path and a cycle is losing and has $\Delta_{\text{DCJ}}^{\lambda} \geq 0$ (cannot decrease the DCJ-indel distance)

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

- ▶ A DCJ operation ρ applied to a single path P can be:
 - ▶ Gaining, with $\Delta_{\text{DCJ}}^{\lambda}(\rho) \geq 0$ (cannot decrease the DCJ-indel distance)
 - ▶ Neutral ($\Delta_{\text{DCJ}}(\rho) = 1$):
 - 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 $\Delta_{\text{DCJ}}^{\lambda} \geq 0$ (cannot decrease the DCJ-indel distance)

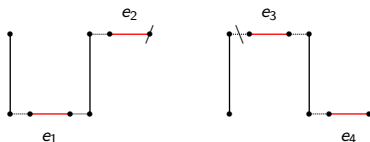
Path recombinations can have $\Delta_{\text{DCJ}}^{\lambda} \leq -1$

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

Sources

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

AA	+	BB
2 runs	+	2 runs

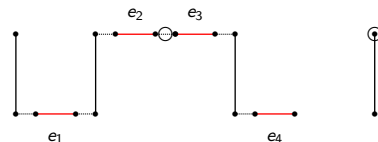


→
gaining
DCJ

Resultants

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

AB	+	AB
3 runs	+	no run



$$AA_{BA} + BB_{AB} = \begin{cases} AB_{BAB} + AB_{\epsilon} \\ (AB_{ABA} + AB_{\epsilon}) \\ (AB_A + AB_B) \end{cases} \quad (\text{all variants have } \Delta_{\text{DCJ}}^{\lambda} = -2)$$

Deducting path recombinations

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

General DCJ-indel distance formula:

$$d_{\text{DCJ}}^{\text{ID}}(A, B) = n - |C| - \frac{|P_{AB}|}{2} + \sum_{C \in RG} \lambda(C) - \delta,$$

where δ is the value obtained by optimizing deducting path recombinations

Optimizing deducing path recombinations (for computing δ)

Run-type of a path

$$\left\{ \begin{array}{ll} \varepsilon & \equiv \varepsilon \text{ (empty)} \\ \textcolor{teal}{A}B\textcolor{red}{A}B \dots \textcolor{teal}{A} & \equiv \textcolor{teal}{A} \text{ (odd)} \\ \textcolor{red}{B}A\textcolor{red}{B}A \dots \textcolor{red}{B} & \equiv \textcolor{red}{B} \text{ (odd)} \\ \textcolor{teal}{A}B\textcolor{red}{A}B \dots \textcolor{teal}{A}B & \equiv \textcolor{teal}{A}B \text{ (even)} \\ \textcolor{red}{B}A\textcolor{red}{B}A \dots \textcolor{red}{B}A & \equiv \textcolor{red}{B}A \text{ (even)} \end{array} \right.$$

Path types

$$\left\{ \begin{array}{l} AA_\varepsilon, AA_A, AA_B, AA_{AB} (\equiv AA_{BA}) \\ BB_\varepsilon, BB_A, BB_B, BB_{AB} (\equiv BB_{BA}) \\ \textcolor{red}{AB}_\varepsilon, \textcolor{red}{AB}_A, \textcolor{red}{AB}_B, \textcolor{red}{AB}_{AB}, \textcolor{red}{AB}_{BA} \end{array} \right.$$

\Rightarrow an $\textcolor{red}{AB}$ -path is always read from A to B

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

sources	resultants	Δ_λ	Δ_{DCJ}	Δ_{DCJ}^λ
$AA_{AB} + BB_{AB}$	$\bullet + \bullet$	-2	0	-2
$AA_{AB} + BB_A$	$\bullet + \textcolor{red}{AB}_{BA}$	-1	0	-1
$AA_{AB} + BB_B$	$\bullet + \textcolor{red}{AB}_{AB}$	-1	0	-1
$AA_A + BB_{AB}$	$\bullet + \textcolor{red}{AB}_{AB}$	-1	0	-1
$AA_B + BB_{AB}$	$\bullet + \textcolor{red}{AB}_{BA}$	-1	0	-1
$AA_A + BB_A$	$\bullet + \bullet$	-1	0	-1
$AA_B + BB_B$	$\bullet + \bullet$	-1	0	-1

sources	resultants	Δ_λ	Δ_{DCJ}	Δ_{DCJ}^λ
$AA_{AB} + AA_{AB}$	$AA_A + AA_B$	-2	+1	-1
$BB_{AB} + BB_{AB}$	$BB_A + BB_B$	-2	+1	-1
$AA_{AB} + \textcolor{red}{AB}_{AB}$	$\bullet + AA_A$	-2	+1	-1
$AA_{AB} + \textcolor{red}{AB}_{BA}$	$\bullet + AA_B$	-2	+1	-1
$BB_{AB} + \textcolor{red}{AB}_{AB}$	$\bullet + BB_B$	-2	+1	-1
$BB_{AB} + \textcolor{red}{AB}_{BA}$	$\bullet + BB_A$	-2	+1	-1
$\textcolor{red}{AB}_{AB} + \textcolor{red}{AB}_{BA}$	$\bullet + \bullet$	-2	+1	-1

Sources:

$\bar{W} : AA_{AB}$

$\bar{W} : AA_A$

$\bar{W} : AA_B$

$\bar{M} : BB_{AB}$

$\bar{M} : BB_A$

$\bar{M} : BB_B$

$Z : \textcolor{red}{AB}_{AB}$

$N : \textcolor{red}{AB}_{BA}$

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

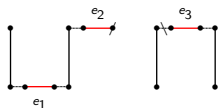
sources	resultants	Δ_λ	Δ_{DCJ}	Δ_{DCJ}^λ
$AA_A + \textcolor{red}{AB}_{BA}$	$\bullet + AA_{AB}$	-1	+1	0
$AA_B + \textcolor{red}{AB}_{AB}$	$\bullet + AA_{AB}$	-1	+1	0
$BB_A + \textcolor{red}{AB}_{AB}$	$\bullet + BB_{AB}$	-1	+1	0
$BB_B + \textcolor{red}{AB}_{BA}$	$\bullet + BB_{AB}$	-1	+1	0

sources	resultants	Δ_λ	Δ_{DCJ}	Δ_{DCJ}^λ
$AA_A + BB_B$	$\bullet + \textcolor{red}{AB}_{AB}$	0	0	0
$AA_B + BB_A$	$\bullet + \textcolor{red}{AB}_{BA}$	0	0	0
$\textcolor{red}{AB}_{AB} + \textcolor{red}{AB}_{AB}$	$AA_A + BB_B$	-2	+2	0
$\textcolor{red}{AB}_{BA} + \textcolor{red}{AB}_{BA}$	$AA_B + BB_A$	-2	+2	0

Optimizing deducing path recombinations (for computing δ)

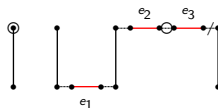
Deducing chain of path recombinations $\left\{ \begin{array}{ll} \text{transforming} & 2 \times \mathbb{A}\mathbb{A}_{\mathbb{A}\mathbb{B}} + \mathbb{B}\mathbb{B}_{\mathbb{A}} + \mathbb{B}\mathbb{B}_{\mathbb{B}} \\ \text{into} & 3 \times \mathbb{A}\mathbb{B}_{\varepsilon} + \mathbb{A}\mathbb{B}_{\mathbb{B}} \\ \text{with} & \text{overall } \Delta_{\text{DCJ}}^{\lambda} = -3 \end{array} \right.$

$\mathbb{A}\mathbb{A}_{\mathbb{A}\mathbb{B}} + \mathbb{B}\mathbb{B}_{\mathbb{A}}$
2 runs + 1 run
 $\lambda = 2 + \lambda = 1$

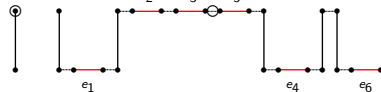


$(\Delta_{\text{DCJ}}^{\lambda} = -1)$
gaining
DCJ

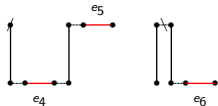
$\mathbb{A}\mathbb{B}_{\varepsilon} + \mathbb{A}\mathbb{B}_{\mathbb{B}\mathbb{A}}$
no run + 2 runs
 $\lambda = 0 + \lambda = 2$



$\mathbb{A}\mathbb{B}_{\varepsilon} + \mathbb{A}\mathbb{B}_{\mathbb{B}}$
no run + 3 runs
 $\lambda = 0 + \lambda = 2$

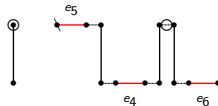


$\mathbb{A}\mathbb{A}_{\mathbb{A}\mathbb{B}} + \mathbb{B}\mathbb{B}_{\mathbb{B}}$
2 runs + 1 run
 $\lambda = 2 + \lambda = 1$



$(\Delta_{\text{DCJ}}^{\lambda} = -1)$
gaining
DCJ

$\mathbb{A}\mathbb{B}_{\varepsilon} + \mathbb{A}\mathbb{B}_{\mathbb{A}\mathbb{B}}$
no run + 2 runs
 $\lambda = 0 + \lambda = 2$



$(\Delta_{\text{DCJ}}^{\lambda} = -1)$
neutral \rightarrow
DCJ



id		sources			resultants			$\Delta_{\text{DCJ}}^\lambda$	scr
\mathcal{P}	WM	$\mathbb{A}\mathbb{A}_{\mathcal{AB}}$	$\mathbb{B}\mathbb{B}_{\mathcal{AB}}$	—	—	—	$2 \times \bullet$	-2	-1
\mathcal{Q}	WWM	$2 \times \mathbb{A}\mathbb{A}_{\mathcal{AB}}$	$\mathbb{B}\mathbb{B}_{\mathcal{A}} + \mathbb{B}\mathbb{B}_{\mathcal{B}}$	—	—	—	$4 \times \bullet$	-3	-3/4
	MMW	$\mathbb{A}\mathbb{A}_{\mathcal{A}} + \mathbb{A}\mathbb{A}_{\mathcal{B}}$	$2 \times \mathbb{B}\mathbb{B}_{\mathcal{AB}}$	—	—	—	$4 \times \bullet$	-3	-3/4
\mathcal{T}	WZM	$\mathbb{A}\mathbb{A}_{\mathcal{AB}}$	$\mathbb{B}\mathbb{B}_{\mathcal{A}}$	$\mathbb{A}\mathbb{B}_{\mathcal{AB}}$	—	—	$3 \times \bullet$	-2	-2/3
	WWM	$2 \times \mathbb{A}\mathbb{A}_{\mathcal{AB}}$	$\mathbb{B}\mathbb{B}_{\mathcal{A}}$	—	$\mathbb{A}\mathbb{A}_{\mathcal{B}}$	—	$2 \times \bullet$	-2	-2/3
	WNM	$\mathbb{A}\mathbb{A}_{\mathcal{AB}}$	$\mathbb{B}\mathbb{B}_{\mathcal{B}}$	$\mathbb{A}\mathbb{B}_{\mathcal{BA}}$	—	—	$3 \times \bullet$	-2	-2/3
	WWM	$2 \times \mathbb{A}\mathbb{A}_{\mathcal{AB}}$	$\mathbb{B}\mathbb{B}_{\mathcal{B}}$	—	$\mathbb{A}\mathbb{A}_{\mathcal{A}}$	—	$2 \times \bullet$	-2	-2/3
	MNW	$\mathbb{A}\mathbb{A}_{\mathcal{A}}$	$\mathbb{B}\mathbb{B}_{\mathcal{AB}}$	$\mathbb{A}\mathbb{B}_{\mathcal{BA}}$	—	—	$3 \times \bullet$	-2	-2/3
	MMW	$\mathbb{A}\mathbb{A}_{\mathcal{A}}$	$2 \times \mathbb{B}\mathbb{B}_{\mathcal{AB}}$	—	—	$\mathbb{B}\mathbb{B}_{\mathcal{B}}$	$2 \times \bullet$	-2	-2/3
	MZW	$\mathbb{A}\mathbb{A}_{\mathcal{B}}$	$\mathbb{B}\mathbb{B}_{\mathcal{AB}}$	$\mathbb{A}\mathbb{B}_{\mathcal{AB}}$	—	—	$3 \times \bullet$	-2	-2/3
	MMW	$\mathbb{A}\mathbb{A}_{\mathcal{B}}$	$2 \times \mathbb{B}\mathbb{B}_{\mathcal{AB}}$	—	—	$\mathbb{B}\mathbb{B}_{\mathcal{A}}$	$2 \times \bullet$	-2	-2/3
\mathcal{S}	ZN	—	—	$\mathbb{A}\mathbb{B}_{\mathcal{AB}} + \mathbb{A}\mathbb{B}_{\mathcal{BA}}$	—	—	$2 \times \bullet$	-1	-1/2
	WM	$\mathbb{A}\mathbb{A}_{\mathcal{A}}$	$\mathbb{B}\mathbb{B}_{\mathcal{A}}$	—	—	—	$2 \times \bullet$	-1	-1/2
	WM	$\mathbb{A}\mathbb{A}_{\mathcal{B}}$	$\mathbb{B}\mathbb{B}_{\mathcal{B}}$	—	—	—	$2 \times \bullet$	-1	-1/2
	WM	$\mathbb{A}\mathbb{A}_{\mathcal{AB}}$	$\mathbb{B}\mathbb{B}_{\mathcal{A}}$	—	—	$\mathbb{A}\mathbb{B}_{\mathcal{BA}}$	\bullet	-1	-1/2
	WM	$\mathbb{A}\mathbb{A}_{\mathcal{AB}}$	$\mathbb{B}\mathbb{B}_{\mathcal{B}}$	—	—	$\mathbb{A}\mathbb{B}_{\mathcal{AB}}$	\bullet	-1	-1/2
	WZ	$\mathbb{A}\mathbb{A}_{\mathcal{AB}}$	—	$\mathbb{A}\mathbb{B}_{\mathcal{AB}}$	$\mathbb{A}\mathbb{A}_{\mathcal{A}}$	—	\bullet	-1	-1/2
	WN	$\mathbb{A}\mathbb{A}_{\mathcal{AB}}$	—	$\mathbb{A}\mathbb{B}_{\mathcal{BA}}$	$\mathbb{A}\mathbb{A}_{\mathcal{B}}$	—	\bullet	-1	-1/2
	WW	$2 \times \mathbb{A}\mathbb{A}_{\mathcal{AB}}$	—	—	$\mathbb{A}\mathbb{A}_{\mathcal{A}} + \mathbb{A}\mathbb{A}_{\mathcal{B}}$	—	—	-1	-1/2
	MW	$\mathbb{A}\mathbb{A}_{\mathcal{A}}$	$\mathbb{B}\mathbb{B}_{\mathcal{AB}}$	—	—	$\mathbb{A}\mathbb{B}_{\mathcal{AB}}$	\bullet	-1	-1/2
	MW	$\mathbb{A}\mathbb{A}_{\mathcal{B}}$	$\mathbb{B}\mathbb{B}_{\mathcal{AB}}$	—	—	$\mathbb{A}\mathbb{B}_{\mathcal{BA}}$	\bullet	-1	-1/2
	MZ	—	$\mathbb{B}\mathbb{B}_{\mathcal{AB}}$	$\mathbb{A}\mathbb{B}_{\mathcal{AB}}$	—	$\mathbb{B}\mathbb{B}_{\mathcal{B}}$	\bullet	-1	-1/2
	MN	—	$\mathbb{B}\mathbb{B}_{\mathcal{AB}}$	$\mathbb{A}\mathbb{B}_{\mathcal{BA}}$	—	$\mathbb{B}\mathbb{B}_{\mathcal{A}}$	\bullet	-1	-1/2
	MM	—	$2 \times \mathbb{B}\mathbb{B}_{\mathcal{AB}}$	—	—	$\mathbb{B}\mathbb{B}_{\mathcal{A}} + \mathbb{B}\mathbb{B}_{\mathcal{B}}$	—	-1	-1/2

id	sources			resultants				$\Delta_{\text{DCJ}}^\lambda$	scr	
\mathcal{M}	$\text{ZZ}\overline{\text{WM}}$	$\text{AA}_{\mathcal{B}}$	$\text{BB}_{\mathcal{A}}$	$2 \times \text{AB}_{\mathcal{AB}}$	—	—	$4 \times \bullet$	−2	−1/2	
	$\text{NN}\overline{\text{WM}}$	$\text{AA}_{\mathcal{A}}$	$\text{BB}_{\mathcal{B}}$	$2 \times \text{AB}_{\mathcal{BA}}$	—	—	$4 \times \bullet$	−2	−1/2	
\mathcal{N}	$\text{ZW}\overline{\text{M}}$	$\text{AA}_{\mathcal{B}}$	$\text{BB}_{\mathcal{A}}$	$\text{AB}_{\mathcal{AB}}$	—	—	$\text{AB}_{\mathcal{BA}}$	$2 \times \bullet$	−1	−1/3
	$\text{ZZ}\overline{\text{W}}$	$\text{AA}_{\mathcal{B}}$	—	$2 \times \text{AB}_{\mathcal{AB}}$	$\text{AA}_{\mathcal{A}}$	—	—	$2 \times \bullet$	−1	−1/3
	$\text{ZZ}\overline{\text{M}}$	—	$\text{BB}_{\mathcal{A}}$	$2 \times \text{AB}_{\mathcal{AB}}$	—	$\text{BB}_{\mathcal{B}}$	—	$2 \times \bullet$	−1	−1/3
	$\text{NW}\overline{\text{M}}$	$\text{AA}_{\mathcal{A}}$	$\text{BB}_{\mathcal{B}}$	$\text{AB}_{\mathcal{BA}}$	—	—	$\text{AB}_{\mathcal{AB}}$	$2 \times \bullet$	−1	−1/3
	$\text{NN}\overline{\text{W}}$	$\text{AA}_{\mathcal{A}}$	—	$2 \times \text{AB}_{\mathcal{BA}}$	$\text{AA}_{\mathcal{B}}$	—	—	$2 \times \bullet$	−1	−1/3
	$\text{NN}\overline{\text{M}}$	—	$\text{BB}_{\mathcal{B}}$	$2 \times \text{AB}_{\mathcal{BA}}$	—	$\text{BB}_{\mathcal{A}}$	—	$2 \times \bullet$	−1	−1/3

Sources:

$\text{W} : \text{AA}_{\mathcal{AB}}$

$\overline{\text{W}} : \text{AA}_{\mathcal{A}}$

$\underline{\text{W}} : \text{AA}_{\mathcal{B}}$

$\text{M} : \text{BB}_{\mathcal{AB}}$

$\overline{\text{M}} : \text{BB}_{\mathcal{A}}$

$\underline{\text{M}} : \text{BB}_{\mathcal{B}}$

$\text{Z} : \text{AB}_{\mathcal{AB}}$

$\text{N} : \text{AB}_{\mathcal{BA}}$

DCJ-indel distance formula:

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

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

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

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 $C \in \text{RG}(\mathbb{A}, \mathbb{B})$:
 - 2.1 Split C with **gaining** DCJs (that have $\Delta_{\lambda} = 0$) 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 $\Delta_{\lambda} = 0$).
 - 2.3 Apply **gaining** DCJ operations (that have $\Delta_{\lambda} = 0$) 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

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

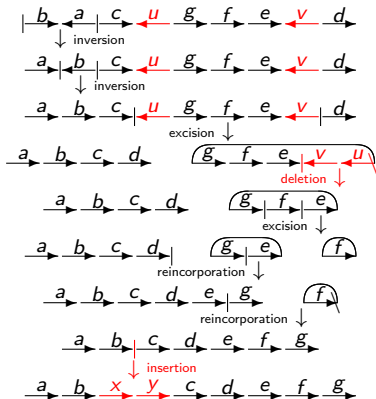
However, these gaining DCJs can often be replaced by $\begin{cases} \text{neutral DCJs with } \Delta_\lambda = -1 \\ \text{losing DCJs with } \Delta_\lambda = -2 \end{cases}$



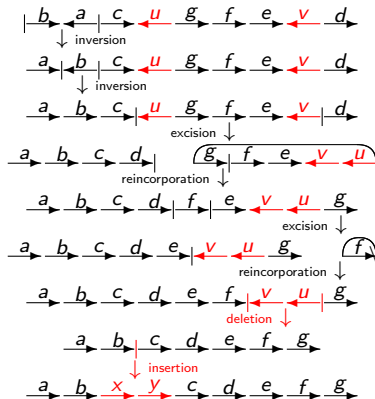
There is a big range of possibilities between the presented sorting algorithm and a sorting algorithm that minimizes gaining DCJs with $\Delta_\lambda = 0$ (maximizing indels)

Restricted DCJ-indel-distance (singular linear genomes)

general DCJ-indel sorting



restricted DCJ-indel sorting



In any sorting sequence, it is always possible to

- move **deletions** down
- move **insertions** up

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

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

Quiz 3

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.
- ☒ C 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)