Survey

1 The contents of the lecture are for me...

- A Interesting
- B Neutral
- C Uninteresting

2 I find the pace of the lecture...

- A Too fast
- B Adequate
- C Too slow

3 The level of the exercise sheets is...

- A Too difficult
- B Adequate
- C Too easy

4 The length of the exercise sheets is...

- A Too long
- B Adequate
- C Too short

Topics of today:

Canonical inversion distance and sorting:

1. Relational / Breakpoint diagram
2. Split / Neutral / Joining inversions
3. Good / bad components
4. Hurdles and fortress

Canonical inversion model - circular chromosomes

(Unichromosomal genomes $\equiv$ chromosomes)

Given two canonical circular chromosomes $\mathbb{A}$ and $\mathbb{B}$,...

Canonical Inversion Distance Problem: Compute the minimum number of inversions required to transform $\mathbb{A}$ into $\mathbb{B}$.

Denote by $d_{INV}(\mathbb{A}, \mathbb{B})$ the inversion distance of $\mathbb{A}$ and $\mathbb{B}$.

Canonical Inversion Sorting Problem: Find a sequence of $d_{INV}(\mathbb{A}, \mathbb{B})$ inversions that transform $\mathbb{A}$ into $\mathbb{B}$.

Relational diagram of canonical circular chromosomes

Given canonical circular chromosomes $\mathbb{A}$ and $\mathbb{B}$, their **relational diagram** $RD(\mathbb{A}, \mathbb{B}) = (V, E)$ 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 $\mathbb{B}$

   The vertices corresponding to $\xi(\mathbb{A})$ are drawn in an upper line, while the vertices corresponding to $\xi(\mathbb{B})$ are drawn in a lower line.

   In each line, the vertices must follow the same (circular) order of the corresponding extremities in the respective chromosome, according to one of the two reading directions.

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

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

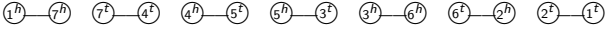
   - **Adjacency edges:** $\begin{cases} E_\alpha(\mathbb{A}) = \{uv : u, v \in V(\xi(\mathbb{A})) \text{ and } \ell(u)\ell(v) \in \alpha(\mathbb{A})\} \\ E_\alpha(\mathbb{B}) = \{uv : u, v \in V(\xi(\mathbb{B})) \text{ and } \ell(u)\ell(v) \in \alpha(\mathbb{B})\} \end{cases}$
   - **Extremity edges:** $E_\xi = \{uv : u \in V(\xi(\mathbb{A})) \text{ and } v \in V(\xi(\mathbb{B})) \text{ and } \ell(u) = \ell(v)\}$

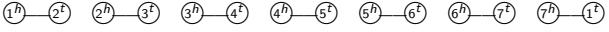
Note that:

- Let $n = |\mathcal{G}_\star|$. The number of edges in $E_\alpha(\mathbb{A}) \cup E_\alpha(\mathbb{B})$ is $2n$ (n adjacency edges per chromosome).

Relational diagram of canonical circular chromosomes

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





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

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

Every vertex has degree two:

$RD(\mathbb{A}, \mathbb{B})$ is a collection of (even) cycles (alternating edes in E_ξ and in $E_\alpha(\mathbb{A}) \cup E_\alpha(\mathbb{B})$)

cycle with k edges in $E_\alpha(\mathbb{A}) \cup E_\alpha(\mathbb{B})$: k-cycle

$\mathcal{C}$ = set of cycles in $RD(\mathbb{A}, \mathbb{B})$

If $\mathbb{A} = \mathbb{B}$,
$RG(\mathbb{A}, \mathbb{B})$ has only 2-cycles:

$2n = 2|\mathcal{C}| \quad \Rightarrow \quad n = |\mathcal{C}|$

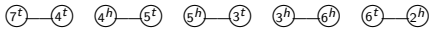
Otherwise, if $\mathbb{A} \neq \mathbb{B}$:

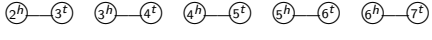
$n > |\mathcal{C}|$

Types of inversion and lower bound for the inversion distance

Assign one (arbitrary) direction to each cycle of $RD(\mathbb{A}, \mathbb{B})$

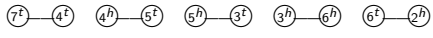
$\mathbb{A} = (1\ \bar{7}\ 4\ 5\ 3\ \bar{6}\ \bar{2})$

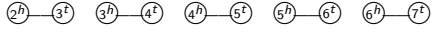




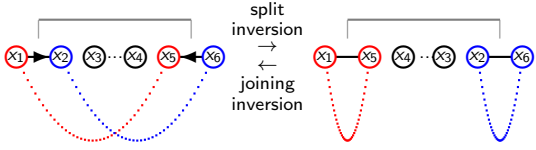
$\mathbb{B} = (1\ 2\ 3\ 4\ 5\ 6\ 7)$

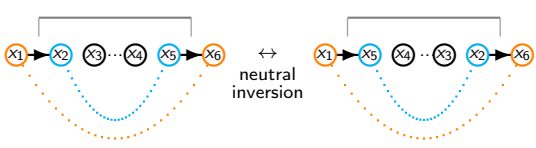
$\mathbb{A} = (1\ \bar{7}\ 4\ 5\ 3\ \bar{6}\ \bar{2})$





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





Lower bound for the inversion distance: $d_{INV}(\mathbb{A}, \mathbb{B}) \geq n - |\mathcal{C}|$

Types of cycles

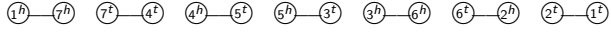
Trivial cycle: one adjacency in each chromosome

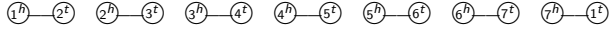
2-cycle (sorted)

Good cycle: $\begin{cases} \text{at least one pair of adjacencies with opposite directions in chromosome } \mathbb{A} \\ \text{at least one pair of adjacencies with opposite directions in chromosome } \mathbb{B} \end{cases}$

Can be split into two cycles by applying an inversion in $\mathbb{A}$ or in $\mathbb{B}$

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





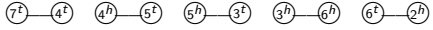
$(1\ 2\ 3\ 4\ 5\ 6\ 7)$

Types of cycles

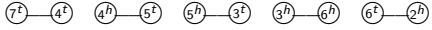
Semi-good cycle: $\begin{cases} \text{at least one pair of adjacencies with opposite directions in one of the two chromosomes} \\ \text{all adjacencies have the same direction in the other chromosome} \end{cases}$

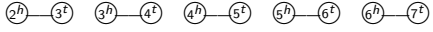
Can be split into two cycles by applying an inversion only in $\mathbb{A}$ or only in $\mathbb{B}$

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

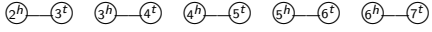


$(1\ \bar{7}\ 4\ 5\ 3\ \bar{6}\ \bar{2})$





$(1\ 2\ 3\ 4\ 5\ 6\ 7)$



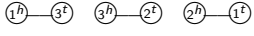
$(1\ 2\ 3\ 4\ 5\ 6\ 7)$

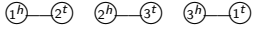
Types of cycles

Bad cycle: $\begin{cases} \text{all adjacencies in chromosome } \mathbb{A} \text{ have the same direction} \\ \text{all adjacencies in chromosome } \mathbb{B} \text{ have the same direction} \end{cases}$

Cannot be split into two cycles

$(1\ 3\ 2)$

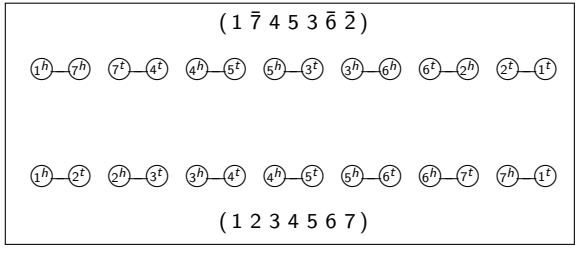




$(1\ 2\ 3)$

Relational diagram ≅ Breakpoint diagram

Relational diagram:



Looking either at the top line or at the bottom line of the diagram:

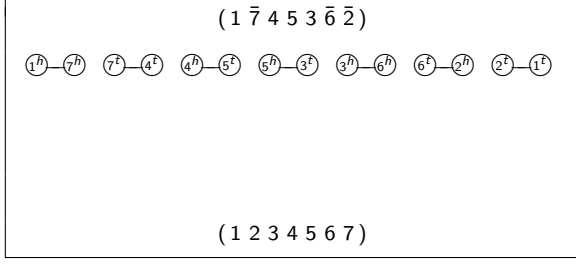
Two interleaving cycles: $c \ldots c' \ldots c \ldots c'$

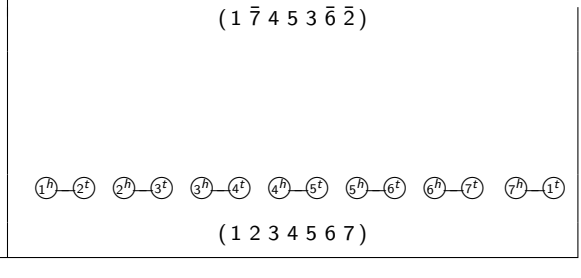
Interleaving sequence of cycles: $c_1, c_2, \ldots, c_k$ such that c_i and c_{i+1} are interleaving for all $1 \leq i \leq k-1$

Interleaving component or simply **component** K:

- for each pair of cycles $c, c' \in K$ there is an interleaving sequence from c to c'
- K is maximal

Breakpoint diagrams:



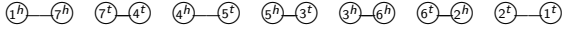


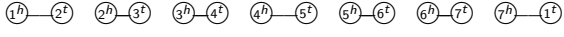
Types of (interleaving) components

Trivial component: only one trivial 2-cycle

Good component: at least one good or semi-good cycle

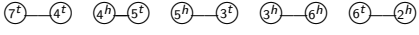
$(1\ \bar{7}\ 4\ 5\ 3\ \bar{6}\ \bar{2})$

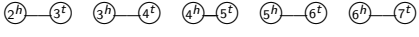




$(1\ 2\ 3\ 4\ 5\ 6\ 7)$

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





$(1\ 2\ 3\ 4\ 5\ 6\ 7)$

Types of (interleaving) components

Bad component: only bad cycles

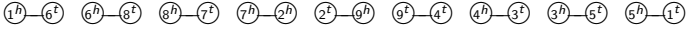
$(1\ 4\ 3\ 2)$

$(1\ 2\ 3\ 4)$

Unsafe inversions

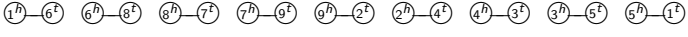
A split inversion applied to a cycle of a good component can create bad components

$(1\ 6\ 8\ 7\ \bar{2}\ \bar{9}\ 4\ 3\ 5)$



$(1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9)$

$(1\ 6\ 8\ 7\ 9\ 2\ 4\ 3\ 5)$



$(1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9)$

Sorting a good component - finding safe split inversions

Target adjacency: $\begin{cases} \text{good} \\ \text{bad} \end{cases}$

Overlapping target adjacencies

Overlap graph of a good component

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

$$(\ 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \ )$$

Sorting a good component - finding safe split inversions

Sorting a bad component with a neutral inversion

Overlap graph of a bad component

(1 4 3 2)

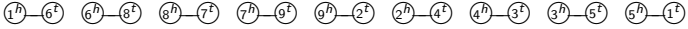
(1 2 3 4)

Sorting bad components - hurdles

K_1, K_2 and K_3 are three distinct components in $RD(\mathbb{A}, \mathbb{B})$ so that $K_3 \dots K_1 \dots K_1 \dots K_3 \dots K_2 \dots K_2$

$\Rightarrow$ K_3 **separates** K_1 and K_2

$$(1\ 6\ 8\ 7\ 9\ 2\ 4\ 3\ 5)$$



$$(1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9)$$

By joining with an inversion two cycles c_1 and c_2, that belong to two distinct components K_1 and K_2 respectively, we merge not only the components K_1 and K_2, but also all components that separate K_1 and K_2 , into a single **good** component K.

Sorting bad components - simple hurdles and super hurdles

h : number of hurdles in $RD(\mathbb{A}, \mathbb{B})$

Sorting bad components - simple hurdles and super hurdles

Sorting bad components - fortress

$$f : \begin{cases} 0 & RD(\mathbb{A}, \mathbb{B}) \text{ is not a fortress} \\ 1 & RD(\mathbb{A}, \mathbb{B}) \text{ is a fortress} \end{cases}$$

Canonical inversion distance of circular chromosomes

$$\mathsf{d}_{\mathrm{INV}}(\mathbb{A}, \mathbb{B}) = n - |\mathcal{C}| + h + f$$

References

Transforming Cabbage into Turnip: Polynomial Algorithm for Sorting Signed Permutations by Reversals

(Sridhar Hannenhalli and Pavel A. Pevzner)

Journal of the ACM, Vol. 46, No. 1, pages. 1–27 (1999)

The Inversion Distance Problem

(Anne Bergeron, Julia Mixtacki and Jens Stoye)

In: Mathematics of Evolution and Phylogeny. Gascuel O (Ed); (2005)