In silico methods for genome rearrangement analysis: from identification of common markers to ancestral reconstruction

Géraldine Jean

LaBRI - Université Bordeaux 1

9 décembre 2008

Sous la direction de :
Macha Nikolski et Serge Dulucq

Comparative genomics

What ?

Study of genome structure by the comparison of different species

Why ?

- Understanding function processes
- Understanding evolution processes

How ?

Huge amount of data in contemporary genomes

- Automation of analyses
- Development of *in silico* methods

Interests ?

- Short-term : scientific answers
- On the long run : medicinal or therapeutic solutions

Comparative genomics

Genetic material comparison : DNA

- **First approach : study of local mutations in genes**
  Gene comparison by sequence alignment

```
sequence 1 : CAGCA-CGTGGATTCTCGG
               | ||  ||||| ||  |
sequence 2 : TATCAGCGTGG-CACTAGC
```

Observation (Palmer & Herbon 1988)

The major part of *Bassica oleracea*'s and *Bassica canpestris*'s genes is quasi identical (99%), but only the gene order differ significantly.

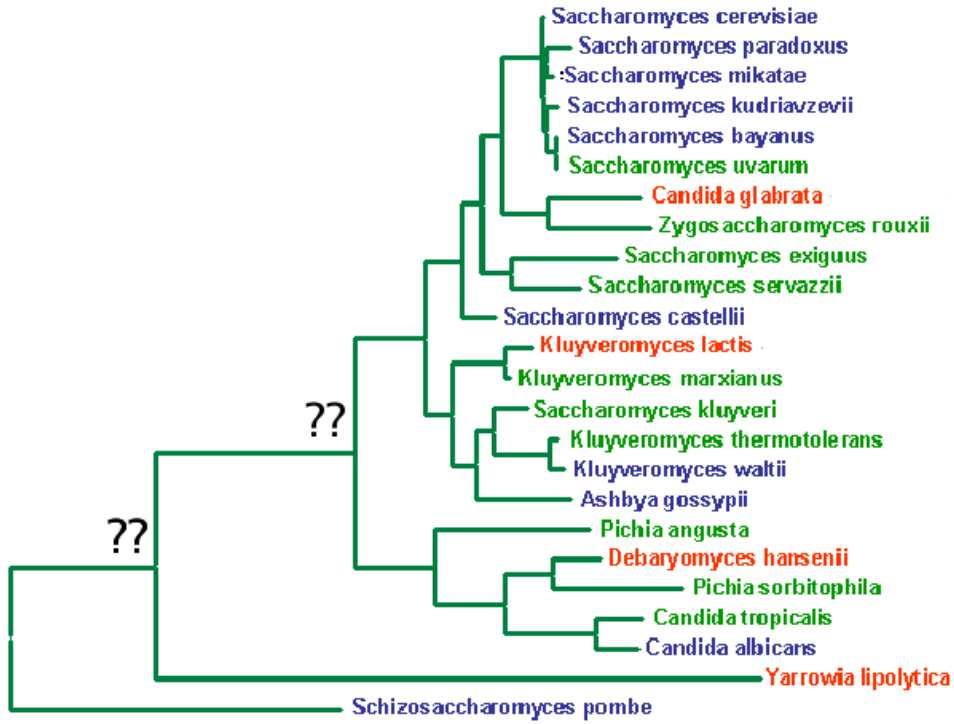
- **Second approach : study of genomic rearrangements**
  Whole genome architecture comparison by gene order and content study

Comparative genomics

- **Challenge** : emitting hypotheses on the history of contemporary genomes and the general mechanisms of their formation

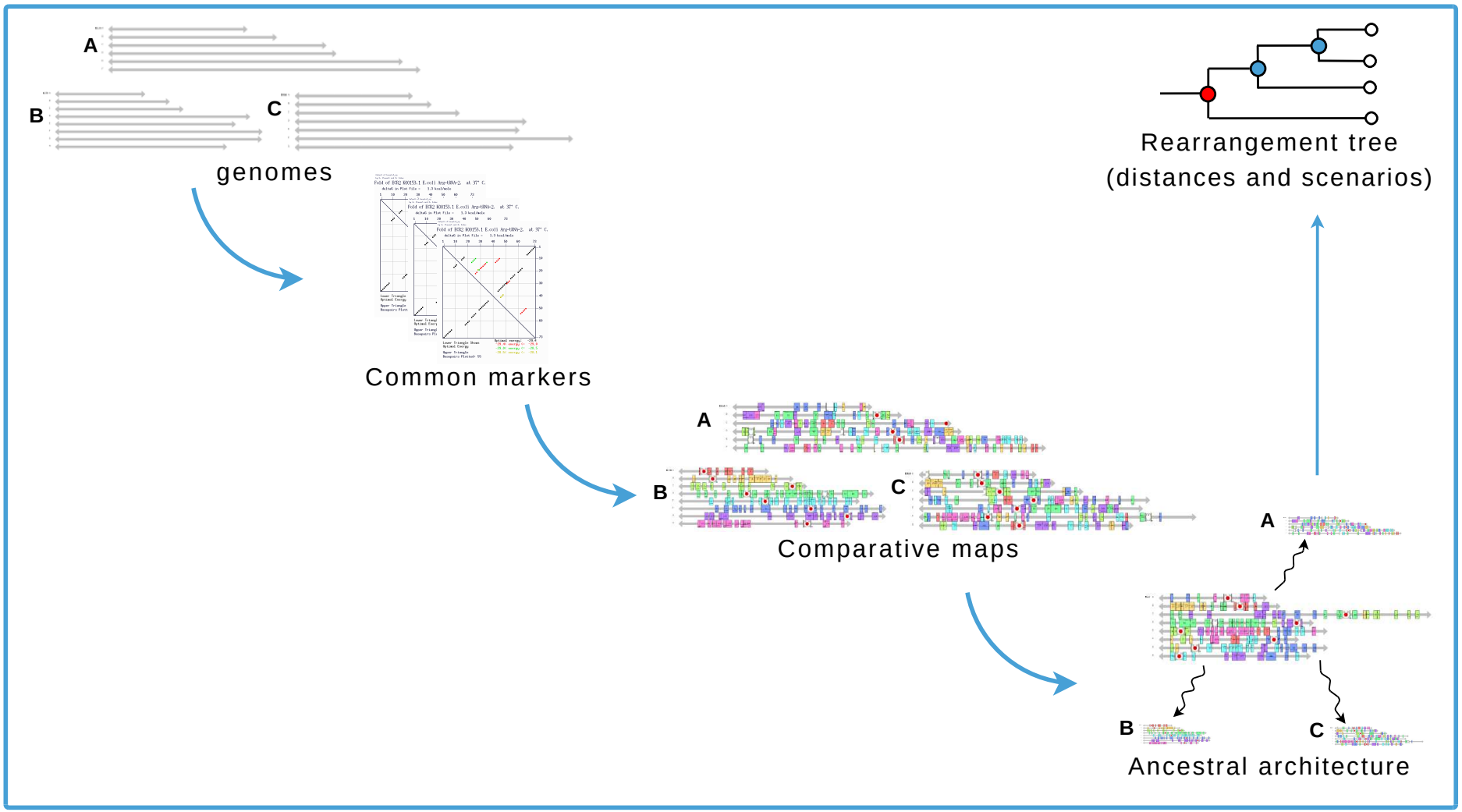
- **Problem** : impossibility of knowing with certainty the architecture of the common ancestral genomes

- **Solution** : developing methods for constructing a plausible architecture of ancestral genomes



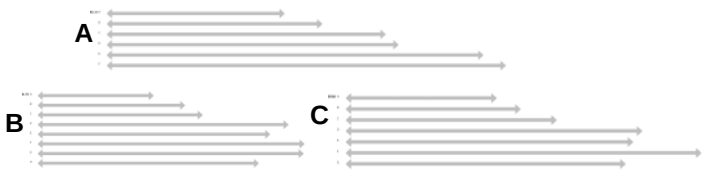
From common markers to ancestral reconstruction

Constructing ancestral architecture from the comparison of contemporary genomes requires 3 basic steps :

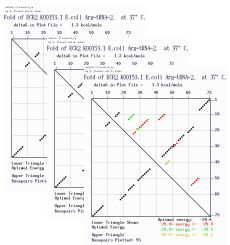


In this thesis

Combinatorial and algorithmic approach

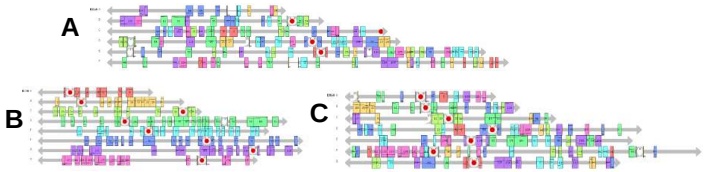


genomes



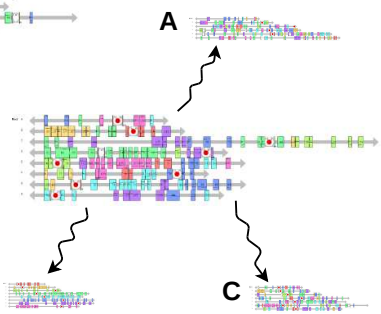
Common markers

Orthology and synteny study:
SyDiG algorithm (in prep)

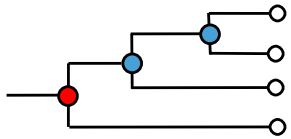


Comparative maps

Partial reconstructions of a median genome:
Super-block method (Jean et al., JCB, revision)



Ancestral architecture



Rearrangement tree
(distances and scenarios)

- Computation of a parsimonious scenario (Jean et al., IPL, 2007)
- Interative vizualization of scenarios: VIRAGE tool
(Jean et al. , ESF EMBO, 2007)

Chromosomal rearrangements in Yeasts

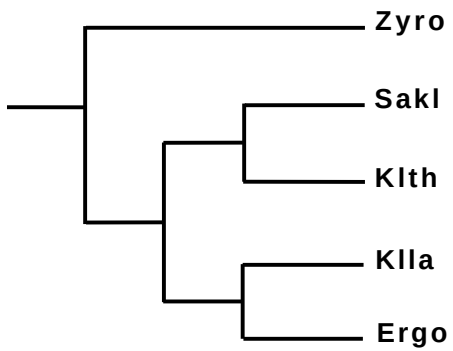
Is it **possible** to study ?

- Uniqueness of Génolevures data : complete genome sequences, availability of protein families
- Hemiascomycete yeasts : weak redundancy, synteny
- Additional information : positions of centromeres

⇒ We can specify markers for Hemiascomycetes

Hemiascomycete genomes

Species	Mnemonic
Kluyveromyces thermotolerans	Klth
Ashbya gossypii	Ergo
Kluyveromyces lactis	Klla
Saccharomyces kluyveri	Sakl
Zygosaccharomyces rouxii	Zyro



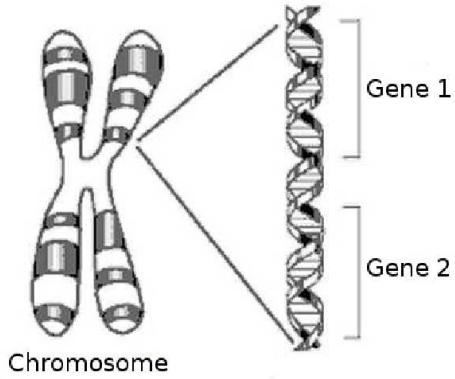
Cladogram

Outline

- Modeling a genome and evolutionary mechanisms
- Partial ancestral reconstructions : super-block method
- Constructing and visualizing parsimonious scenarios
- Conclusion and perspectives

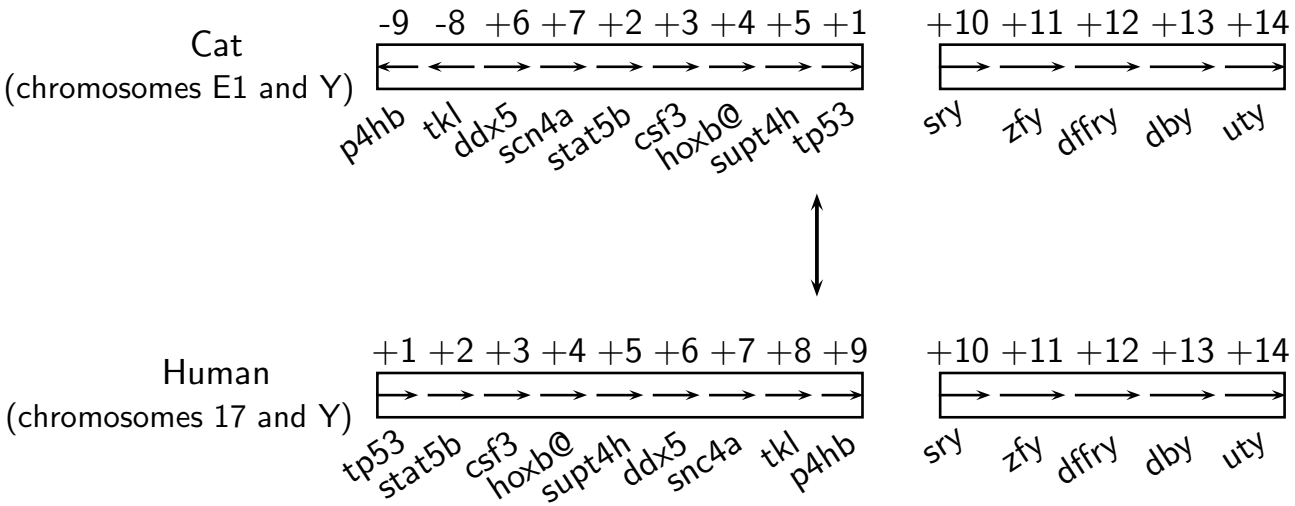
Signed permutation model

- **Gene** : functional unity composed of DNA
- **Chromosome** : sequence of genes
- **Genome** : set of chromosomes



Model :

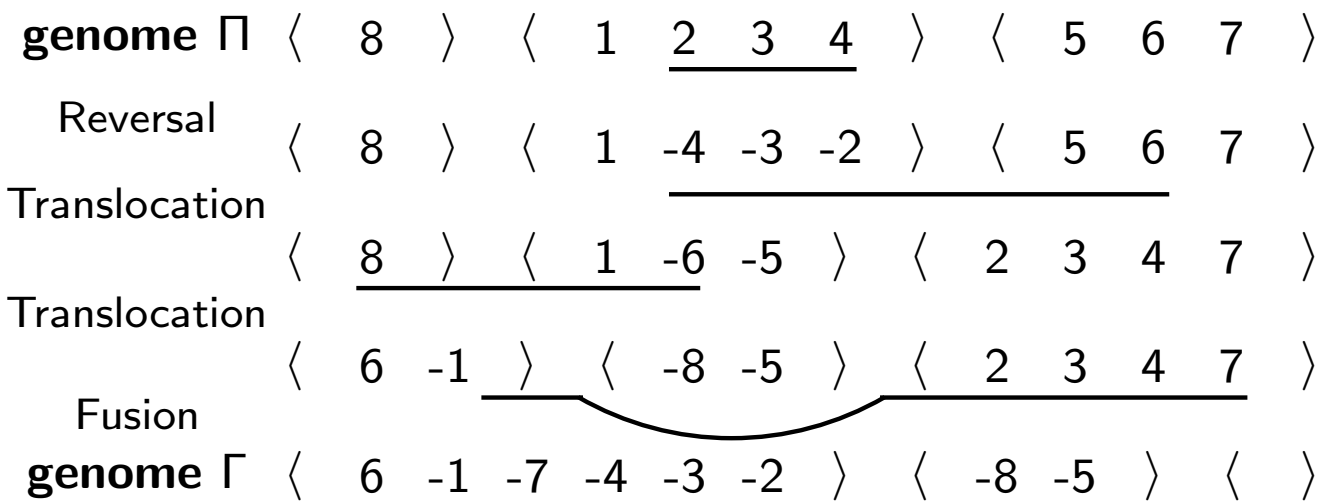
- **Marker (a gene or a set of genes) :** signed ordinal
- **Chromosome :** signed permutation
- **Genome :** set of signed permutations



Rearrangement operations

- Evolution mechanisms :

- Mathematical operations on permutations :



Distance and scenario

Two closely related problems :

- **Rearrangement distance** $d(\Pi, \Gamma)$:
  minimum number of rearrangements that transform Π into Γ
- **Parsimonious rearrangement scenario** :
  sequence of rearrangements respecting rearrangement distance

Distance and scenario

Two closely related problems :

- **Rearrangement distance** $d(\Pi, \Gamma)$:
  minimum number of rearrangements that transform Π into Γ
- **Parsimonious rearrangement scenario** :
  sequence of rearrangements respecting rearrangement distance

Π : $\langle 8 \rangle \langle 1 \ \ 2 \ \ 3 \ \ 4 \rangle \langle 5 \ \ 6 \ \ 7 \rangle$

$d(\Pi, \Gamma) = 4$

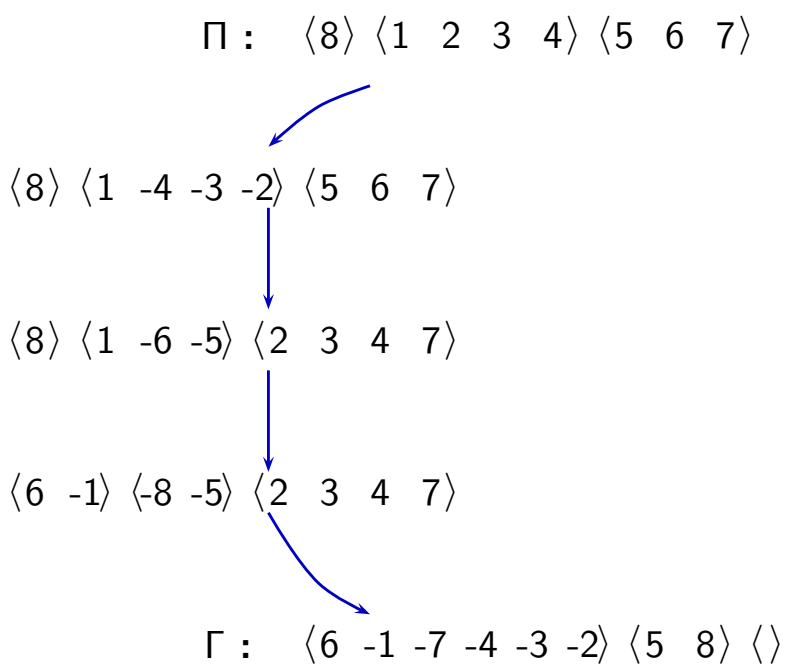
Γ : $\langle 6 \ \ \text{-}1 \ \ \text{-}7 \ \ \text{-}4 \ \ \text{-}3 \ \ \text{-}2 \rangle \langle 5 \ \ 8 \rangle \langle \rangle$

Distance and scenario

Two closely related problems :

- **Rearrangement distance** $d(\Pi, \Gamma)$:
  minimum number of rearrangements that transform Π into Γ
- **Parsimonious rearrangement scenario** :
  sequence of rearrangements respecting rearrangement distance

$d(\Pi, \Gamma) = 4$

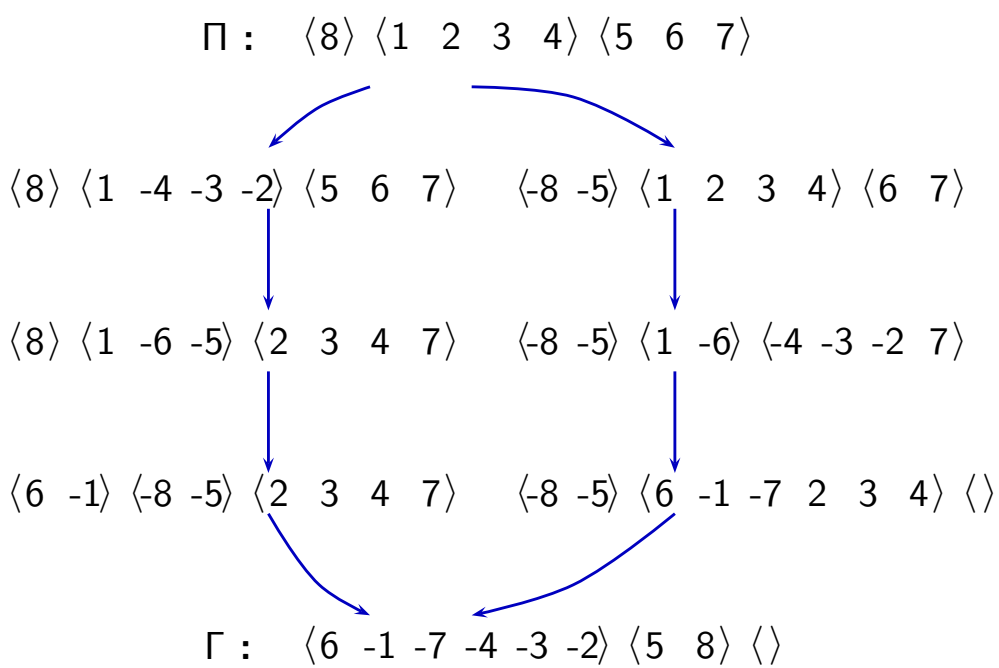


Distance and scenario

Two closely related problems :

- **Rearrangement distance** $d(\Pi, \Gamma)$:
  minimum number of rearrangements that transform Π into Γ
- **Parsimonious rearrangement scenario** :
  sequence of rearrangements respecting rearrangement distance

$d(\Pi, \Gamma) = 4$

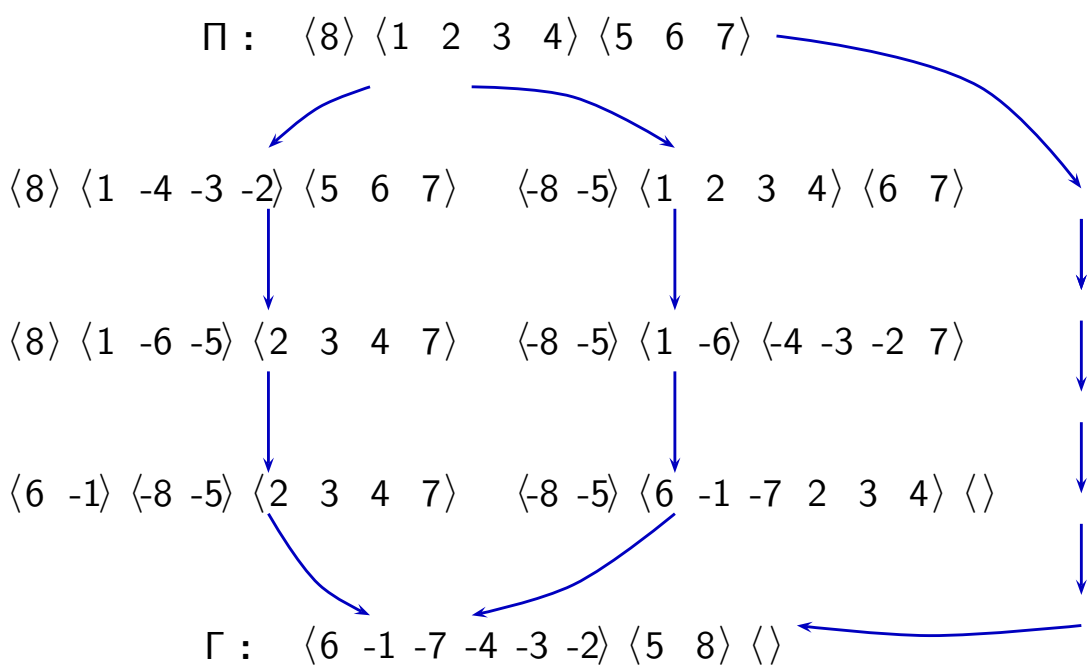


Distance and scenario

Two closely related problems :

- **Rearrangement distance** $d(\Pi, \Gamma)$:
  minimum number of rearrangements that transform Π into Γ
- **Parsimonious rearrangement scenario** :
  sequence of rearrangements respecting rearrangement distance

$d(\Pi, \Gamma) = 4$

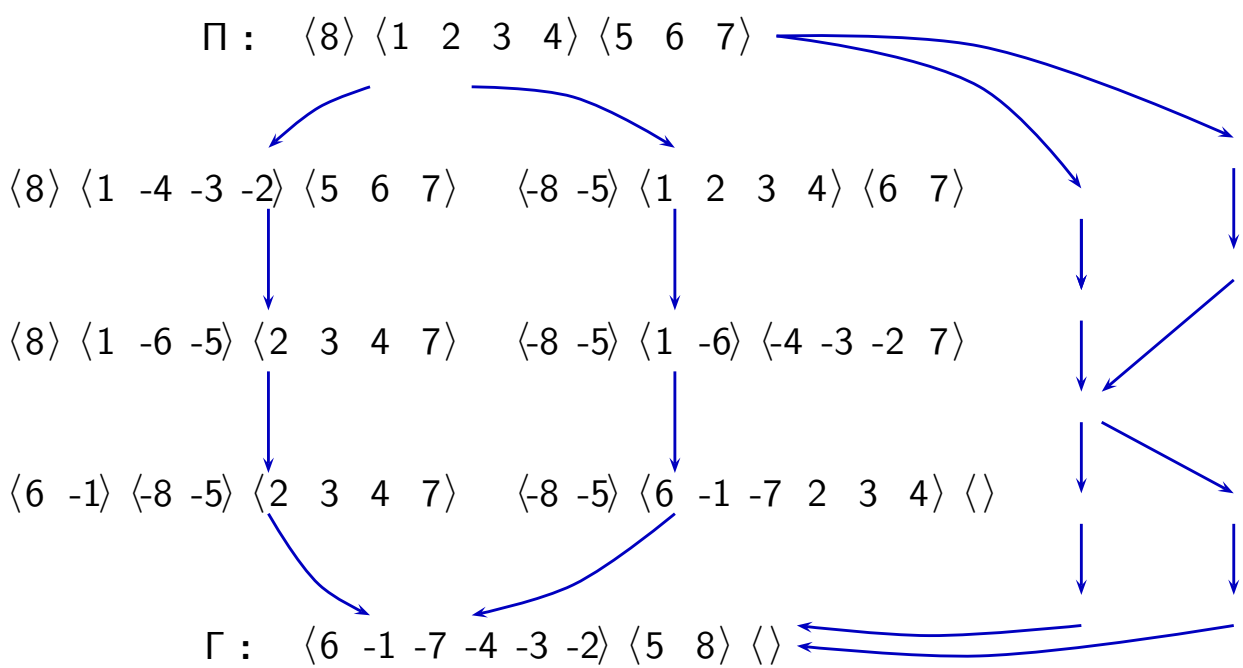


Distance and scenario

Two closely related problems :

- **Rearrangement distance** $d(\Pi, \Gamma)$:
  minimum number of rearrangements that transform Π into Γ
- **Parsimonious rearrangement scenario** :
  sequence of rearrangements respecting rearrangement distance

$d(\Pi, \Gamma) = 4$

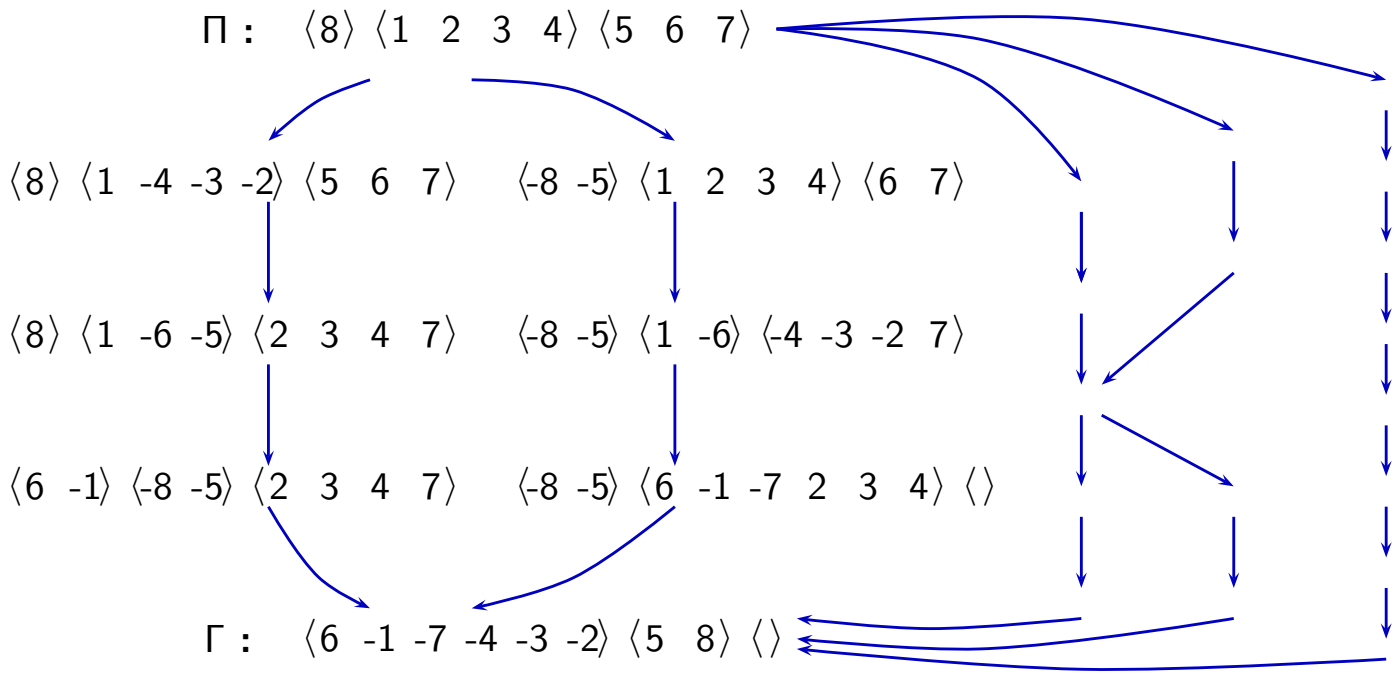


Distance and scenario

Two closely related problems :

- **Rearrangement distance** $d(\Pi, \Gamma)$:
  minimum number of rearrangements that transform Π into Γ
- **Parsimonious rearrangement scenario** :
  sequence of rearrangements respecting rearrangement distance

$d(\Pi, \Gamma) = 4$



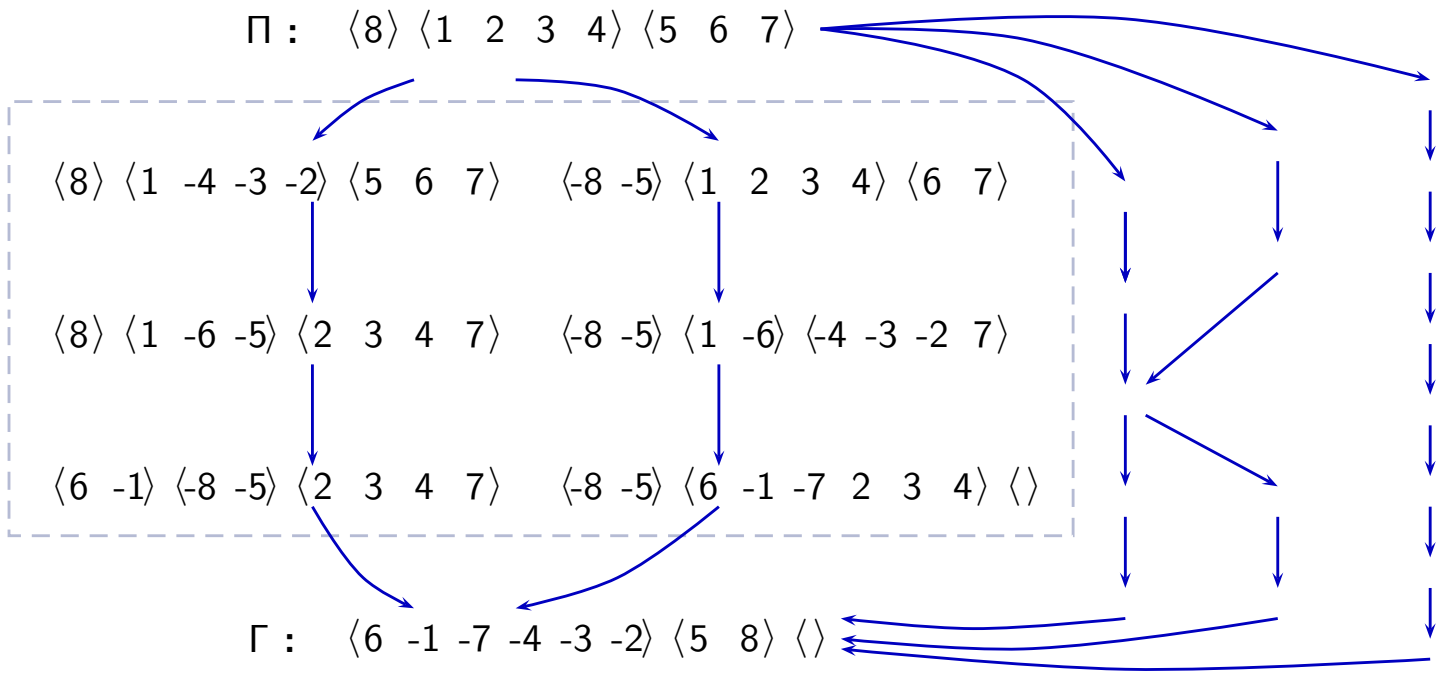
Distance and scenario

Two closely related problems :

- **Rearrangement distance** $d(\Pi, \Gamma)$:
  minimum number of rearrangements that transform Π into Γ
- **Parsimonious rearrangement scenario** :
  sequence of rearrangements respecting rearrangement distance

$d(\Pi, \Gamma) = 4$

Two parsimonious scenarios



Breakpoints

Definition (Nadeau and Taylor 1984)

Two consecutive elements π_i and π_{i+1} of a chromosome π are said to be **adjacent** in the genome Π.
If two elements π_i and π_{i+1} are adjacent in Π but neither $\pi_i.\pi_{i+1}$ nor $-\pi_{i+1}.-\pi_i$ are present in Γ, then $\pi_i.\pi_{i+1}$ forms a **breakpoint** in Π.

$$\Pi = \langle \text{-9} \quad \text{-8} \quad +6 \quad +7 \quad +2 \quad +3 \quad +4 \quad +5 \quad +1 \rangle$$
$$\Gamma = \langle +1 \quad +2 \quad +3 \quad +4 \quad +5 \quad +6 \quad +7 \quad +8 \quad +9 \rangle$$

Breakpoints in Π are :

$$\Pi = \langle \text{-9} \quad \text{-8} \bullet +6 \quad +7 \bullet +2 \quad +3 \quad +4 \quad +5 \bullet +1 \rangle \bullet$$

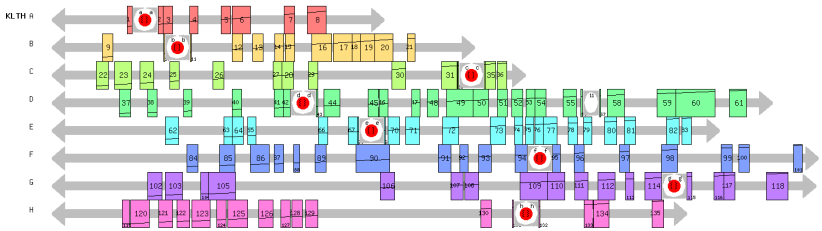
- Breakpoint distance $b(\Pi, \Gamma)$: Number of breakpoints between Π and Γ
- $d(\Pi, \Gamma)$ and $b(\Pi, \Gamma)$ closely related : finding rearrangements within Π that eliminate breakpoints between Π and Γ

Signed permutations for Yeasts

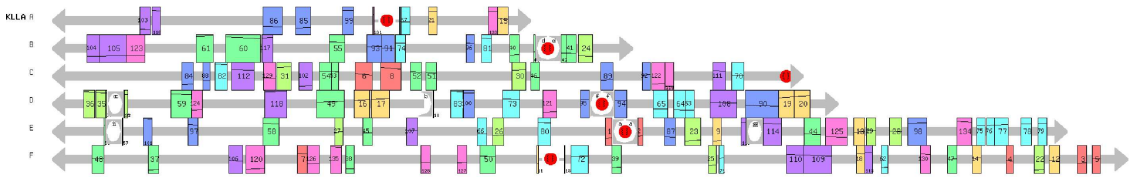
640 common markers computed with **SyDiG algorithm** :

- Same marker content without duplication, centromere positions
- Conservation of the 120 longest markers for study

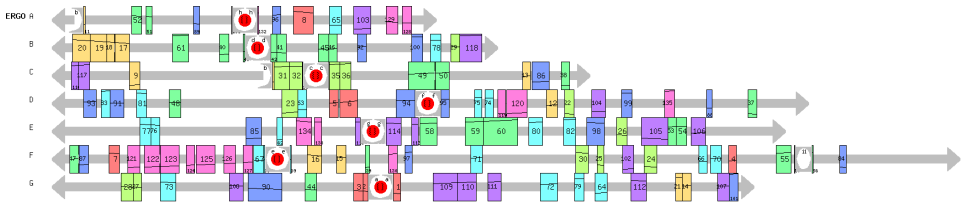
Klth



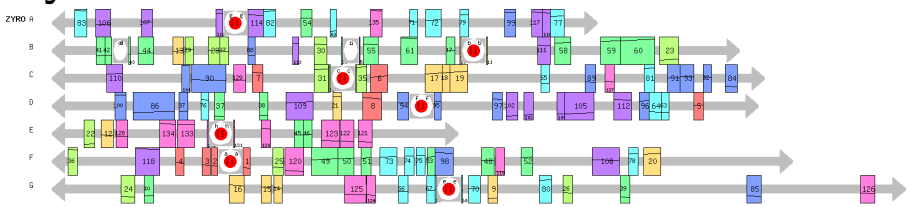
Klla



Ergo



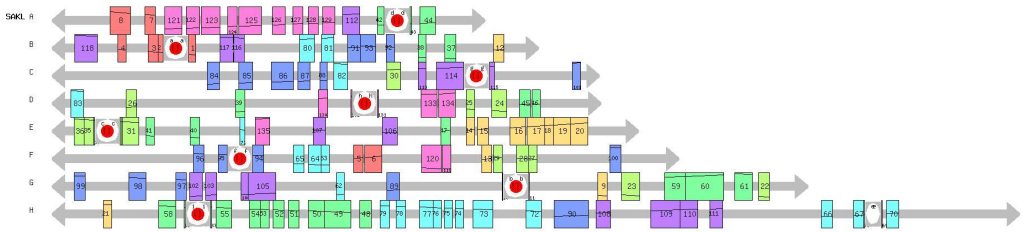
Zyro



	Klth	Ergo	Klla	Sakl	Zyro
Klth	0	88	105	45	84
Ergo		0	109	85	101
Klla			0	98	115
Sakl				0	79
Zyro					0

Pairwise distances

Sakl



Partial ancestral reconstructions : super-block method

Interest of ancestral reconstruction

Discovery of biological clues by the analysis of computed results :

- History of contemporary genomes
- General mechanisms of their formation

But results from two techniques do not necessarily agree

In Silico	**In Vitro**
Rearrangement distance	Chromosomal painting
(Bourque & Pevzner 2002), (Bourque, Tesler & Pevzner 2004)	(Wienberg et al. 1990)
Complete genomes ($\approx$ 5 species)	Eutherian clade ($\approx$ 80 species)
$\approx$ 3 Kb ($\approx$ yeast gene length)	$\approx$ 4 Mb ($>$ yeast chromosome length)

Possible solution : integrate more **biological knowledge** into the mathematical approach (Rocchi 2006)

Ancestors as median genomes

Definition

Median genome problem : given $G_1, ..., G_N$, find M such that for a distance d

$$\sum_{i=1}^{N} d(M, G_i) \text{ is minimal}$$

Different distances : **rearrangement, breakpoint**, double cut and join

- NP-complete even for $N = 3$
  breakpoint distance (Bryant 1998, Pe'er & Shamir 1998)
  rearrangement distance (Caprara 1999 and 2003)
- High number of equivalent solutions (Eriksen 2007)
- Minimal solutions can be highly divergent (Eriksen 2007)

$\Rightarrow$ Giving a unique global solution is biologically misleading

Piece-wise ancestral reconstruction

Looking for **common features** present in ancestral genome architecture

CARs (Contiguous Ancestral Regions) method (Ma et al. 2006) :

- Based on phylogenetic tree
- Analogous to Fitch's parsimony method (Fitch 1971)

Phylogeny relationships	**Chromosome dynamics**
Rate of punctual mutations in genomic sequences	Rearrangement of breakpoint distance
Temporal notion	No time-scale of the rearrangement events

⇒ We propose a new piece-wise method without phylogenetic consideration

A new method for reconstructing ancestral architecture

Breakpoint-based method (Sankoff & Blanchette 1997) : more an adjacency is frequent in contemporary genomes, more it should appear in the ancestral genome

An adjacency which appears in only one genome is not informative !

New method :

- Piece-wise reconstruction
- Based on breakpoints and rearrangement distance
- Adding biological constraints is possible

Adjacency frequencies

Definition

The frequency of an adjacency a**,** $u(a)$, is the number of genomes where a is present

Particular adjacencies in $\pi = \{\pi_1, .., \pi_n\}$: telomeres $0.\pi_1$ and $\pi_n.0$

Example

$$G_1 = \{1\ 2\ 3\ 4,\ 5\ 6\} \qquad G_2 = \{1\ 2\ 3\ 4,\ -5\ 6\}$$
$$G_3 = \{3\ 1\ 4\ 2\ -5,\ 6\} \qquad G_4 = \{2\ 1\ 3\ 4,\ 5\ 6\}$$

frequency	adjacencies
4	6.0
3	3.4, 0.5, 4.0
2	5.6, 2.3, 1.2, 0.1
1	−5.6, 2.-5, 4.2, 1.4, 1.3, 3.1, 2.1, 0.6, 5.0, 0.3, 0.2

Adjacency graph

Unsigned representation of signed permutation (Hannenhalli & Pevzner, 1995) :

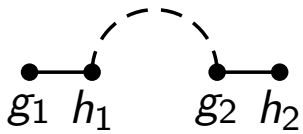
$g = 2\pi_i - 1 \quad h = 2\pi_i$ (a) π_i is positive

$h = 2\pi_i \quad g = 2\pi_i - 1$ (b) π_i is negative

Denoted : $\pi_i.\pi_j$ by $(g_i\ h_i).(g_j\ h_j)$ and $\pi_i.-\pi_j$ by $(g_i\ h_i).(h_j\ g_j)$

Example

The **adjacency graph** for a set $A = \{(g_1\ h_1).(g_2\ h_2)\}$:



- 4 vertices g_1, h_1, g_2 and h_2
- two edges stand for markers $e_1 = \{g_1, h_1\}$ and $e_2 = \{g_2, h_2\}$.
- one edge stands for the adjacency $e_3 = \{h_1, g_2\}$

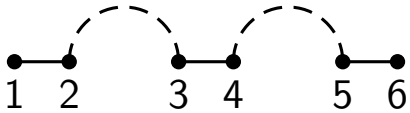
Intuition

Hypothesis

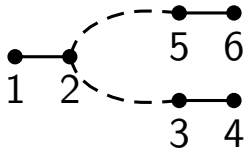
For a set of genomes $\{G_i\}$, the higher is the frequency of an adjacency, the higher is the probability that it should be present in a median genome.

Build **partial assemblies** of median genomes

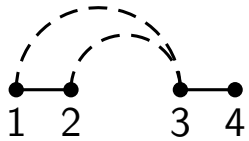
1. Build a partition $\mathcal{P}$ of adjacencies where each part is composed of inter-dependent adjacencies.



complementary adjacencies



vertex contradiction



cycle contradiction

2. Inspect $\mathcal{P}$ in decreasing frequency of its parts, and construct the partial assemblies by favoring adjacencies with higher frequency.

Assemble these partial assemblies into **potential medians**

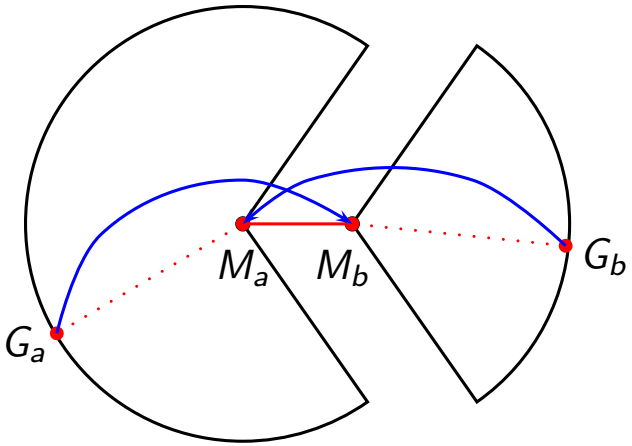
Adjacencies and distances

N contemporary genomes $\{G_i\}$, d rearrangement distance
C the set of all pairs of contradictory adjacencies

Theorem (Jean et al.)

For any pair $\{a, b\} \in C$ and two genomes M_a and M_b identical up to 2 adjacencies with $a \in M_a$ and $b \in M_b$, it holds that

$$\left| \sum_i^N d(M_a, G_i) - \sum_i^N d(M_b, G_i) \right| \leq N.$$



If $u(a) > u(b)$

$$\sum_i^N d(M_a, G_i) - \sum_i^N d(M_b, G_i) \ll N$$

Similarly for the breakpoint distance

Groups of adjacencies

$\mathcal{P}(A)$ be a partition of $\mathcal{A}$, set of all adjacencies.
$\mathcal{P}_0(\mathcal{A})$: adjacencies in cycle contradiction and singletons

Merging of parts : $\sqcup$ defines a partition of $\mathcal{A}$ such that for any $p \in \sqcup(\mathcal{P}(\mathcal{A}))$

- $\exists\ p_1 \in \mathcal{P}(A)$ s.t. $p = p_1$ or
- $\exists\ p_1, p_2 \in \mathcal{P}(\mathcal{A})$ s.t. $p = p_1 \cup p_2$ and moreover $\exists\ a \in p_1$ and $\exists\ b \in p_2$ s.t. $u(a) = u(b) = u(p_1) = u(p_2)$ and either a and b are dependent or a and b participate in a cycle $c \in G$ without vertex $v = 0$ s.t. $\forall v \in c$ we have $u(v) \geq u(a)$.

Definition

A **group** g is a part of $\sqcup^n(\mathcal{P}_0(A))$, where $\sqcup^n(\mathcal{P}_0(A))$ is the fixed point of $\sqcup$.

Groups of adjacencies, continued

Example

$G_1 = \{$ **1 2 3 4, 5 6** $\}$ $G_2 = \{$ **1 2 3 4, -5 6** $\}$

$G_3 = \{$ **3 1 4 2 -5, 6** $\}$ $G_4 = \{$ **2 1 3 4, 5 6** $\}$

Groups of adjacencies, continued

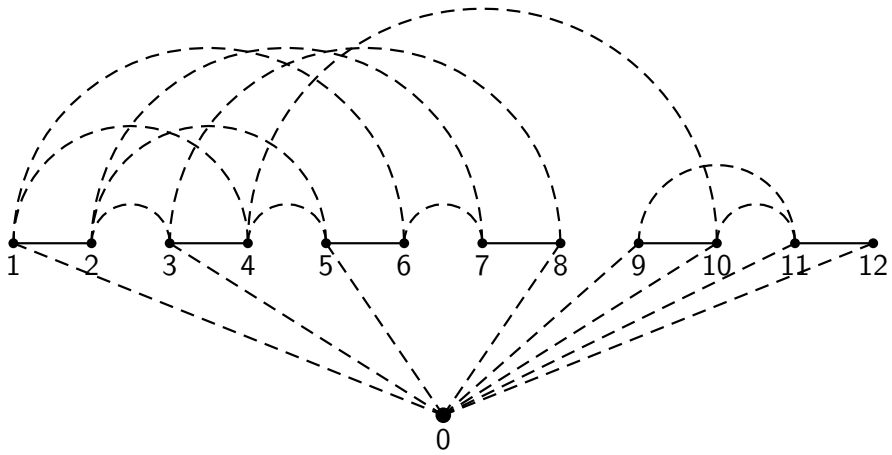
Example

$G_1 = \{ \underset{1\ 2}{\mathbf{1}} \quad \underset{3\ 4}{\mathbf{2}} \quad \underset{5\ 6}{\mathbf{3}} \quad \underset{7\ 8,}{\mathbf{4,}} \quad \underset{9\ 10}{\mathbf{5}} \quad \underset{11\ 12}{\mathbf{6}} \}$

$G_2 = \{ \underset{1\ 2}{\mathbf{1}} \quad \underset{3\ 4}{\mathbf{2}} \quad \underset{5\ 6}{\mathbf{3}} \quad \underset{7\ 8,}{\mathbf{4,}} \quad \underset{10\ 9}{\mathbf{-5}} \quad \underset{11\ 12}{\mathbf{6}} \}$

$G_3 = \{ \underset{5\ 6}{\mathbf{3}} \quad \underset{1\ 2}{\mathbf{1}} \quad \underset{7\ 8}{\mathbf{4}} \quad \underset{3\ 4}{\mathbf{2}} \quad \underset{10\ 9,}{\mathbf{-5,}} \quad \underset{11\ 12}{\mathbf{6}} \}$

$G_4 = \{ \underset{3\ 4}{\mathbf{2}} \quad \underset{1\ 2}{\mathbf{1}} \quad \underset{5\ 6}{\mathbf{3}} \quad \underset{7\ 8,}{\mathbf{4,}} \quad \underset{9\ 10}{\mathbf{5}} \quad \underset{11\ 12}{\mathbf{6}} \}$



Adjacency graph

Groups of adjacencies, continued

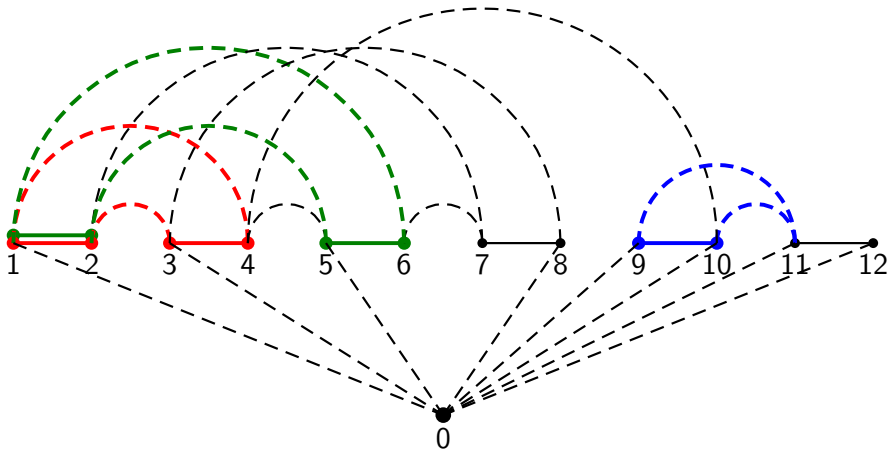
Example

$G_1 = \{ \underset{1\ 2}{\mathbf{1}} \ \underset{3\ 4}{\mathbf{2}} \ \underset{5\ 6}{\mathbf{3}} \ \underset{7\ 8,}{\mathbf{4,}} \ \underset{9\ 10}{\mathbf{5}} \ \underset{11\ 12}{\mathbf{6}} \}$

$G_2 = \{ \underset{1\ 2}{\mathbf{1}} \ \underset{3\ 4}{\mathbf{2}} \ \underset{5\ 6}{\mathbf{3}} \ \underset{7\ 8,}{\mathbf{4,}} \ \underset{10\ 9}{\mathbf{-5}} \ \underset{11\ 12}{\mathbf{6}} \}$

$G_3 = \{ \underset{5\ 6}{\mathbf{3}} \ \underset{1\ 2}{\mathbf{1}} \ \underset{7\ 8}{\mathbf{4}} \ \underset{3\ 4}{\mathbf{2}} \ \underset{10\ 9,}{\mathbf{-5,}} \ \underset{11\ 12}{\mathbf{6}} \}$

$G_4 = \{ \underset{3\ 4}{\mathbf{2}} \ \underset{1\ 2}{\mathbf{1}} \ \underset{5\ 6}{\mathbf{3}} \ \underset{7\ 8,}{\mathbf{4,}} \ \underset{9\ 10}{\mathbf{5}} \ \underset{11\ 12}{\mathbf{6}} \}$



Partition $\mathcal{P}_0$: 3 cycle contradictions and singletons

Part freq.	Adjacencies
4	6.0(4)
3	3.4(3)
3	4.0(3)
3	0.5(3)
2	**-5.6(1), 5.6(2)**
2	2.3(2)
2	0.1(2)
2	**1.2(2), 2.1(1)**
1	**1.3(1), 3.1(1)**
1	0.3(1)
1	1.4(1)
1	4.2(1)
1	2.-5(1)
1	0.6(1)
1	0.-5(1)
1	0.2(1)

Groups of adjacencies, continued

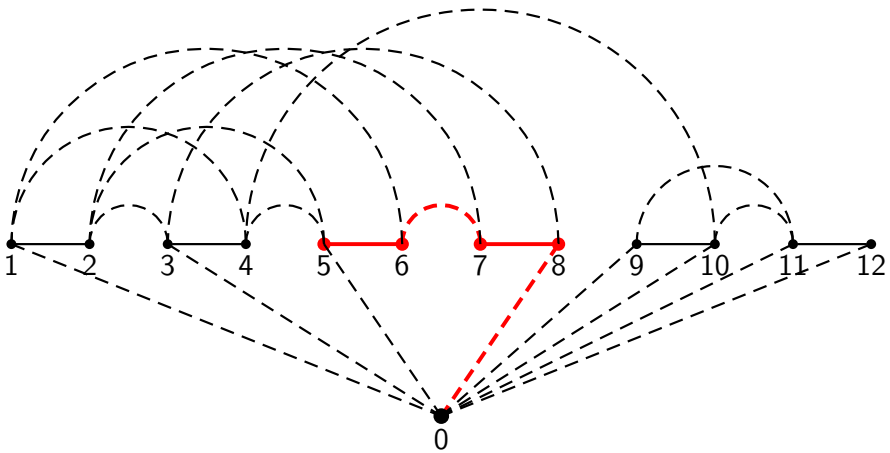
Example

$G_1 = \{ \underset{1\ 2}{1} \ \underset{3\ 4}{2} \ \underset{5\ 6}{3} \ \underset{7\ 8,}{4,} \ \underset{9\ 10}{5} \ \underset{11\ 12}{6} \}$

$G_2 = \{ \underset{1\ 2}{1} \ \underset{3\ 4}{2} \ \underset{5\ 6}{3} \ \underset{7\ 8,}{4,} \ \underset{10\ 9}{-5} \ \underset{11\ 12}{6} \}$

$G_3 = \{ \underset{5\ 6}{3} \ \underset{1\ 2}{1} \ \underset{7\ 8}{4} \ \underset{3\ 4}{2} \ \underset{10\ 9,}{-5,} \ \underset{11\ 12}{6} \}$

$G_4 = \{ \underset{3\ 4}{2} \ \underset{1\ 2}{1} \ \underset{5\ 6}{3} \ \underset{7\ 8,}{4,} \ \underset{9\ 10}{5} \ \underset{11\ 12}{6} \}$



Complementary adjacencies 3.4 and 4.0 (same part frequencies)

Part freq.	Adjacencies
4	6.0(4)
3	**3.4(3)**
3	**4.0(3)**
3	0.5(3)
2	-5.6(1), 5.6(2)
2	2.3(2)
2	0.1(2)
2	1.2(2), 2.1(1)
1	1.3(1), 3.1(1)
1	0.3(1)
1	1.4(1)
1	4.2(1)
1	2.-5(1)
1	0.6(1)
1	0.-5(1)
1	0.2(1)

Groups of adjacencies, continued

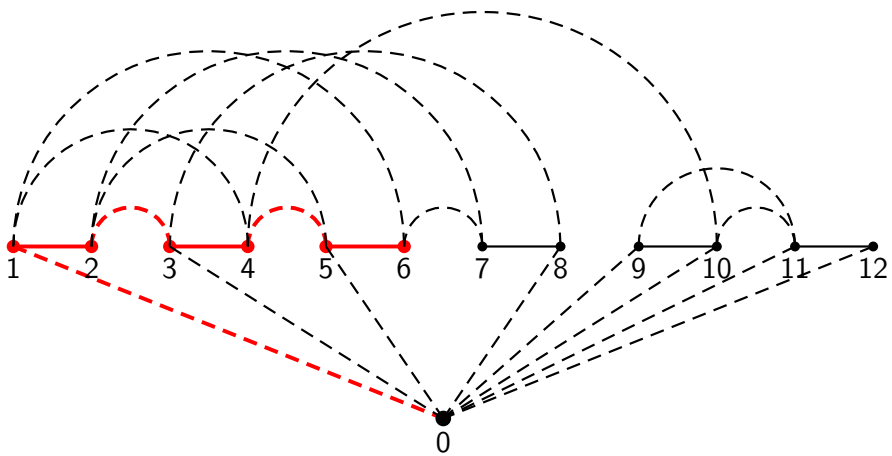
Example

$G_1 = \{ \underset{1\ 2}{\mathbf{1}} \ \underset{3\ 4}{\mathbf{2}} \ \underset{5\ 6}{\mathbf{3}} \ \underset{7\ 8,}{\mathbf{4,}} \ \underset{9\ 10}{\mathbf{5}} \ \underset{11\ 12}{\mathbf{6}} \}$

$G_2 = \{ \underset{1\ 2}{\mathbf{1}} \ \underset{3\ 4}{\mathbf{2}} \ \underset{5\ 6}{\mathbf{3}} \ \underset{7\ 8,}{\mathbf{4,}} \ \underset{10\ 9}{\mathbf{-5}} \ \underset{11\ 12}{\mathbf{6}} \}$

$G_3 = \{ \underset{5\ 6}{\mathbf{3}} \ \underset{1\ 2}{\mathbf{1}} \ \underset{7\ 8}{\mathbf{4}} \ \underset{3\ 4}{\mathbf{2}} \ \underset{10\ 9,}{\mathbf{-5,}} \ \underset{11\ 12}{\mathbf{6}} \}$

$G_4 = \{ \underset{3\ 4}{\mathbf{2}} \ \underset{1\ 2}{\mathbf{1}} \ \underset{5\ 6}{\mathbf{3}} \ \underset{7\ 8,}{\mathbf{4,}} \ \underset{9\ 10}{\mathbf{5}} \ \underset{11\ 12}{\mathbf{6}} \}$



Complementary adjacencies 0.1, 1.2 and 2.3 (same part frequencies)

Part freq.	Adjacencies
4	6.0(4)
3	3.4(3), 4.0(3)
3	0.5(3)
2	-5.6(1), 5.6(2)
2	**2.3(2)**
2	**0.1(2)**
2	**1.2(2)**, 2.1(1)
1	1.3(1), 3.1(1)
1	0.3(1)
1	1.4(1)
1	4.2(1)
1	2.-5(1)
1	0.6(1)
1	0.-5(1)
1	0.2(1)

Groups of adjacencies, continued

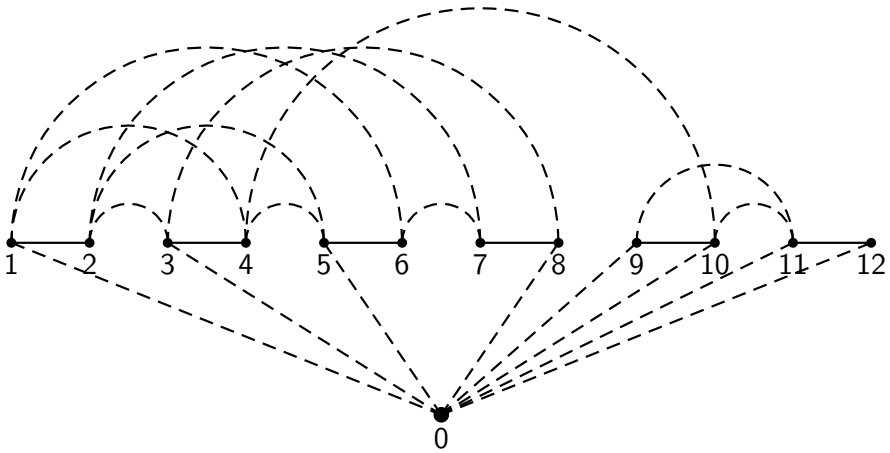
Example

$G_1 = \{ \underset{1\ 2}{\mathbf{1}} \ \underset{3\ 4}{\mathbf{2}} \ \underset{5\ 6}{\mathbf{3}} \ \underset{7\ 8,}{\mathbf{4,}} \ \underset{9\ 10}{\mathbf{5}} \ \underset{11\ 12}{\mathbf{6}} \}$

$G_2 = \{ \underset{1\ 2}{\mathbf{1}} \ \underset{3\ 4}{\mathbf{2}} \ \underset{5\ 6}{\mathbf{3}} \ \underset{7\ 8,}{\mathbf{4,}} \ \underset{10\ 9}{\mathbf{-5}} \ \underset{11\ 12}{\mathbf{6}} \}$

$G_3 = \{ \underset{5\ 6}{\mathbf{3}} \ \underset{1\ 2}{\mathbf{1}} \ \underset{7\ 8}{\mathbf{4}} \ \underset{3\ 4}{\mathbf{2}} \ \underset{10\ 9,}{\mathbf{-5,}} \ \underset{11\ 12}{\mathbf{6}} \}$

$G_4 = \{ \underset{3\ 4}{\mathbf{2}} \ \underset{1\ 2}{\mathbf{1}} \ \underset{5\ 6}{\mathbf{3}} \ \underset{7\ 8,}{\mathbf{4,}} \ \underset{9\ 10}{\mathbf{5}} \ \underset{11\ 12}{\mathbf{6}} \}$



Other complementary adjacencies and adjacencies in contradiction...

Part freq.	Adjacencies
4	6.0(4)
3	3.4(3), 4.0(3)
3	0.5(3)
2	-5.6(1), 5.6(2)
2	1.2(2), 2.1(1), 2.3(2), 0.1(2)
1	**1.3(1), 3.1(1)**
1	**0.3(1)**
1	**1.4(1)**
1	**4.2(1)**
1	**2.-5(1)**
1	0.6(1)
1	**0.-5(1)**
1	**0.2(1)**

Groups of adjacencies, continued

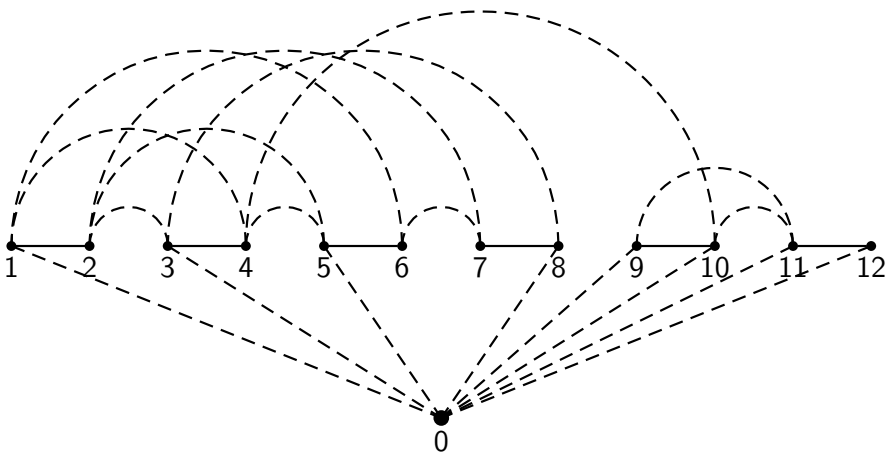
Example

$G_1 = \{ \underset{1\ 2}{\mathbf{1}} \ \underset{3\ 4}{\mathbf{2}} \ \underset{5\ 6}{\mathbf{3}} \ \underset{7\ 8,}{\mathbf{4,}} \ \underset{9\ 10}{\mathbf{5}} \ \underset{11\ 12}{\mathbf{6}} \}$

$G_2 = \{ \underset{1\ 2}{\mathbf{1}} \ \underset{3\ 4}{\mathbf{2}} \ \underset{5\ 6}{\mathbf{3}} \ \underset{7\ 8,}{\mathbf{4,}} \ \underset{10\ 9}{\mathbf{-5}} \ \underset{11\ 12}{\mathbf{6}} \}$

$G_3 = \{ \underset{5\ 6}{\mathbf{3}} \ \underset{1\ 2}{\mathbf{1}} \ \underset{7\ 8}{\mathbf{4}} \ \underset{3\ 4}{\mathbf{2}} \ \underset{10\ 9,}{\mathbf{-5,}} \ \underset{11\ 12}{\mathbf{6}} \}$

$G_4 = \{ \underset{3\ 4}{\mathbf{2}} \ \underset{1\ 2}{\mathbf{1}} \ \underset{5\ 6}{\mathbf{3}} \ \underset{7\ 8,}{\mathbf{4,}} \ \underset{9\ 10}{\mathbf{5}} \ \underset{11\ 12}{\mathbf{6}} \}$



Group freq.	Adjacencies
4	6.0(4)
3	3.4(3), 4.0(3)
3	0.5(3)
2	-5.6(1), 5.6(2)
2	1.2(2), 2.1(1), 2.3(2), 0.1(2)
1	1.3(1), 3.1(1), 0.3(1), 1.4(1), 4.2(1), 2.-5(1), 0.-5(1), 0.2(1)
1	0.6(1)

Final groups of $\sqcup^n(\mathcal{P}_0)$

Super-blocks (intuition part 2)

Definition

A **super-block** is a set S of $n \geq 1$ adjacencies s.t. $\forall a, b \in S$, a does not contradict b, and there exists an order over S such that $\forall i \in [1, n)$, a_i complements a_{i+1}, and a_1, a_n are either independent or $a_1 = a_n = 0$. A **partial assembly** $\mathcal{P} = \{S_k\}$ is a set of superblocks such that $\forall k \neq l$ if $S_k \cap S_l \neq \emptyset \Rightarrow S_k \cap S_l = \{0\}$.

Lemma

The adjacency graph $G = (V, E)$ of a partial assembly $\mathcal{P}$ is a graph such that (1) $\forall v \in V$, $d(v) \leq 2$, except for $v = 0$, and (2) any cycle in G contains 0.

Super-block construction

Example

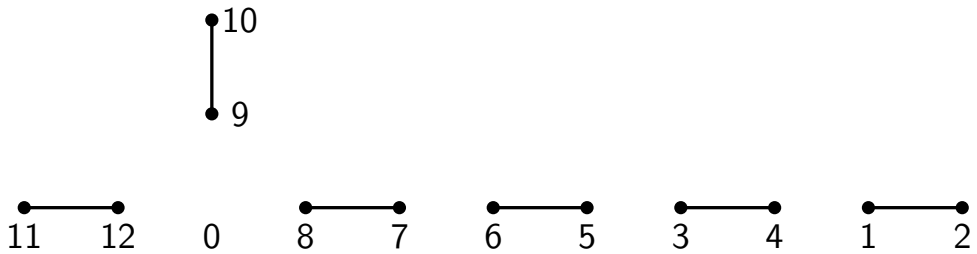
$G_1 = \{$ **1** **2** **3** **4,** **5** **6** $\}$ (1 2 3 4 5 6 7 8, 9 10 11 12)

$G_2 = \{$ **1** **2** **3** **4,** **-5** **6** $\}$ (1 2 3 4 5 6 7 8, 10 9 11 12)

$G_3 = \{$ **3** **1** **4** **2** **-5,** **6** $\}$ (5 6 1 2 7 8 3 4 10 9, 11 12)

$G_4 = \{$ **2** **1** **3** **4,** **5** **6** $\}$ (3 4 1 2 5 6 7 8, 9 10 11 12)

Group freq.	Adjacencies
4	6.0(4)
3	3.4(3), 4.0(3)
3	0.5(3)
2	-5.6(1), 5.6(2)
2	1.2(2), 2.1(1), 2.3(2), 0.1(2)
1	1.3(1), 3.1(1), 0.3(1), 1.4(1), 4.2(1), 2.-5(1), 0.-5(1), 0.2(1)
1	0.6(1)



Initial graph

Super-block construction

Example

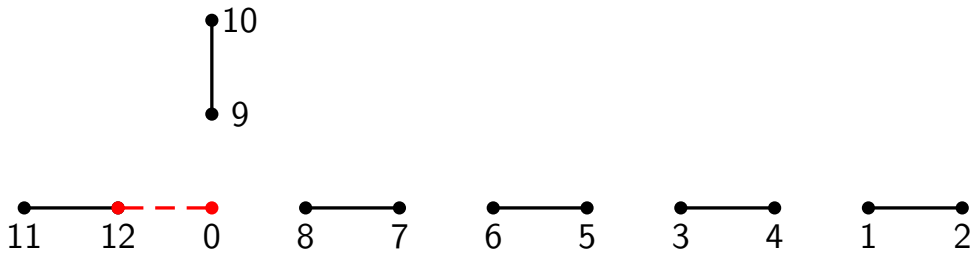
$G_1 = \{ \underset{1\ 2}{\mathbf{1}} \ \underset{3\ 4}{\mathbf{2}} \ \underset{5\ 6}{\mathbf{3}} \ \underset{7\ 8,}{\mathbf{4,}} \ \underset{9\ 10}{\mathbf{5}} \ \underset{11\ 12}{\mathbf{6}} \}$

$G_2 = \{ \underset{1\ 2}{\mathbf{1}} \ \underset{3\ 4}{\mathbf{2}} \ \underset{5\ 6}{\mathbf{3}} \ \underset{7\ 8,}{\mathbf{4,}} \ \underset{10\ 9}{\mathbf{-5}} \ \underset{11\ 12}{\mathbf{6}} \}$

$G_3 = \{ \underset{5\ 6}{\mathbf{3}} \ \underset{1\ 2}{\mathbf{1}} \ \underset{7\ 8}{\mathbf{4}} \ \underset{3\ 4}{\mathbf{2}} \ \underset{10\ 9,}{\mathbf{-5,}} \ \underset{11\ 12}{\mathbf{6}} \}$

$G_4 = \{ \underset{3\ 4}{\mathbf{2}} \ \underset{1\ 2}{\mathbf{1}} \ \underset{5\ 6}{\mathbf{3}} \ \underset{7\ 8,}{\mathbf{4,}} \ \underset{9\ 10}{\mathbf{5}} \ \underset{11\ 12}{\mathbf{6}} \}$

Group freq.	Adjacencies
4	**6.0(4)**
3	3.4(3), 4.0(3)
3	0.5(3)
2	-5.6(1), 5.6(2)
2	1.2(2), 2.1(1), 2.3(2), 0.1(2)
1	1.3(1), 3.1(1), 0.3(1), 1.4(1), 4.2(1), 2.-5(1), 0.-5(1), 0.2(1)
1	0.6(1)



Adding groups by decreasing frequency

$6.0 = (11\ 12).0$

Super-block construction

Example

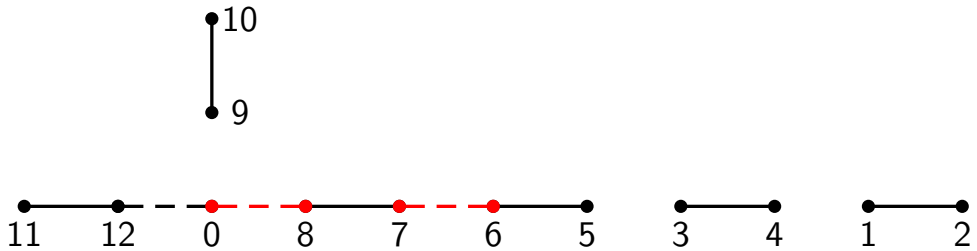
$G_1 = \{$ **1** **2** **3** **4,** **5** **6** $\}$
1 2 3 4 5 6 7 8, 9 10 11 12

$G_2 = \{$ **1** **2** **3** **4,** **-5** **6** $\}$
1 2 3 4 5 6 7 8, 10 9 11 12

$G_3 = \{$ **3** **1** **4** **2** **-5,** **6** $\}$
5 6 1 2 7 8 3 4 10 9, 11 12

$G_4 = \{$ **2** **1** **3** **4,** **5** **6** $\}$
3 4 1 2 5 6 7 8, 9 10 11 12

Group freq.	Adjacencies
4	6.0(4)
3	**3.4(3), 4.0(3)**
3	0.5(3)
2	-5.6(1), 5.6(2)
2	1.2(2), 2.1(1), 2.3(2), 0.1(2)
1	1.3(1), 3.1(1), 0.3(1), 1.4(1), 4.2(1), 2.-5(1), 0.-5(1), 0.2(1)
1	0.6(1)



$3.4 = (5\ 6).(7\ 8)$ and $4.0 = (7\ 8).0$

Super-block construction

Example

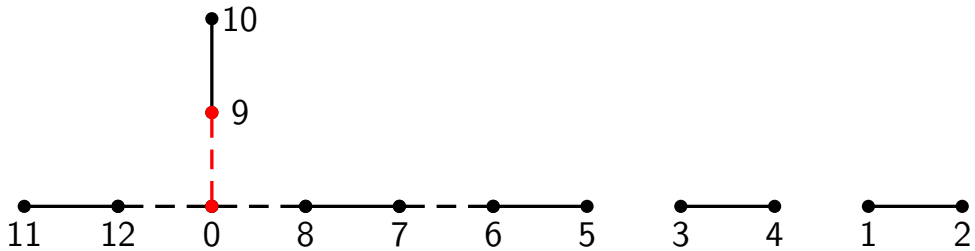
$G_1 = \{$ **1** **2** **3** **4,** **5** **6** $\}$ (1 2 3 4 5 6 7 8, 9 10 11 12)

$G_2 = \{$ **1** **2** **3** **4,** **-5** **6** $\}$ (1 2 3 4 5 6 7 8, 10 9 11 12)

$G_3 = \{$ **3** **1** **4** **2** **-5,** **6** $\}$ (5 6 1 2 7 8 3 4 10 9, 11 12)

$G_4 = \{$ **2** **1** **3** **4,** **5** **6** $\}$ (3 4 1 2 5 6 7 8, 9 10 11 12)

Group freq.	Adjacencies
4	6.0(4)
3	3.4(3), 4.0(3)
3	**0.5(3)**
2	-5.6(1), 5.6(2)
2	1.2(2), 2.1(1), 2.3(2), 0.1(2)
1	1.3(1), 3.1(1), 0.3(1), 1.4(1), 4.2(1), 2.-5(1), 0.-5(1), 0.2(1)
1	0.6(1)



$0.5 = 0.(9\ 10)$

Super-block construction

Example

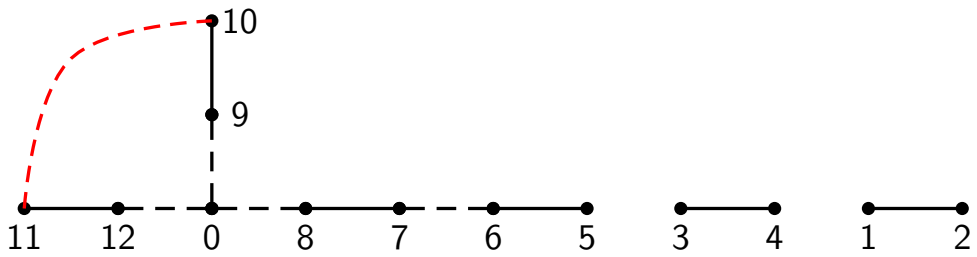
$G_1 = \{ \underset{1\ 2}{\mathbf{1}} \ \underset{3\ 4}{\mathbf{2}} \ \underset{5\ 6}{\mathbf{3}} \ \underset{7\ 8,}{\mathbf{4,}} \ \underset{9\ 10}{\mathbf{5}} \ \underset{11\ 12}{\mathbf{6}} \}$

$G_2 = \{ \underset{1\ 2}{\mathbf{1}} \ \underset{3\ 4}{\mathbf{2}} \ \underset{5\ 6}{\mathbf{3}} \ \underset{7\ 8,}{\mathbf{4,}} \ \underset{10\ 9}{\mathbf{-5}} \ \underset{11\ 12}{\mathbf{6}} \}$

$G_3 = \{ \underset{5\ 6}{\mathbf{3}} \ \underset{1\ 2}{\mathbf{1}} \ \underset{7\ 8}{\mathbf{4}} \ \underset{3\ 4}{\mathbf{2}} \ \underset{10\ 9,}{\mathbf{-5,}} \ \underset{11\ 12}{\mathbf{6}} \}$

$G_4 = \{ \underset{3\ 4}{\mathbf{2}} \ \underset{1\ 2}{\mathbf{1}} \ \underset{5\ 6}{\mathbf{3}} \ \underset{7\ 8,}{\mathbf{4,}} \ \underset{9\ 10}{\mathbf{5}} \ \underset{11\ 12}{\mathbf{6}} \}$

Group freq.	Adjacencies
4	6.0(4)
3	3.4(3), 4.0(3)
3	0.5(3)
2	**-5.6(1), 5.6(2)**
2	1.2(2), 2.1(1), 2.3(2), 0.1(2)
1	1.3(1), 3.1(1), 0.3(1), 1.4(1), 4.2(1), 2.-5(1), 0.-5(1), 0.2(1)
1	0.6(1)



$-5.6 = (10\ 9).(11\ 12)$ in conflict with $0.5 = 0.(9\ 10)$

$\Rightarrow$ Only addition of $5.6 = (9\ 10).(11\ 12)$

Super-block construction

Example

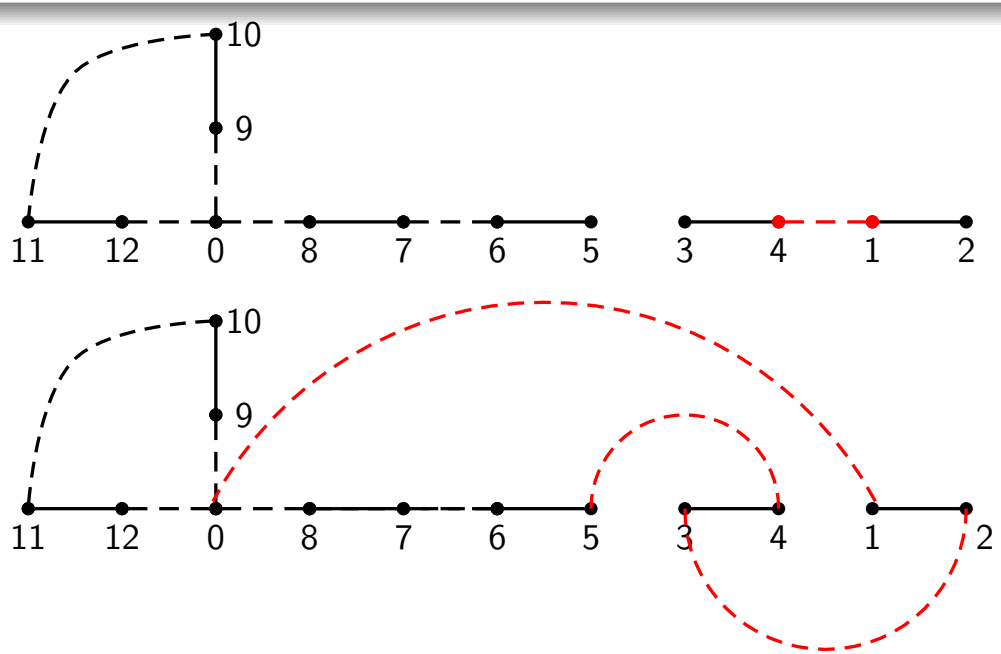
$G_1 = \{$ 1 2 3 4, 5 6 $\}$ (1 2 3 4 5 6 7 8, 9 10 11 12)

$G_2 = \{$ 1 2 3 4, -5 6 $\}$ (1 2 3 4 5 6 7 8, 10 9 11 12)

$G_3 = \{$ 3 1 4 2 -5, 6 $\}$ (5 6 1 2 7 8 3 4 10 9, 11 12)

$G_4 = \{$ 2 1 3 4, 5 6 $\}$ (3 4 1 2 5 6 7 8, 9 10 11 12)

Group freq.	Adjacencies
4	6.0(4)
3	3.4(3), 4.0(3)
3	0.5(3)
2	-5.6(1), 5.6(2)
2	**1.2(2), 2.1(1), 2.3(2), 0.1(2)**
1	1.3(1), 3.1(1), 0.3(1), 1.4(1), 4.2(1), 2.-5(1), 0.-5(1), 0.2(1)
1	0.6(1)



2 possible sets :
$\{2.1\}$ or $\{0.1, 1.2, 2.3\}$

Super-block construction

Example

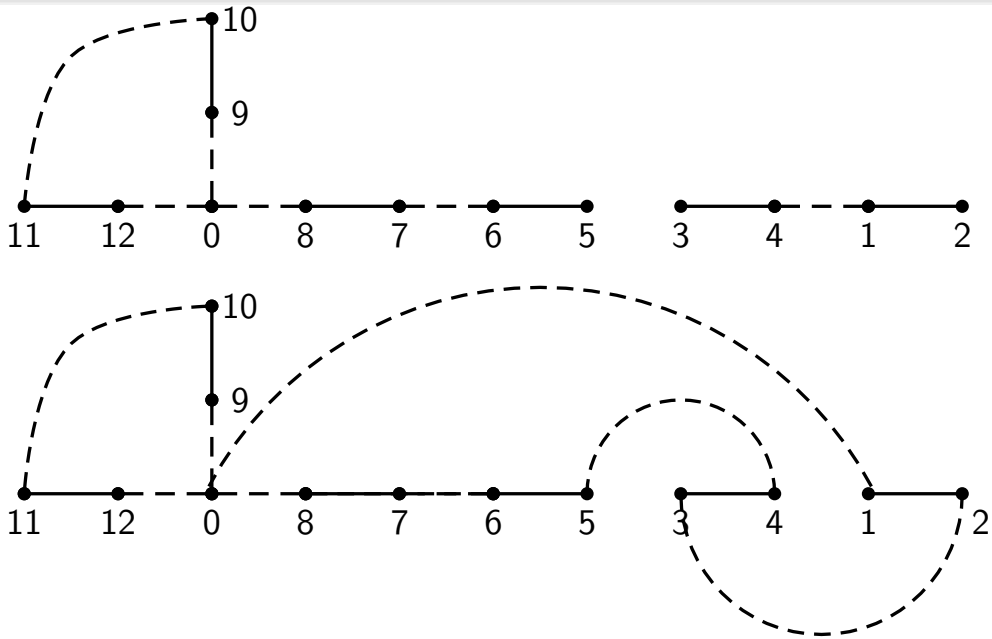
$G_1 = \{ \underset{1\ 2}{\mathbf{1}} \ \underset{3\ 4}{\mathbf{2}} \ \underset{5\ 6}{\mathbf{3}} \ \underset{7\ 8,}{\mathbf{4,}} \ \underset{9\ 10}{\mathbf{5}} \ \underset{11\ 12}{\mathbf{6}} \}$

$G_2 = \{ \underset{1\ 2}{\mathbf{1}} \ \underset{3\ 4}{\mathbf{2}} \ \underset{5\ 6}{\mathbf{3}} \ \underset{7\ 8,}{\mathbf{4,}} \ \underset{10\ 9}{\mathbf{-5}} \ \underset{11\ 12}{\mathbf{6}} \}$

$G_3 = \{ \underset{5\ 6}{\mathbf{3}} \ \underset{1\ 2}{\mathbf{1}} \ \underset{7\ 8}{\mathbf{4}} \ \underset{3\ 4}{\mathbf{2}} \ \underset{10\ 9,}{\mathbf{-5,}} \ \underset{11\ 12}{\mathbf{6}} \}$

$G_4 = \{ \underset{3\ 4}{\mathbf{2}} \ \underset{1\ 2}{\mathbf{1}} \ \underset{5\ 6}{\mathbf{3}} \ \underset{7\ 8,}{\mathbf{4,}} \ \underset{9\ 10}{\mathbf{5}} \ \underset{11\ 12}{\mathbf{6}} \}$

Group freq.	Adjacencies
4	6.0(4)
3	3.4(3), 4.0(3)
3	0.5(3)
2	-5.6(1), 5.6(2)
2	1.2(2), 2.1(1), 2.3(2), 0.1(2)
1	**1.3(1), 3.1(1), 0.3(1), 1.4(1), 4.2(1), 2.-5(1), 0.-5(1), 0.2(1)**
1	**0.6(1)**



Not enough support in groups of frequency 1

Super-block construction

Example

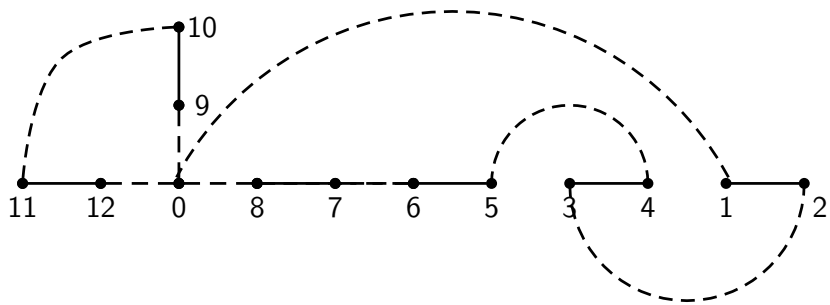
$G_1 = \{$ **1 2 3 4, 5 6** $\}$ (1 2 3 4 5 6 7 8, 9 10 11 12)

$G_2 = \{$ **1 2 3 4, -5 6** $\}$ (1 2 3 4 5 6 7 8, 10 9 11 12)

$G_3 = \{$ **3 1 4 2 -5, 6** $\}$ (5 6 1 2 7 8 3 4 10 9, 11 12)

$G_4 = \{$ **2 1 3 4, 5 6** $\}$ (3 4 1 2 5 6 7 8, 9 10 11 12)

Two partial assemblies :

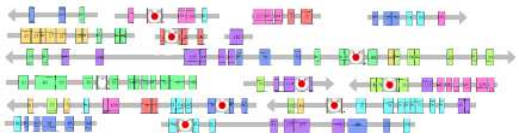


$M_1 = \{2\ 1,\ 3\ 4,\ 5\ 6\}$ having 3 super-blocks and $\sum d(M_1, G_i) = 10$

$M_2 = \{1\ 2\ 3\ 4,\ 5\ 6\}$ having 2 super-blocks and $\sum d(M_2, G_i) = 9$

Final assembly

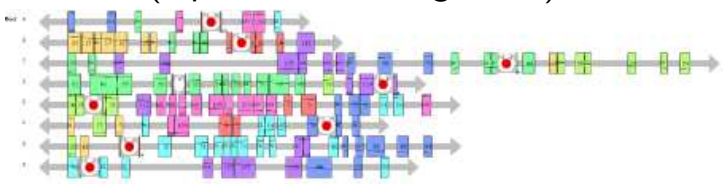
A partial assembly
(a set of super-blocks)



Fusions of super-blocks
=
solving extremity adjacencies
of super-blocks

biological constraint :
one and only one centromere
by chromosome

A final assembly
(a possible median genome)



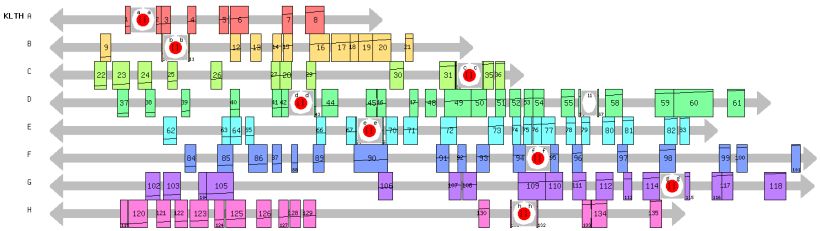
Theorem (Jean et al.)

For any $\mathcal{P} \in \{\mathcal{P}\}$ of $G_1, ..., G_N$ such that $\mathcal{P} = \{S_k\}$, there exists a genome M such that for any chromosome π of M either $\exists\, S_k \in \mathcal{P}$ such that $\pi = S_k$, or $\exists\, \{S_k\} \subseteq \mathcal{P}$ such that π is formed by a series of fusions $\pi = S_1 ... S_k$. Moreover,

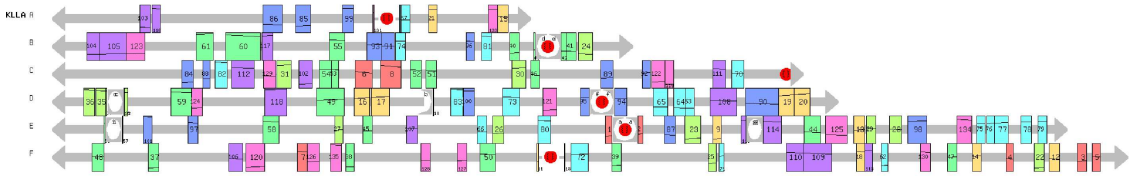
$$\sum_i^N d(M, G_i) - \sum_i^N d(P, G_i) \leq 0 \text{ and } \sum_i^N b(M, G_i) - \sum_i^N b(P, G_i) \leq 0.$$

Comparative maps

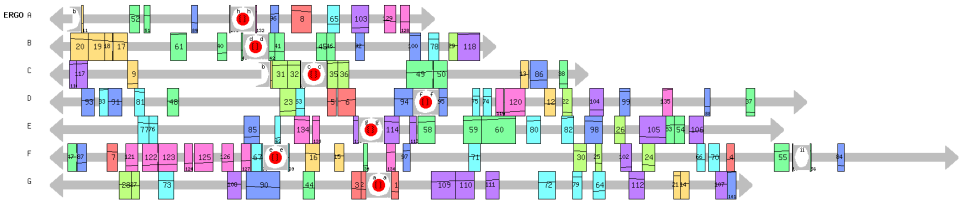
Klth



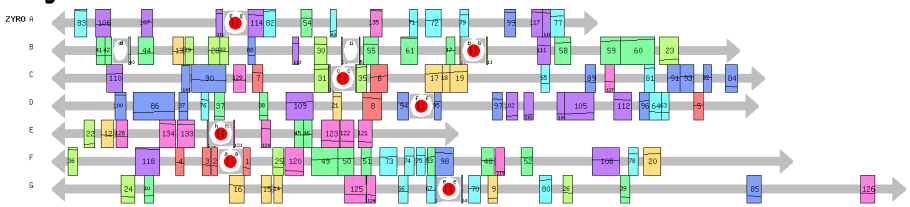
Klla



Ergo



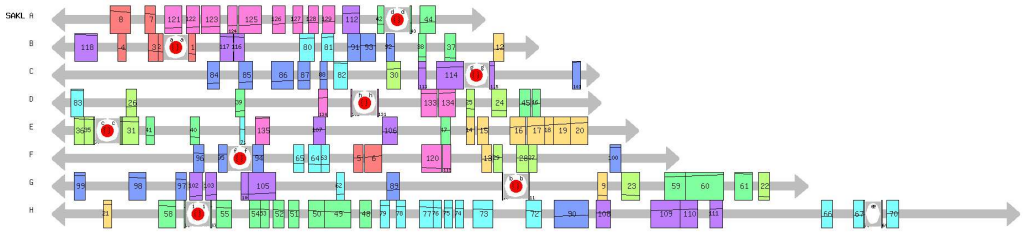
Zyro



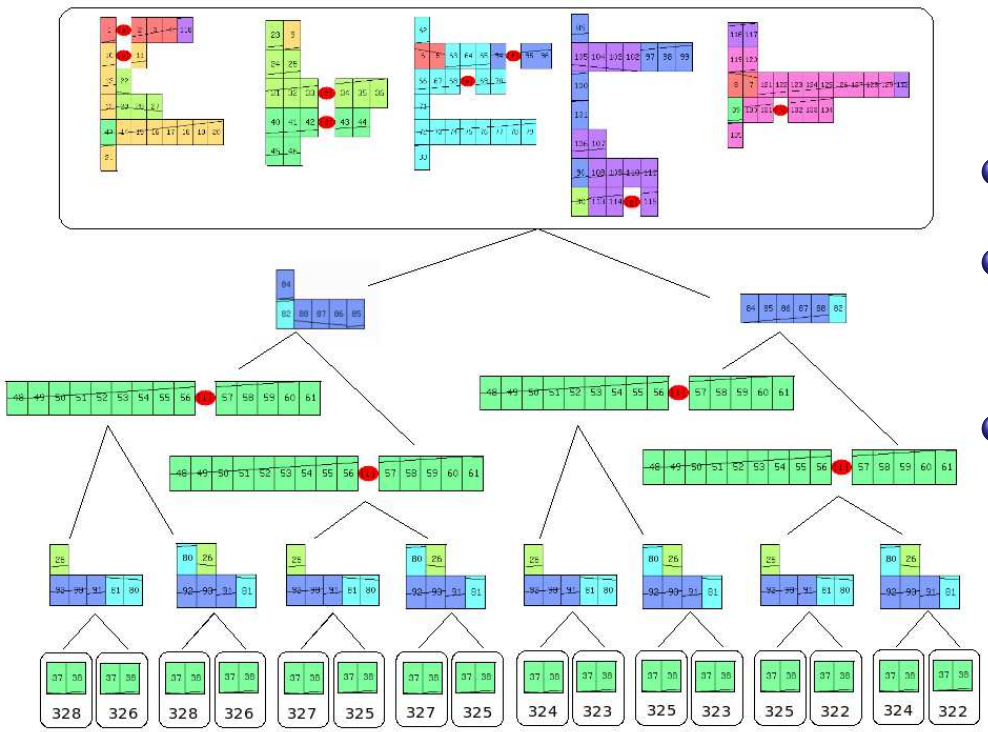
	Klth	Ergo	Klla	Sakl	Zyro
Klth	0	88	105	45	84
Ergo		0	109	85	101
Klla			0	98	115
Sakl				0	79
Zyro					0

Pairwise distances

Sakl

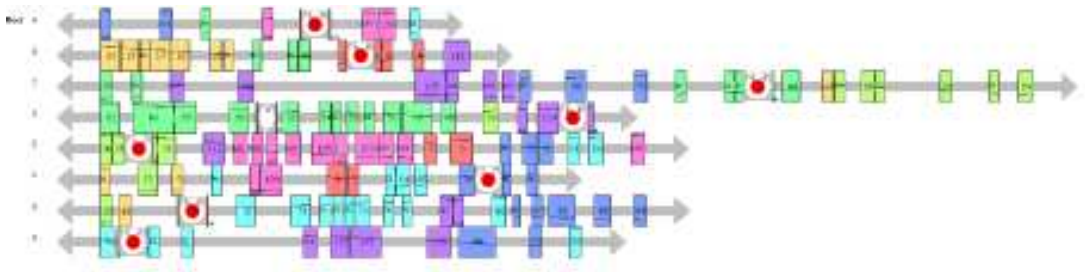


Sharing tree of super-blocks



- 1 branch = 1 partial assembly
- 16 partial assemblies $\longrightarrow$ 4 unsolved conflicts
- 29 common super-blocks on 34 or 35

- A possible final assembly
- $\sum_{i=1}^{5} d(M, G_i) = 284$



Constructing and visualizing parsimonious scenarios

HP theory (Hannenhalli & Pevzner 1995)

Definition

A parsimonious scenario is a sequence of rearrangements respecting rearrangement distance

- Unichromosomal case : exact reversal distance and parsimonious scenario in polynomial time
- Multichromosomal case : using unichromosomal theory by mimicking multichromosomal rearrangements by reversals on a single permutation

$$\Pi = \{< -1\ 2 >, < 3\ 4 >, < 5\ 8\ 7\ 6 >\}$$
$$\Gamma = \{< 1\ 2\ 3\ 4 >, < 5\ 6\ 7\ 8 >\}$$

Capping (caps)

Concatenate

$$\hat{\pi} = 9\ -1\ 2\ 10\ 11\ 3\ 4\ 12\ 13\ 5\ 8\ 7\ 6\ 14$$
$$\hat{\gamma} = 9\ 1\ 2\ 3\ 4\ 10\ 11\ 5\ 6\ 7\ 8\ 12\ 13\ 14$$

- Errors in distance and scenario : corrected by (Tesler, 2002) and by (Ozery-Flato & Shamir, 2003)
- Still one error to delineate chromosomes

Breakpoint graph

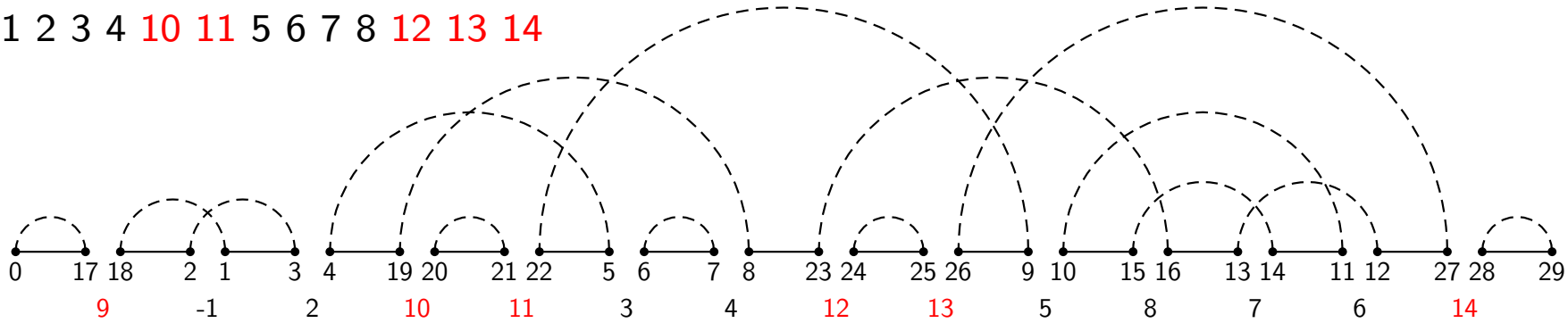
HP theory based on the **breakpoint graph** $G(\Pi, \Gamma)$

- Two vertices per signed identifier :

$2\pi_i - 1$ $+\pi_i$ $2\pi_i$ $\qquad$ $2\pi_i$ $-\pi_i$ $2\pi_i - 1$

- edges : adjacencies in Π (black edges) or Γ (dashed edges)

$\hat{\pi} = 9\ -1\ 2\ 10\ 11\ 3\ 4\ 12\ 13\ 5\ 8\ 7\ 6\ 14$
$\hat{\gamma} = 9\ 1\ 2\ 3\ 4\ 10\ 11\ 5\ 6\ 7\ 8\ 12\ 13\ 14$



Aim : increase the number of cycles

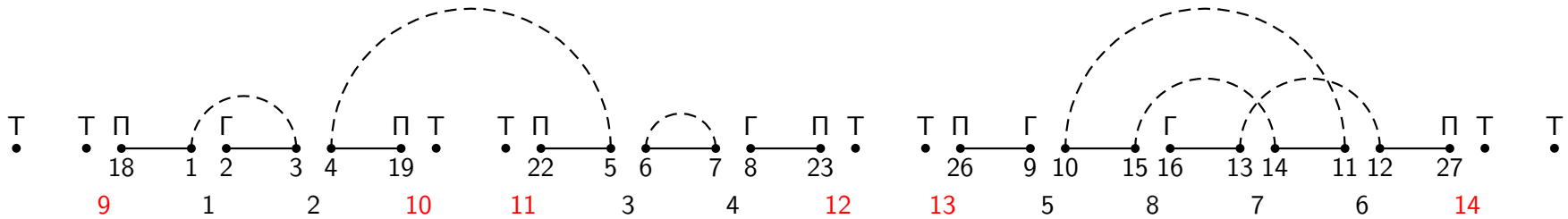
Breakpoint graph

HP theory based on the **breakpoint graph** $G(\Pi, \Gamma)$

- Two vertices per signed identifier :

- edges : adjacencies in Π (black edges) or Γ (dashed edges)

$\hat{\pi} = 9\ -1\ 2\ 10\ 11\ 3\ 4\ 12\ 13\ 5\ 8\ 7\ 6\ 14$
$\hat{\gamma} = 9\ 1\ 2\ 3\ 4\ 10\ 11\ 5\ 6\ 7\ 8\ 12\ 13\ 14$



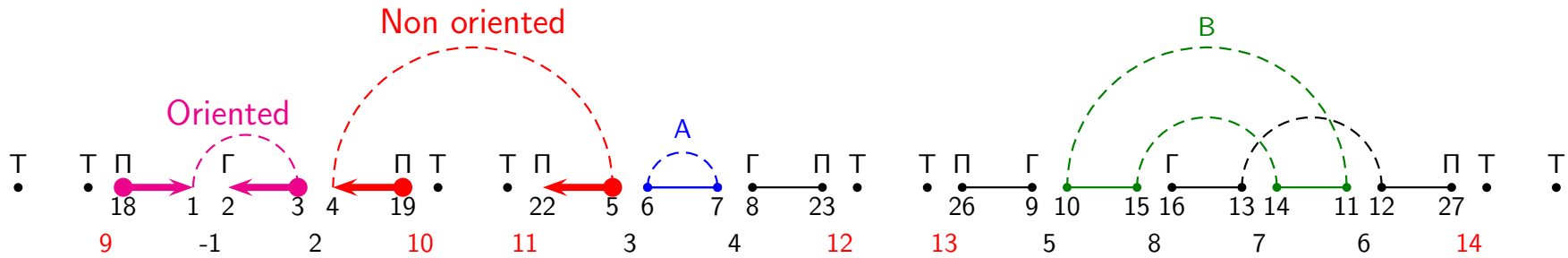
Distance computation independent of capping and concatenate
Particular vertices : Tails (T), Π-caps (Π) and Γ-tails (Γ)

Breakpoint graph

Cycle and path decomposition :

- ΠΠ-path, ΠΓ-path, ΓΓ-path.
- trivial cycles or paths (A) vs proper cycles or paths (B)

Orientation : arbitrary orientation of dark edges within a cycle or path



Number of rearrangements compared to eliminated breakpoints

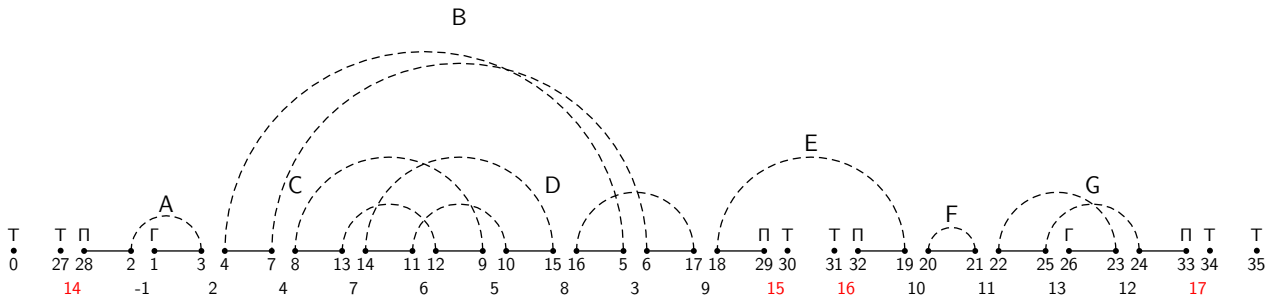
- 1 rearrangement for 1 or 2 breakpoints in oriented cycle
- 2 rearrangements for 1 or 2 breakpoints in non oriented cycle

Interleaving graph component classifications

Distance based on the breakpoint graph decomposition

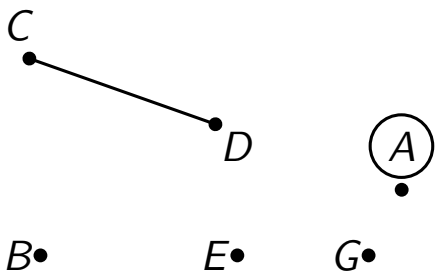
Incoherences and new notions in the literature

Clarification into a double classification (Jean & Nikolski 2007)



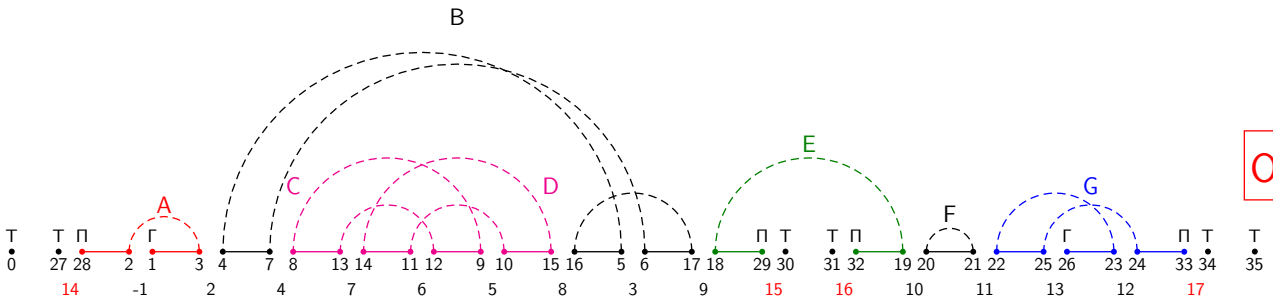
Connected components

Interleaving graph

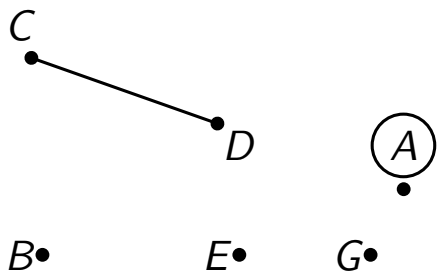


Interleaving graph component classifications

Proposition of a double classification

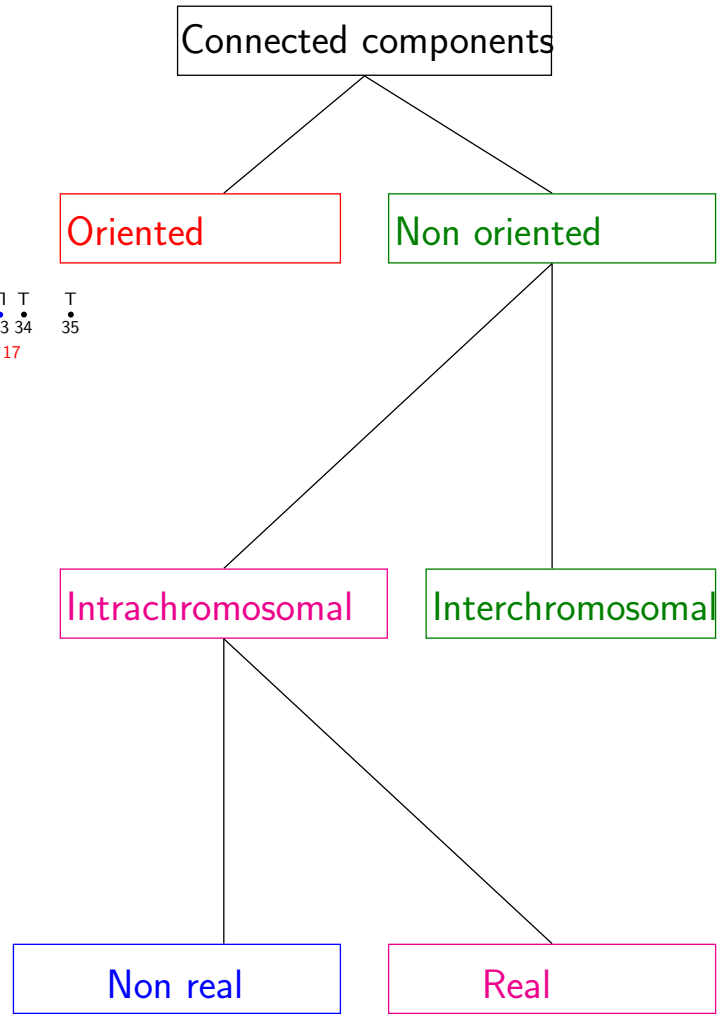


Interleaving graph



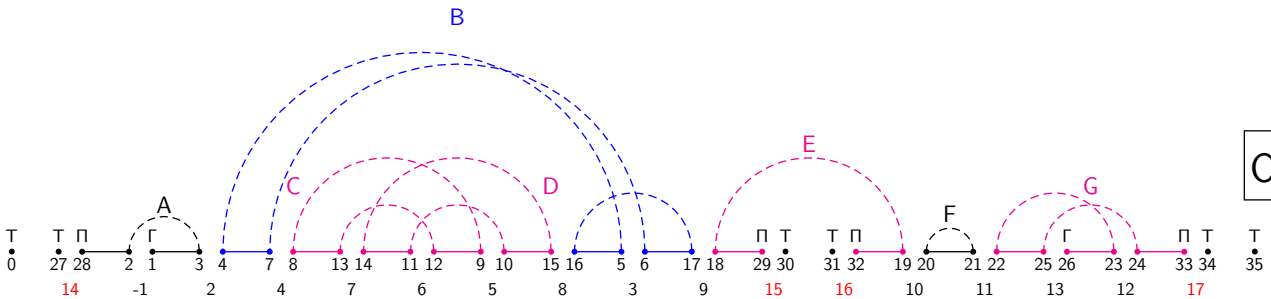
A component is real if it does not cover a Π-cap or a Γ-tail.

Intrinsic classification



Interleaving graph component classifications

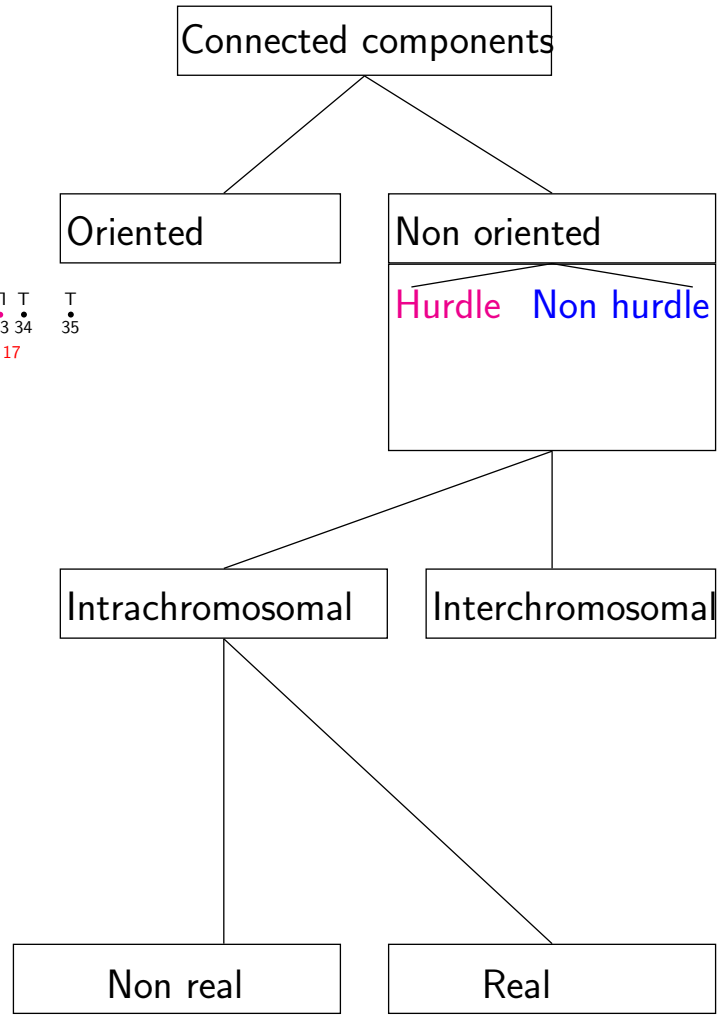
Proposition of a double classification



B separates $\{C, D\}$ from E or G

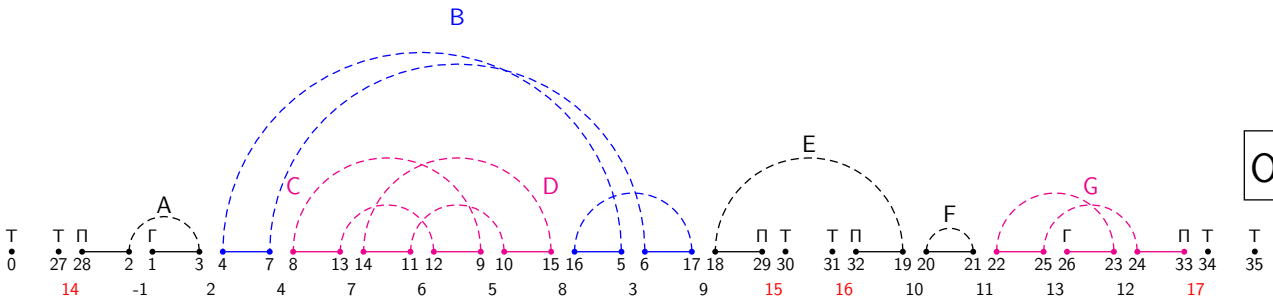
$\{C, D\}$ is a super hurdle
E and G are simple hurdles

Extrinsic classification



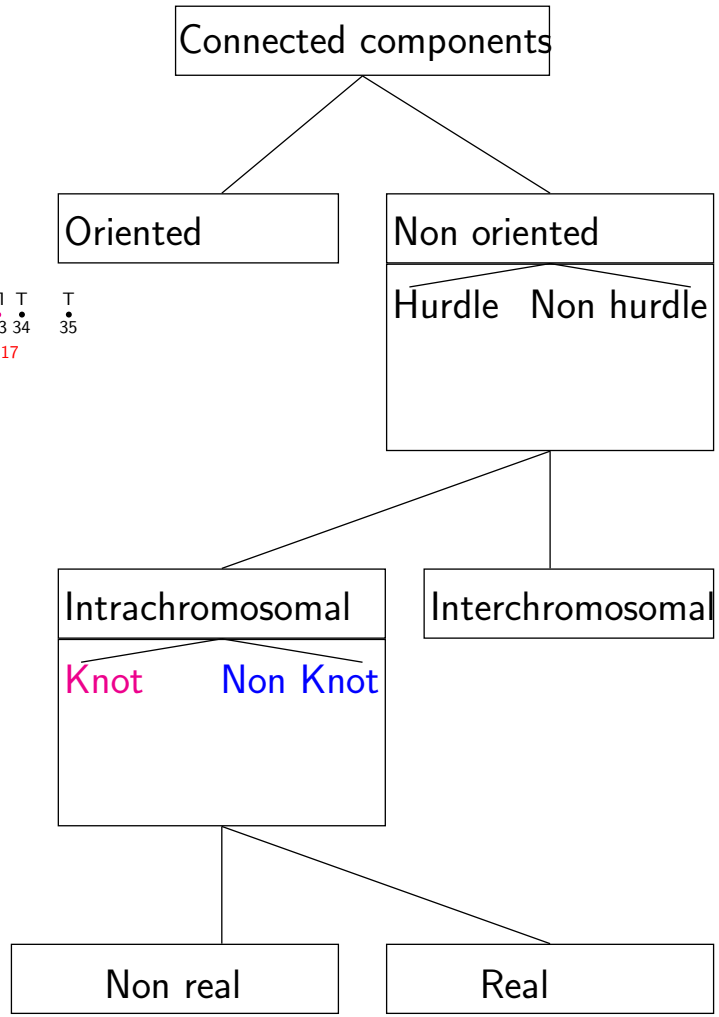
Interleaving graph component classifications

Proposition of a double classification



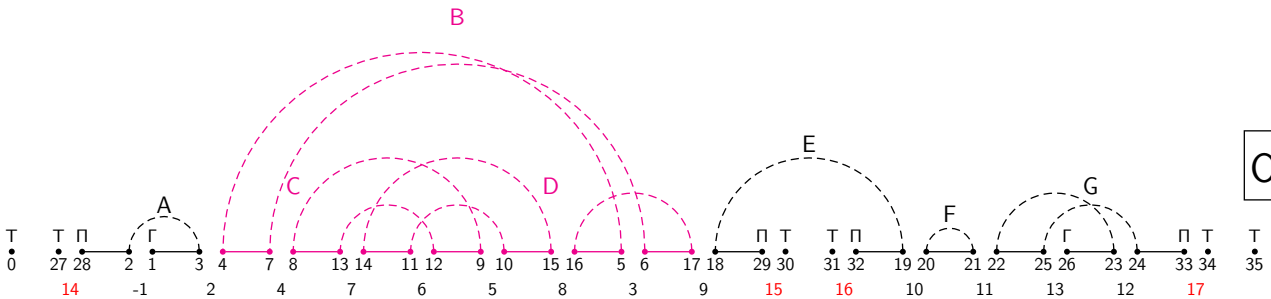
B separates $\{C, D\}$ from G

Extrinsic classification



Interleaving graph component classifications

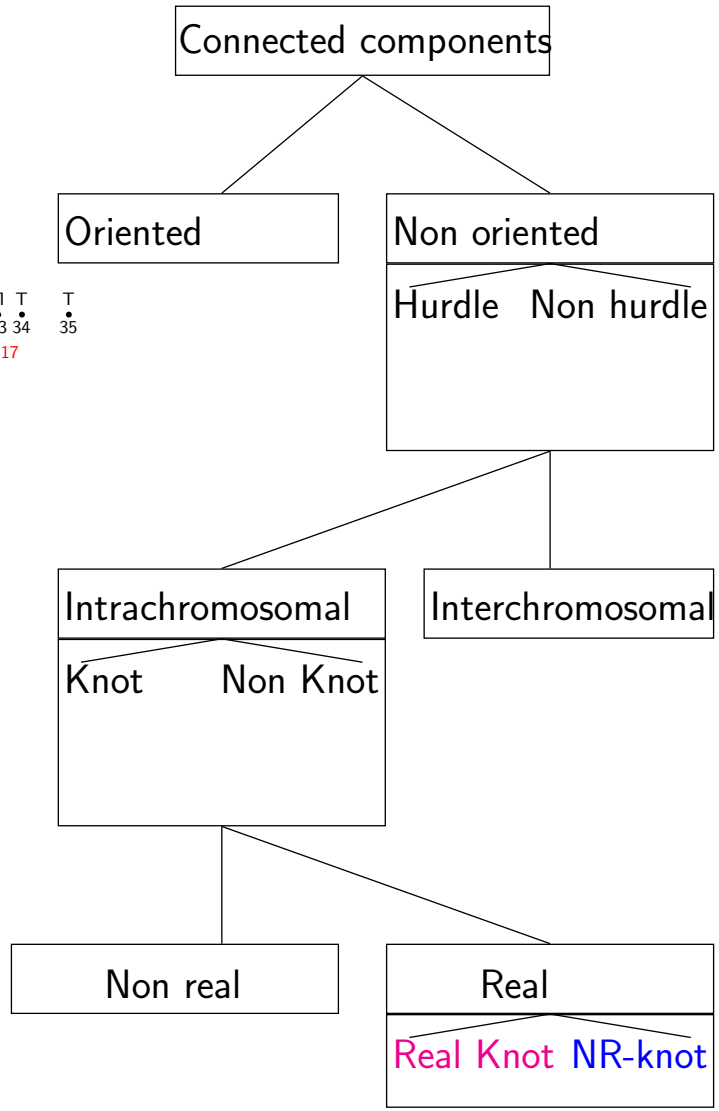
Proposition of a double classification



B is the greatest real-knot

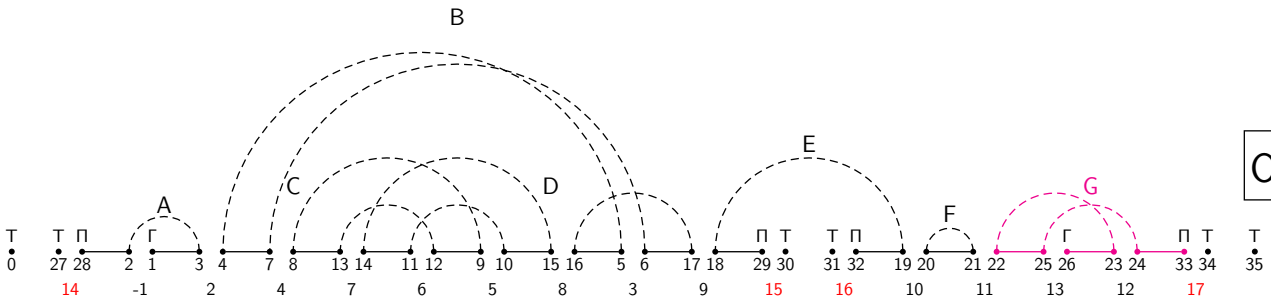
$\{C, D\}$ is a minimal real-knot

Extrinsic classification



Interleaving graph component classifications

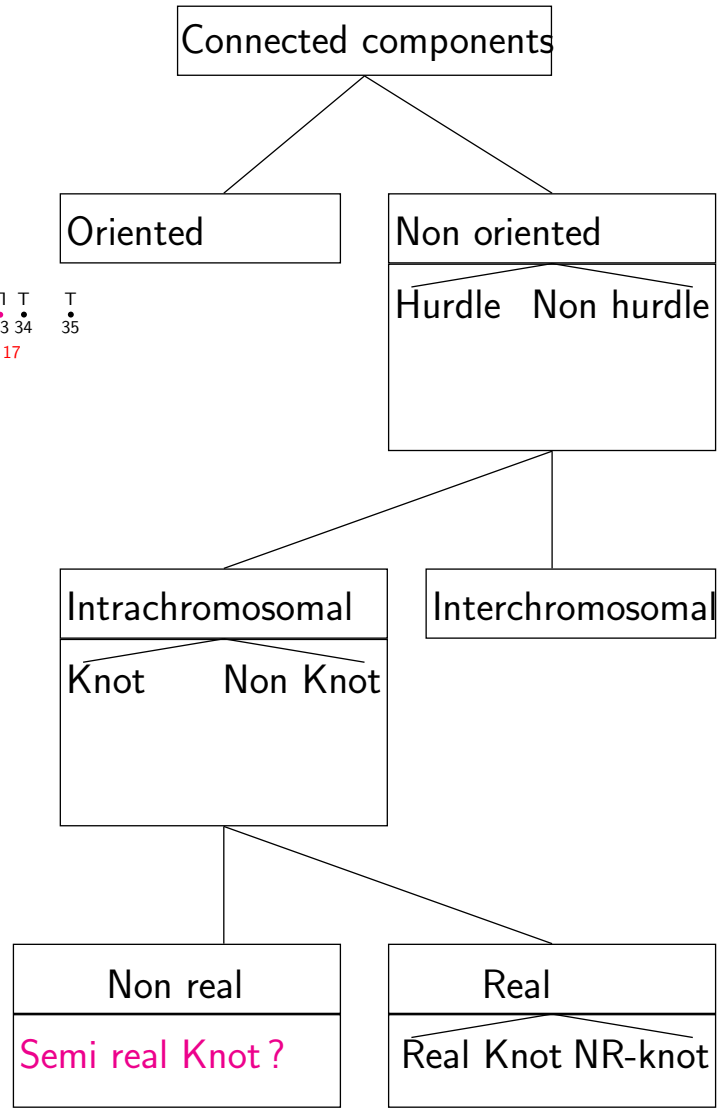
Proposition of a double classification



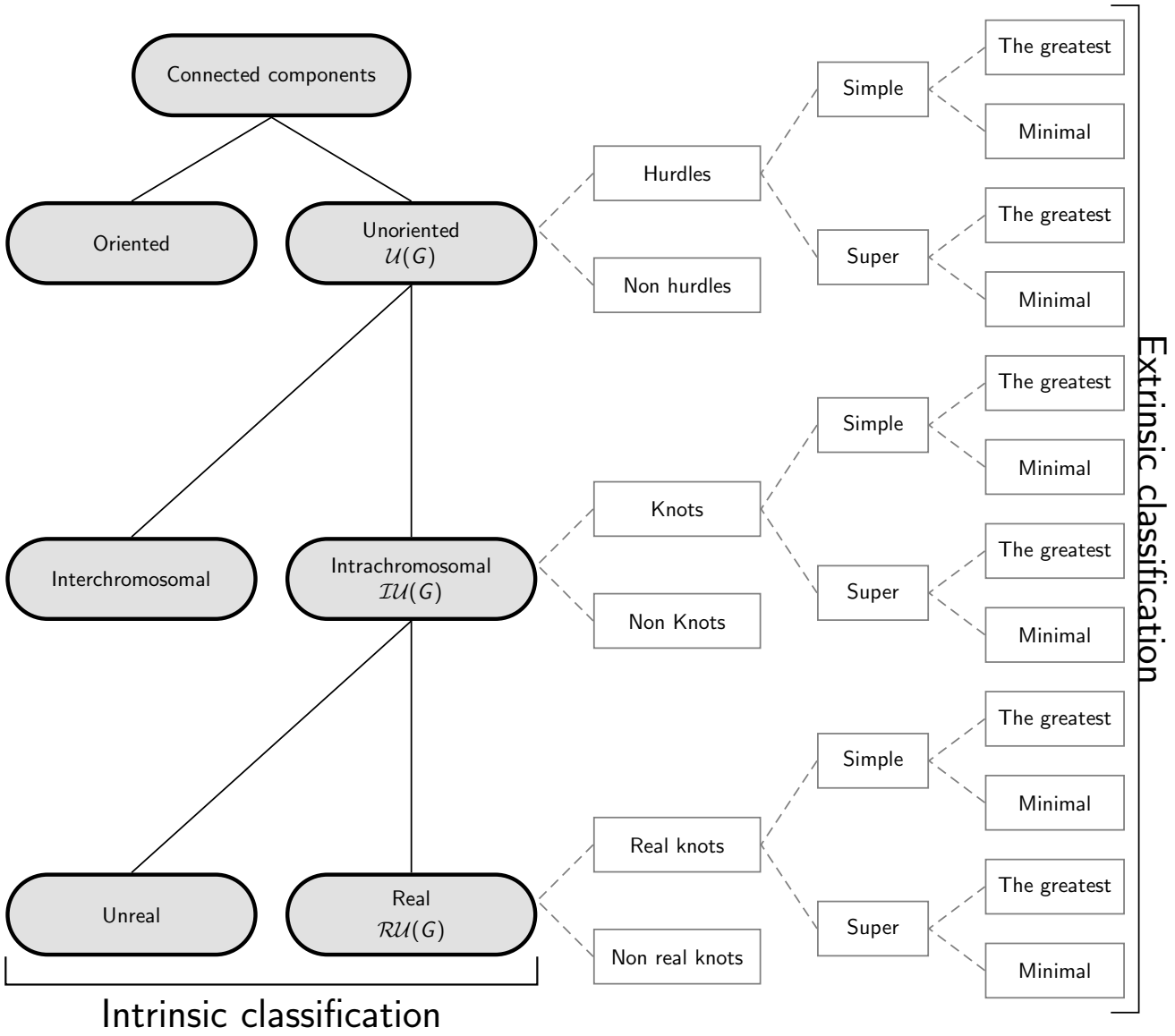
A **semi-real-knot** does not cover a ΓΓ-path and becomes a minimal real-knot or the simple greatest real knot after closing its ΠΓ-paths.

A **simple component** contains at least one ΠΓ-path and is not a semi-real-knot.

Extrinsic classification



Double classification



All of the configurations are encountered in distance and scenario computations

(Jean & Nikolski 2007)

Why optimal capping and concatenation are needed ?

Distance

arbitrary capping and concatenation

Scenario

sequence of rearrangements respecting minimal distance

optimal capping and concatenation

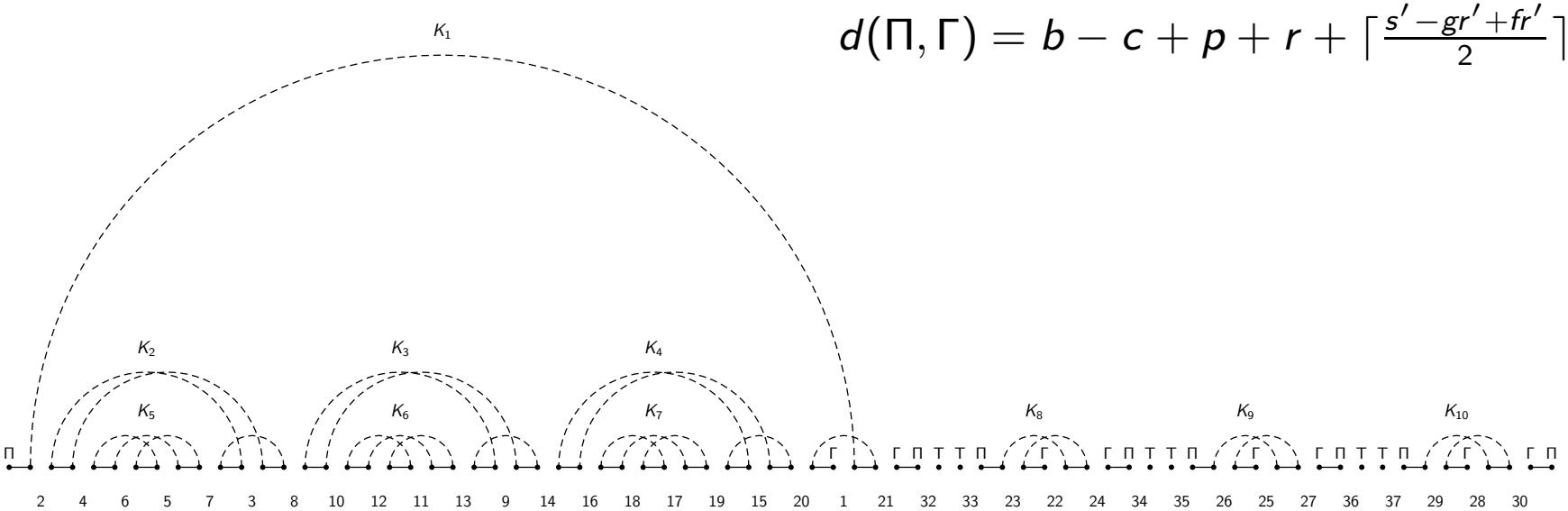
Reduction to the unichromosomal problem

- Reversals only
- Nonsense operations : chromosome flipping and cap reposition

Counterexample to the Ozery-Flato's and Shamir's algorithm (2003)

(Ozery-Flato & Shamir 2003)

$$d(\Pi,\Gamma) = b - c + p + r + \lceil \frac{s' - gr' + fr'}{2} \rceil$$

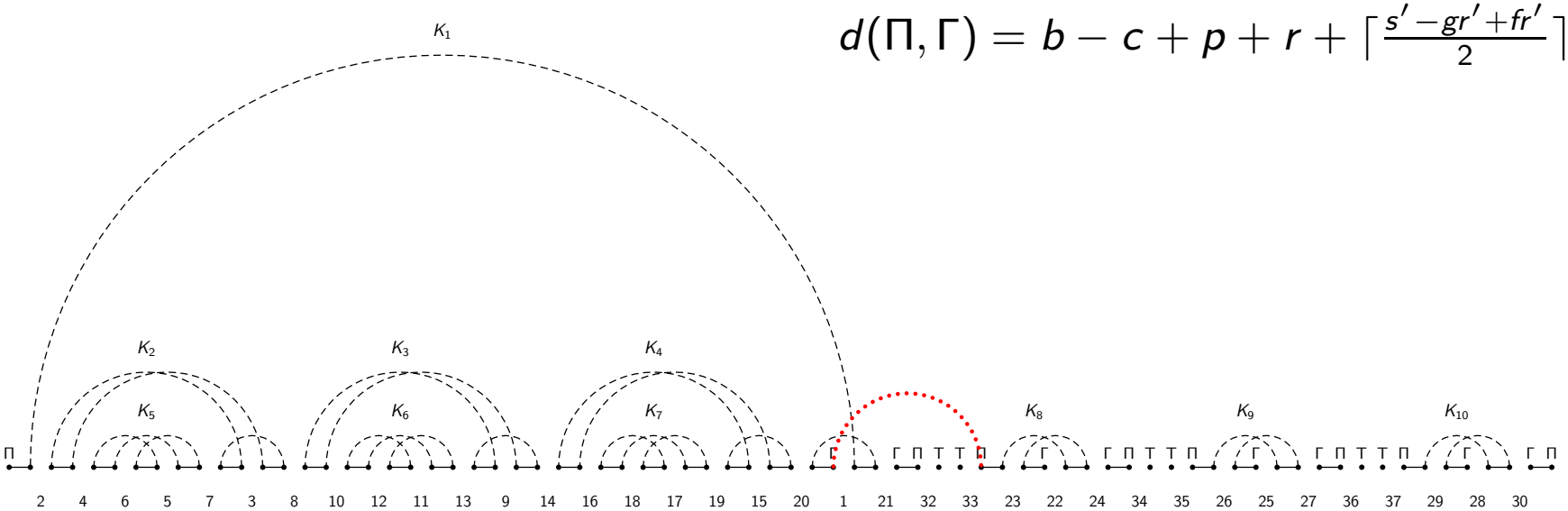


$$d(\Pi,\Gamma) = 34 - 14 + 0 + 3 + \lceil \frac{4-0+0}{2} \rceil = 25$$

Counterexample to the Ozery-Flato's and Shamir's algorithm (2003)

(Ozery-Flato & Shamir 2003)

$$d(\Pi,\Gamma) = b - c + p + r + \lceil \frac{s' - gr' + fr'}{2} \rceil$$



$$d(\Pi,\Gamma) = 34 - 14 + 0 + 3 + \lceil \frac{4-0+0}{2} \rceil = 25$$

$\Rightarrow$ cycle closure modifies the rearrangement distance so that it is no longer optimal

$$d = 34 - 13 + 0 + 3 + \lceil \frac{2-0+1}{2} \rceil = 26$$

A correct algorithm for optimal capping

Correct_Optimal_Capping

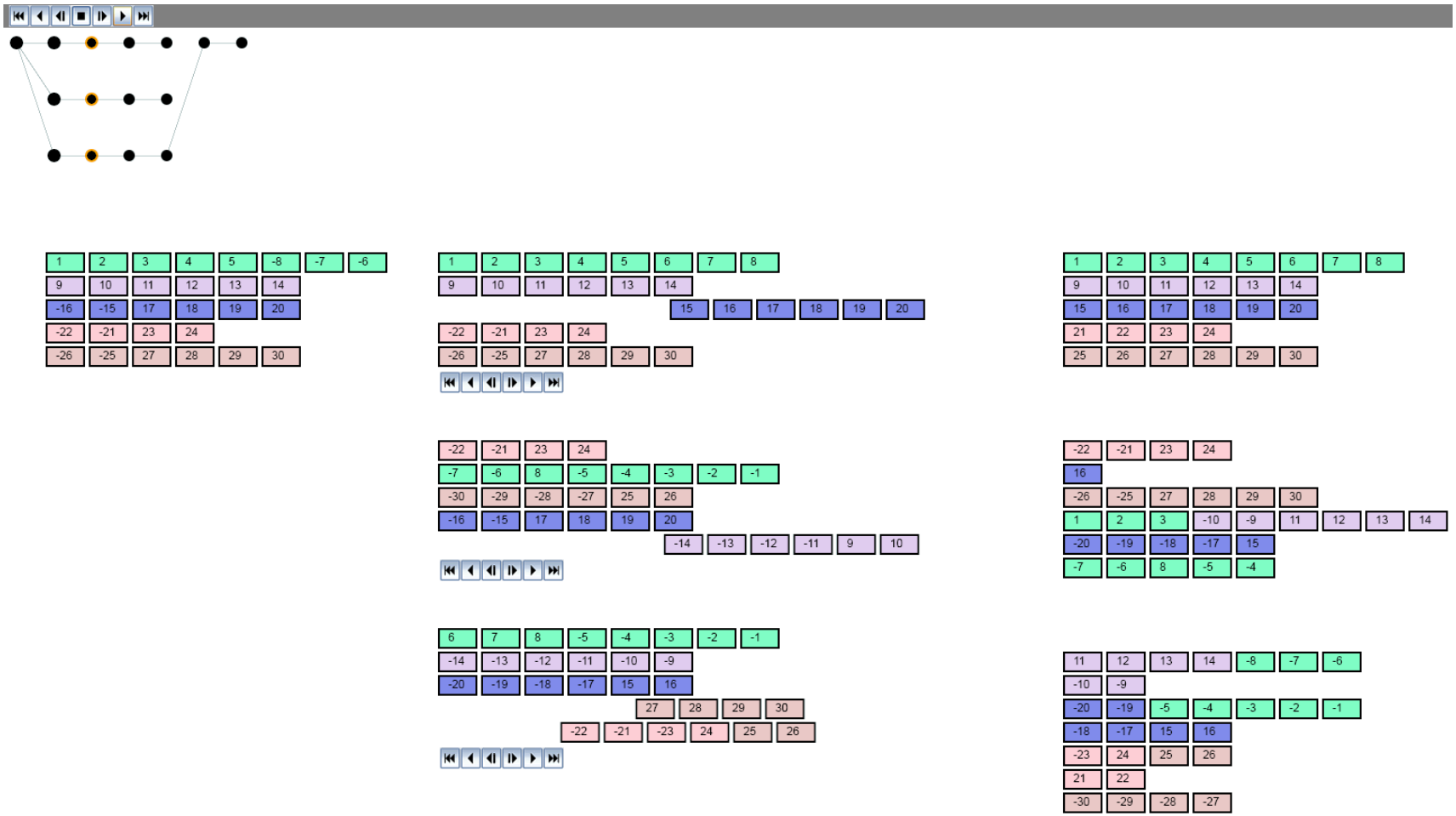
Construct the graph $G = G(\Pi, \Gamma)$
while there is a $\Gamma\Gamma$-path in G **do**
    Find an interchromosomal or oriented edge joining this $\Gamma\Gamma$-path with a $\Pi\Pi$-path and add it to G
end while
Close all remaining $\Pi\Pi$-paths in G
Close all $\Pi\Gamma$-paths in simple components in G
if s' is even and $s' \geq 2$ and G is a fortress-of-real-knots **then**
    if G has the semi-greatest-real-knot **then**
        Close all $\Pi\Gamma$-paths in the semi-greatest-real-knot
    else
        Close all $\Pi\Gamma$-paths in any one semi-real-knot
    end if
end if
while G has more than one semi-real-knot **do**
    Find an interchromosomal or oriented edge joining $\Pi\Gamma$-paths in any two semi-real-knot and add it to G
end while
Close all remaining $\Pi\Gamma$-paths in G
Find a capping $\hat{\Gamma}$ defined by the graph $G(\hat{\Pi}, \hat{\Gamma})$

Theorem (Jean & Nikolski 2007)

Let $d = d(\Pi, \Gamma)$ be the distance between two multichromosomal genomes Π and Γ. Algorithm Correct_Optimal_Capping constructs an optimal capping $\hat{\Gamma}$ for any arbitrary capping $\hat{\Pi}$, such that $d(\hat{\Pi}, \hat{\Gamma}) = d$.

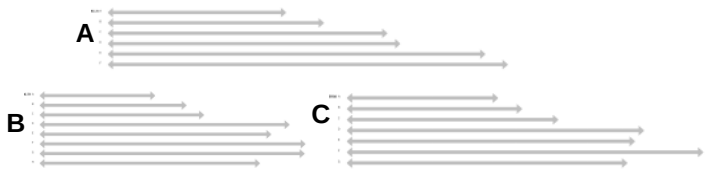
VIRAGE

A new interactive tool for the visualization of rearrangements from scenarios

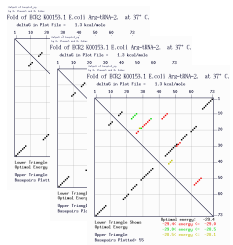


Conclusions and perspectives

Conclusion

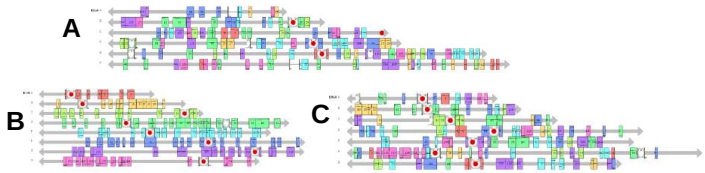


genomes



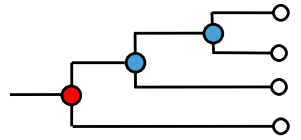
Common markers

SyDiG algorithm:
Identification of common markers in distant genomes



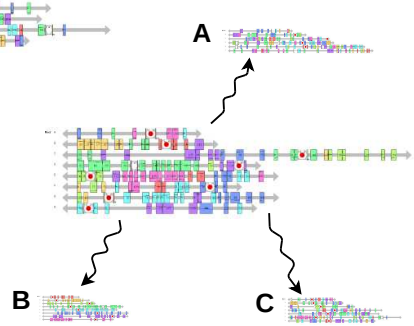
Comparative maps

Parsimonious scenario:
Suggestion of a synthesis for component classification
Introduction of a correct algorithm for optimal capping with its proof
Development of a visualization tool: VIRAGE



Rearrangement tree

Super-block method:
Based on an adjacency analysis and a rearrangement study
Results: all the sets of possible partial assemblies
Centromeres positions can be take into the account



Ancestral architecture

Conclusion

Development and validation of the methods on real data :

Complete framework for Hemiascomycetes yeasts

- Identification of common markers
- Ancestral hypothesis from comparative maps
- Interactive visualization of possible scenarios

Perspectives

Theoretical perspectives :

- Method extension to genomes with duplicates and different gene content
- Adding centromere position constraint in parsimonious scenarios
- Whole reconstruction of the species tree by recovering the root and internal nodes

Biological perspectives :

- Application to the *Drosophilia* twelve (Stark et al. 2007)
- Medical interest : application to 5 species phylogenetically close to *Candida glabrata* (available soon)

"In Amelie's rearrangement world"

Acknowledgments








INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

INRIA FUTURS

Genevures Génoplante

CNRS

ACI IMPBIO

Special thanks to ENSEIRB students for the first version of VIRAGE
And thank you for your attention !

Distance formula

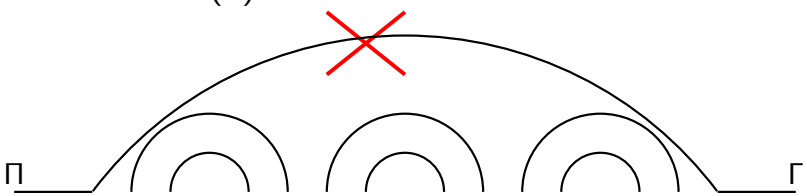
Theorem (Ozery-Flato & Shamir 2003)

$$d(\Pi, \Gamma) = b - c + p + r + \lceil \frac{s' - gr' + fr'}{2} \rceil$$

b is the number of genes plus the number of chromosomes (max)
c is the number of cycles and paths
p is the number of ΓΓ-paths
r is the number of real-knots
s' is the number of semi-real-knots

gr' $= 1$ if the greatest-real-knot exists and $s' > 0$, $gr' = 0$ else.
fr'$= 1$ if (a) or (b), fr'$=0$ else.

(a) fortress of real-knots



(b) weak fortress of real-knots