

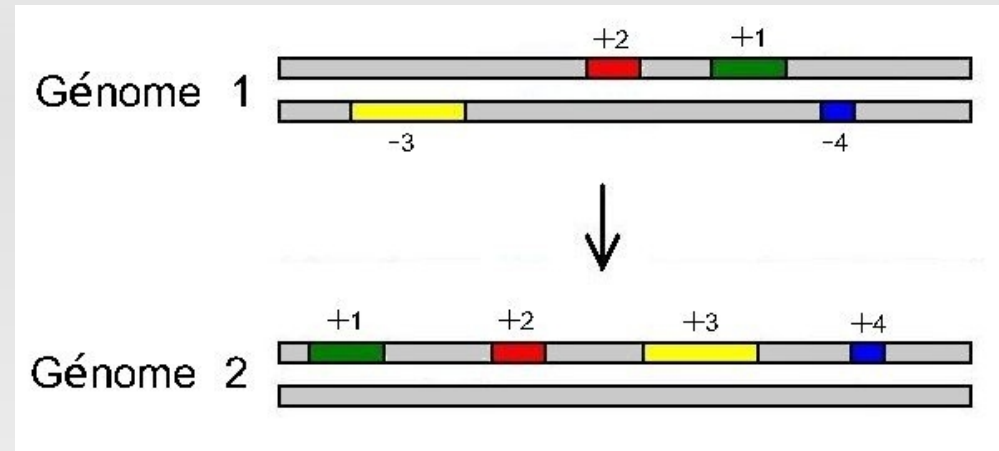
Le tri par inversions

Marília D. V. Braga

Une permutation représente un génome

9 février 2007
Marília D. V. Braga

- Les génomes sont représentés par des permutations signées
- Chaque valeur représente un marqueur (gène)



- Exemple:

La permutation initiale est $\pi = (-3, +2, +1, -4)$

La permutation triée est $\pi' = (+1, +2, +3, +4)$

- Une **inversion** inverse l'ordre des valeurs affectées et leurs signes (orientations)

- Exemple :

Une solution optimale pour la permutation

$\pi = (-3, +2, +1, -4)$:

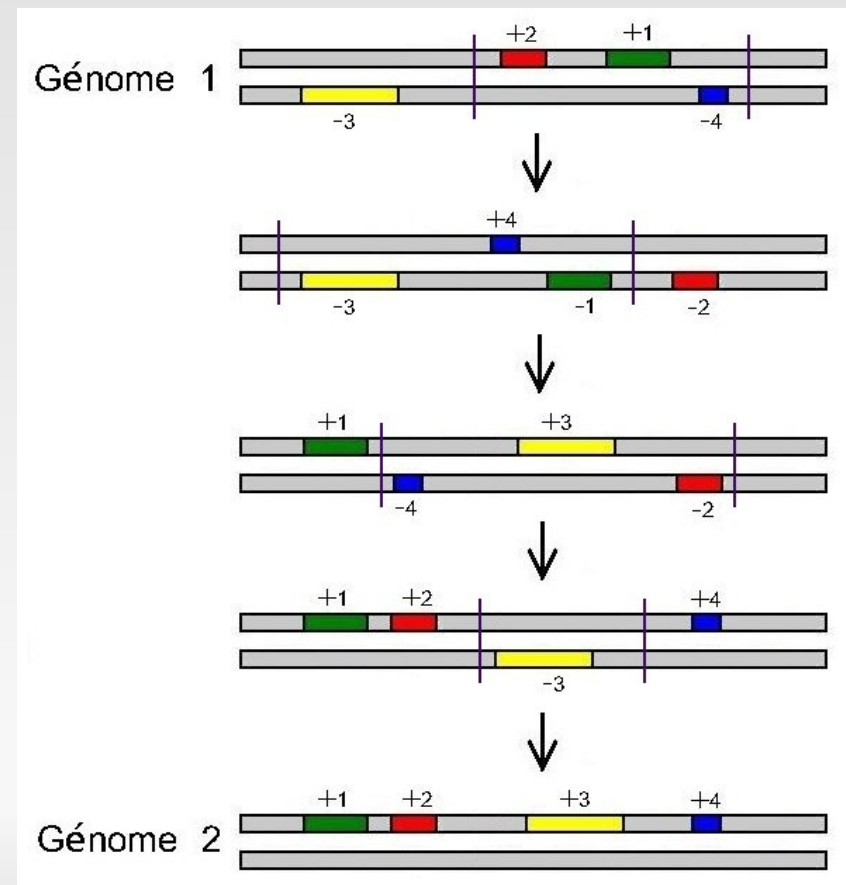
(-3, +2, +1, -4)

(-3, +4, -1, -2)

(+1, -4, +3, -2)

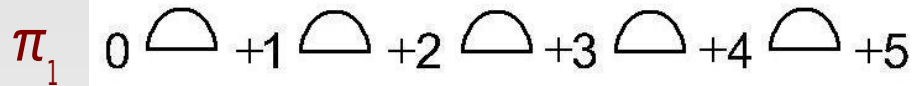
(+1, +2, -3, +4)

(+1, +2, +3, +4)



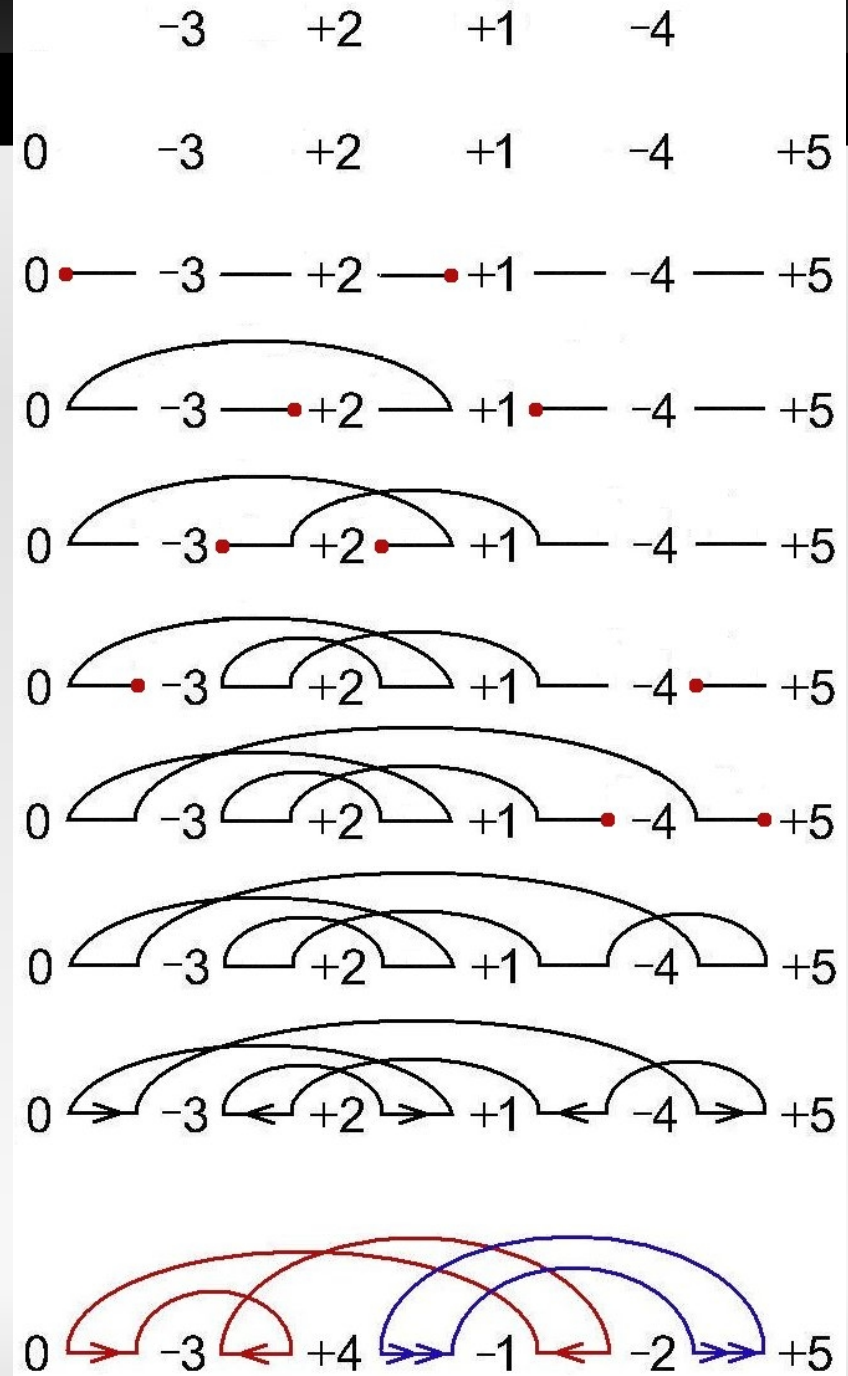
Le graphe de points de cassure (1)

- Le graphe pour la permutation triée $\pi_1 = (+1, +2, +3, +4)$ a cinq cycles (adjacences)



- Le graphe pour la permutation $\pi_2 = (-3, +2, +1, -4)$ a un cycle avec cinq points de cassure

- Le graphe pour la permutation $\pi_3 = (-3, +4, -1, -2)$ a deux cycles (un avec trois et l'autre avec deux points de cassure)

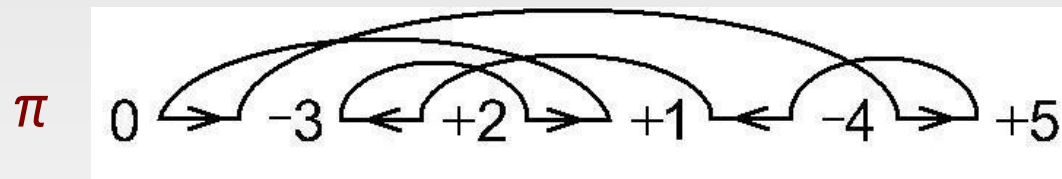


Le graphe de points de cassure (2)

9 février 2007
Marília D. V. Braga

Étant donnée une permutation π :

- . $\mathbf{n(\pi)}$ est le nombre de valeurs de π
- . $\mathbf{c(\pi)}$ est le nombre de cycles du graphe de points de cassure pour π



Pour la permutation $\pi = (-3, +2, +1, -4)$:

$$\mathbf{n(\pi) = 4 \text{ et } c(\pi) = 1}$$

Les effets d'une inversion sur les cycles

9 février 2007
Marília D. V. Braga

- Si les points de cassure font partie d'un même cycle et leurs directions sont opposées : le cycle est coupé en deux (**inversion orientée**).

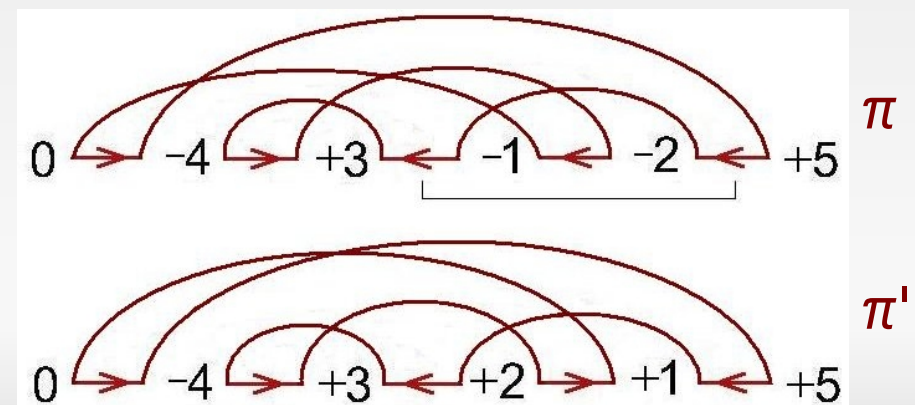
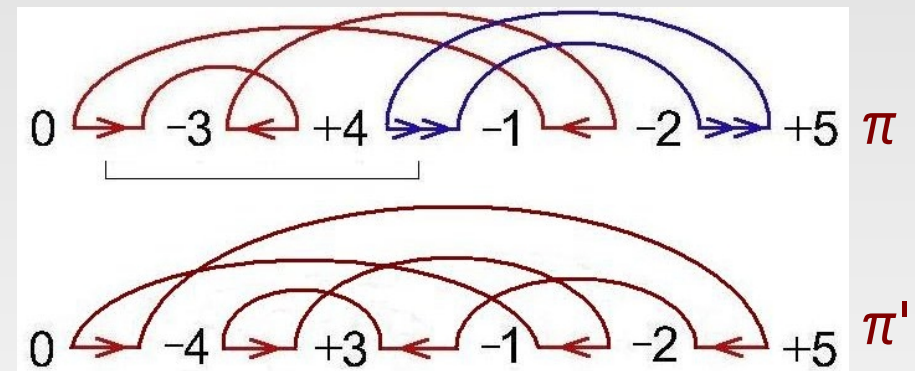
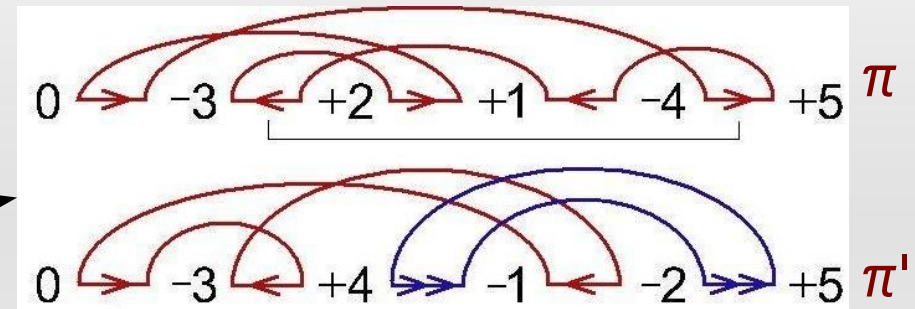
$$c(\pi') = c(\pi) + 1$$

- Si les points de cassure font partie de deux cycles différents : les deux cycles sont regroupés dans un seul cycle

$$c(\pi') = c(\pi) - 1$$

- Si les points de cassure font partie d'un même cycle et ont la même direction : le nombre de cycles ne change pas

$$c(\pi') = c(\pi)$$



Le tri par inversions

9 février 2007
Marília D. V. Braga

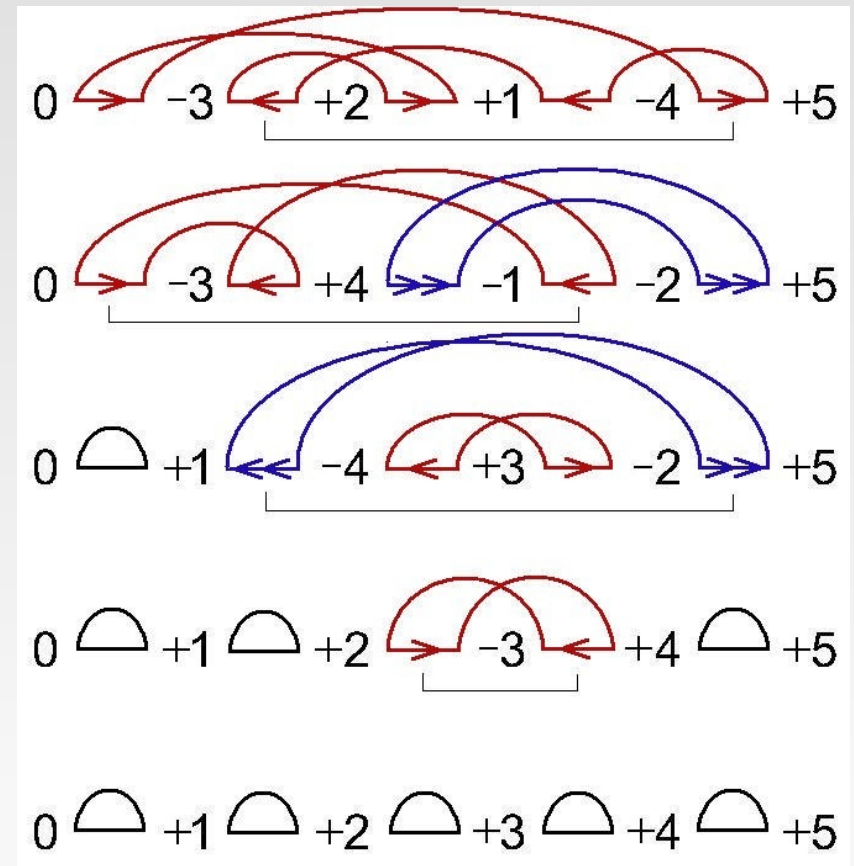
- Seules les inversions orientées sont capables d'augmenter le nombre de cycles du graphe (chaque inversion orientée augmente le nombre de cycles de un)

- $d(\pi)$ est la distance d'inversions pour une permutation π (le nombre d'inversions d'une solution optimale du tri par inversions pour π)

- $d(\pi) \geq n(\pi) + 1 - c(\pi)$

- Exemple :

Une solution optimale pour la permutation $\pi = (-3, +2, +1, -4)$



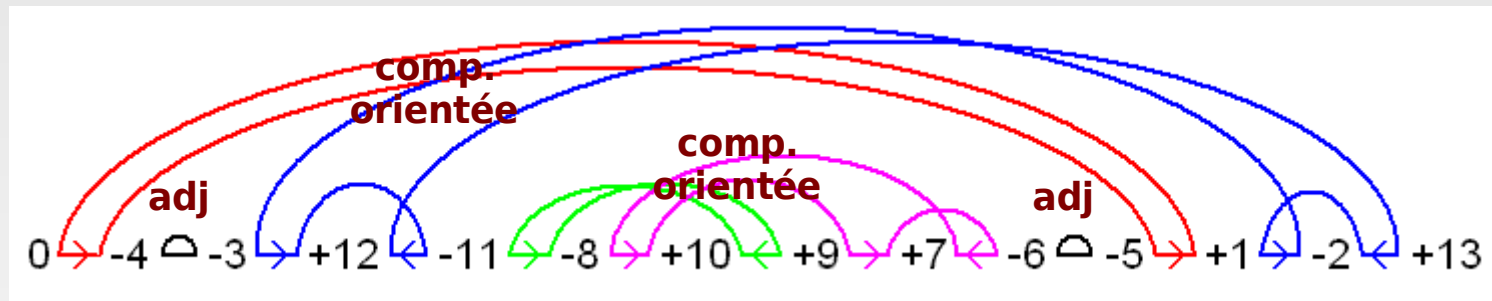
$$d(\pi) = n(\pi) + 1 - c(\pi) = 4 + 1 - 1 = 4$$

Les composantes du graphe (1)

9 février 2007
Marília D. V. Braga

Une composante du graphe de points de cassure est un groupe de cycles qui se chevauchent

Le graphe pour la permutation $\pi = (-4, -3, +12, -11, -8, +10, +9, +7, -6, -5, +1, -2, +13)$ a six cycles et quatre composantes (deux adjacences)



Une composante orientée est une composante du graphe pour laquelle il existe une inversion orientée.

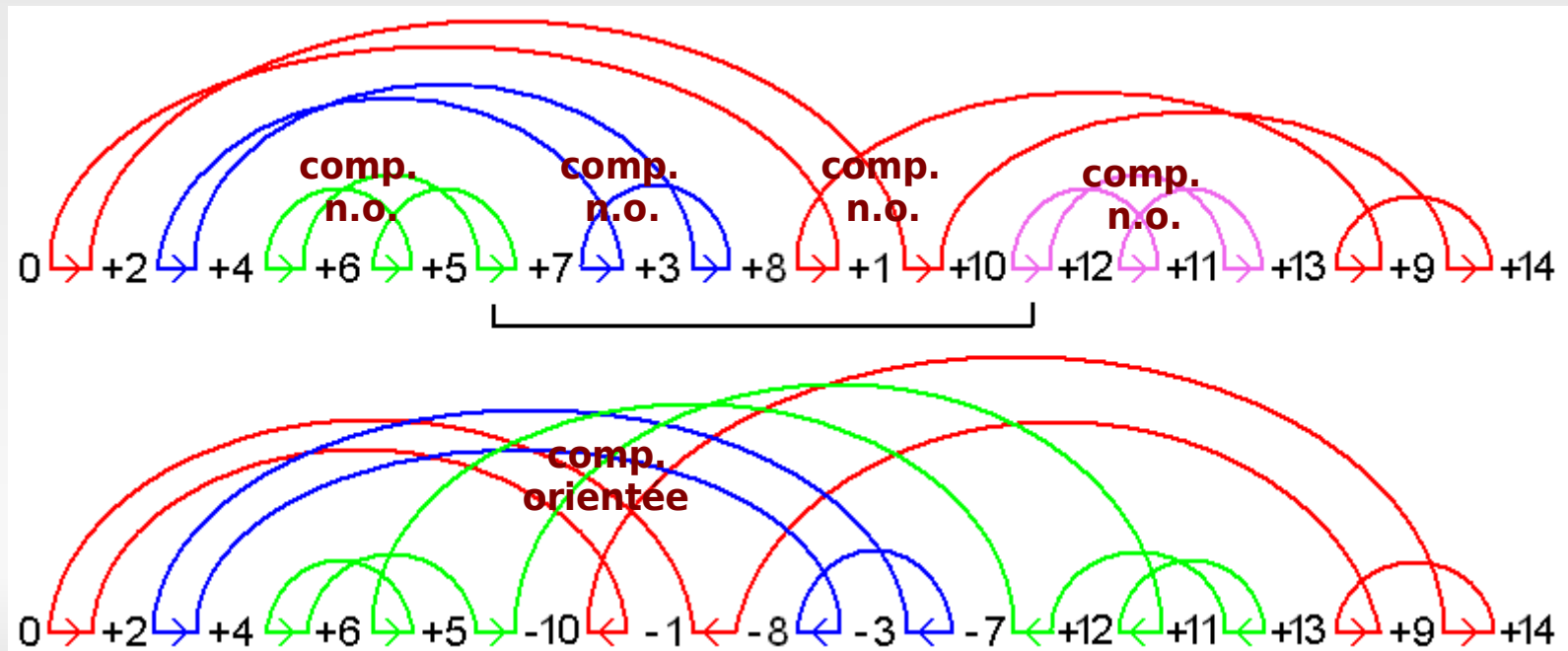
Une composante non-orientée est une composante du graphe qui n'est pas une adjacence et pour laquelle il n'existe pas d'inversion orientée.

Les composantes du graphe (2)

9 février 2007
Marília D. V. Braga

Une inversion regroupe toutes les composantes affectées par cette inversion

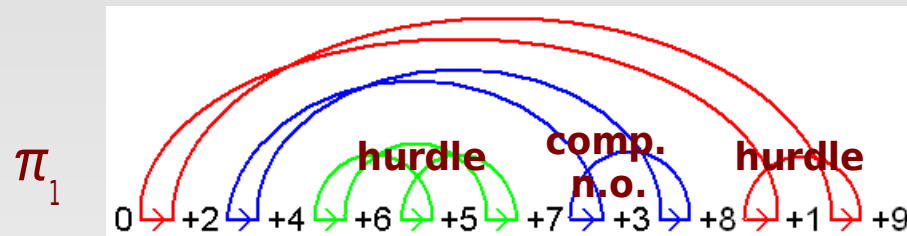
(une composante est affectée par une inversion si elle contient au moins une extrémité de l'inversion ou si elle a des points de cassure dans et en dehors de l'inversion)



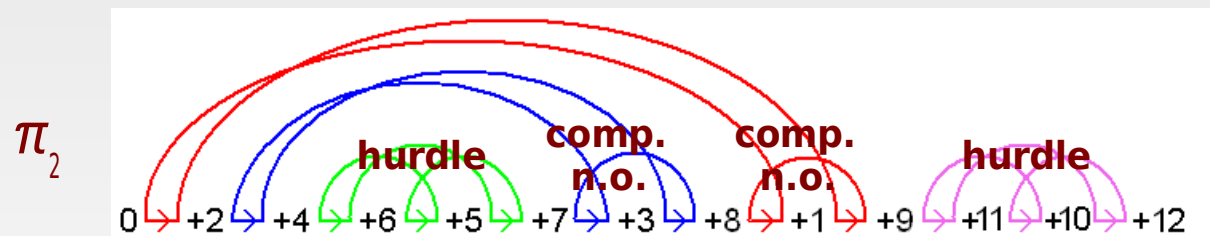
Les composantes non-orientées – *hurdles* (1)

9 février 2007
Marília D. V. Braga

Une composante non-orientée qui ne sépare pas deux autres composantes non-orientées est une ***hurdle***.

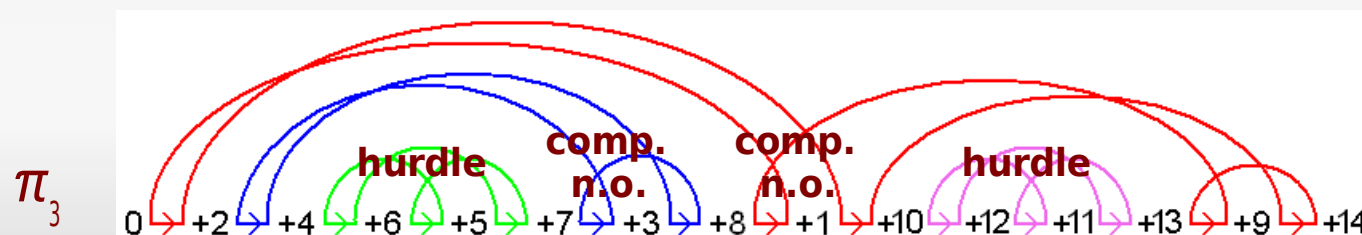


$h(\pi)$ est le nombre de *hurdles* d'une permutation π



$$h(\pi_1) = 2$$

$$h(\pi_2) = 2$$



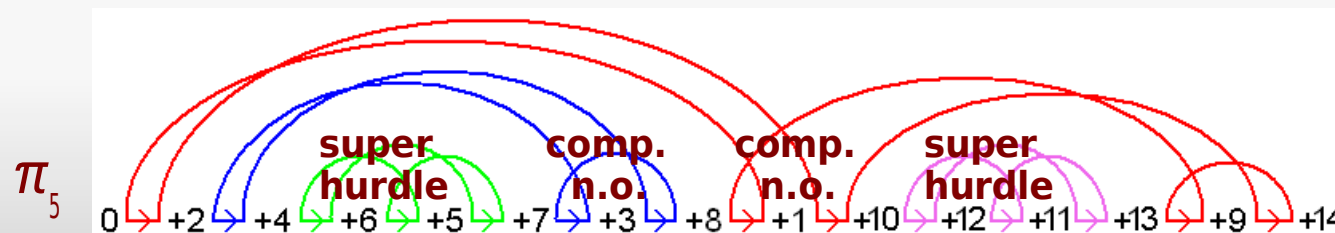
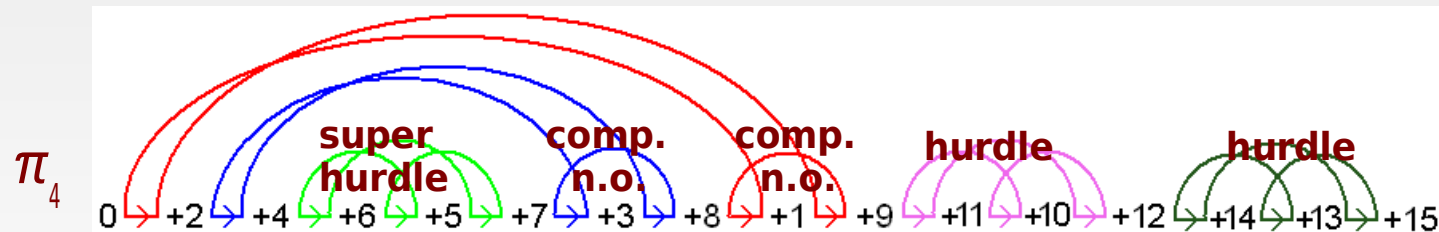
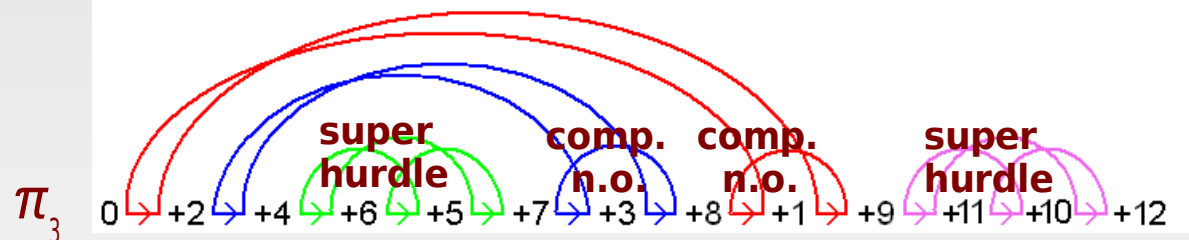
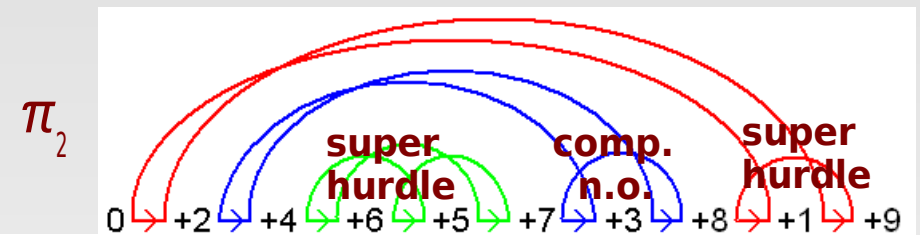
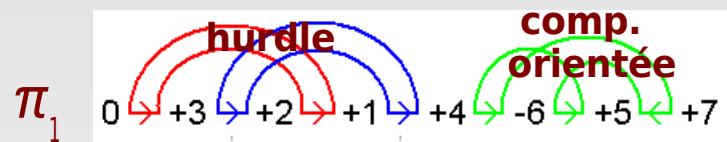
$$h(\pi_3) = 2$$

Les composantes non-orientées – *hurdles* (2)

9 février 2007
Marília D. V. Braga

Une *hurdle* qui protège une composante non-orientée qui n'est pas aussi une *hurdle* est une **super-hurdle**.

sh(π) est le nombre de *super-hurdles* d'une permutation π



$$h(\pi_1) = 1, \text{sh}(\pi_1) = 0$$

$$h(\pi_2) = 2, \text{sh}(\pi_2) = 2$$

$$h(\pi_3) = 2, \text{sh}(\pi_3) = 2$$

$$h(\pi_4) = 3, \text{sh}(\pi_4) = 1$$

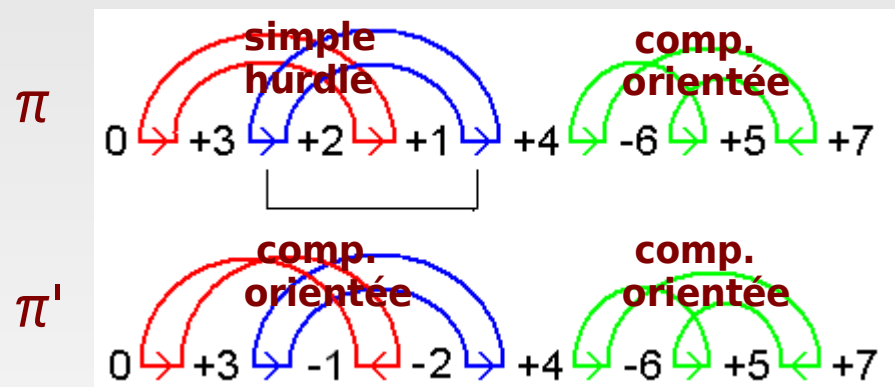
$$h(\pi_5) = 2, \text{sh}(\pi_5) = 1$$

Les composantes non-orientées – *hurdles* (3)

9 février 2007
Marília D. V. Braga

Une *hurdle* qui n'est pas une *super-hurdle* est aussi appelée *simple-hurdle*.

Pour couper une *simple-hurdle*, il faut faire une inversion non-orientée qui agit sur un seul de ses cycles :



$$h(\pi') = h(\pi) - 1$$

$$c(\pi') = c(\pi)$$

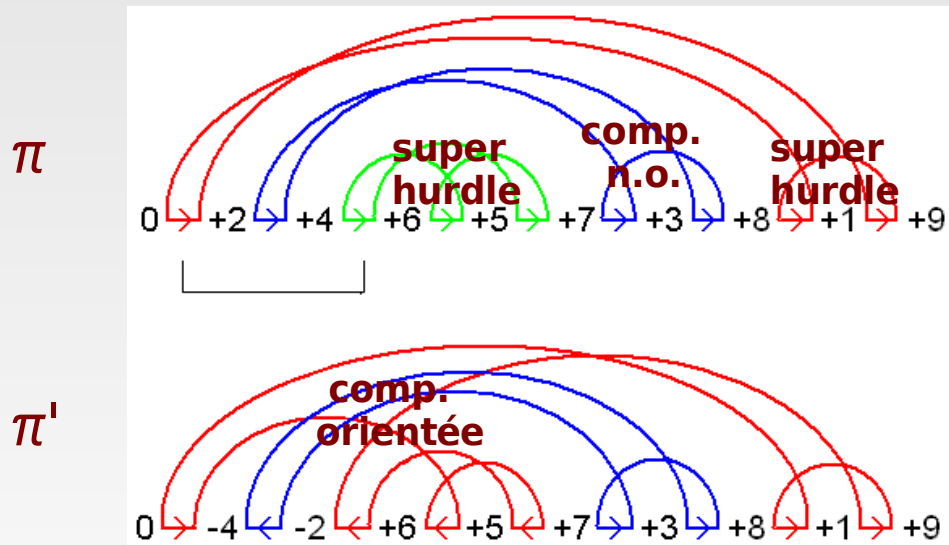
$$d(\pi) \geq n(\pi) + 1 + h(\pi) - c(\pi)$$

Les composantes non-orientées - *super-hurdles*

9 février 2007
Marília D. V. Braga

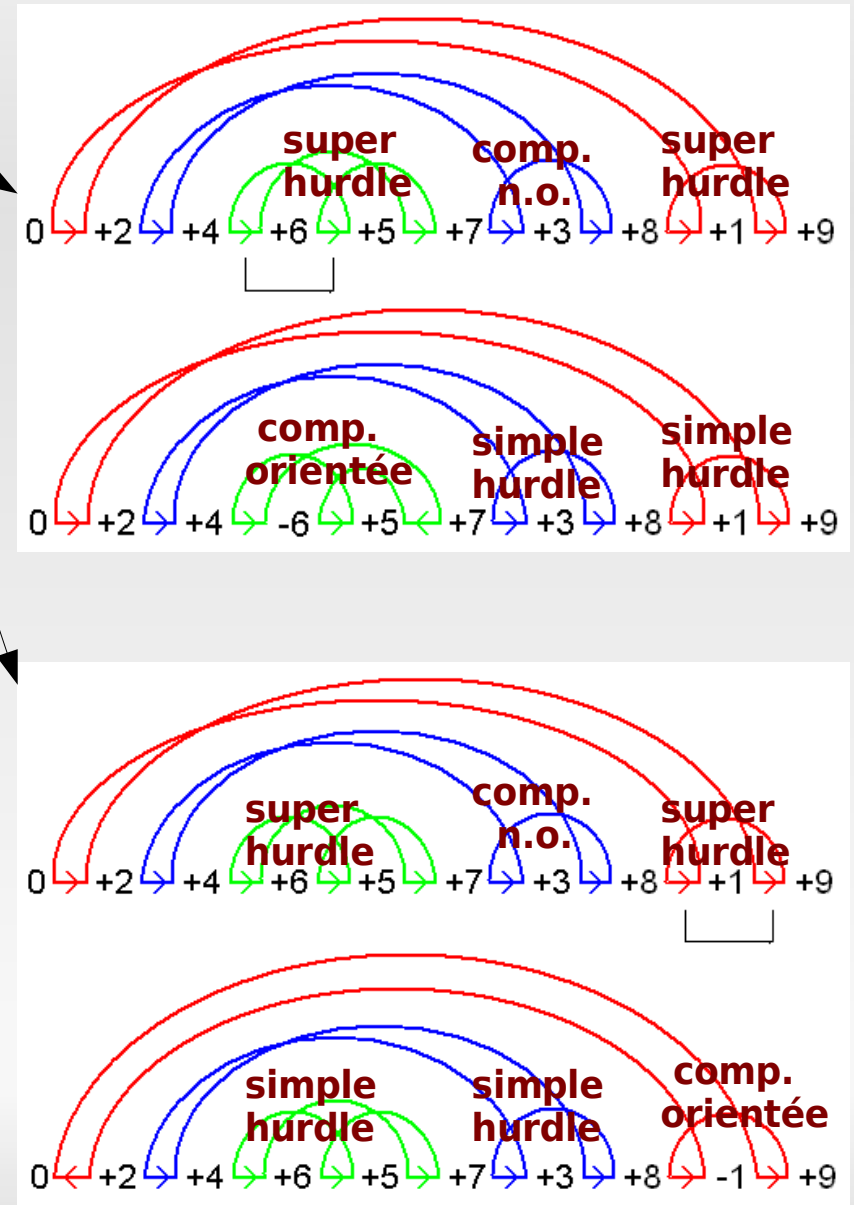
Couper une *super-hurdle* produit l'apparition d'une nouvelle *hurdle*

Pour éliminer les *super-hurdles*, il faut les regrouper :



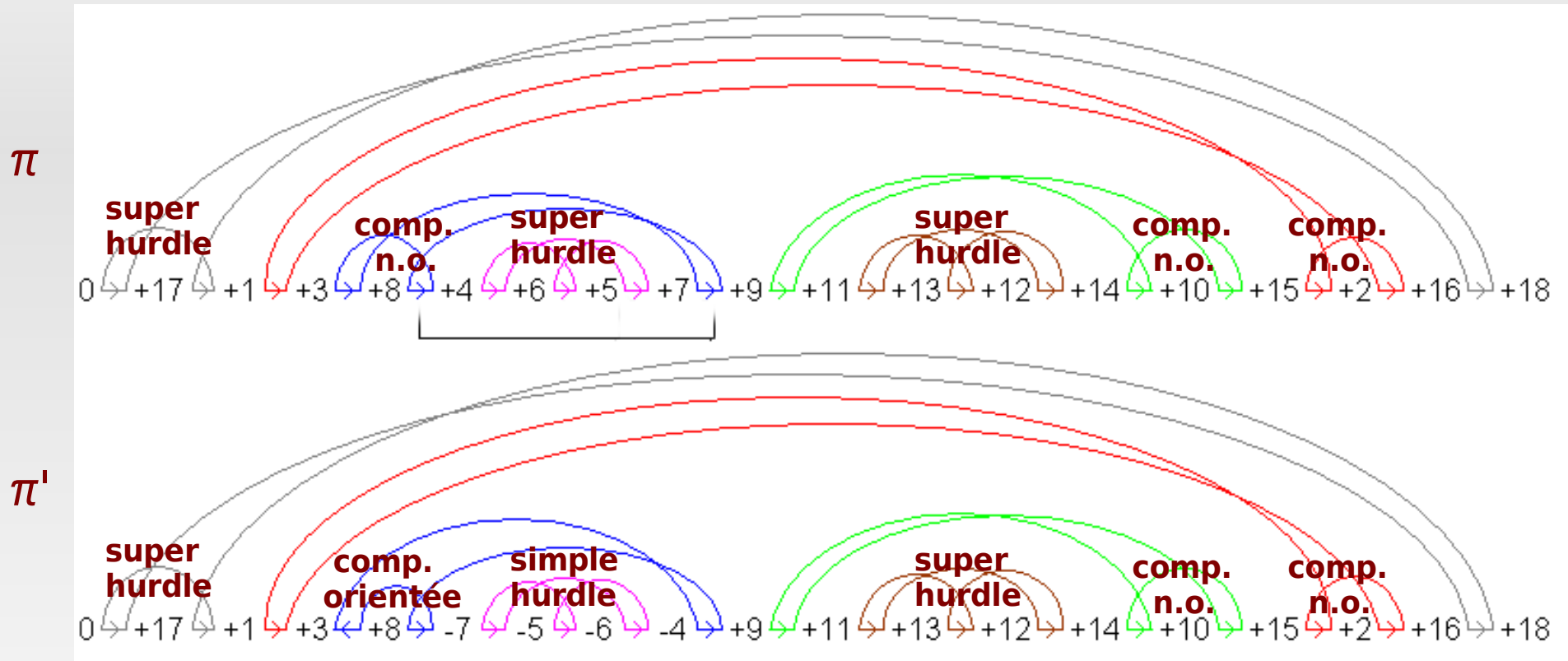
$$h(\pi') = h(\pi) - 2 \text{ et } c(\pi') = c(\pi) - 1$$

$$d(\pi) \geq n(\pi) + 1 + h(\pi) - c(\pi)$$



Élimination de la forteresse (1)

9 février 2007
Marília D. V. Braga



$$f(\pi) = 1$$

$$c(\pi) = 6$$

$$h(\pi) = 3$$

$$(sh(\pi) = 3)$$

$$f(\pi') = 0$$

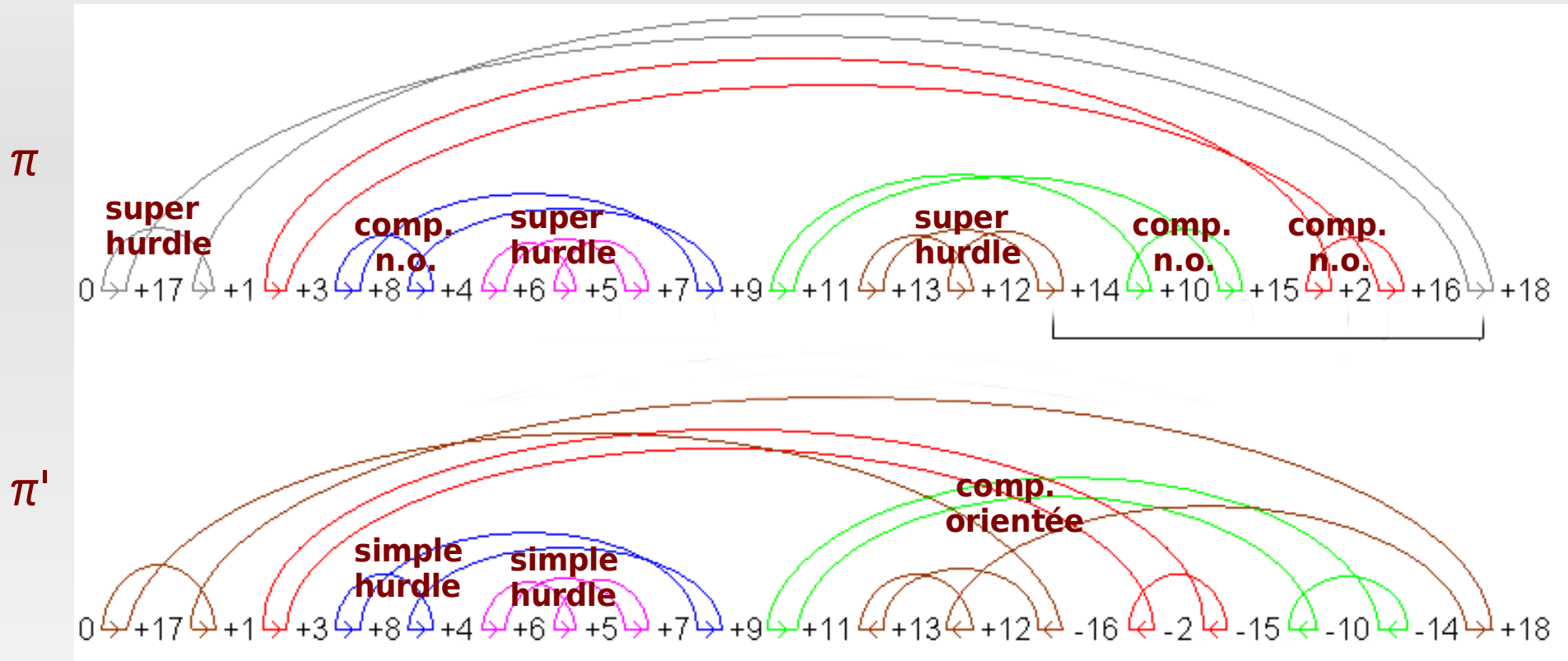
$$c(\pi') = c(\pi) = 6$$

$$h(\pi') = 3$$

$$(sh(\pi') = 2)$$

Élimination de la forteresse (2)

9 février 2007
Marília D. V. Braga



$$f(\pi) = 1$$

$$c(\pi) = 6$$

$$h(\pi) = 3$$

$$(sh(\pi) = 3)$$

$$f(\pi') = 0$$

$$c(\pi') = c(\pi) - 1 = 5$$

$$h(\pi') = 2$$

$$(sh(\pi') = 0)$$

Une inversion est suffisante pour éliminer la forteresse.

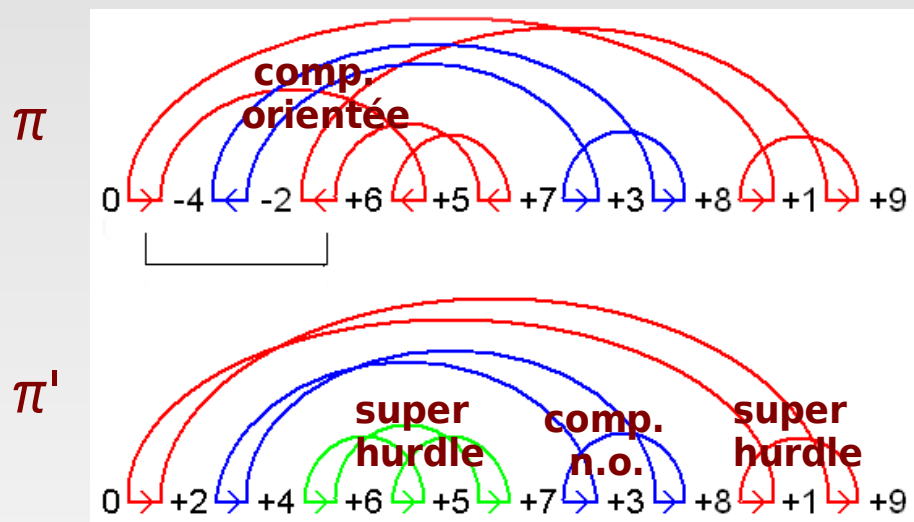
Étant donnée une permutation π :

- . $\mathbf{n(\pi)}$ est le nombre de valeurs de π
- . $\mathbf{c(\pi)}$ est le nombre de cycles du graphe de points de cassure pour π
- . $\mathbf{h(\pi)}$ est le nombre de *hurdles* de π
- . Si π est une forteresse, $\mathbf{f(\pi)=1}$, sinon, $\mathbf{f(\pi)=0}$.
- . $\mathbf{d(\pi) = n(\pi) + 1 + h(\pi) + f(\pi) - c(\pi)}$

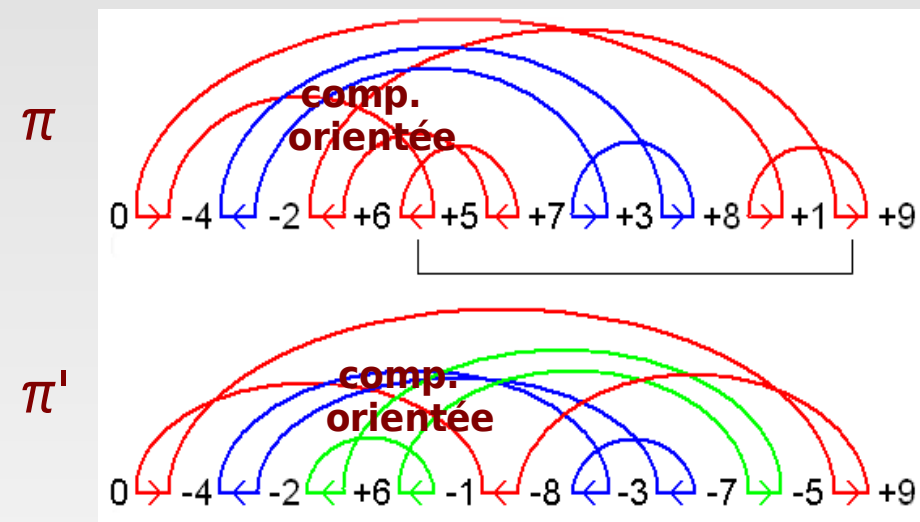
Inversion orientée propre

9 février 2007
Marília D. V. Braga

Une inversion orientée peut produire des nouvelles composantes. Une inversion orientée n'est pas propre si elle produit des *hurdles*.



inversion orientée



inversion orientée propre

Pour chaque composante orientée, il existe toujours une inversion orientée propre.

$$d(\pi) = n(\pi) + 1 + h(\pi) + f(\pi) - c(\pi)$$

Un algorithme pour trier une permutation π :

- . Si π est une forteresse, choisir une inversion qui élimine la forteresse
- . Tant que π a une composante non-orientée:
 - Si π a des *super-hurdles*, regrouper deux *super-hurdles* de π
 - Sinon, couper une *simple-hurdle* de π
- . Trier π par des inversions orientées propres

P. A. Pevzner. Computational Molecular Biology: An Algorithmic Approach, The MIT Press, 2000.

J. C. Setubal and J. Meidanis. Introduction to Computational Molecular Biology, Boston: PWS Publishing Company, 1997.