# Nouveaux résultats sur le tri par transpositions

Anthony Labarre[1]

23 Janvier 2006

Université Libre de Bruxelles
alabarre@ulb.ac.be

ALPHY/GTGC 2006

[1]Subventionné par le "Fonds pour la Formation à la Recherche dans l'Industrie et dans l'Agriculture" (F.R.I.A.).

# Contexte et motivation

- Réarrangements génomiques: mesure de divergence entre les espèces;
- Nos hypothèses:
  - ordre des gènes connu;
  - tous les génomes partagent le même ensemble/nombre de gènes;
  - un seul type de mutation pris en compte;

# Tri par transpositions

- *Transposition* = échange de deux intervalles adjacents:

$$( 2\ 1\ \boxed{7\ 3\ 5\ 6}\ \boxed{4\ 10\ 9}\ 8 )$$
$$\downarrow$$
$$( 2\ 1\ \boxed{4\ 10\ 9}\ \boxed{7\ 3\ 5\ 6}\ 8 )$$

- *Distance des transpositions* $d(\pi, \sigma)$ = nombre minimal de transpositions transformant $\pi$ en $\sigma$;
- *Tri par transpositions*: ramener une permutation à l'identité $\iota = (1\ 2\ \cdots\ n)$ par un nombre minimal de transpositions;

# Exemple

- Le tri de $\pi = (3\ 1\ 4\ 2)$ nécessite deux transpositions:

$$\pi = \quad (\ 3\ \boxed{1}\ \boxed{4}\ 2\ )$$
$$\downarrow$$
$$(\ \boxed{3\ 4}\ \boxed{1\ 2}\ )$$
$$\downarrow$$
$$\iota = \quad (\ 1\ 2\ 3\ 4\ )$$

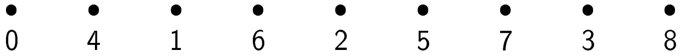
- Donc $d(\pi, \iota) = d(\pi) = 2$.

# Status du problème

- Complexité:
  - du tri: inconnue;
  - du calcul de la distance: inconnue;
- Valeur maximale: inconnue (entre $\left\lfloor \frac{n+1}{2} \right\rfloor + 1$ et $\left\lfloor \frac{2n-2}{3} \right\rfloor$);
- Meilleur taux d'approximation: $\frac{11}{8}$ [Elias and Hartman, 2005];

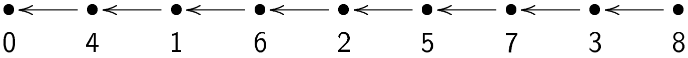
# Le graphe des cycles [Bafna and Pevzner, 1998]

- Soit $\pi = (4\ 1\ 6\ 2\ 5\ 7\ 3)$;



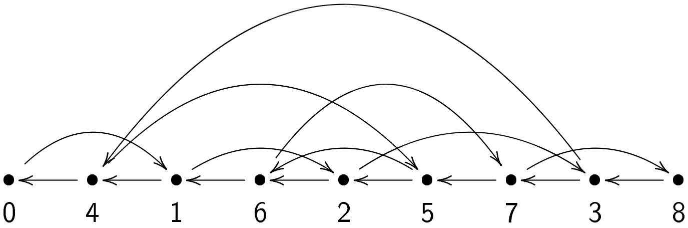
# Le graphe des cycles [Bafna and Pevzner, 1998]

- Soit $\pi = (4\ 1\ 6\ 2\ 5\ 7\ 3)$;



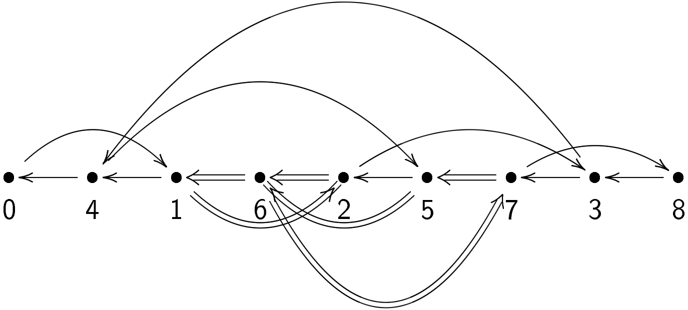
# Le graphe des cycles [Bafna and Pevzner, 1998]

- Soit $\pi = (4\ 1\ 6\ 2\ 5\ 7\ 3)$;



# Cycles alternés

- Décomposition unique en cycles alternés:

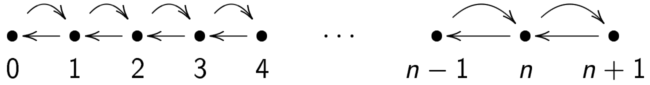



- *Parité* d'un cycle = celle du nombre d'arcs horizontaux qu'il contient; ici, $c(G(\pi)) = 2 = c_{odd}(G(\pi))$.

# Cycles alternés

- Nombre maximal de cycles alternés pour $(1\ 2\ \cdots\ n)$:

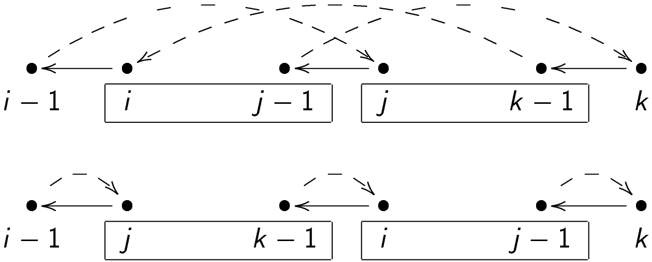
$G(\iota)$ :




- Donc trier par transpositions revient à créer des cycles alternés impairs “le plus vite possible”;

# Une minoration

Au mieux: deux nouveaux cycles impairs en une transposition:




## Théorème

*[Bafna and Pevzner, 1998]* $\forall\ \pi \in S_n : d(\pi) \geq \frac{n+1-c_{odd}(G(\pi))}{2}$.

# Résultats ([Labarre, 2005] et [Labarre, 2006])

1. Liens entre la théorie "traditionnelle" et une structure plus classique;
2. Caractérisation de classes de permutations dont la distance est calculable en temps polynomial, sans graphe;
3. Nouvelles majorations et, dans certains cas, la distance exacte.

# Permutations réduites

- *Point de rupture* : couple $(\pi_i, \pi_{i+1})$ avec $\pi_{i+1} \neq \pi_i + 1$;
- $b(\pi)$ = nombre de points de rupture de $\pi$;
- $\pi$ est *réduite* si $b(\pi) = n - 1$, $\pi_1 \neq 1$, et $\pi_n \neq n$.

| réduite | non réduites |
|---|---|
| (4 2 1 3) | $(\underline{1}\ 4\ 3\ 2)$ |
| | $(3\ 2\ 1\ \underline{4})$ |
| | $(4\ \underline{2\ 3}\ 1)$ |

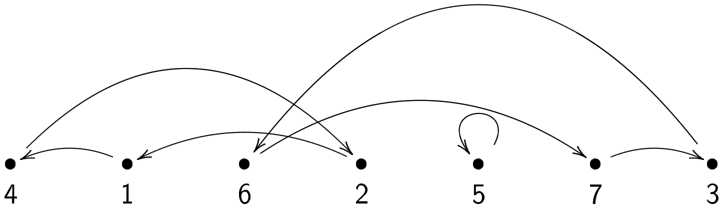
## Théorème

*[Christie, 1998] Toute permutation est "réductible", et ceci n'affecte pas sa distance.*

# Le Γ-graphe

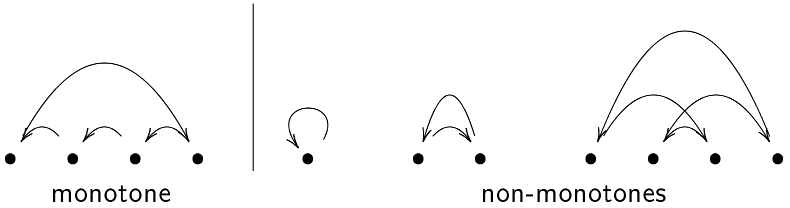
- Toute permutation est le produit de cycles disjoints;
- Le Γ-*graphe* est le *graphe de la permutation* muni d'un ordre total sur les sommets.

Soit $\pi = (4\ 1\ 6\ 2\ 5\ 7\ 3) = (1,\ 4,\ 2)\ (3,\ 6,\ 7)\ (5)$; alors $\Gamma(\pi)$ est:



# Quelques définitions

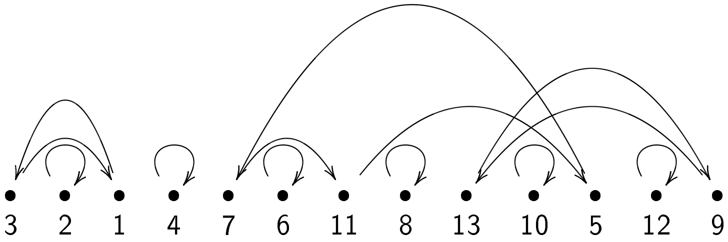
- Un *$k$-cycle* de $\Gamma(\pi)$ contient $k$ sommets;
- Un $k$-cycle est *(im)pair* si $k$ est (im)pair;
- *Monotonicité* d'un cycle:



# Les $\gamma$-permutations

## Définition

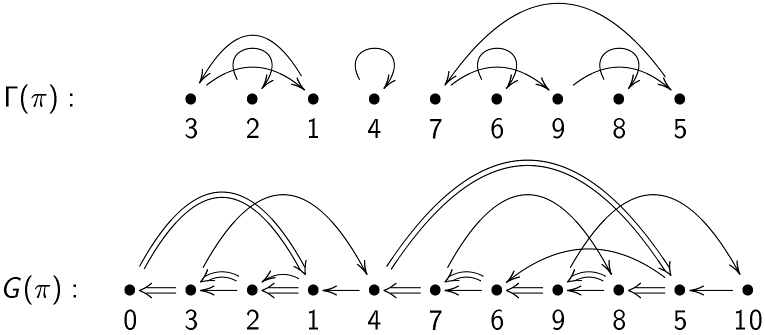
*Une $\gamma$-permutation est une permutation réduite fixant tous les éléments pairs.*



# Correspondance entre $G$ et $\Gamma$

## Proposition

*Si $\pi$ est une $\gamma$-permutation, alors tout $k$-cycle de $\Gamma(\pi)$ ($k \geq 2$) apparaît exactement deux fois dans $G(\pi)$.*



# Minoration sur la distance des $\gamma$-permutations

- Rappel: $d(\pi) \geq \frac{n+1-c_{odd}(G(\pi))}{2}$ (Théorème 1);
- Donc, par la proposition précédente:

## Lemme

*Pour toute $\gamma$-permutation $\pi$ de $S_n$, on a: $d(\pi) \geq n - c_{odd}(\Gamma(\pi))$ .*

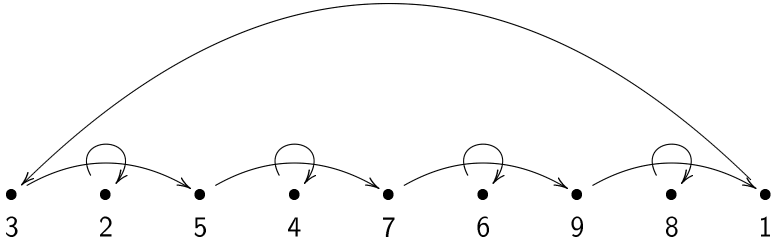
- Stratégie: traiter séparément chaque cycle de $\Gamma(\pi)$;

# Cycles monotones

## Proposition

*Si $\pi$ est une $\gamma$-permutation de $S_n$ ne contenant qu'un seul "grand cycle" monotone $C$, alors:*

$$d(\pi) = n - c_{odd}(\Gamma(\pi)) = |C| - (|C| \bmod 2)\,.$$



# Exemple

$$
\begin{array}{lr}
(\ 3\ \boxed{2\ 5}\ \boxed{4\ 7\ 6\ 9\ 8\ 1}\ ) & 1 \\
(\ \boxed{3\ 4}\ \boxed{7\ 6\ 9\ 8\ 1\ 2}\ 5\ ) & 2 \\
(\ 7\ \boxed{6\ 9}\ \boxed{8\ 1\ 2\ 3\ 4\ 5}\ ) & 3 \\
(\ \boxed{7\ 8}\ \boxed{1\ 2\ 3\ 4\ 5\ 6}\ 9\ ) & 4 \\
(\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9\ ) &
\end{array}
$$

$\Rightarrow d(\pi) = 4 = 5 - (5 \bmod 2).$

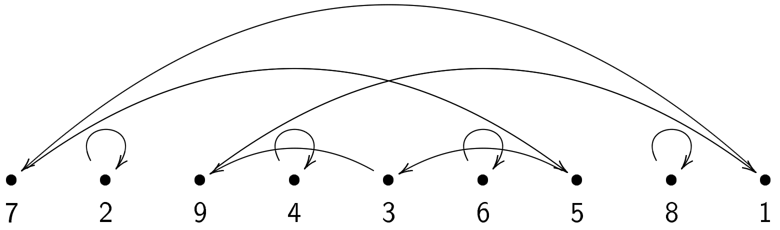
# Cycles non-monotones

Même résultat pour les cycles non-monotones:

## Proposition

*Si $\pi$ est une $\gamma$-permutation de $S_n$ ne contenant qu'un seul "grand cycle" non-monotone $C$, alors:*

$$d(\pi) = n - c_{odd}(\Gamma(\pi)) = |C| - (|C| \bmod 2)\,.$$



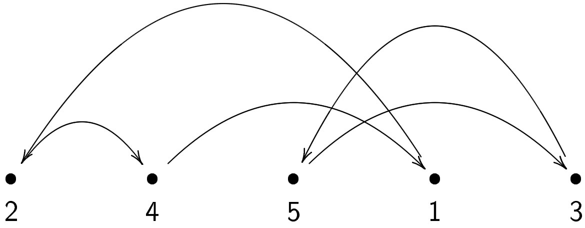
# Distance des $\gamma$-permutations

- Récapitulons; si $\pi$ est une $\gamma$-permutation de $S_n$, alors:
  - $d(\pi) \geq n - c_{odd}(\Gamma(\pi))$;
  - tout cycle de $\Gamma(\pi)$ peut être traité indépendamment des autres;
  - la distance de chaque cycle atteint la minoration;
- Donc:

$$d(\pi) = n - c_{odd}(\Gamma(\pi)).$$

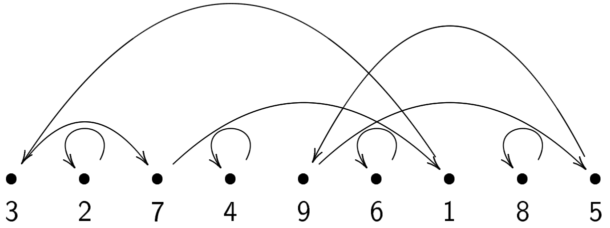
# Une nouvelle majoration

- Toute permutation (sauf $\iota$) peut être obtenue à partir d'une $\gamma$-permutation.



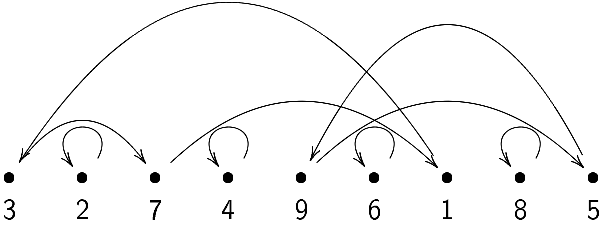
# Une nouvelle majoration

- Toute permutation (sauf $\iota$) peut être obtenue à partir d'une $\gamma$-permutation.



# Une nouvelle majoration

- Toute permutation (sauf $\iota$) peut être obtenue à partir d'une $\gamma$-permutation.

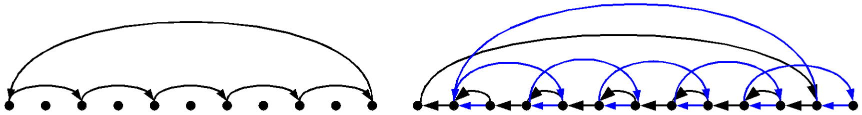



## Théorème

$$\forall\ \pi \in S_n : d(\pi) \leq n - c_{odd}(\Gamma(\pi)).$$

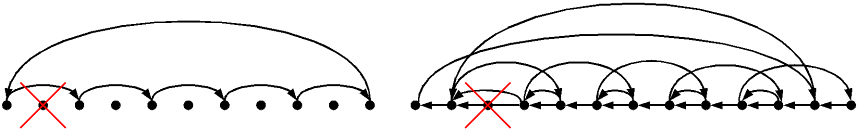
# Autres instances solubles

- Que se passe-t'il quand on supprime des 1-cycles dans Γ?



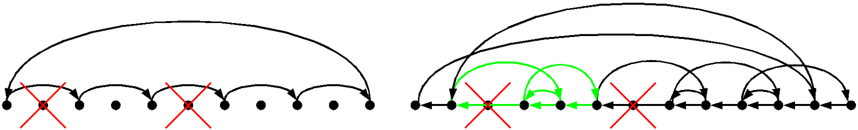
# Autres instances solubles

- Que se passe-t'il quand on supprime des 1-cycles dans Γ?



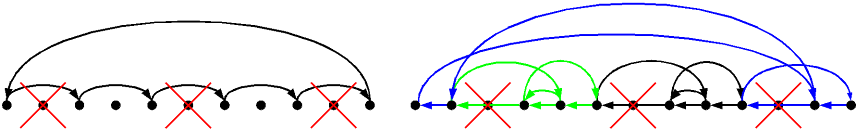
# Autres instances solubles

- Que se passe-t'il quand on supprime des 1-cycles dans Γ?



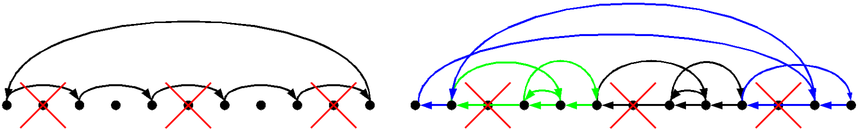
# Autres instances solubles

- Que se passe-t'il quand on supprime des 1-cycles dans Γ?



# Autres instances solubles

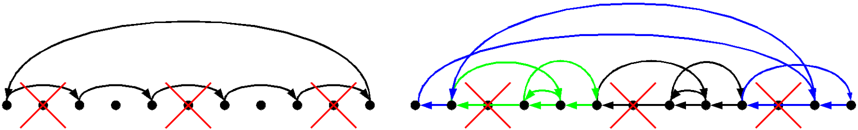
- Que se passe-t'il quand on supprime des 1-cycles dans $\Gamma$?



- $k$ suppressions $\Rightarrow c_{odd}(G(\pi)) = k$

# Autres instances solubles

- Que se passe-t'il quand on supprime des 1-cycles dans Γ?



- $k$ suppressions $\Rightarrow c_{odd}(G(\pi)) = k$
- dans ce cas:

$$d(\pi) = n - c_{odd}(\Gamma(\pi)) - k + \left(\frac{n+k+1}{2} \bmod 2\right).$$

# Suppression de points fixes entre deux cycles monotones

- Si aucun "grand" cycle n'en croise d'autre, alors:
  - la stratégie utilisée pour les $\gamma-$permutations est encore optimale;
  - la distance de chaque cycle est connue;
- On a:

$$d(\pi) = n - c_{odd}(\Gamma(\pi)) - K + \sum_{i=1}^{t}\left(\frac{n_i + k_i + 1}{2} \mod 2\right).$$

- Autres cas solubles dérivables;

# Problèmes ouverts

- Complexité?
- Diamètre?
- Meilleurs résultats:
  - pour les cycles non-monotones?
  - pour les cycles se croisant?

# Références

Bafna, V. and Pevzner, P. A. (1998).
Sorting by transpositions.
*SIAM J. Discrete Math.*, 11(2):224–240 (electronic).

Christie, D. A. (1998).
*Genome Rearrangement Problems.*
PhD thesis, University of Glasgow, Scotland.

Elias, I. and Hartman, T. (2005).
A 1.375−Approximation Algorithm for Sorting by Transpositions.
In *Proceedings of WABI*, LNCS 3692, pages 204–215.

Labarre, A. (2005).
A new tight upper bound on the transposition distance.
In *Proceedings of WABI*, LNCS 3692, pages 216–227.

Labarre, A. (2006).
New bounds and tractable instances for the transposition distance.
Soumis.