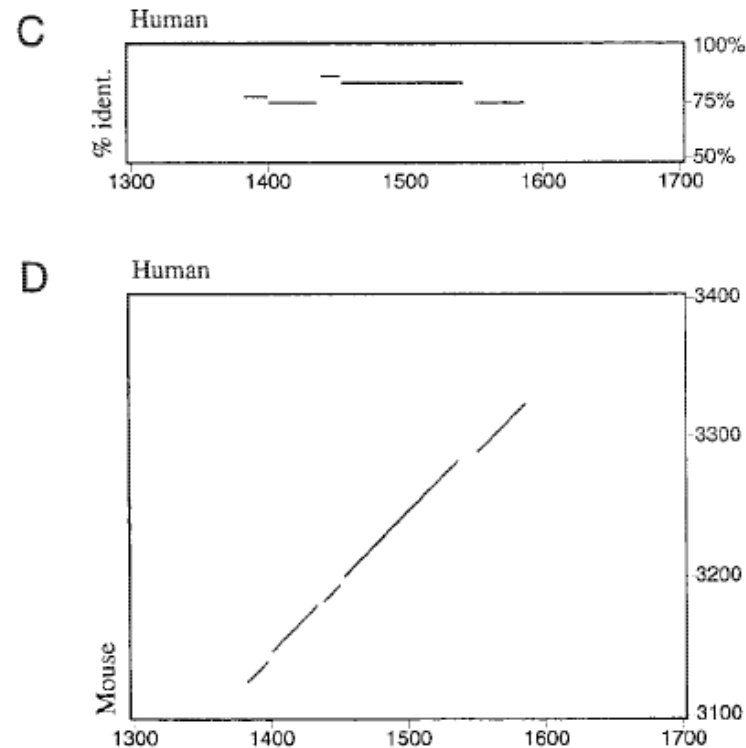


ALIGNEMENT DE GÉNOMES

VISUALISATION DES ALIGNEMENTS : PIPs

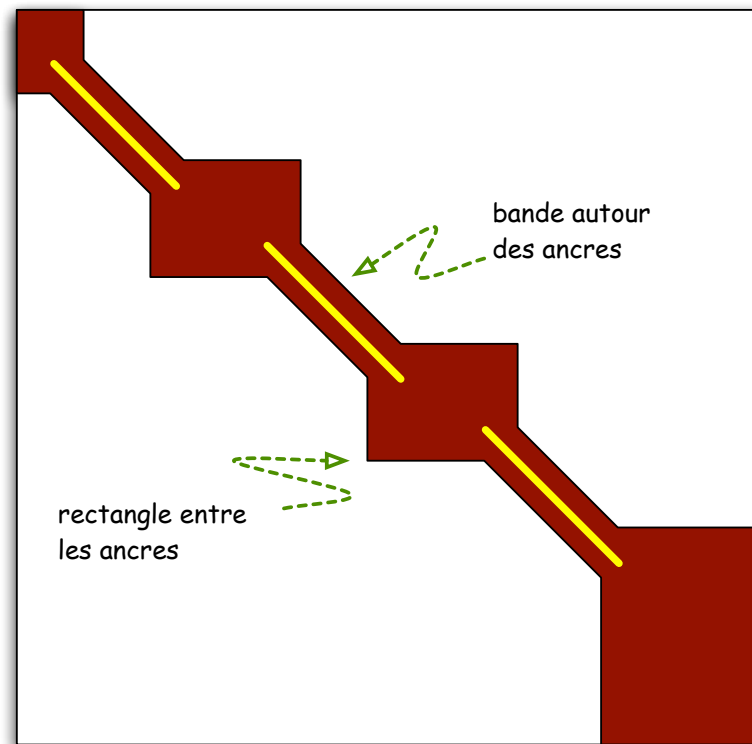
PIP — *percent identity plot*



(on considère les blocs entre les trous, et on affiche le pourcentage de matchs dans chaque bloc)

Schwartz & al. *Genome Res.* 10 : 577

LAGAN : ALIGNEMENT GLOBAL AVEC ANCRERES

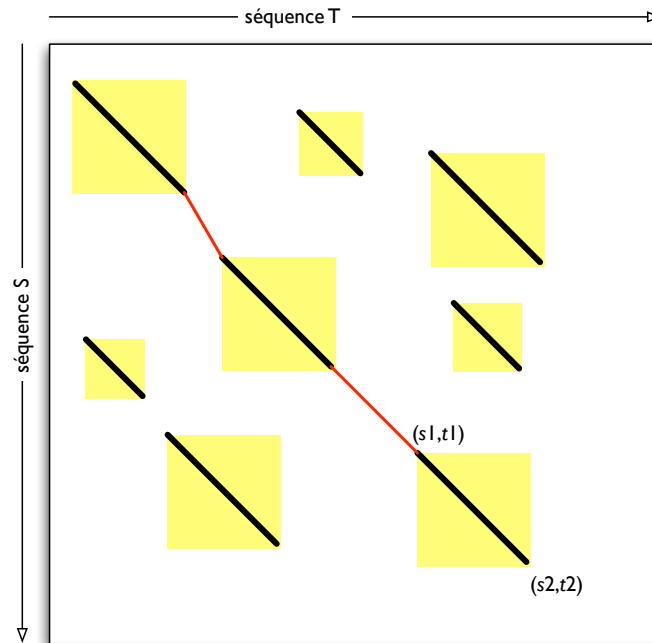


sélection d'ancres : chaîne optimale d'alignements locaux

QUELQUES AUTRES IDÉES

- recalcul de scores (alignement local avec trous : trouver le placement optimal du trou entre deux hits)
- génération de nouveaux ancrés plus épais dans des régions sans ancrés par hachage plus sensible

PROGRAMMATION DYNAMIQUE ÉPARSE



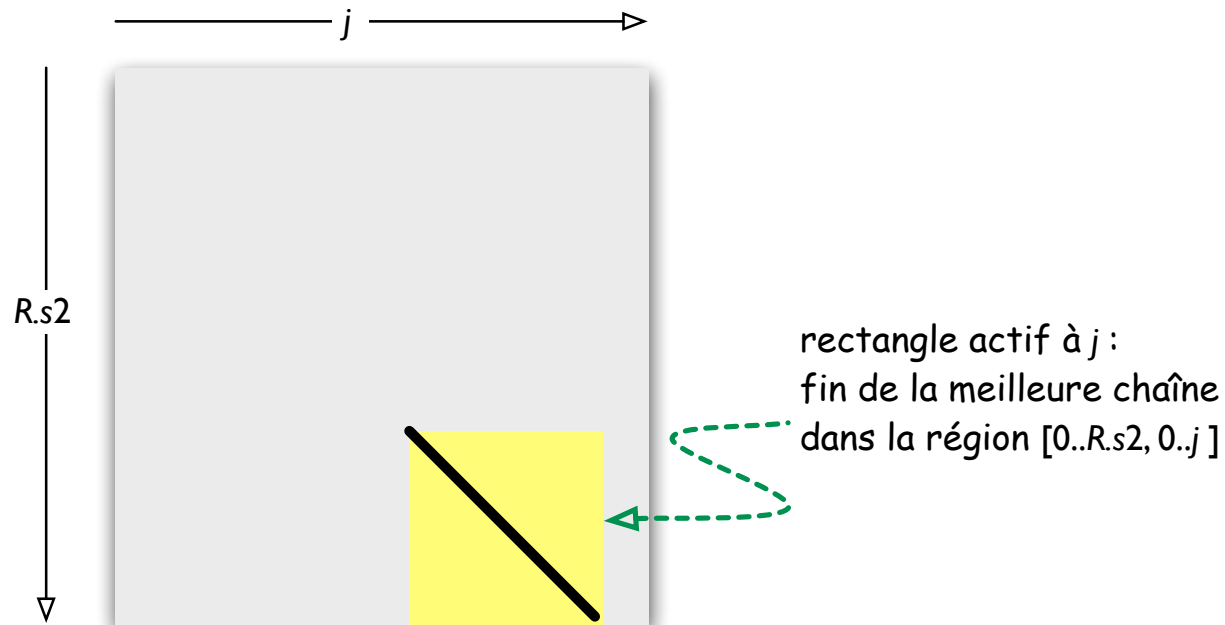
Liste de rectangles (alignements locaux) :

rectangle R en position $(R.s1, R.t1), (R.s2, R.t2)$ avec score $R.v$

chaîne : ensemble de rectangles *non-chevauchant*

PD ÉPARSE 2

Récurrance pour calculer $V(R)$: score maximum d'une chaîne qui finit par rectangle R



Clé : maintenir la liste \mathcal{L}_j de rectangles *actifs* pour $j = 0, \dots, |T|$

PD ÉPARSE 3

Déf. Rectangle R est **actif** pour j ssi

(1) $R.t_2 \leq j$; et

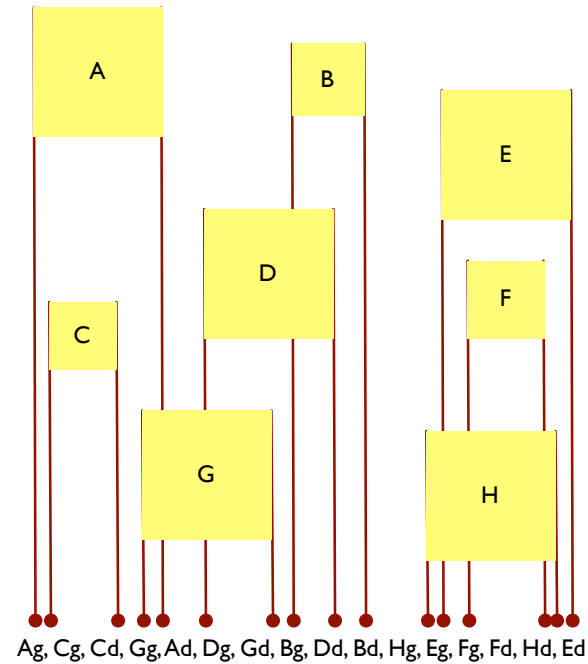
(2) $\forall R'$ avec $R'.t_2 \leq j$: si $R'.s_2 \leq R.s_2$ alors $V(R') \leq V(R)$.

Lemme. Si $R'.s_2 = R.s_2$ et tous les deux sont actifs, alors $V(R') = V(R)$.

$\Rightarrow \mathcal{L}$ contient des rectangles actifs avec $.s_2$ différents — choisir R avec $R.t_2$ minimum si > 1 rectangles actifs avec le même $.s_2$.

PD ÉPARSE 4

Trier les $.t1, .t2$: liste \mathcal{J} , calculer \mathcal{L}_j pour $j \in \mathcal{J}$ en ordre ascendant



PD ÉPARSE 5

en arrivant au côté gauche du rectangle R :

- trouver $R' \in \mathcal{L}_{R.t1}$ t.q. $R'.s2 \leq R.s1$ avec $V(R')$ maximal
- $V(R) \leftarrow V(R') + R.v$ (si aucun tel R' , alors $V(R) \leftarrow R.v$)

Lemme. Si $R, R' \in \mathcal{L}_j$, alors $R.s2 \neq R'.s2$ et $R.s2 < R'.s2$ ssi $V(R) < V(R')$.

Preuve. Par définition de \mathcal{L}_j .

\Rightarrow stocker \mathcal{L}_j dans l'ordre de $.s2$ et $V(\cdot)$ en même temps

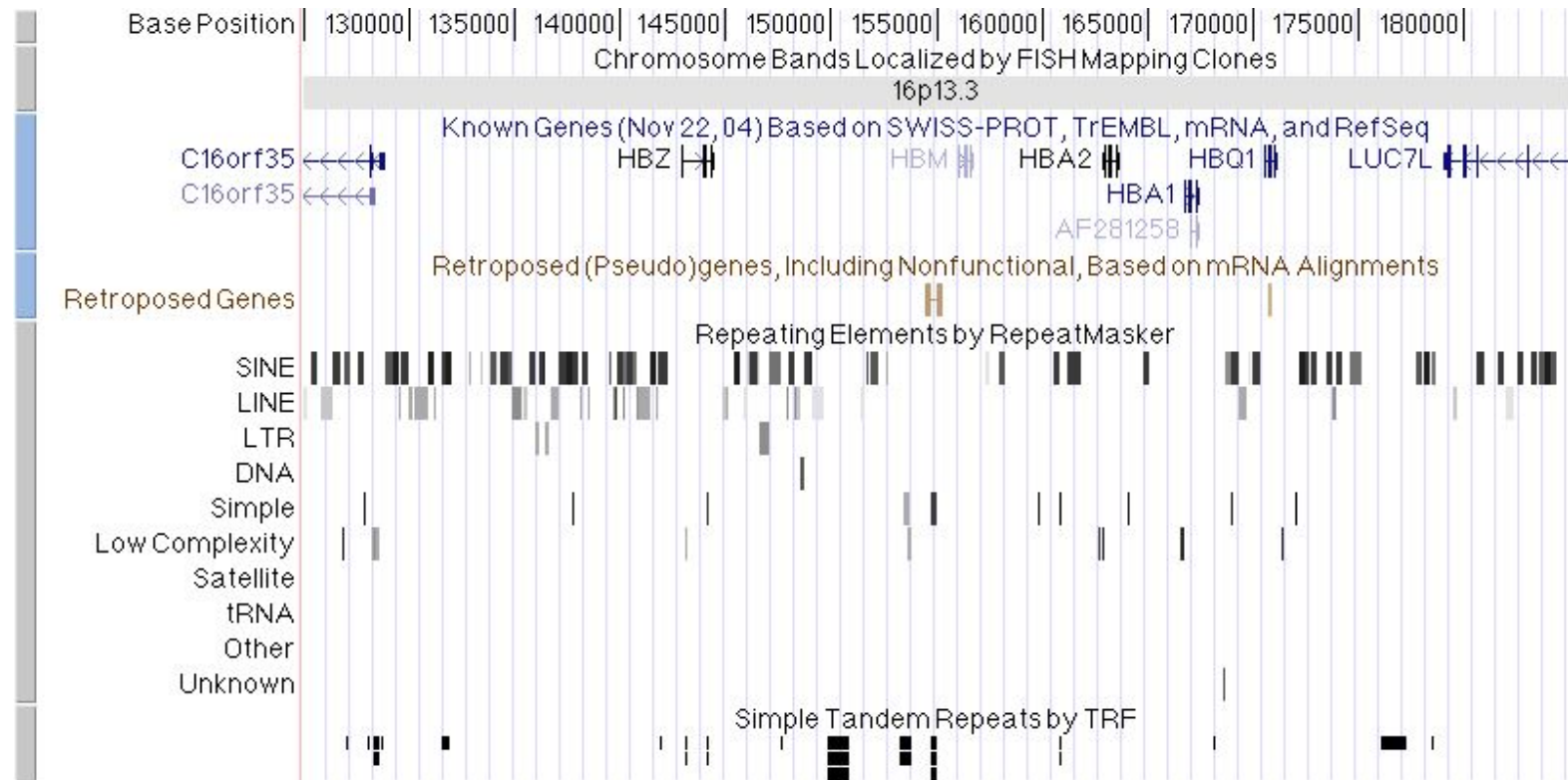
PD ÉPARSE 6

en arrivant au côté droit du rectangle R — mise à jour de \mathcal{L} :

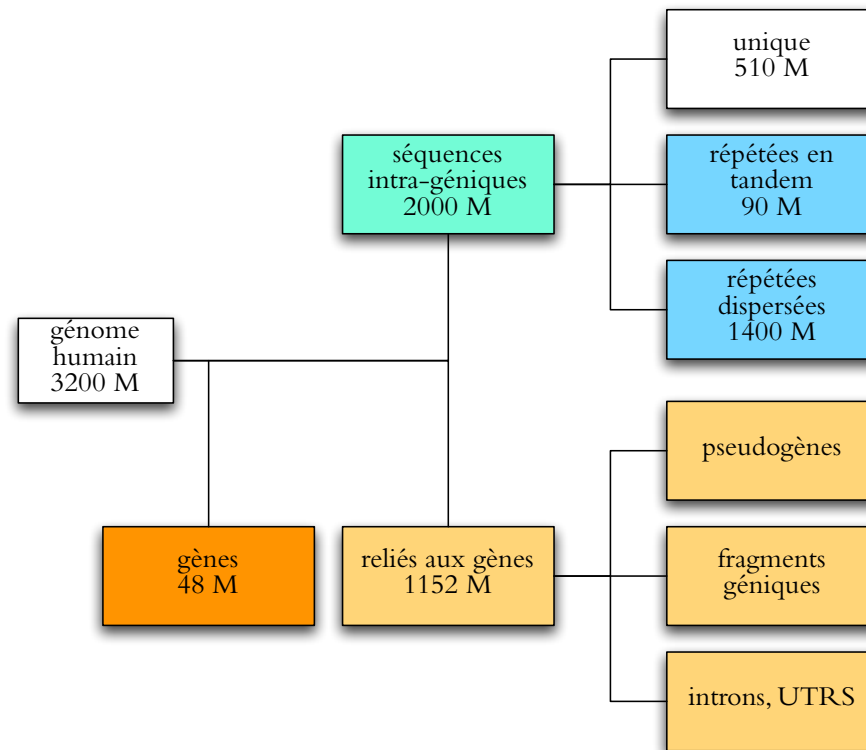
- supprimer $R' \in \mathcal{L}$ t.q. $R'.s_2 \geq R.s_2$ et $V(R') < V(R)$
- trouver $R' \in \mathcal{L}$ t.q. $R'.s_2 \leq R.s_2$ et $V(R')$ max. : si $V(R') = V(R)$ et $R'.s_2 < R.s_2$, ou si $V(R') < V(R)$, alors insérer R dans \mathcal{L} juste après R' (ou s'il n'y a pas de R' avec $R'.s_2 \leq R.s_2$, insérer au début)

Temps de calcul : $O(r \log r)$ pour r rectangles (utiliser un arbre binaire balancé ; chaque rectangle est inclus à ou est supprimé de \mathcal{L} une fois au plus)

CONSTITUTION D'UN GÉNOME



CONSTITUTION D'UN GÉNOME 2



en Procaryotes : 90–97% codant

après Watson et al *Molecular Biology of the Gene* (2004)

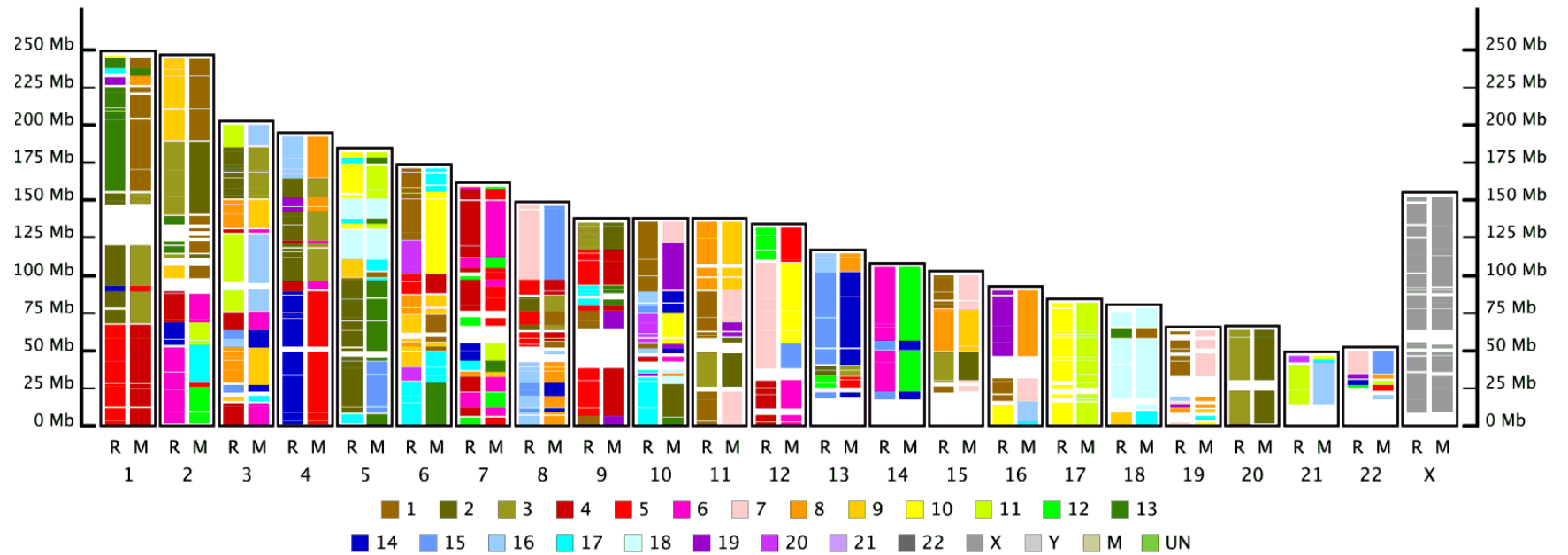
ALIGNEMENT DE GÉNOMES

Problème : beaucoup de réarrangements, insertions indépendantes dans les deux génomes

→ n'utiliser que les chaînes d'alignements locaux

on veut détecter des régions **synténiques** — descendantes de la même région dans l'ancêtre commun

SYNTÉNIÉS

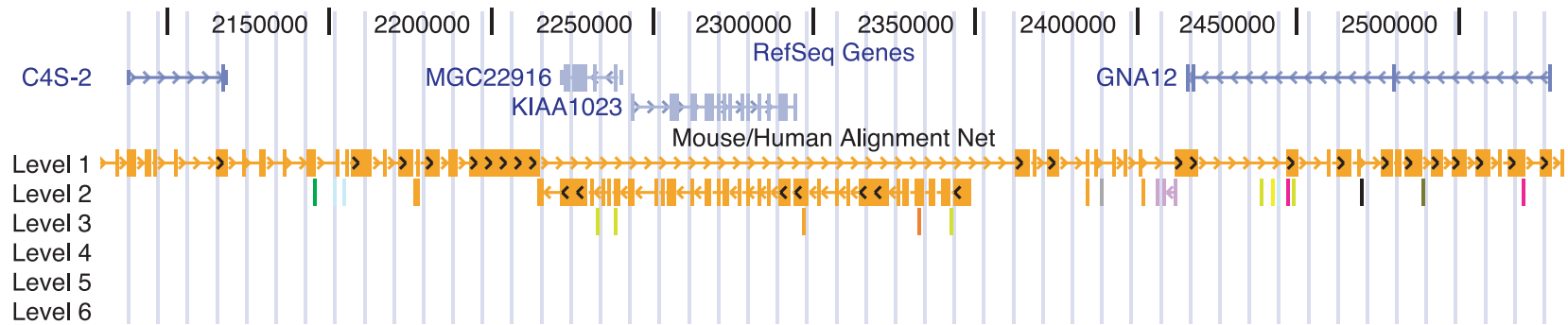


(chromosomes humains coloriés par synténies dans souris et rat)

<http://www.genboree.org/>

MICRO-RÉARRANGEMENTS

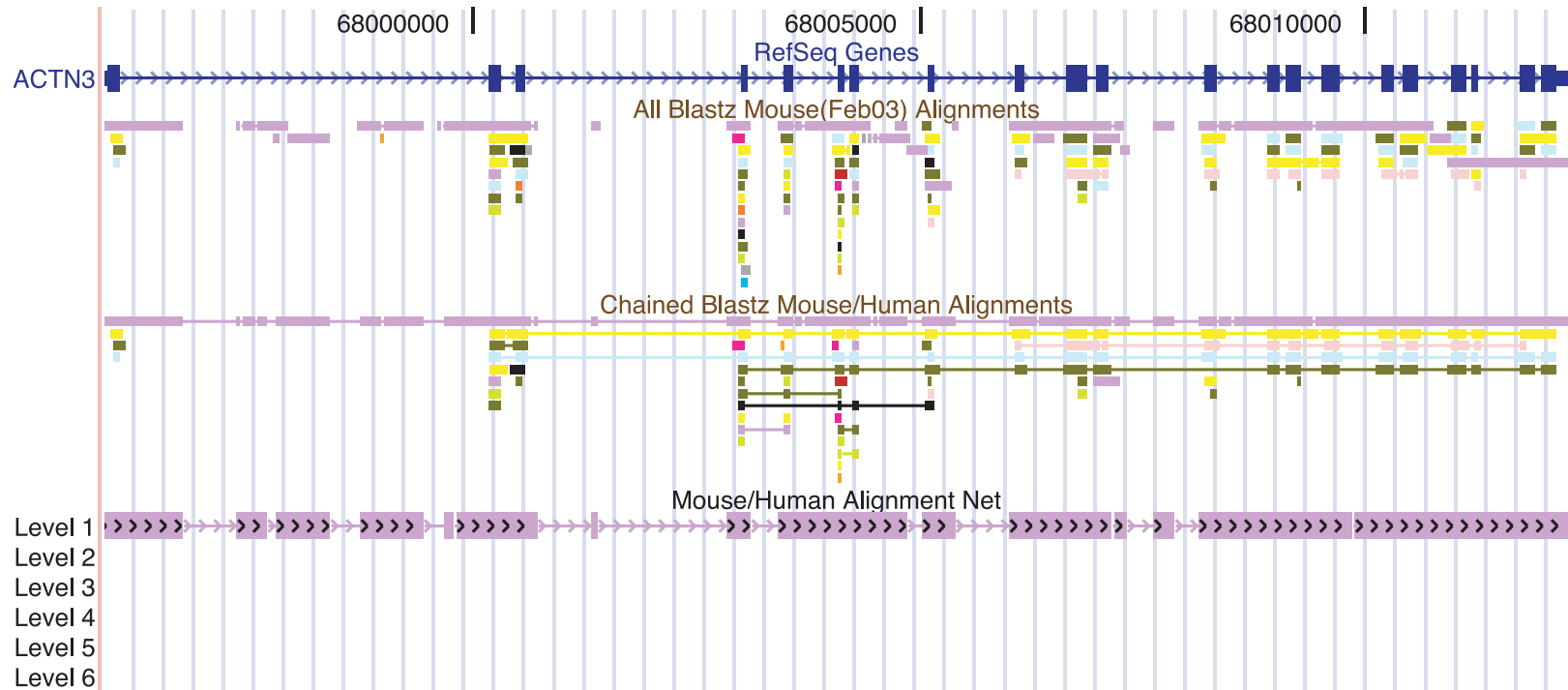
beaucoup de réarrangements à tous les niveaux



Kent & al. *PNAS* 100 : 11484

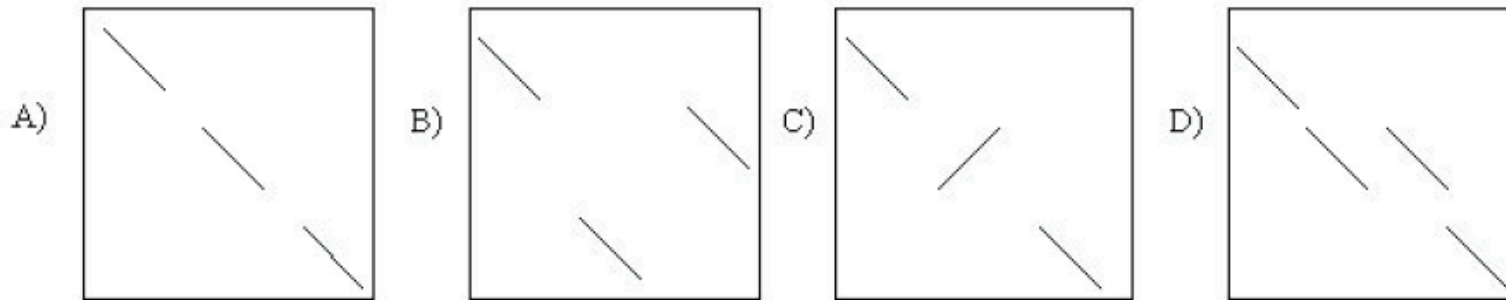
CHAÎNES ET NETS

Chaînage récursive — chaînes imbriquées



Kent & al. *PNAS* 100 : 11484

ALIGNEMENT GLOBAL — RÉARRANGEMENTS

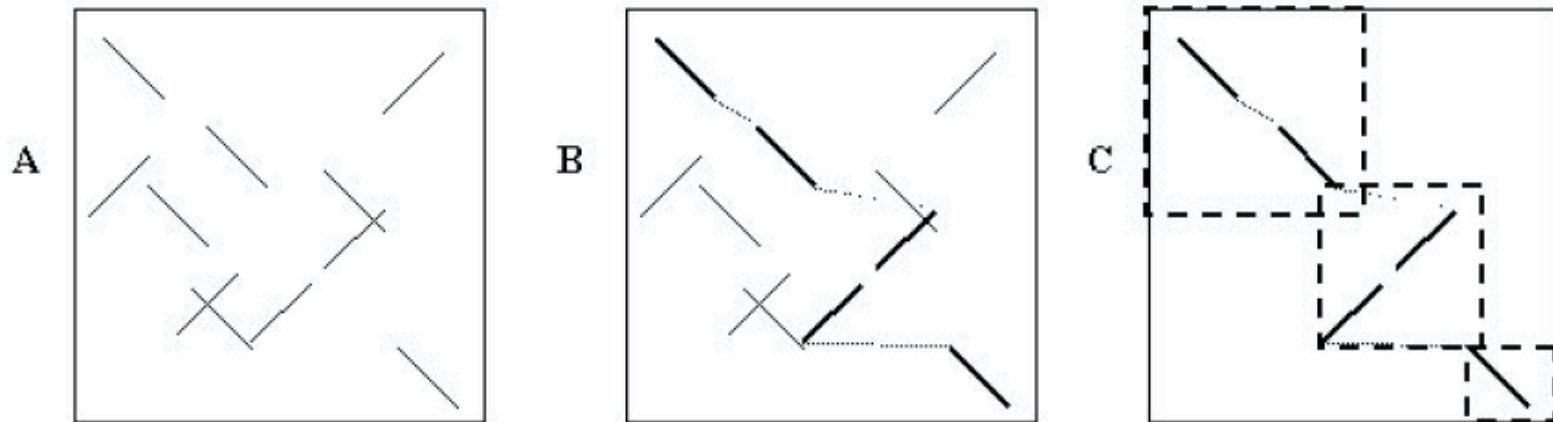


(B. transposition, C. inversion, D. duplication)

Dans des synténies humain-souris : par 1 Mbp on a en moyenne
2 inversions, 17 duplications, 7 transpositions, 200 deletions de longueur > 100 pb,

Brudno & al. *Bioinformatics* 19 :i54; Kent & al. *PNAS* 100 : 11484

ALIGNEMENT GLOCAL



réarrangements : inversions et translocations sont permises

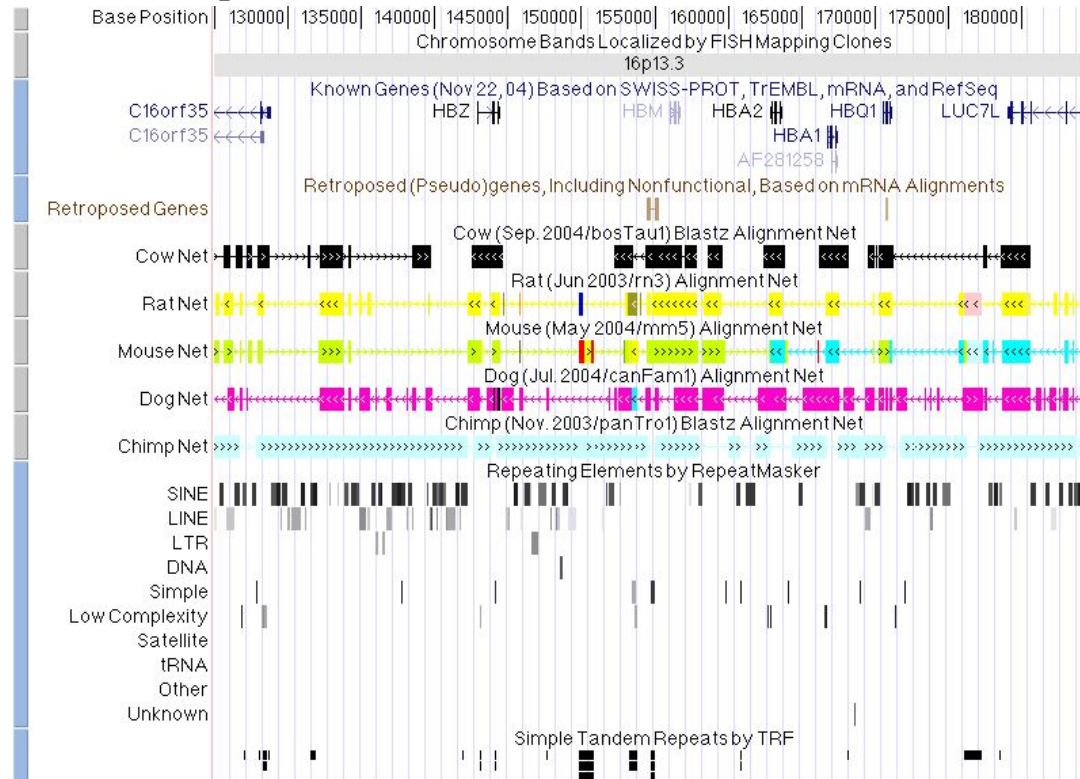
A. alignements locaux

B. chaîne 1-monotone

C. LAGAN dans les blocs sans inversions

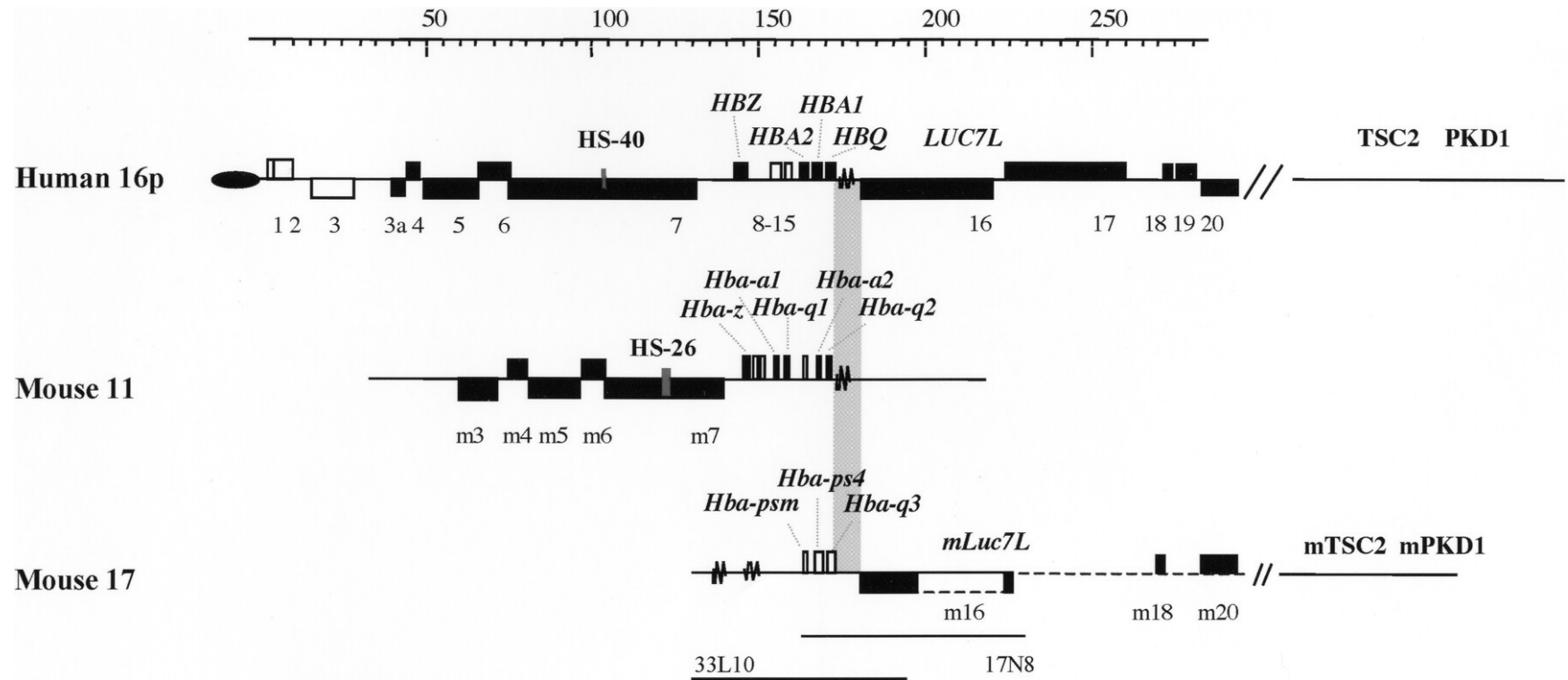
EXEMPLE : α -GLOBINES DANS LE SOURIS

[breakpoint dans le souris]



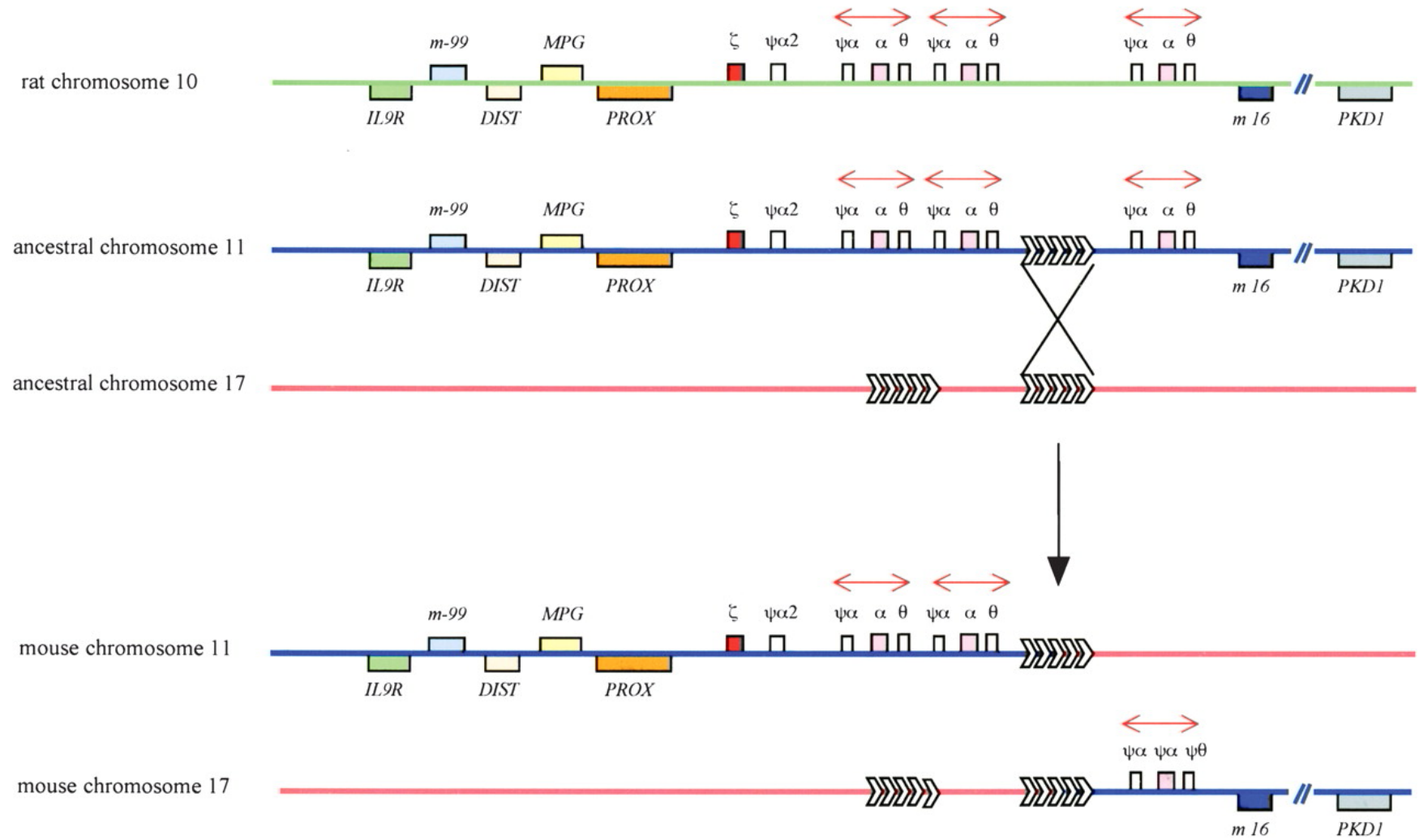
Tufarelli & al. *Genome Res.* 14 :623 (2004)

α-GLOBINES DANS LE SOURIS



Tufarelli & al. *Genome Res.* 14 :623 (2004)

α -GLOBINES

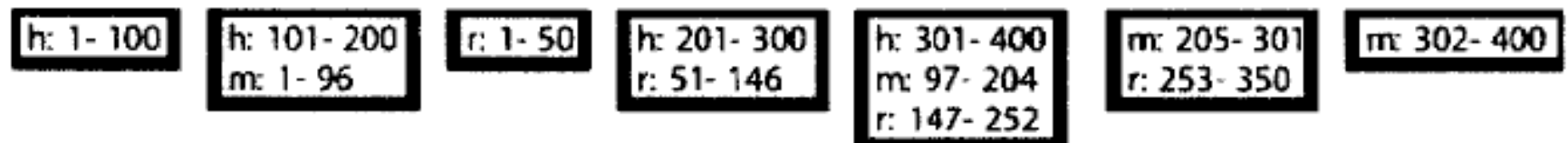


Tufarelli & al. *Genome Res.* 14 :623 (2004)

ALIGNMENT COMME ANNOTATION

Annotation d'un génome : segments avec des propriétés (gène, promoteur, exon, intron, ...)

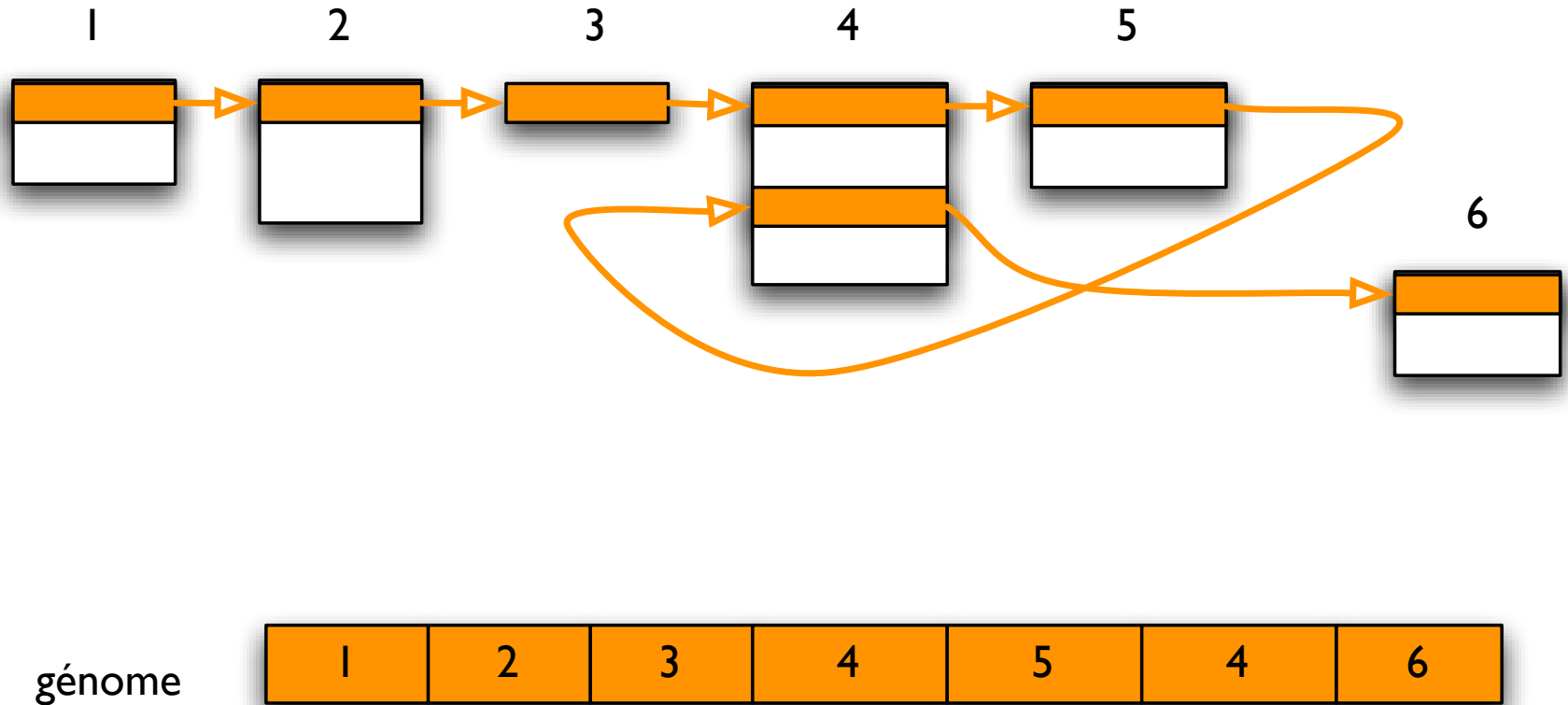
Alignement de deux ou plusieurs génomes : peut être considéré comme l'annotation d'un génome par les autres



Alignement : **ensemble de blocs** où chaque bloc est une région d'un seul génome, ou l'alignement local de multiples génomes

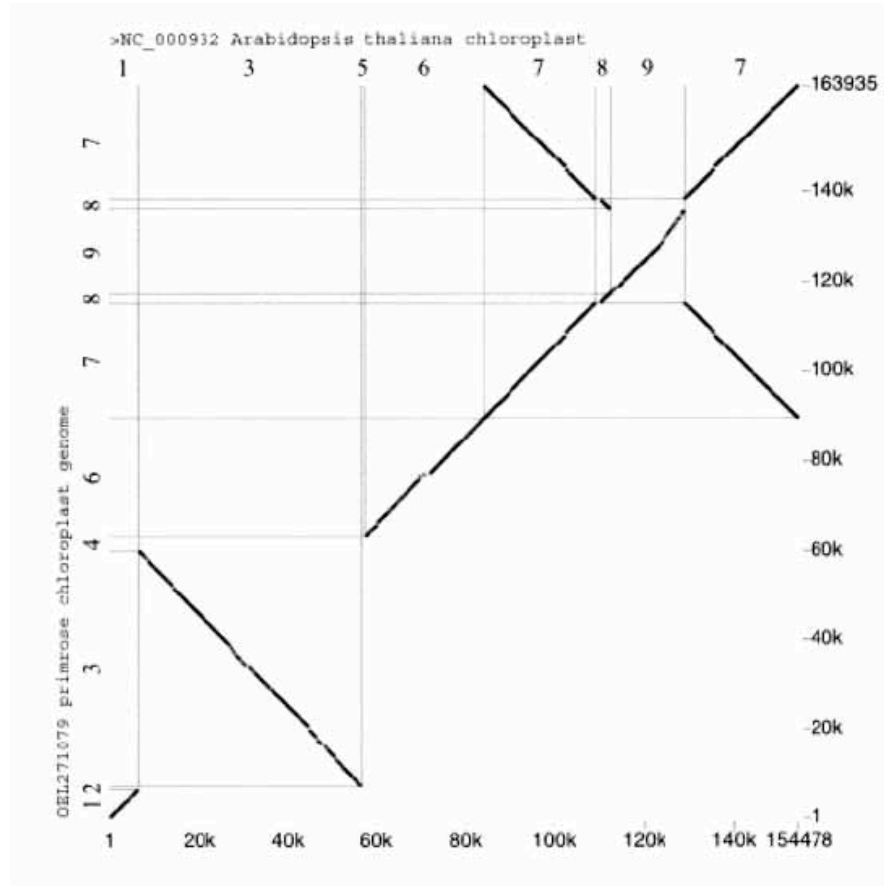
ENSEMBLE DE BLOCS

Projection des blocs sur un génome



Blanchette & al. *Genome Res.* 14 : 708

EXEMPLE



Blanchette & al. *Genome Res.* 14 : 708