

TP cartographie de marqueurs

Thomas Schiex et Simon de Givry, INRA UBIA Toulouse

11 octobre 2013

La documentation de Carthagene se trouve ici : <http://www.inra.fr/mia/T/CarthaGene/>
Les fichiers de données sont ici : <http://www.inra.fr/mia/T/schiex/Doc/teaching.shtml>
Sous Windows, sauvegardez les sans changer leur extension sur le Bureau (C:/Users/nom/Desktop).
Attention à ne pas utiliser de caractères accentués ni espaces dans les noms de vos répertoires!

1 Cartes génétiques de la mouche *Drosophila*

1.1 Protocole de croisement

Sachant deux lignées pures *Drosophila Simulans* et *Drosophila Mauritiana* dont le croisement donne des femelles fertiles et des males steriles, proposer deux schémas possibles de croisement de type Backcross F2.

Les fichiers `bm4zb.cg` et `bs4zb.cg` contiennent les typages au format Carthagene du résultat de chacun de ces deux croisements pour 39 marqueurs moléculaires et 162 échantillons. Charger le fichier `bm4zb.cg` dans Carthagene (commandes `cd`, `ls` et `dsload`).

1.2 Nettoyer les données

Calculer manuellement la distortion des deux premiers marqueurs *acr* et *ald* dans le fichier `bm4zb.cg`. Que peut on conclure? Par la suite, on ne prend pas en compte la distortion.

Y-a-t-il des marqueurs confondus? (commande `mrkdouble`)

Si oui, les retirer avec la commande `mrkmerges`.

1.3 Trouver les groupes de liaison

Combien y a t il de groupes de liaison? (`group` et `mrknames [groupget 1]`)

A quels chromosomes se réfèrent-ils? Pour cela, chercher par exemple la position des marqueurs *ninaE* et *prd* sur <http://flybase.org> (champ *Jump to Gene*).

Calculer le LOD score entre *ninaE* et *prd* (`mrksetset [mrkids {ninaE prd}]`, `mrklod2p`, `mrksetset [mrkallget]`). Que peut on en conclure?

1.4 Construire les cartes pour chaque groupe de liaison

Calculer le taux de recombinaison entre les marqueurs *efi* et *jan*. Quel est leur distance de Haldane en cM? (`mrksetset [mrkids {efi jan}]`, `mrkfr2p`, `mrkdist2p`)

Comparer en terme de vraisemblance les cartes construites avec différentes heuristiques (`nicemapd`, `nicemapl`, `mfmapped`, `mfmapl`). Peut-on les améliorer par recherche exhaustive globale? locale (flips)? Peut-on avoir confiance dans la meilleure carte trouvée (Taboo/greedy et

heaprint)? Donner la distance de Haldane en cM de la meilleure carte pour chaque groupe de liaison (bestprintd). Et sauvegarder son image (copie d'écran ou clic-droit et commande imprimer dans un fichier au format postscript).

1.5 Drosophila Simulans et Mauritiana

Refaire la construction des cartes en utilisant conjointement les deux sources d'information issues des deux protocoles backcross (fichiers bs4zb.cg et bm4zb.cg) (commande dsmergen).

Comparer les résultats obtenus avec les cartes précédentes.

2 Carte d'hybrides irradiés du chien aidé du génome humain

On dispose du génome de référence de l'homme et on cherche à cartographier le chromosome 2 du chien. Pour cela, on a génotypé à l'aide de 88 hybrides irradiés des marqueurs associés à des gènes présents à la fois sur l'homme et sur le chien (gènes orthologues). On dispose ainsi d'un ordre de référence de 427 marqueurs sur les chromosomes humains. Le fichier order02_human.cg contient pour chaque marqueur le numéro du chromosome humain et sa position en paire de bases. Le fichier rh02_dog.cg contient les 88 motifs de retention (A=marqueur absent, H=here=marqueur retenu).

2.1 Distance évolutive entre génomes

Le principe de la cartographie comparée est que des génomes proches en distance évolutive (distance phylogénétique) doivent avoir peu d'événements de remaniements chromosomiques (*inversion, transposition, transposition inverse*) entre eux. Une mesure simplifiée de la distance évolutive consiste à compter le nombre de points de cassure (*breakpoint* en anglais) entre deux ordres de marqueurs. Un point de cassure est une paire de marqueurs consécutifs dans un ordre et non consécutif dans l'autre (ou sur des chromosomes différents). Par exemple, il y a 2 breakpoints entre les ordres 12345 et 14325. Combien y a t il de breakpoints entre un ordre par défaut (i.e. dans l'ordre de numérotation des marqueurs) et l'ordre de référence humain? (dsbp) Combien y a t il de breakpoints obligatoires quelque soit la carte proposée?

2.2 Construction de carte comparée

Le modèle probabiliste consiste à combiner vraisemblance des données RH et probabilité a priori d'un ordre sachant le nombre de breakpoints (commandes dsmergor et dsbplambda). En utilisant l'ordre par défaut, donner la succession des segments chromosomiques humains. Comparer avec la carte de synténie sur <http://www.ensembl.org> (*Species/Dog, Chr02 overview, Comparative genomics, Synteny, Human*).

Les données de la mouche proviennent de l'article

Genetic Architecture of a Morphological Shape Difference Between Two Drosophila Species
Zhao-Bang Zeng, Jianjun Liu, Lynn F. Stamb, Chen-Hung Kao, John M. Mercer, Cathy C. Laurie
Genetics, Vol. 154, 299-310, January 2000

Les données du chien proviennent de l'article

T. Faraut, S. de Givry, P. Chabrier, T. Derrien, F. Galibert, C. Hitte, and T. Schiex
A comparative genome approach to marker ordering
Bioinformatics, 23(2):50-56, 2007