

# Comparaison de minisatellites

S everine B erard et  eric Rivals

L.I.R.M.M., UMR CNRS 5506 - 161, rue Ada, 34392 Montpellier Cedex 5

courriel : {berard, rivals}@lirmm.fr

Comme d'autres s equences r ep et ees en tandem, les minisatellites sont polymorphes. En effet, ils subissent des fluctuations de leur nombre de motifs. Ce ph enom ene est d u  a des  ev enements  evolutifs dont le r esultat est soit une *duplication en tandem*, soit une *d el etion en tandem*. Nous entendons par duplication en tandem, un  ev enement qui ajoute un motif copi e  a c ot e de l'original, et par d el etion en tandem, l' ev enement inverse. Comme pour les microsattellites, la variation de longueur des minisatellites est impliqu ee dans plusieurs maladies comme le diab ete, l' pilepsie et le cancer. Du fait de leur polymorphisme, les minisatellites sont utilis es pour la cartographie g en etique et les  tudes m edico-l egales.

En 1991, Jeffreys et ses collaborateurs ont mis au point une technique PCR sp ecifique aux minisatellites, la MVR-PCR [2], o u MVR signifie Minisatellites Variant Repeat. Cette m ethode fournit une *carte du minisatellite*, c ad une s equence de symboles, o u chaque symbole est associ e avec une variante du motif. Pour exploiter pleinement les informations  evolutives contenues dans ces cartes, il est crucial de pouvoir les comparer de mani ere automatique. Pour comparer ces cartes, nous proposons ici une m ethode nouvelle qui prend en compte les  ev enements de duplication et de d el etion en tandem. Nous nous placons dans le cadre d'un mod ele  evolutif sym etrique et unaire. Notre mod ele consid ere cinq  ev enements  evolutifs mod elis es s'appliquant aux motifs, la mutation (M), l'insertion (I), la suppression (S), l'amplification (A) et la contraction (C). Mutation, insertion et suppression sont les  ev enements que l'on consid ere traditionnellement dans l'alignement de s equences mis  a part qu'ils s'appliquent ici  a des motifs. Les deux  ev enements sp ecifiques sont l'amplification et la contraction, ils correspondent respectivement  a la duplication et  a la d el etion en tandem. Par exemple, la s equence *abc* subit l'amplification du motif *b*, puis sa contraction :

**amplification** :  $abc \longrightarrow abbc$

**contraction** :  $abbc \longrightarrow abc$ .

Notre mod ele est unaire car l'amplification (resp. la contraction) ajoute (resp. retire) un seul motif  a la fois. Pour compl eter ce mod ele, nous avons besoin d'un crit ere quantitatif permettant d'estimer la similarit e des cartes. Aussi,  a chaque op eration est associ e un co ut r eel positif. Une s equence d' ev enements qui transforme une carte *s* en une carte *r* est appel ee un *alignement*. Le co ut d'un alignement est la somme des co uts des op erations qui le composent. Nous traduisons le fait que les amplifications et les contractions ont une fr equance plus  elev ee que les autres op erations en leur donnant un co ut plus faible.

L'algorithme que nous avons mis au point pour trouver l'alignement optimal entre deux cartes est un algorithme de programmation dynamique. Il donne un *score* d'alignement qui est une distance m etrique pour les cartes. Cela permet de reconstruire des relations  evolutives entre individus ou populations,  a partir de la matrice de distances. Pour plus de d etails vous pouvez vous r ef erer  a [1].

Nous avons appliqu e notre algorithme au minisatellite humain MSY1 qui est situ e sur le chromosome Y [3]. Le jeu de donn ees contient 609 cartes d'individus r eparties en 27 haplogroupes. La figure 1 repr esente l'arbre des individus de l'haplogroupe 4 reconstruit  a partir de notre distance. On constate la s eparation entre les japonais d'un c ot e, et les tib etains et un mongol de l'autre. D'autres r esultats sont en cours d'investigation.

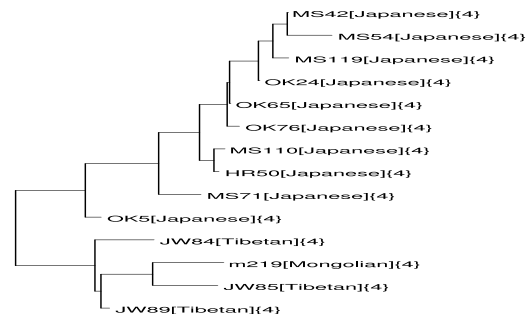


FIG. 1 – Arbre obtenu sur l'haplogroupe 4.

## R ef erences

- [1] S. B erard and E. Rivals. In *Proc. of the 6th RECOMB*, pages 67–76, Washington, USA, 2002. ACM Press.
- [2] A. J. Jeffreys, A. MacLeod, K. Tamaki, D. L. Neil, and D. G. Monckton. *Nature*, 354(6350) :204–9, 1991.
- [3] M. A. Jobling, N. Bouzekri, and P. G. Taylor. *Hum Mol Genet*, 7(4) :643–53, 1998.