

Bioinformatique pour le traitement des données de séquençage

Détection de variants

Maud Gautier, Annabelle Haudry, Thibault Latrille

11 Septembre 2018



Qu'est ce que la détection de variants ?



Données de séquençage
d'un individu diploïde

Détection
de variants

Variant
homozygote

Variant
hétérozygote

AATTTATTATTAGGCGATACGGAGGCCGGAGCAGAGACAGC
ATTTTATTATTAGGCGATATGGAGGCAGAGCAGAGTCAGC

Individu diploïde

AATTTATTTTTAGGCGATACGGAGGCCAGAGCAGAGTCAGC

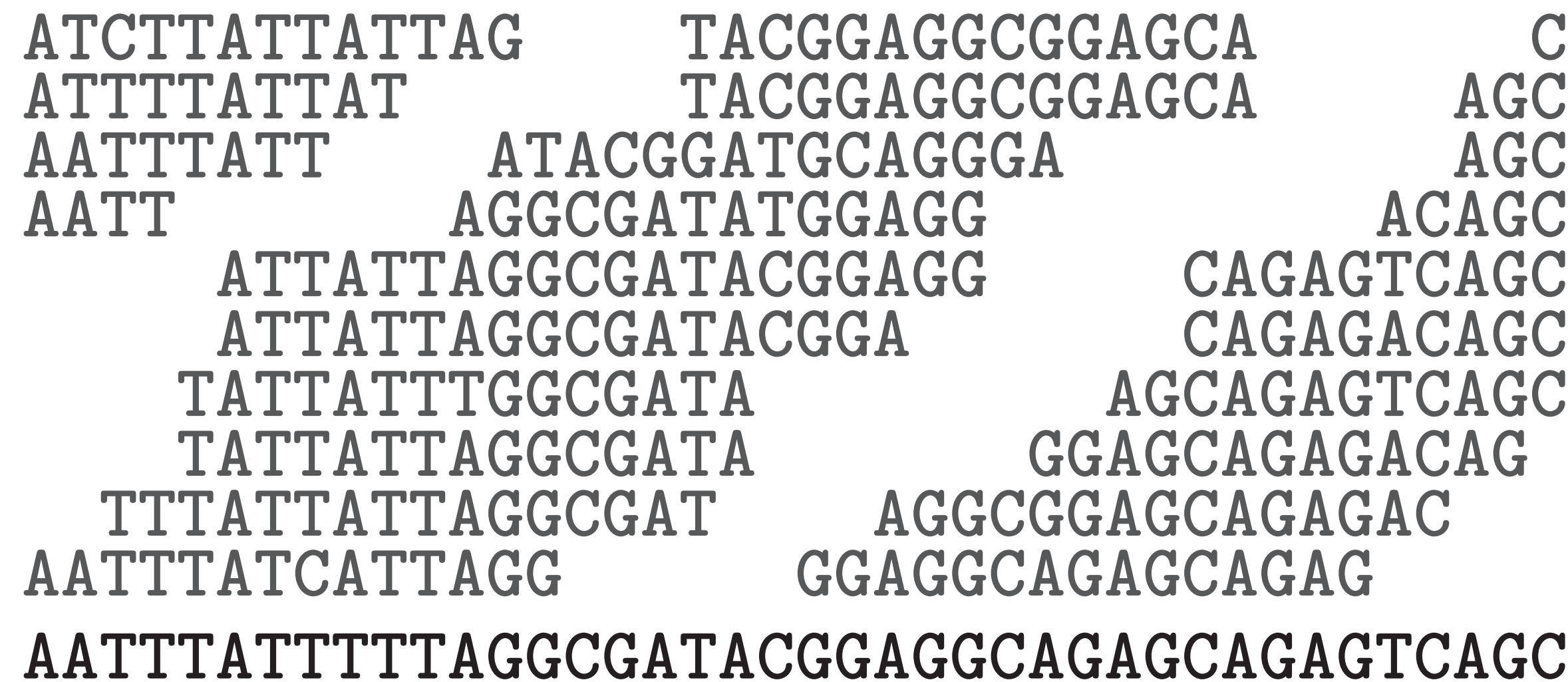
Genome de référence

Qu'est ce que la détection de variants ?



Données de séquençage d'un individu diploïde (.fastq)

Alignement (matinée)



Données de séquençage alignées (.bam et .sam)

Genome de référence (.fasta)

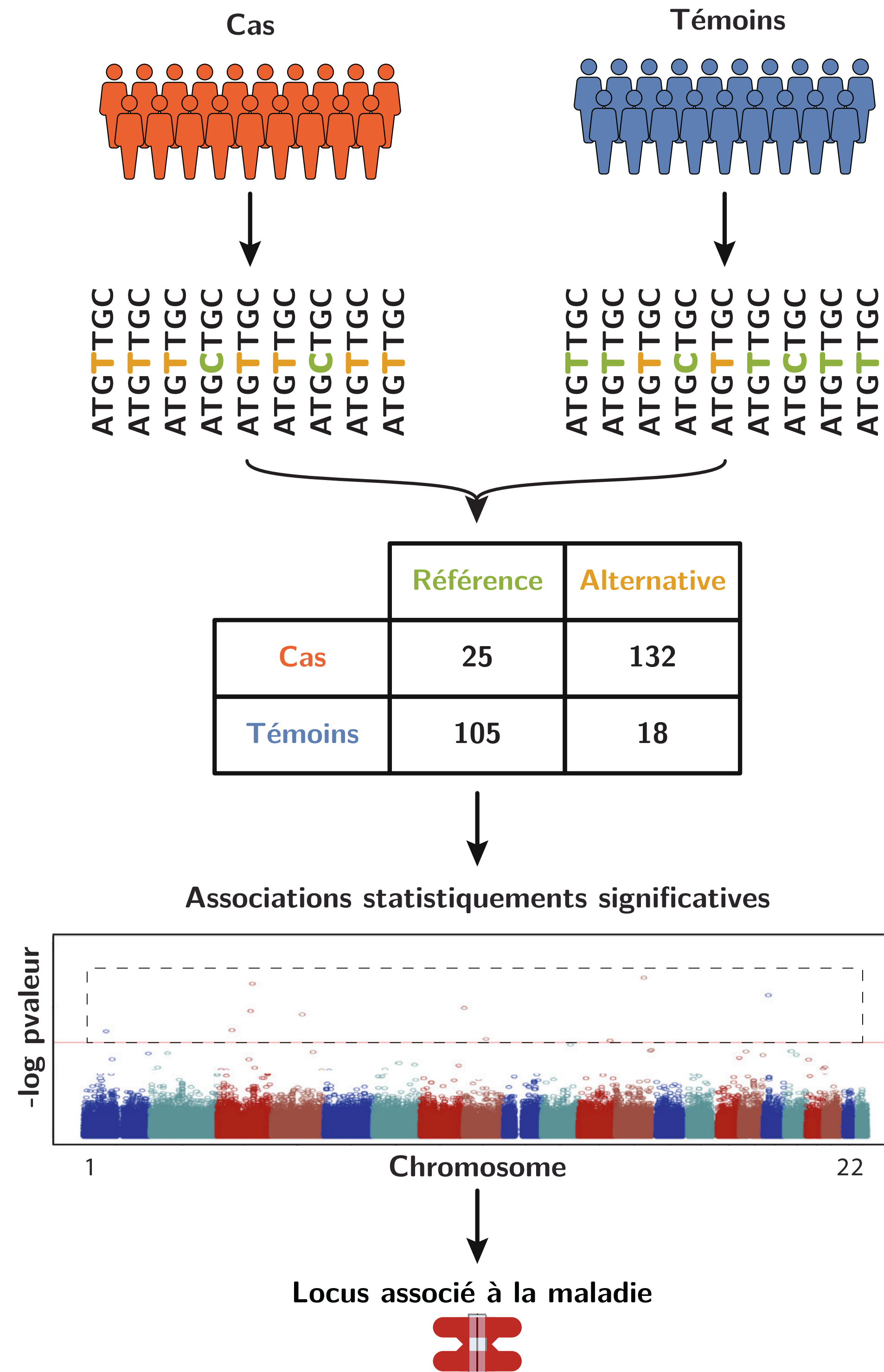
Détection de variants (après-midi)



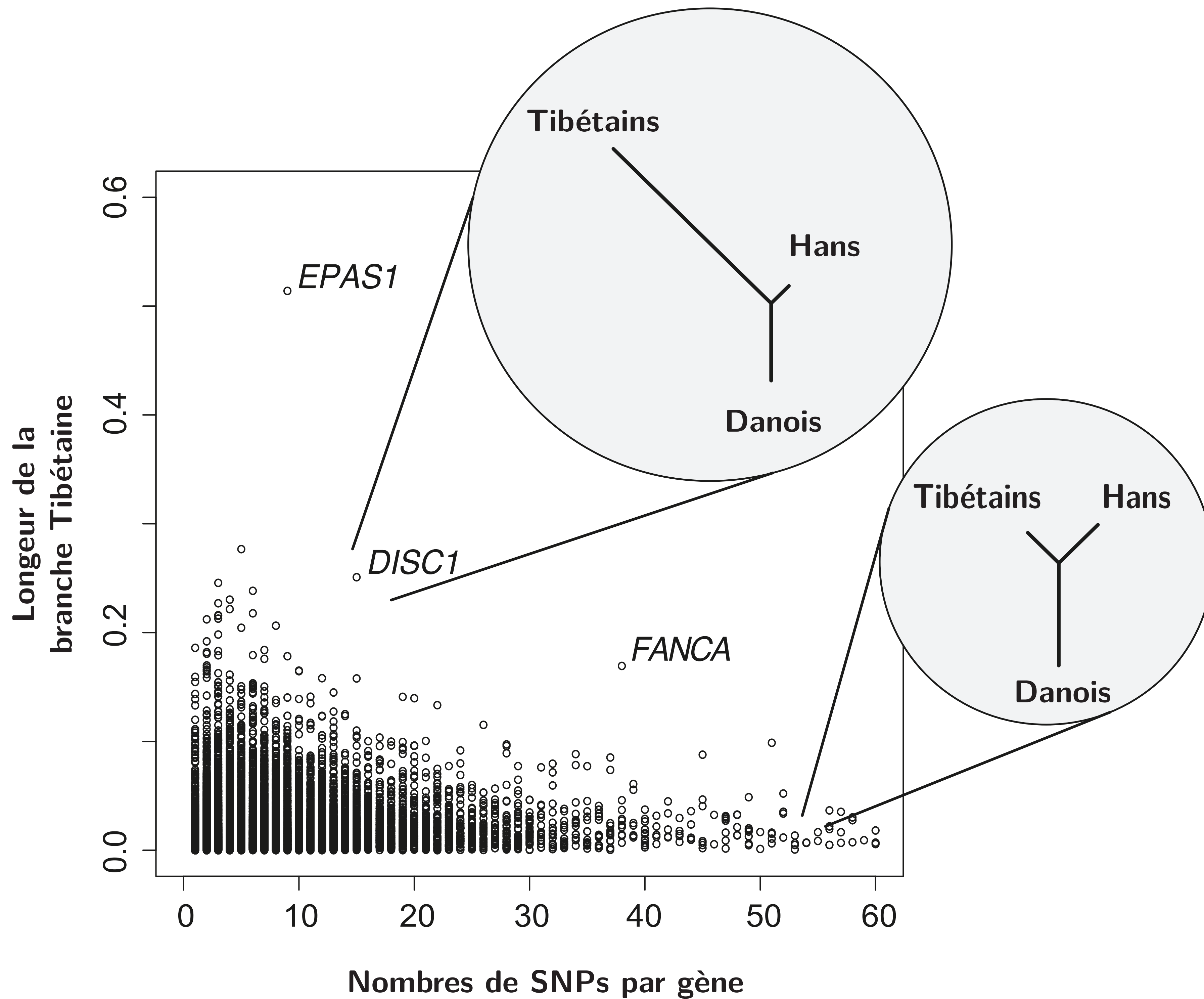
Individu diploïde (.vcf)

Genome de référence (.fasta)

Pourquoi cherche-t-on à détecter des variants ?



Quels gènes sont sélectionnés ?

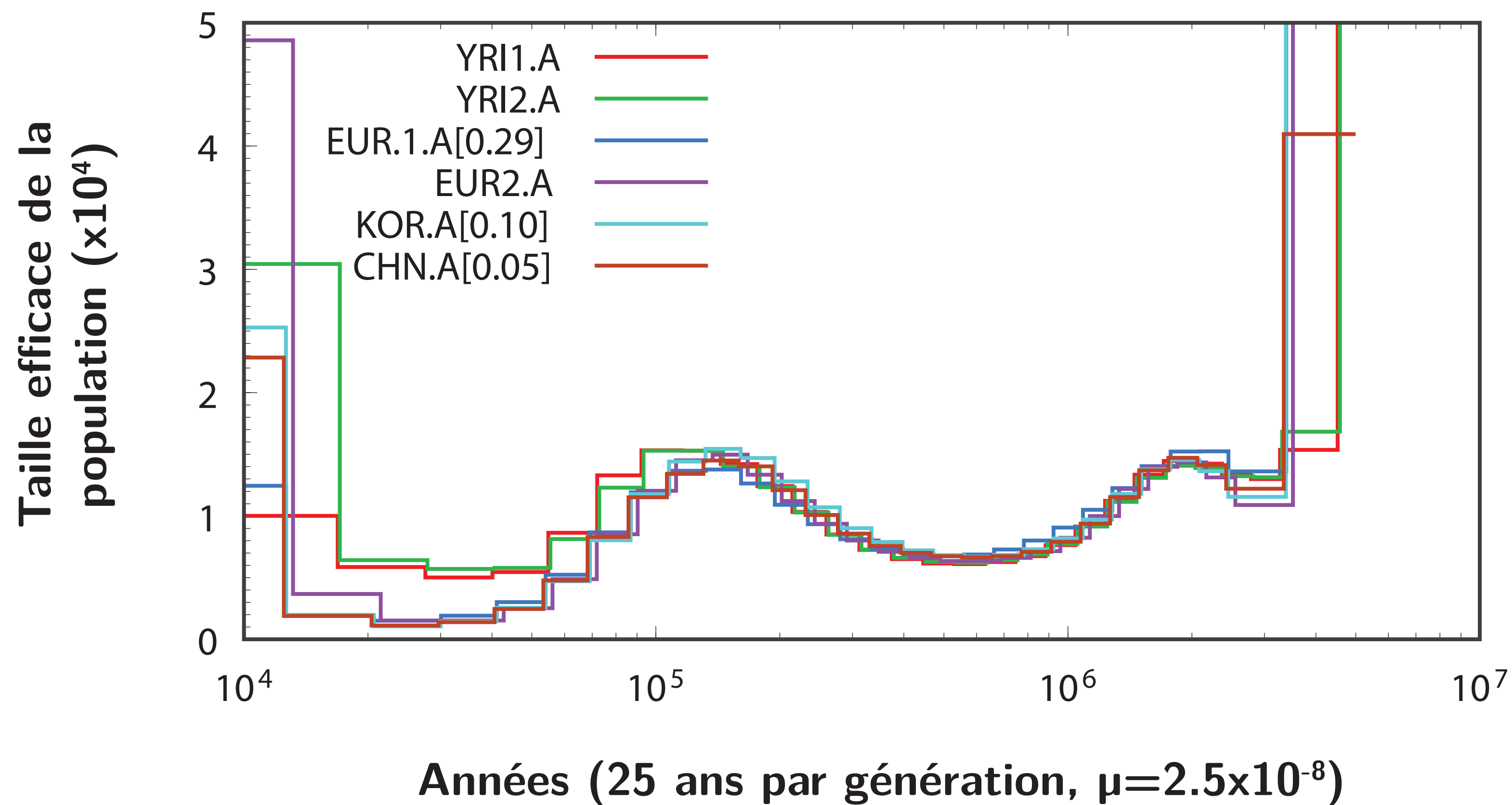


Yi *et al*, Science (2009)

Nos ancêtres étaient-ils nombreux ?

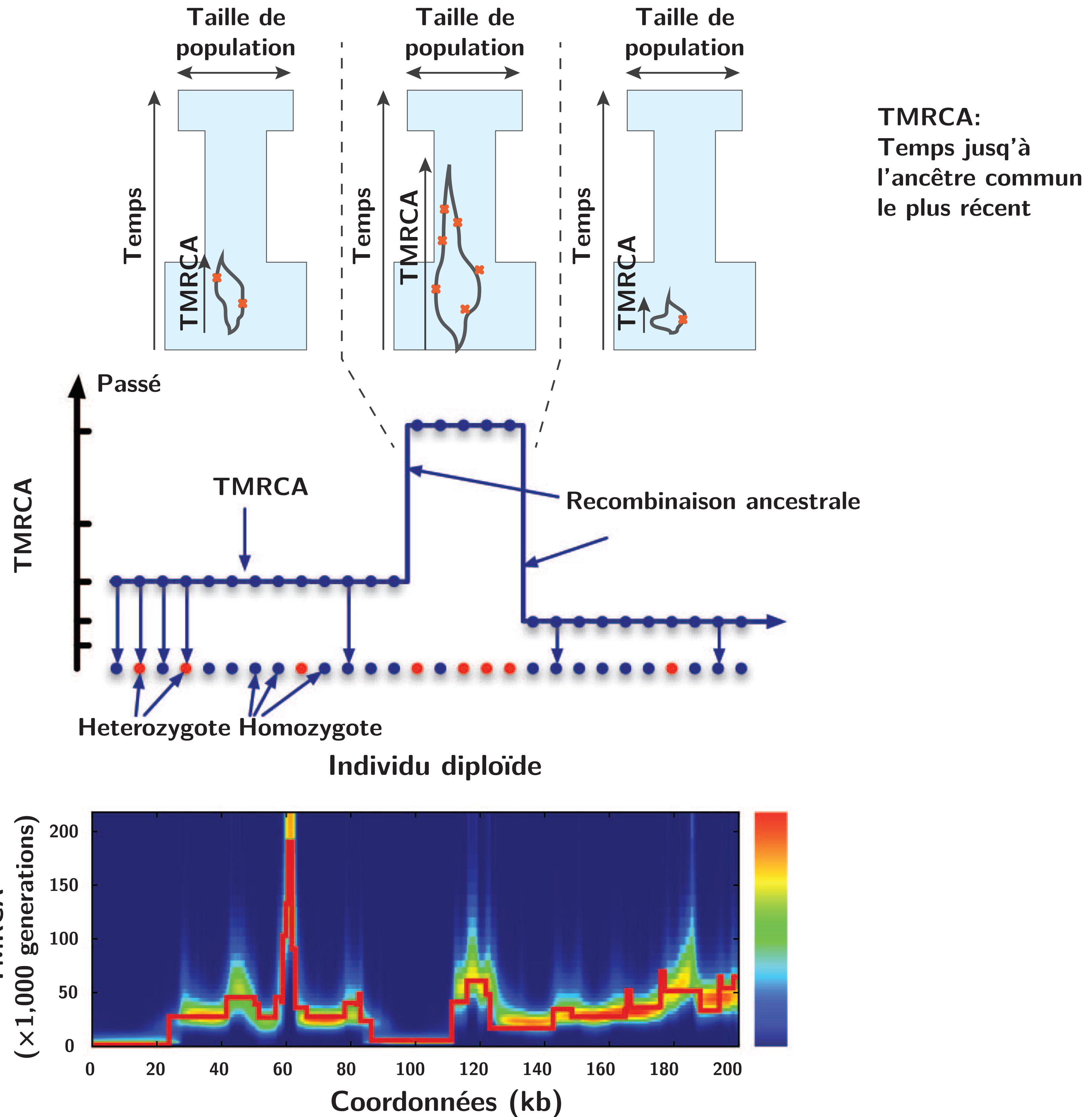
-----AATTATTATTAGGCGATACGGAGGCGGAGCAGAGACAGC----- Individu diploïde
-----ATTTTATTATTAGGCGATATGGAGGCGAGCAGAGTCAGC-----

Inference of human population history
from individual whole-genome sequences,
H. Li & R. Durbin, *Nature* (2011)

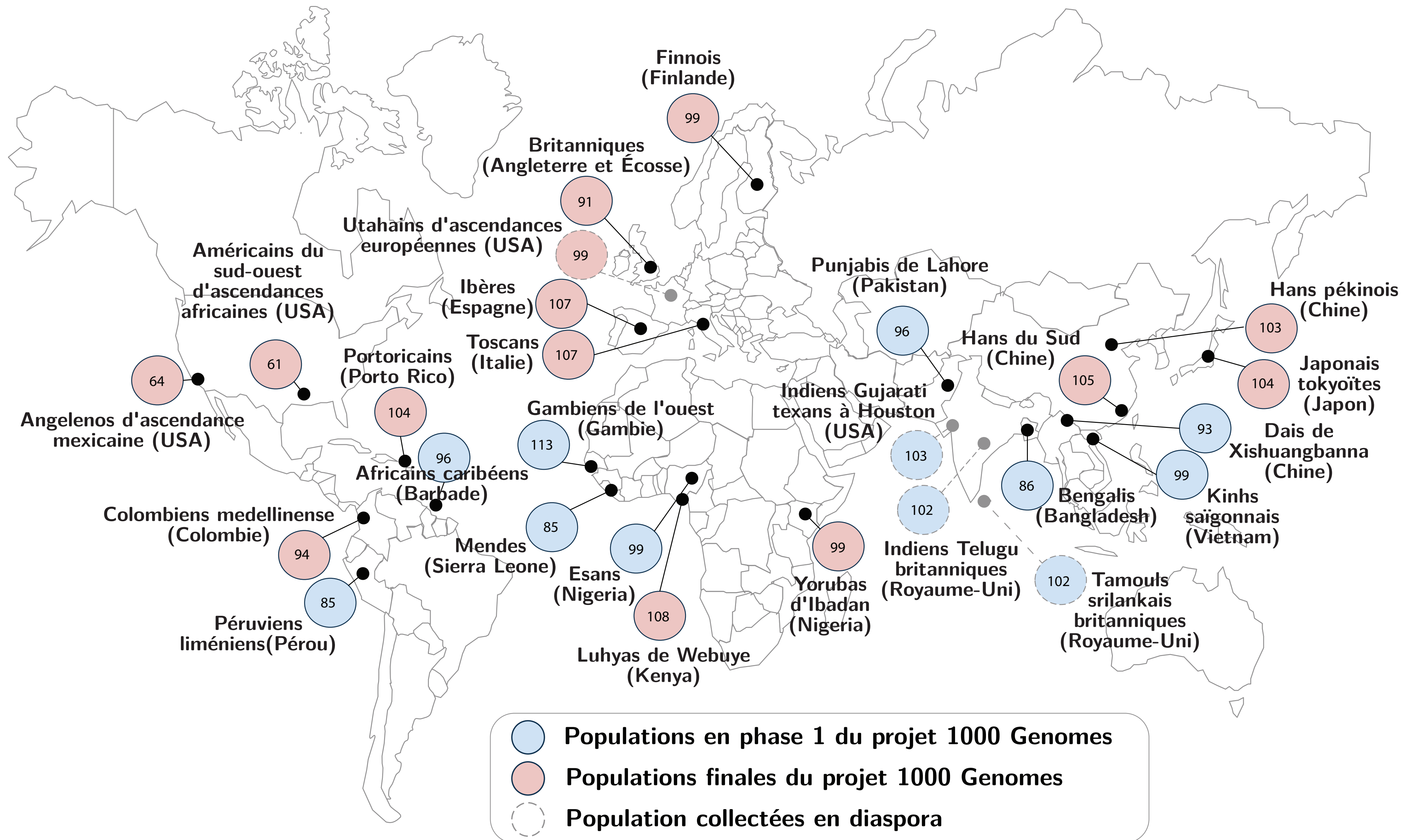


Li & Durbin, *Nature* (2011)

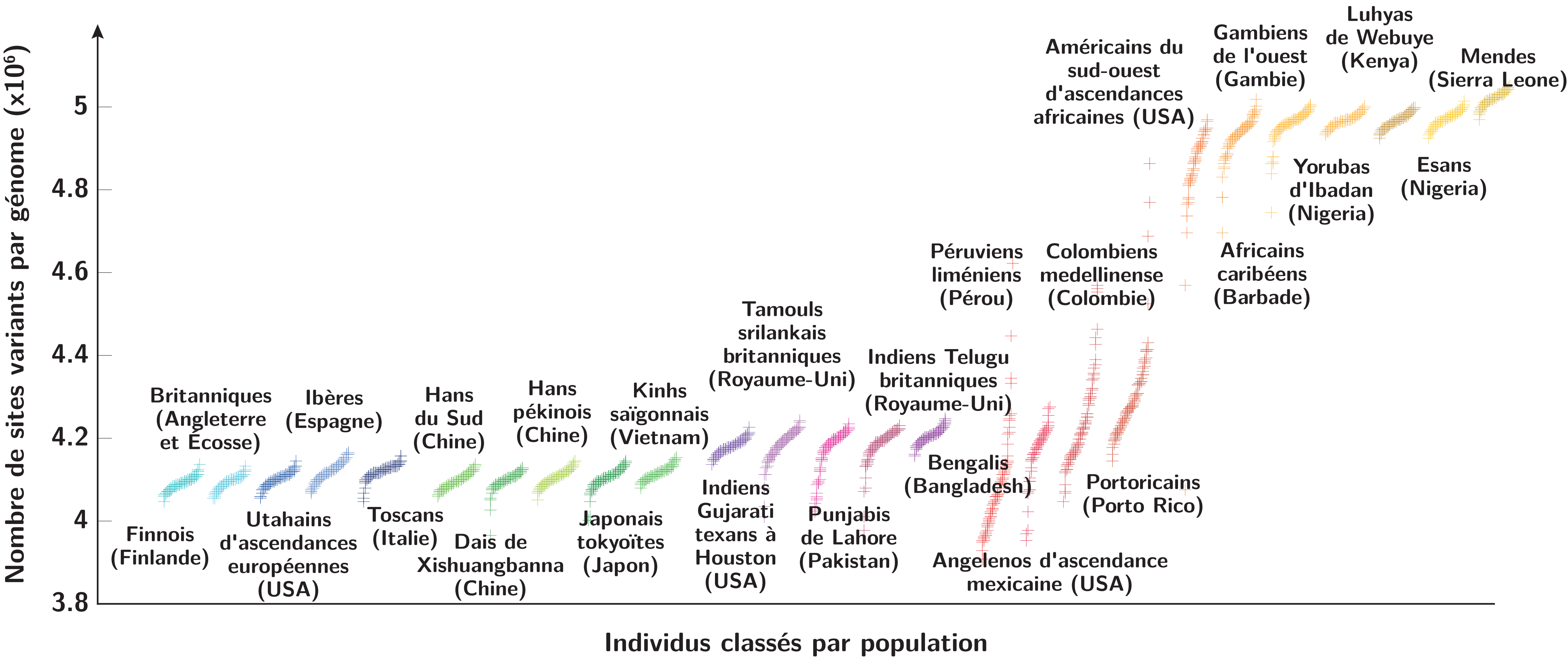
Nos ancêtres étaient-ils nombreux ?



Quelles données va-t-on utiliser ?

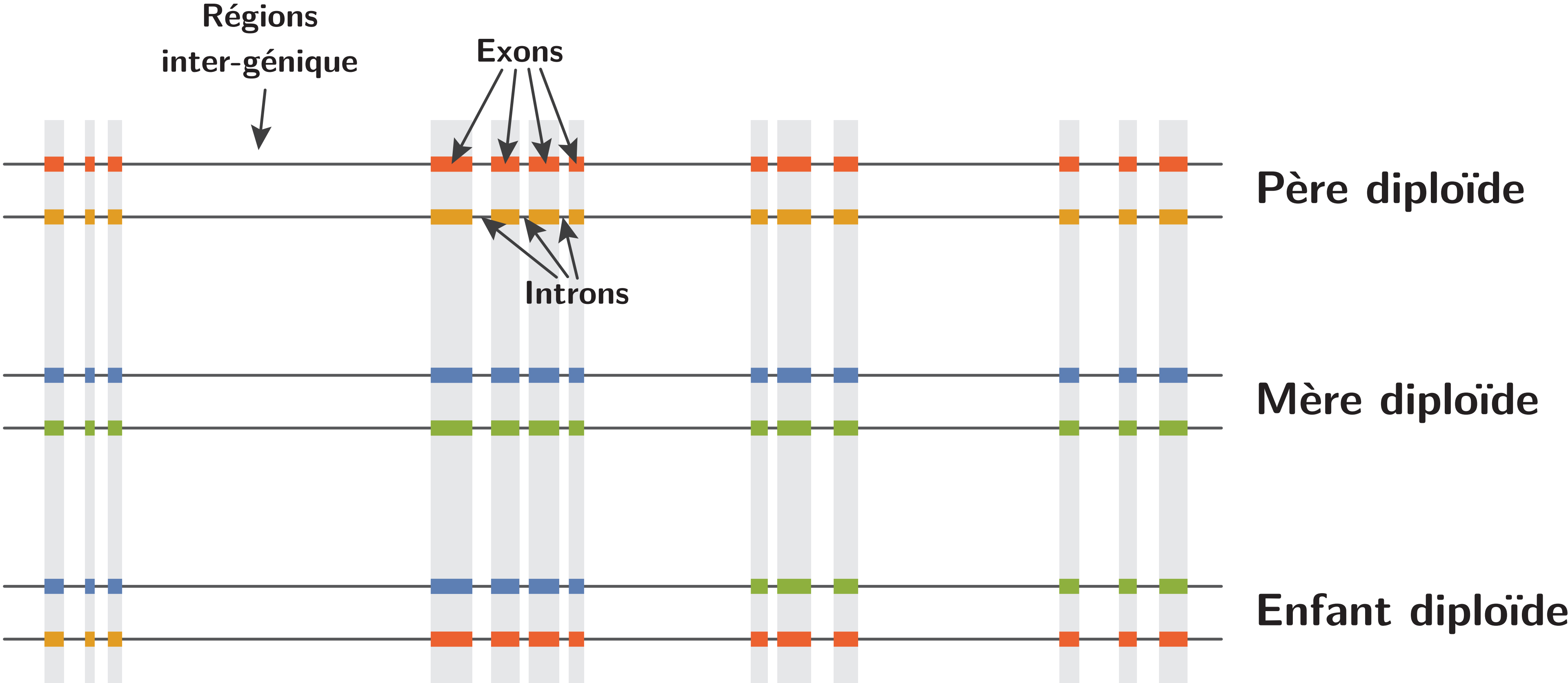


Quels sont les résultats du projet 1000 Génomes ?



The 1000 Genomes Project Consortium, Nature (2015)

Va-t-on utiliser tout le jeu de données ?



Séquençage de l'exome pour un trio père-mère-enfant

Comment est obtenu l'exome ?

