

Chapitre 5. Du génome aux génomes : génomique des populations

Introduction à la bioinformatique (SSV3U15, L2 SV AMU)

Jacques van Helden
Aix-Marseille Université
orcid.org/0000-0002-8799-8584

Plan du cours

- Variations génétiques au sein des populations humaines
- Génotypage des individus
- Retracer les origines d'un individu dans son ADN
- Retracer l'évolution de l'espèce humaine dans l'ADN

Variations génétiques au sein des populations humaines

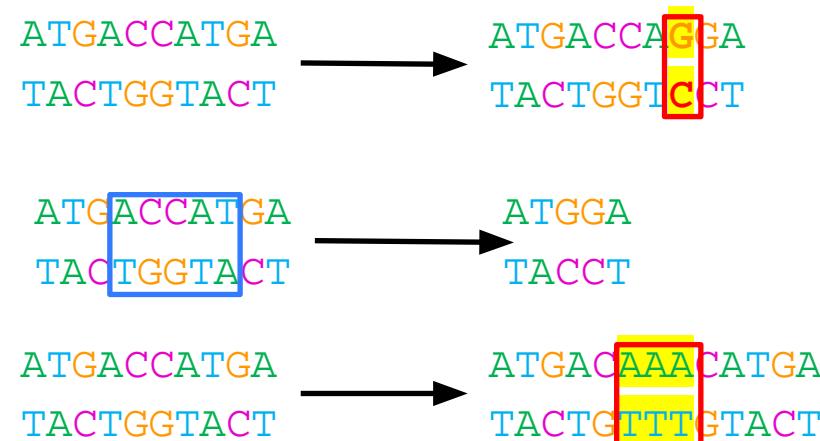
Du génome humain aux génomes humains

- Les projets initiaux de séquençage portaient sur « le » génome d'une espèce, ce qui consistait à séquencer le génome d'un individu particulier, ou d'un petit nombre d'individus.
- Depuis 2008, ces projets ont été étendus pour s'attaquer au séquençage de milliers, et bientôt de millions d'individus;
 - 2008-2015 : projet « **1000 génomes** » (en pratique, 2500 génomes à l'issue du projet)
 - www.internationalgenome.org
 - But: étudier la diversité génétique des populations humaines
 - 2016 : plan France Médecine Génomique 2025 (FMG2025)
 - Séquençage complet du génome de 200.000 individus / an
 - Applications au soin: diagnostic et thérapies individuelles (cancers + maladies rares)
 - Application à la recherche : découverte de gènes impliqués dans la maladie et la santé
 - 2018 : Plan européen “**One plus million genomes**” ([1+MG](#))
 - Regroupement de plusieurs plans nationaux de séquençage de génomes.
 - 2022 : **Genome Data Infrastructure (GDI)**
 - Partage international des données génomiques humaines
- Les buts sont d'une part d'étudier la diversité génétique de l'espèce humaine (génomique des populations), et d'autre part d'identifier des gènes associés à des pathologies.

Rappel: typologie des mutations

● Mutations ponctuelles

- Substitution : remplacement d'un résidu (une lettre) par un autre
- Délétion : suppression d'un fragment de la molécule
- Insertion : ajout d'un fragment de molécule



● Variations structurelles (résultent de réarrangement chromosomiques)

- Délétions
- Insertions
- Translocations
- Variations du nombre de copies

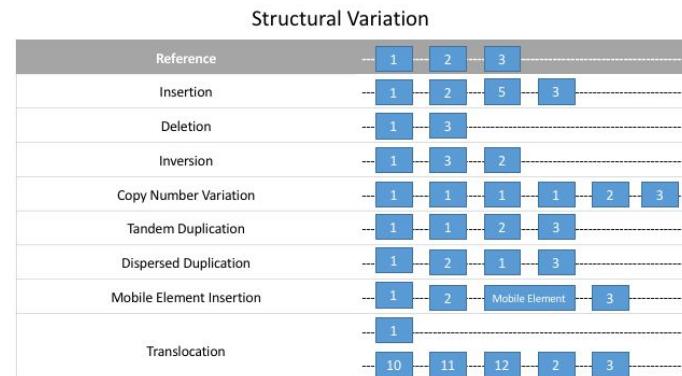


Figure 1: Depiction of different types of structural variants compared to the reference genome. Each different number represents a different gene.

Variabilité du génome humain

- **Polymorphisme génétique / variation génétique** : position génomique où l'on observe des différences (variations) de séquence entre les individus d'une population.
- **Variant** : l'une des modalités d'une variation. Par exemple, un SNP peut présenter un A chez 84% (**variant majoritaire**) des individus, un T chez 16% (**variant minoritaire**)
- **Variations « communes »**
 - Variations dont l'allèle mineur a une fréquence suffisante ($\geq 5\%$) dans au moins une population.
 - Estimations: 15 millions de variants (SNPs et microsatellites) dans l'espèce humaine (Ginsburg and Willard, 2009).
- **Variants rares / individuels**
 - Variants trouvés chez un individu unique ou quelques individus.
 - Des mutations se produisent à chaque division cellulaire, à un taux de ~1 mutation ponctuelle par 50 millions de nucléotides (Boulouri, 2010) → sur un génome diploïde, on s'attend à observer ~120 mutations par division cellulaire.
- **Haplotypes**
 - Les variations ne se transmettent pas de façon indépendante, mais par groupes contigus de variations liées génétiquement: les haplotypes.
 - On estime qu'il existe entre 250.000 et 500.000 haplotypes (Topol, 2007).

Ginsburg and Willard. Genomic and personalized medicine: foundations and applications. Transl Res (2009) vol. 154 (6) pp. 277-87.

Boulouri, H. (2010). Personal Genomics and Personalized Medicine, 1st edn. Imperial College Press.

Topol et al. The genomics gold rush. JAMA (2007) vol. 298 (2) pp. 218-21.

Exemple de fichier de description de variants

Chaque ligne caractérise une variation

Certaines sont connues préalablement à l'étude → on indique leur identifiant.

Les informations incluent une indication de la fréquence de l'allèle alternatif dans l'échantillon analysé.

CHROM	POS	ID	Identifiant (si variation connue)	Allèle de référence	Allèle alternatif	QUAL	FILTER	INFO	Fréq. allèle alternatif
			REF	ALT					
22	16050408	.		T	C	.	PASS	DP=2009;AF=0.030;CB=BI,BC;AFR_R2=0.655	
22	16050612	.		C	G	.	PASS	DP=2636;AF=0.012;CB=UM,BI,BC;EUR_R2=0.691;AFR_R2=0.718	
22	16050678	.		C	T	.	PASS	DP=2950;AF=0.010;CB=BI,BC;EUR_R2=0.501;AFR_R2=0.686	
22	16050822	.		G	A	.	PASS	DP=2779;AF=0.041;CB=UM,BI;EUR_R2=0.209;AFR_R2=0.286	
22	16050933	.		G	A	.	PASS	DP=2792;AF=0.108;CB=UM,BI,BC;EUR_R2=0.45;AFR_R2=0.641	
22	16050967	.		C	G	.	PASS	DP=1045;AF=0.030;CB=BI,BC	
22	16050994	rs7288968		G	C	.	PASS	DP=1232;AF=0.010;CB=BI,BC;AFR_R2=0.618	
22	16051107	rs6518357		C	A	.	PASS	DP=1993;AF=0.020;CB=BI,BC;AFR_R2=0.955	
22	16051209	rs7292503		A	G	.	PASS	DP=2179;AF=0.010;CB=BI,NCBI;AFR_R2=0.79	
22	16051249	rs62224609		T	C	.	PASS	DP=3184;AF=0.063;CB=UM,BI,BC,NCBI;EUR_R2=0.851;AFR_R2=0.593	
22	16051255	rs113903952		C	T	.	PASS	DP=3426;AF=0.010;CB=BI,BC,NCBI;EUR_R2=0.618;AFR_R2=0.923	
22	16051295	rs76439996		C	G	.	PASS	DP=1901;AF=0.010;CB=BI,BC;AFR_R2=0.967	
22	16051347	rs62224610		G	C	.	PASS	DP=2918;AF=0.231;CB=UM,BI,BC;EUR_R2=0.561;AFR_R2=0.667	

Projet 1000 génomes

Projet 2008-2015

- Séquençage génomique + génotypage par biopuces
- En fin de projet, 2500 individus
- Échantillonnage visant à couvrir tous les continents

Nombre total de variations détectées

- 88 millions de SNPs
- 3,6 millions de délétions/insertions courtes
- 60.000 variants structurels

Variations inter-individuelles moyennes

- ~3 millions de différences entre 2 individus pris au hasard
→ **1 différence / 1000 bp**
- ~4 millions de différences entre un individu et le génome moyen (calculé en retenant pour chaque variation l'allèle majoritaire)

Répartition géographique des variations

- La majorité des variants se retrouvent sur tous les continents (gris foncé), ou dans plusieurs (gris clair)
- Certains variants sont spécifiques d'un continent (couleur claire) ou d'une population (couleur foncée)
- Le nombre de variations par individu est beaucoup plus élevé en Afrique que dans les autres continents. Ceci reflète l'histoire des migrations durant la préhistoire

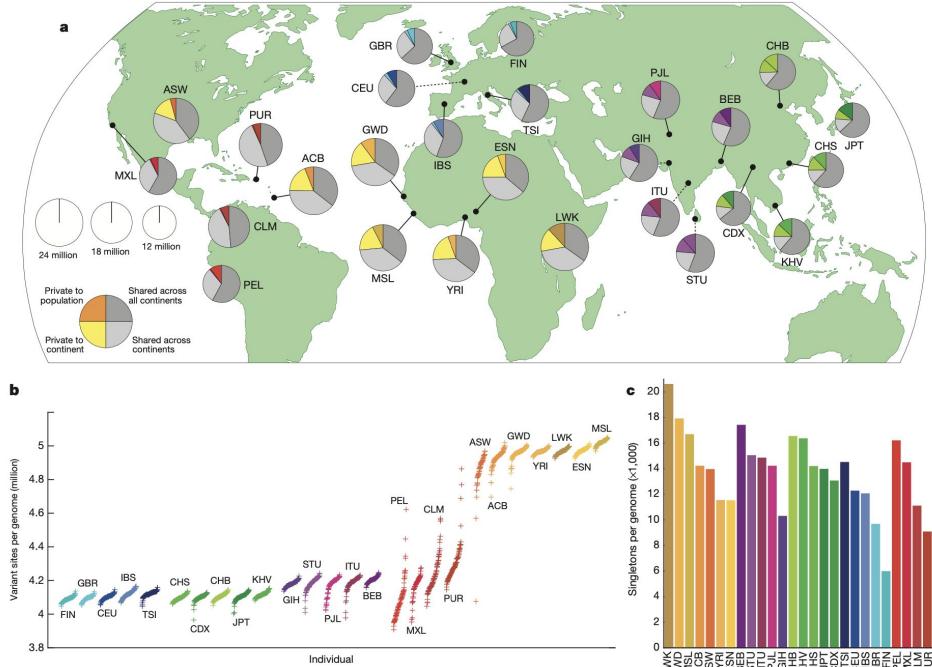


Figure 1 | Population sampling. a, Polymorphic variants within sampled populations. The area of each pie is proportional to the number of polymorphisms within a population. Pies are divided into four slices, representing variants private to a population (darker colour unique to population), private to a continental area (lighter colour shared across continental group), shared

across continental areas (light grey), and shared across all continents (dark grey). Dashed lines indicate populations sampled outside of their ancestral continental region. b, The number of variant sites per genome. c, The average number of singletons per genome.

Grands projets de génomique (volet populationnel + médecine génomique)

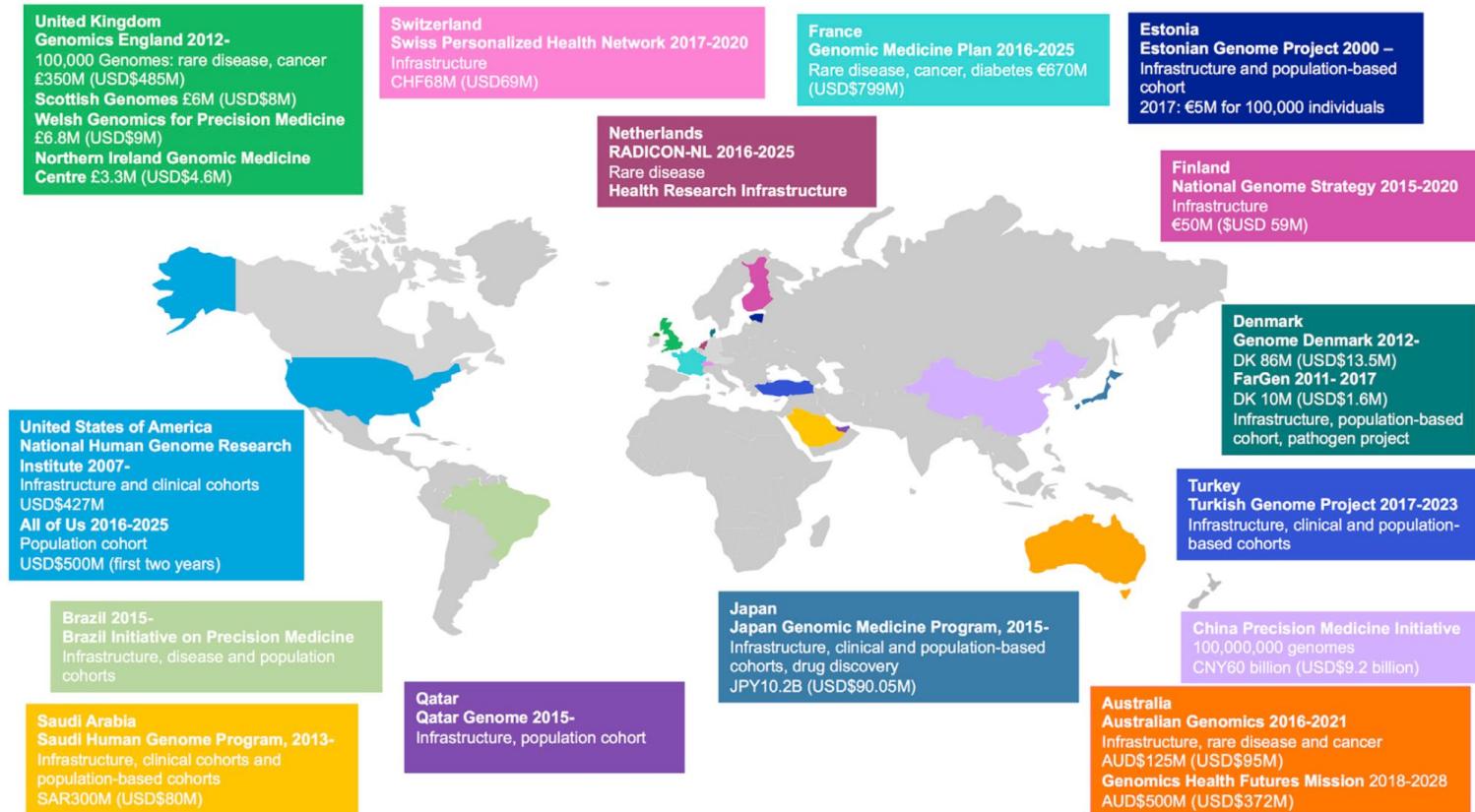


Figure 1. Map of Currently Active Government-Funded National Genomic-Medicine Initiatives

Génotypage des individus

Les marqueurs génotypiques

- Les marqueurs génotypiques sont des petits fragments de génome qui présentent des variations interindividuelles.
 - SNP: nucléotide variable d'un individu à l'autre. Ce type de variation est appelée « **SNP** » pour « **Single-Nucleotide Polymorphism** » (prononcez « *snip* »).

Type 1	CAGTGC <u>G</u> AGGG
Type 2	CAGTGC <u>GG</u> GGGG
 - **Microsatellite**: fragment de séquence génomique comportant une courte séquence (2 à 5 résidus) répétée (typiquement 5 à 50 fois selon les cas). Le nombre de répétitions varie souvent d'un individu à l'autre.

Type 1	CAGCAGCAGCAGCAGCAG
Type 2	CAGCAGCAGCAGCAGCAGCAG
Type 3	CAGCAGCAGCAGCAGCAGCAGCAG

Utilisation des marqueurs génétiques

- Certains marqueurs peuvent être associés à des maladies, ou à des caractères phénotypiques.
- D'autres marqueurs n'ont aucun effet détectable, mais sont utilisés pour identifier des personnes, ou pour analyser les variations génétiques au sein des populations humaines

Exemples de profils génétiques basés sur les microsatellites

- L'identification des personnes ne nécessite pas de connaître l'entièreté de leur génome.
- La police scientifique utilise des profils génétiques basés sur une vingtaine de microsatellites. Ce petit nombre de marqueurs suffit à identifier une personne sans équivoque possible.
- Les marqueurs sont choisis pour avoir de bonnes qualités pour l'identification.
 - Ils permettent par exemple d'identifier une personne sur base de traces biologiques laissées sur les lieux d'un délit, à condition de disposer du profil de cette même personne (ou d'un parent proche) dans la base de données.
- Ces marqueurs n'ont pas de valeur prédictive
 - Pour des raisons éthiques, ces microsatellites ont été choisis pour ne pas être porteurs d'information phénotypiques, afin d'éviter d'exposer des données à caractère personnel (par exemple les risques pour un individu de développer une maladie particulière).

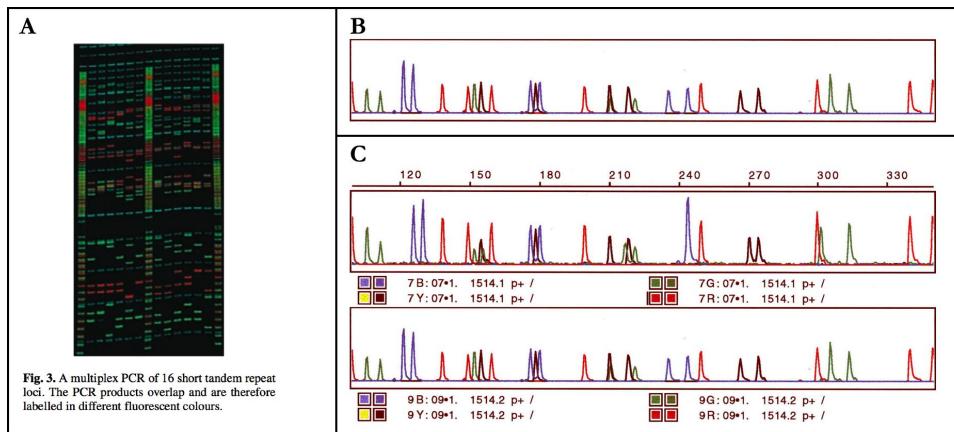


Figure . Exemples de profil génétiques basés sur les microsatellites (source : Pierre Van Renteghem). **A:** multiplex montrant les profils génétiques d'une série d'individus. **B:** profil génétique obtenu à partir d'un échantillon trouvé sur les lieux d'un crime. **C:** profils génétiques de deux suspects.

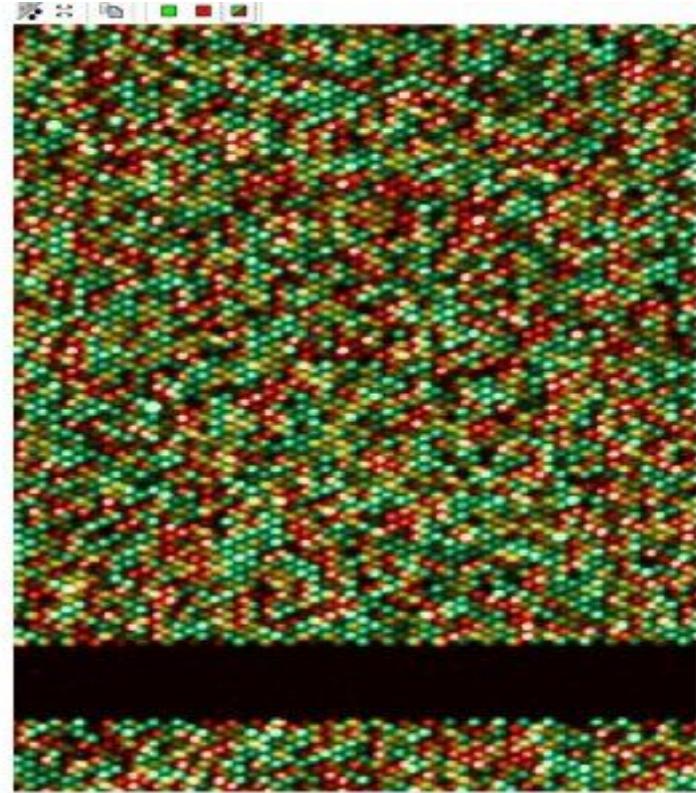
Biopuces de génotypage

- Depuis le début des années 2000, on a développé de nouvelles méthodes de génotypage à haut débit, basées sur la technologie des biopuces.
- Chaque point lumineux correspond à un SNP
- La couleur indique le variant particulier (A, C, G, T) présent chez l'individu considéré.
- Sur une seule biopuce ("microarray"), on peut caractériser entre 500.000 à 6.000.000 SNP différents.

Biopuce



Détail



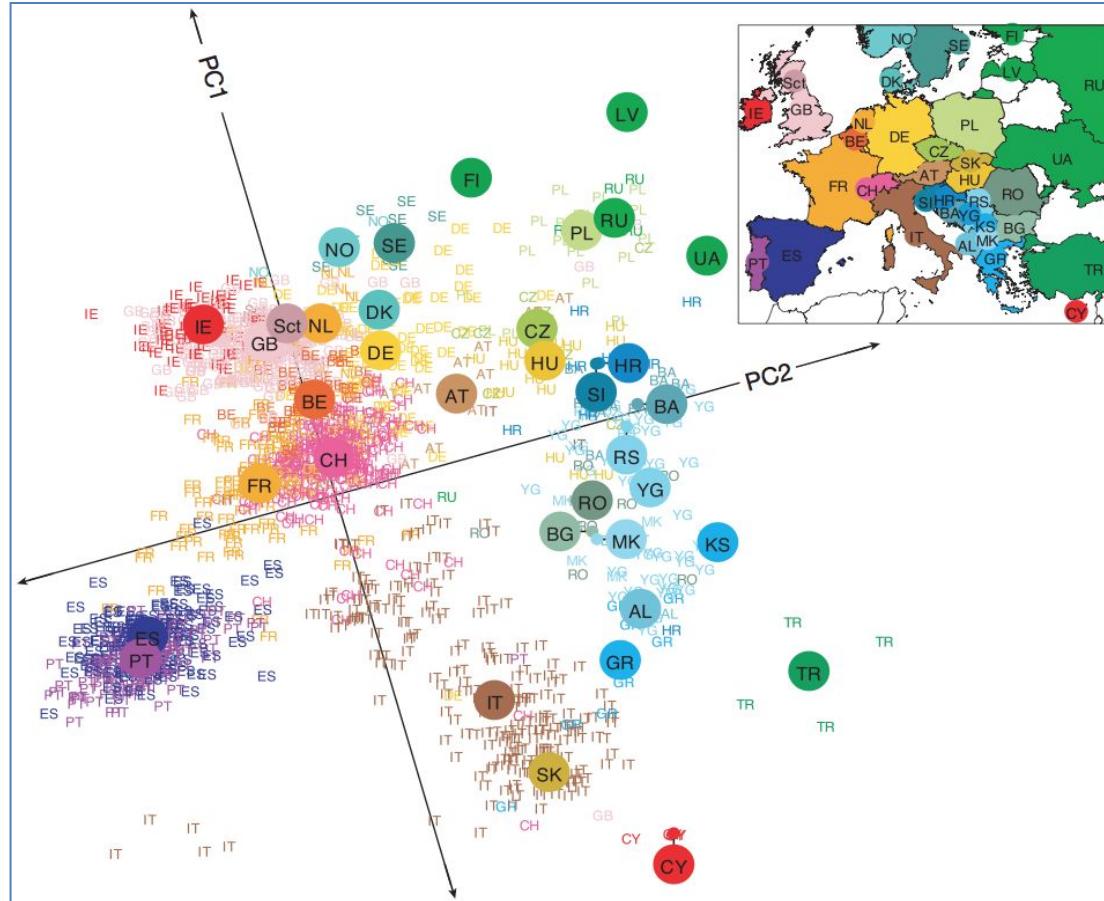
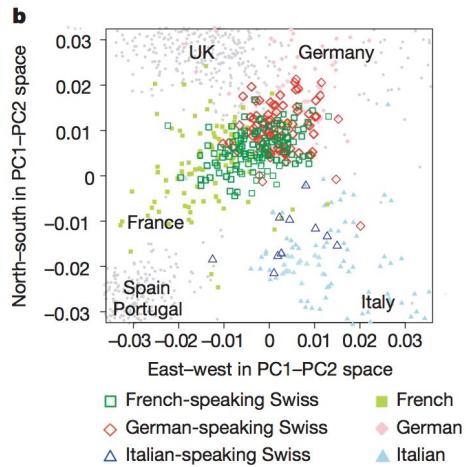
Le génotypage est passé à échelle génomique avec les biopuces à SNP

- Le génotypage consiste à déterminer les particularités génétiques d'un individu à partir de marqueurs prédéfinis.
- Les méthodes traditionnelles reposaient sur un petit nombre de marqueurs génétiques (une trentaine de microsatellites).
- Méthodes de génotypage
- Biopuces
 - Permettent de caractériser 500.000 à 6 millions de SNPs par individu
 - ~80€ par génotypage (novembre 2022)
- Séquençage génomique
 - Ciblé : 1 gène ou un panel de gènes associés à une maladie
 - Exome : l'ensemble des séquences codantes (codant pour des protéines)
 - Coût: ~200€ par génome complet (novembre 2022)

Origines géographiques et ethniques

Quand les gènes révèlent nos origines géographiques

- On peut déterminer l'origine géographique d'un individu européen avec une précision de 400 km, sur simple base de son profil génomique.
 - Au sein de la population suisse on peut distinguer les sous-populations germanophone, francophone et italophone.



Le profilage génomique permet de discriminer les populations juives des non-juives

Une analyse des profils génomiques permet de distinguer, dans un échantillon américain, les personnes juives des non-juives (figure du haut), voire de préciser le nombre de juifs parmi leurs grands-parents (figure du bas).

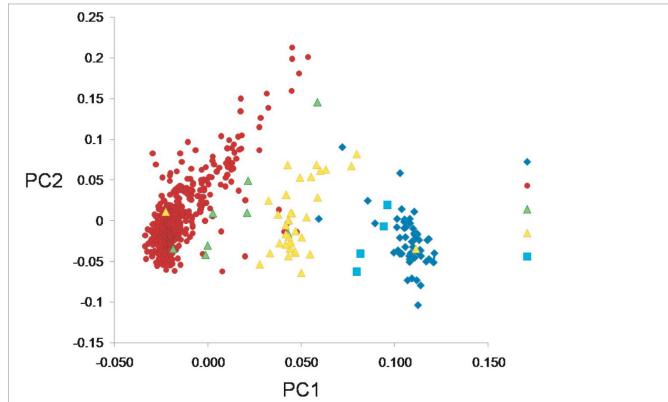
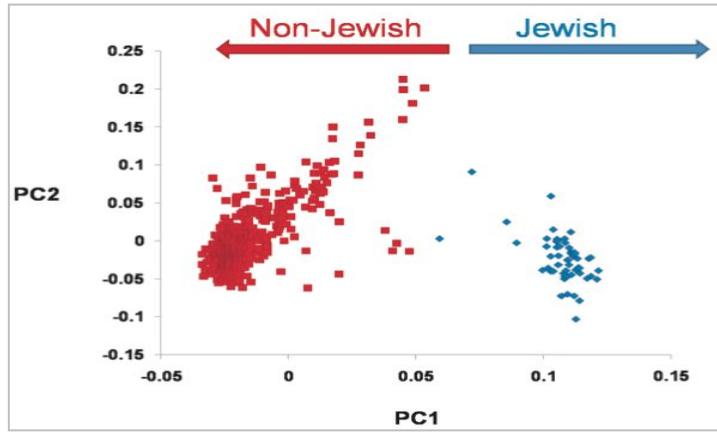


Figure 2
PC1 versus PC2 for people with or without Jewish ancestry. The score on PC1 plotted against the score on PC2 for people with four, three, two, one and no Jewish grandparents.

Need et al. A genome-wide genetic signature of Jewish ancestry perfectly separates individuals with and without full Jewish ancestry in a large random sample of European Americans. *Genome Biol* (2009) vol. 10 (1) pp. R7.

<https://doi.org/10.1186/gb-2009-10-1-r7>

Stratification de populations africaines dans une étude sur la susceptibilité à la malaria

- Les profils génomiques obtenus par biopuces à SNP révèlent la structuration génétique de populations humaines.
- Les graphiques ci-contre montrent une projection sur 2 dimensions (sur base d'une méthode appelée analyse des composantes principales), où chaque point correspond à un individu, et les distances entre points sont (approximativement) proportionnelles au nombre de différences génétiques.
- On colore ensuite les individus selon leur origine géographique (cette information n'est pas utilisée pour le calcul des distances).
- On observe que cette analyse relativement simple permet de distinguer la nationalité et dans une certaine mesure la tribu d'origine au sein des populations africaines.

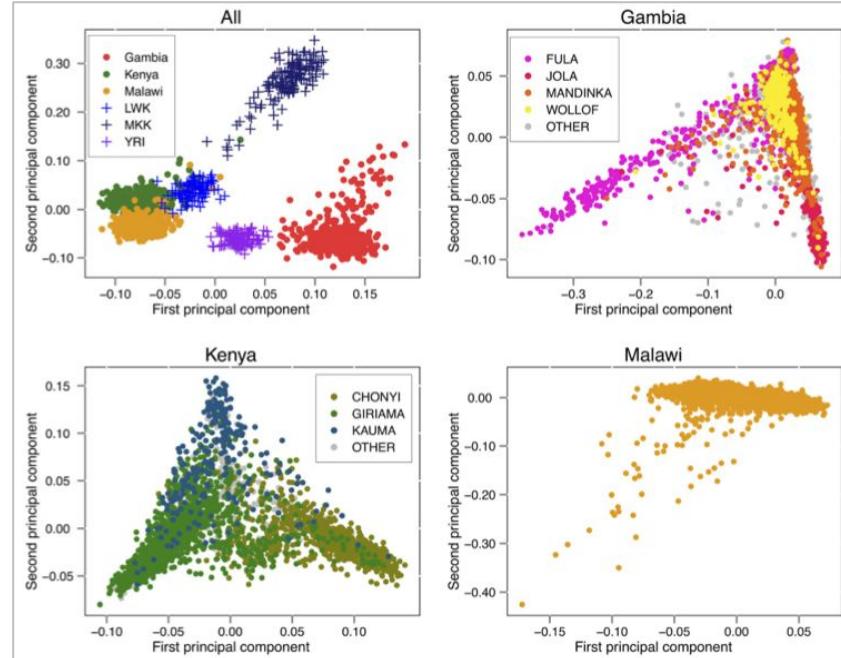


Figure 2. Principle components analysis. Top left: principal components analysis (PCA) of the African populations from Hapmap 3 (LWK = Luhya in Webuye, Kenya, 90 individuals; MKK = Maasai in Kinyawa, Kenya, 143 individuals; YRI = Yoruba in Ibadan, Nigeria, 113 individuals) with 500 randomly selected control samples from each of the three study cohorts. Top right, bottom left, bottom right: PCA of all non-excluded samples in each study cohort, coloured by reported ethnic group. Ethnic group is shown as "OTHER" for groups constituting less than 5% of individuals in the cohort, or where the ethnic group was unreported.
doi:10.1371/journal.pgen.1003509.g002

Band, G., Le, Q.S., Jostins, L., Pirinen, M., Kivinen, K., Jallow, M., Sisay-Joof, F., Bojang, K., Pinder, M., Sirugo, G., et al. (2013) Imputation-Based Meta-Analysis of Severe Malaria in Three African Populations. PLoS Genet, 9, e1003509.

<https://doi.org/10.1371/journal.pgen.1003509>

Retracer les origines d'un individu dans son ADN

Génomique "récréative" : recherche des origines géographiques

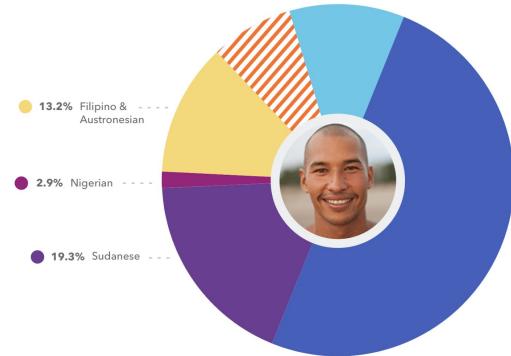
The screenshot shows the 23andMe website homepage. At the top, there's a navigation bar with 'SIGN IN', 'REGISTER KIT', 'HELP', and a 'Shop' button. Below the navigation is a search bar with 'Email address' and a red '→' button. A banner at the top left says 'Our lowest price of the year: Enter your email to unlock £49 Ancestry Service.' On the left, a 'Black Friday Starts Now' section highlights the 'Ancestry Service' for £49, with a 'Buy now' button and a note about the offer ending on October 15th. In the center, there's a large image of a DNA test kit labeled 'Welcome to you' with a barcode and a colorful DNA helix graphic. Below this are three service options: 'LOWEST PRICE OF THE YEAR' (Ancestry Service for £49), 'ESSENTIAL' (Health + Ancestry Service for £159), and 'ENHANCED' (23andMe+ Premium for £159). Each option includes a brief description and a 'Learn more' button.

- La compagnie 23andme (liée à Google) propose depuis le début des années 2000 différents services liés au génotypage.
 - origines géographiques
 - santé (facteurs génétiques de risque)
 - recherche de parents

Génomique "récréative" : recherche des origines géographiques

Where in the world is your DNA from?

Your DNA can tell you where your ancestors lived more than 500 years ago. Explore your ancestry's *breakdown by region*, including East Asia, Sub-Saharan Africa and Europe, with results becoming more refined as our database continues to grow.



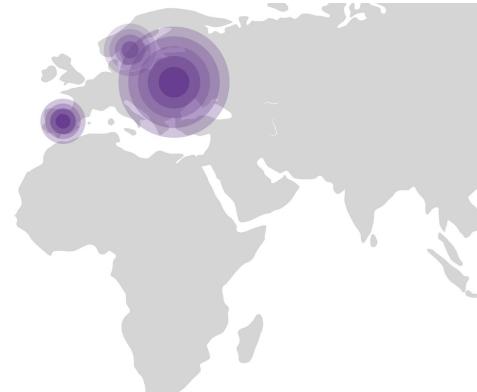
 What if you could travel there?
[Learn more](#) about how you can embark on a DNA travel adventure based on your 23andMe Ancestry Composition results.

Map your ancestry. Go further back.

Discover the origins of your maternal (your mother's mother's mother...) and paternal (your father's father's father...) ancestors and how they moved around the world over thousands of years.

We report on your maternal and paternal lineage by identifying your haplogroups. A haplogroup can trace part of your ancestry back to a specific group of individuals in the distant past.

Women can only trace their maternal haplogroup. This is because the paternal haplogroup is traced through the Y chromosome, which women do not inherit.



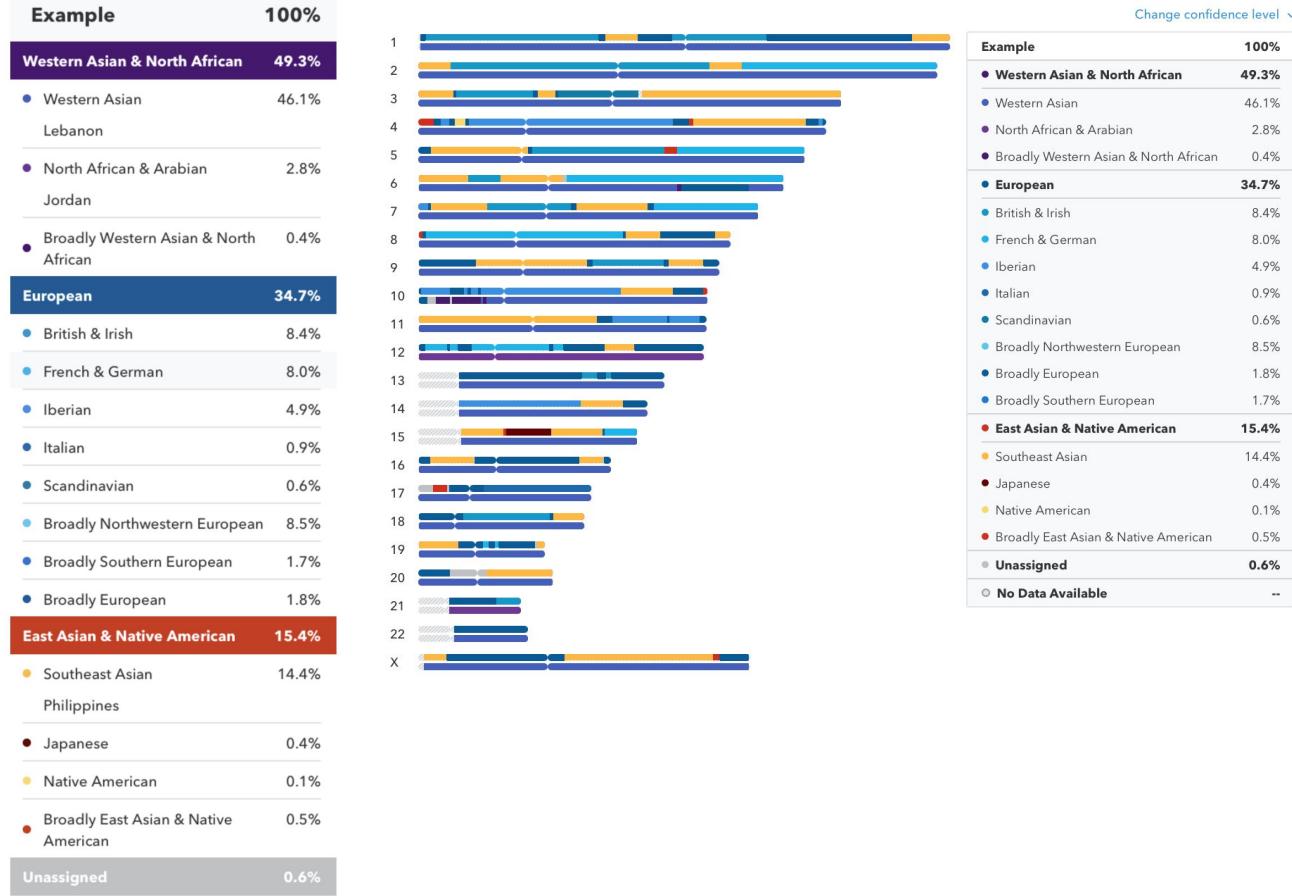
- La compagnie 23andme (liée à Google) propose depuis le début des années 2000 différents services liés au génotypage.
- L'affichage de 23andme met un fort accent sur la recherche des origines géographiques.
- Cette question suscite un intérêt assez marqué aux Etats-Unis, vraisemblablement du fait de l'histoire de ce pays, qui repose sur plusieurs vagues de migrations (colonisation, commerce d'esclaves, ...).

Exemple de résultat de recherche des origines

Liste des pourcentages par région

Coloration chromosomique
(chromosome painting).

Noter le mélange de morceaux
chromosomiques d'origines
différentes.



Retracer l'évolution de l'espèce humaine à partir de séquences d'ADN

Evolution des hominins

- **Hominini** = humain + chimpanzé + bonobo
- **Hominidae** = Hominini + gorille

Divergences

- Hominini vs gorille : 9 Ma
- Humain vs chimpanzé: 6Ma

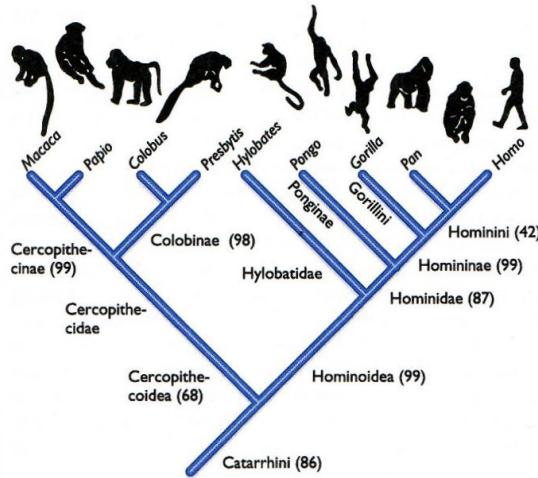
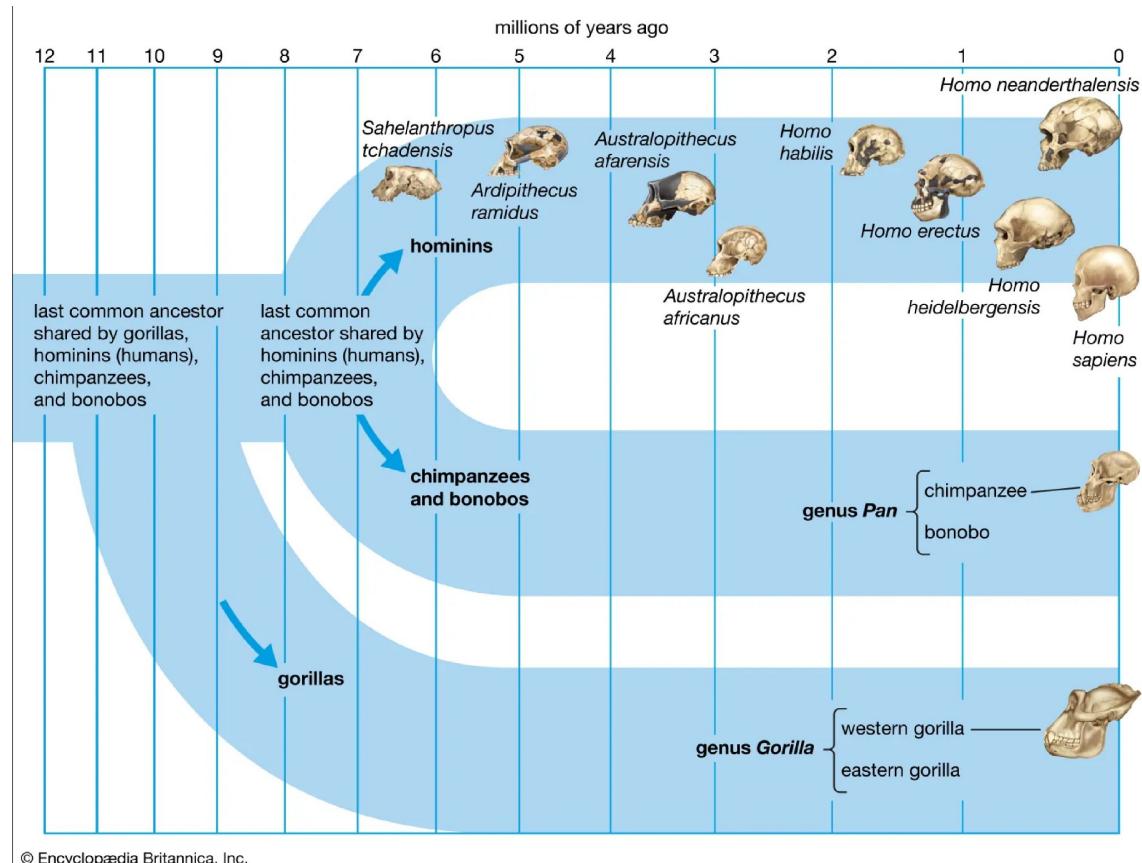


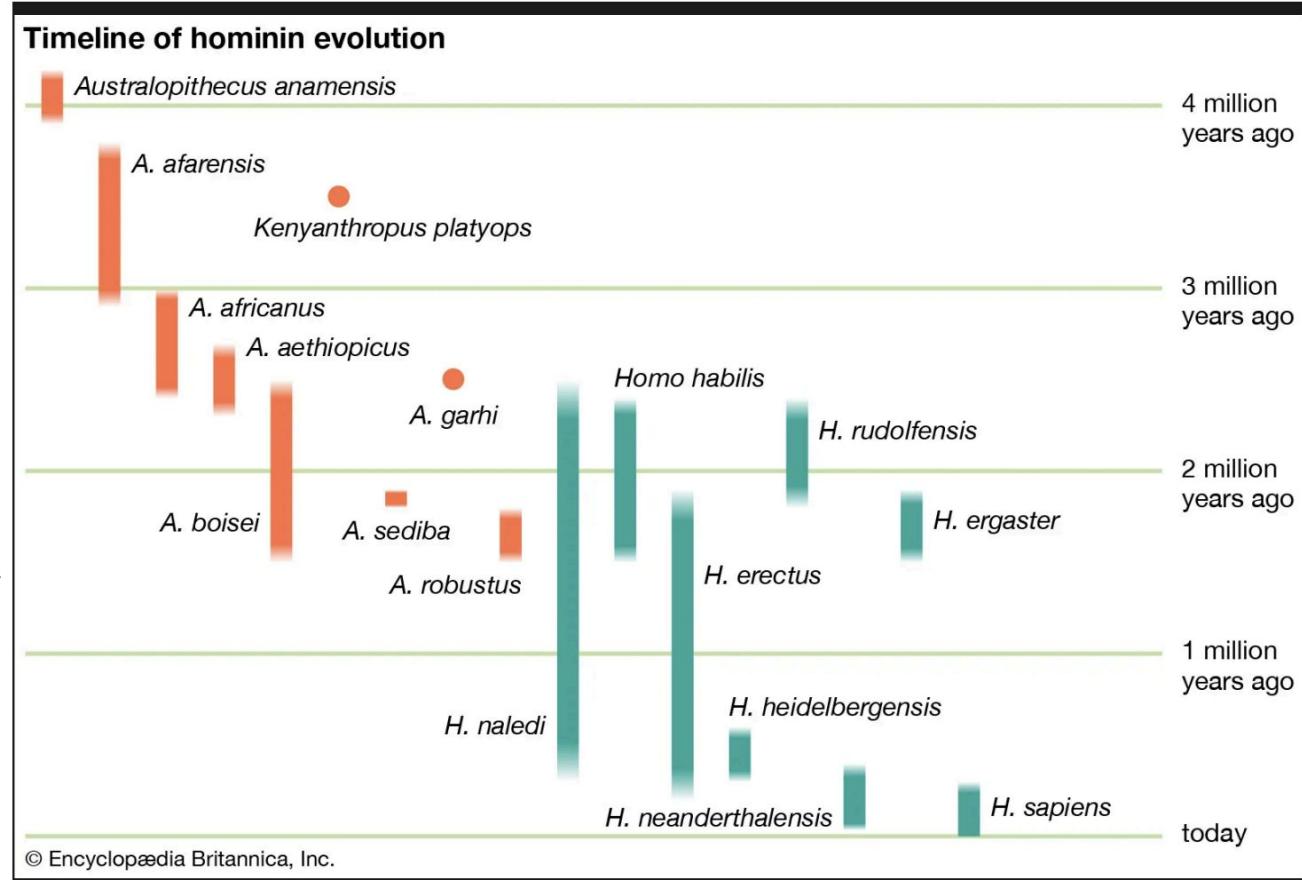
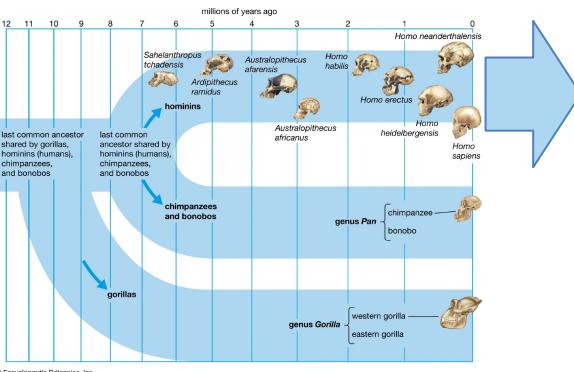
FIGURE 15.4 Cladogram of catarrhine relations: This analysis of 264 morphological characters leads to a chimpanzee/human association as the most parsimonious tree; a tree with a hominoid trichotomy is less parsimonious. This study is one of very few morphological analyses that identifies chimpanzees and humans as one other's closest relatives. (Adapted from Shoshani *et al.*)



- Lewin, R. Human Evolution: An Illustrated Introduction. (Blackwell Publishing Ltd., Hoboken, 2009).
- <https://www.britannica.com/topic/hominin>
- <https://youtu.be/QS8bukoLJTw?t=793>

Evolution de la lignée humaine

- Co-existence de plusieurs genres (Australopithecus et Homo)
- Co-existence d'*Homo sapiens* et *Homo neanderthalensis* pendant une longue période



Migrations des hominidés

- La répartition géographique et la datation des fossiles indiquent plusieurs vagues de migration des différentes espèces d'hominidés depuis l'Afrique vers l'Europe et l'Asie. On regroupe ces migrations sous le nom de **première sortie d'Afrique** ("Out of Africa 1", flèches plus claires pour *Homo erectus* sur la figure du haut). On pense que les sorties d'Afrique se sont produites régulièrement ultérieurement.
- ~200.000 ans : premières traces fossiles de l'espèce *Homo sapiens*
- ~80.000 ans : **deuxième sortie d'Afrique** ("Out of Africa 2", flèches bleu foncé sur la carte du haut), migration d'*Homo sapiens* vers l'Europe, puis l'Asie, puis l'Océanie.
- Le passage au continent américain (via le détroit de Behring) est ultérieur (15 à 35 Ma).
- L'Homme de Neandertal (aire de répartition en ocre sur la figure du bas) s'est éteint il y a ~30 Ma.
- Homo sapiens* et *Homo neanderthalensis* ont donc co-existé en Eurasie pendant plusieurs dizaines de milliers d'années

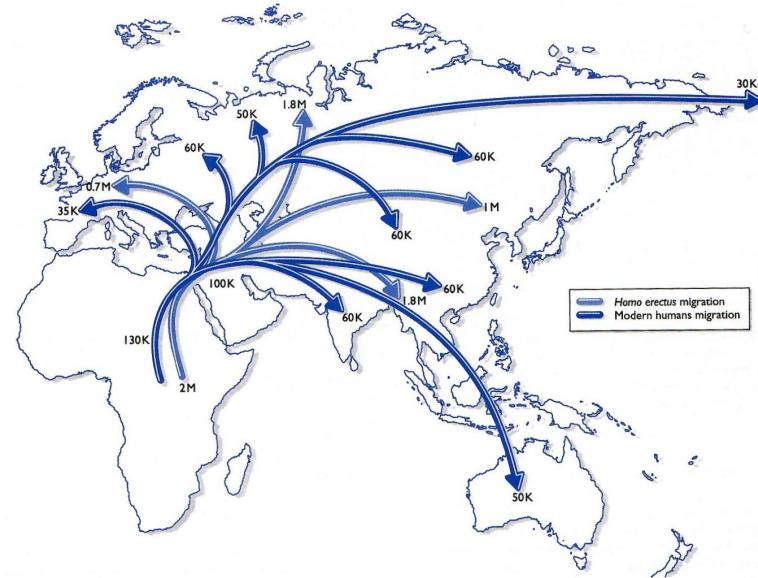
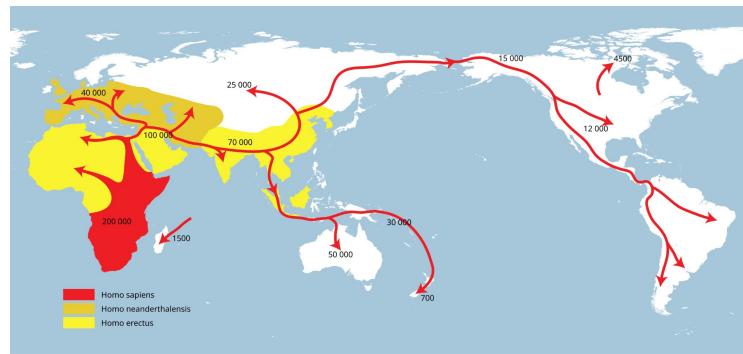


FIGURE 28.2 Two migrations: If the single, recent-origin model is correct, then the original expansion of *Homo erectus* from Africa into the rest of the Old World would have been followed much later by a similar expansion of modern people. This presentation is

certain to be an oversimplification because it implies two discrete events. In fact, multiple population movements must have occurred at different times and in different places. (K, thousand years ago; M, million years ago).



Groucutt, H. S. et al. Rethinking the dispersal of *Homo sapiens* out of Africa. *Evolutionary Anthropology* 24, 149–164 (2015).

Green, R. E. et al. A Draft Sequence of the Neandertal Genome. *Science* 328, 710–722 (2010). <https://doi.org/10.1126/science.1188021>

Richard Edward Green, Neanderthal Genome Project: Insights into Human Evolution

Introgression de gènes néandertaliens dans le génome de l'homme moderne

- En 2010, Svante Pääbo et son équipe séquentent l'ADN de plusieurs espèces humaines fossiles, et comparent chaque région génomique à 5 génomes de référence représentatifs de différentes populations humaines actuelles.
- Cette étude révèle la présence de gènes d'origine Neandertal dans les génomes d'*Homo sapiens*, qui couvrent 1 à 4% du génome des génomes d'Européens ou d'Asiatiques actuels.
- On n'en trouve pas dans les génomes d'Africains.
- On en déduit qu'ils résultent de croisements entre Néandertaliens et Hommes modernes (flèche rouge sur la figure du bas), qui ont eu lieu peu après la deuxième sortie d'Afrique et ont précédé l'expansion eurasienne d'*Homo sapiens*.
- Ces croisements ont provoqué une **introgression** (transfert de gènes entre espèces proches) de Neandertal à *Homo sapiens*.

Echantillons ayant servi aux premiers séquençages génomiques de Neandertal

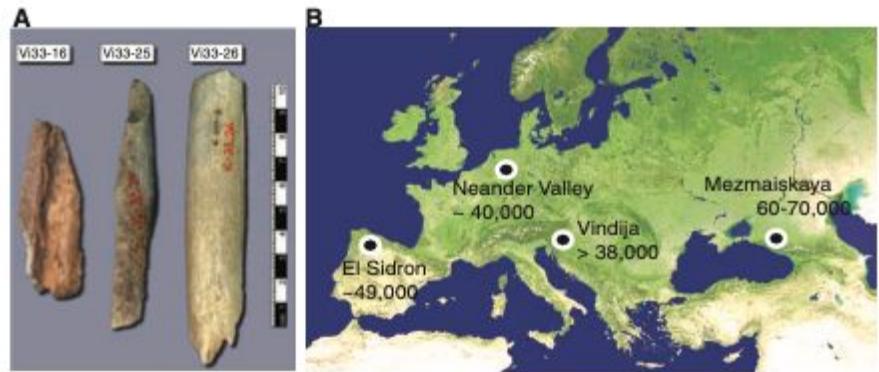
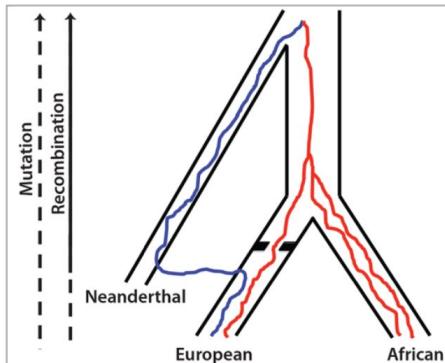


Fig. 1. Samples and sites from which DNA was retrieved. (A) The three bones from Vindija from which Neandertal DNA was sequenced. (B) Map showing the four archaeological sites from which bones were used and their approximate dates (years B.P.).

Introgression de gènes de Neandertal dans le génome d'*Homo sapiens*



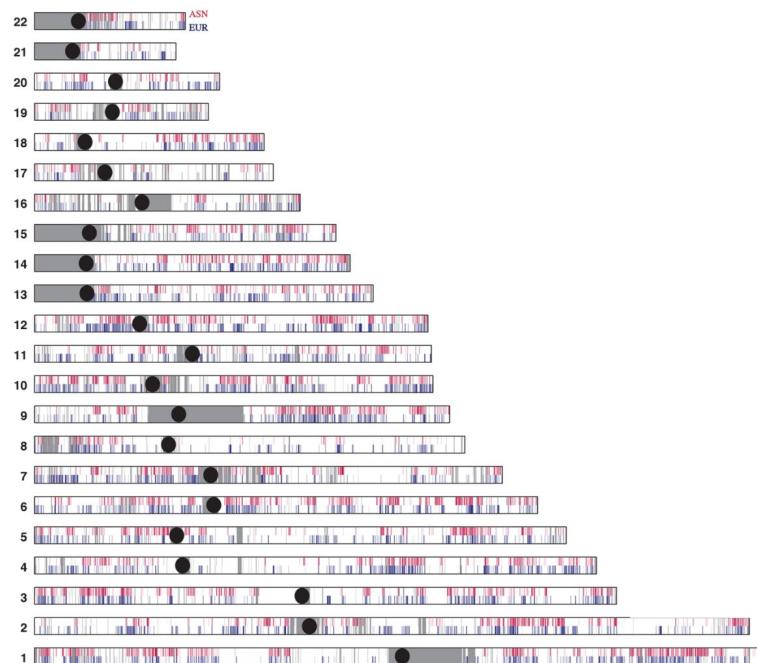
Green, R. E. et al. A Draft Sequence of the Neandertal Genome. Science 328, 710–722 (2010). <https://doi.org/10.1126/science.1188021>

Un excellent exposé par le premier auteur de cette étude : Richard Edward Green. Neanderthal Genome Project: Insights into Human Evolution (2012).

<https://youtu.be/QS8bukoLJTw>

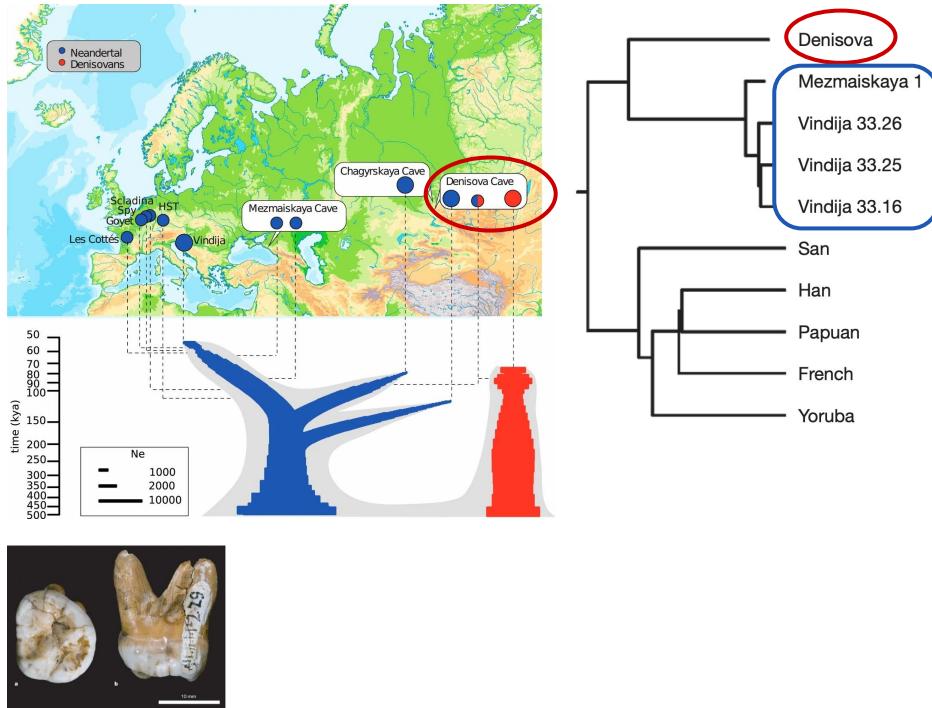
Introgression de gènes néandertaliens dans le génome de l'homme moderne

- Chaque ligne représente un chromosome humain
- Barres verticales : régions génomiques de Neandertal trouvées dans des humains modernes **asiatiques (rouge)** ou **européens (bleu)**
- Certaines régions génomiques se trouvent dans les deux groupes, d'autres spécifiquement dans l'un ou l'autre
- Zones grisées : séquençage de trop faible qualité



Introgression de gènes de l'homme de Denisova dans le génome d'*Homo sapiens*

- L'équipe de Svante Pääbo a également séquencé l'ADN d'échantillons fossiles de la grotte de Denisova, en Sibérie.
- Ceci a révélé
 - une introgression *Homo denisovensis* → *Homo sapiens*
 - un croisement entre *Homo denisovensis* et *Homo neanderthalis*



- Reich, D. et al. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 468, 1053–1060 (2010). doi.org/10.1038/nature09710
- Mafessoni, F. et al. A high-coverage Neandertal genome from Chagyrskaya Cave. *Proc. Natl. Acad. Sci. U.S.A.* 117, 15132–15136 (2020). doi.org/10.1073/pnas.2004944117
- Svante Pääbo. Archaic Genomics. https://youtu.be/726Wwh_P9U4?t=399

Introgression de gènes de l'homme de Denisova dans le génome d'*Homo sapiens*

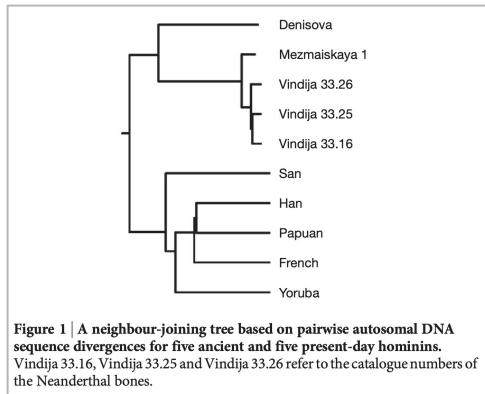
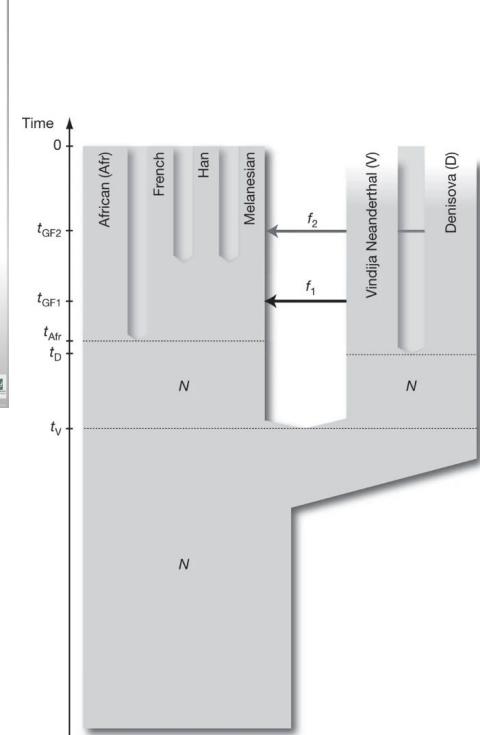
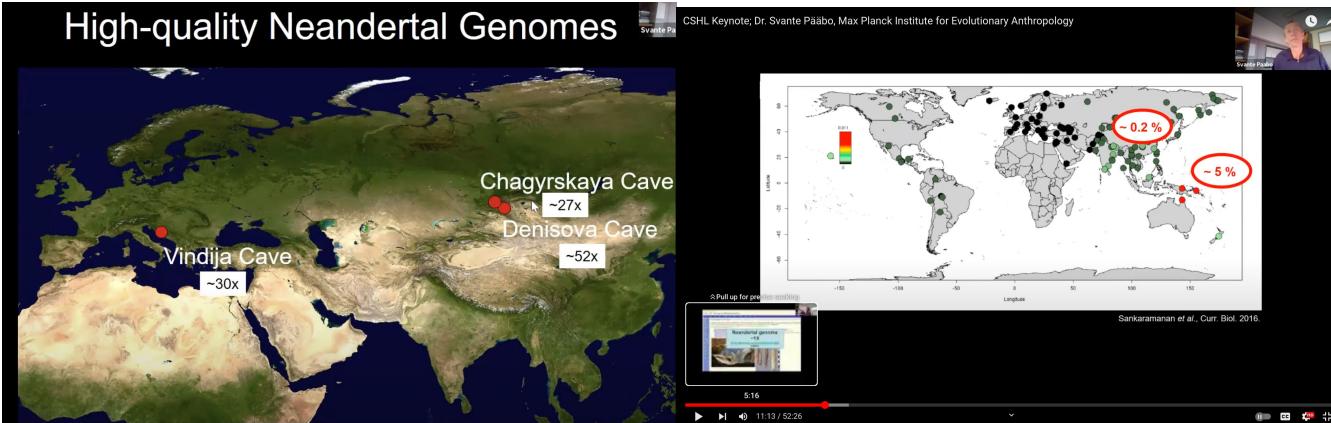


Figure 1 | A neighbour-joining tree based on pairwise autosomal DNA sequence divergences for five ancient and five present-day hominins. Vindija 33.16, Vindija 33.25 and Vindija 33.26 refer to the catalogue numbers of the Neanderthal bones.



High-quality Neandertal Genomes



Reich, D. et al. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 468, 1053–1060 (2010). <https://doi.org/10.1038/nature09710>

Svante Pääbo, Archaic Genomes CSHL Keynote (2020). https://youtu.be/726Wwh_P9U4

Richard Edward Green. Neanderthal Genome Project: Insights into Human Evolution (2012). <https://youtu.be/QS8bukoLJTw>

Fille de Neandertal et de Denisova

Cette carte génétique montre le résultat du séquençage d'un spécimen trouvé dans la grotte de Denisova.

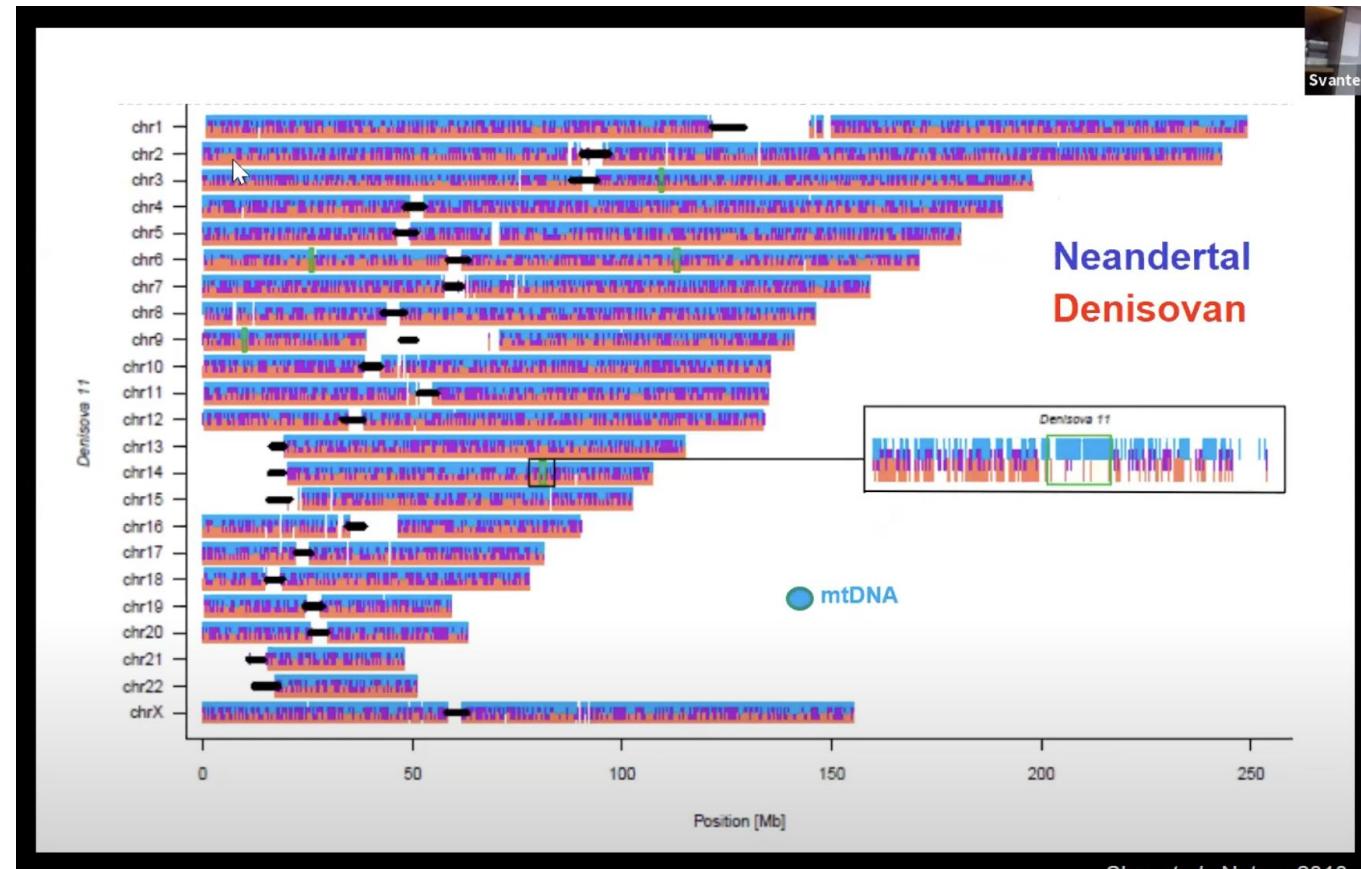
La coloration indique les morceaux de chromosome qui sont génétiquement plus proches du génome de Denisova (rouge) ou Neandertal (bleu).

Deux chromosomes X → il s'agissait d'une femme

On constate que chaque paire de chromosome comporte un exemplaire Neandertal et un Denisova.

Le chromosome mitochondrial (mtDNA) est de type Neandertal.

On conclut que cette femme avait une mère Neandertal et un père Denisovien.



Histoire des migrations humaines d'après les indices génomiques (biopuces à SNP)

Flèches bleues épaisses: expansion initiale

Flèches bleues continues fines: migrations préhistoriques et brassage génétique (croisements entre individus des différents groupes).

Flèches bleues pointillées : migrations et brassages génétiques historiques.

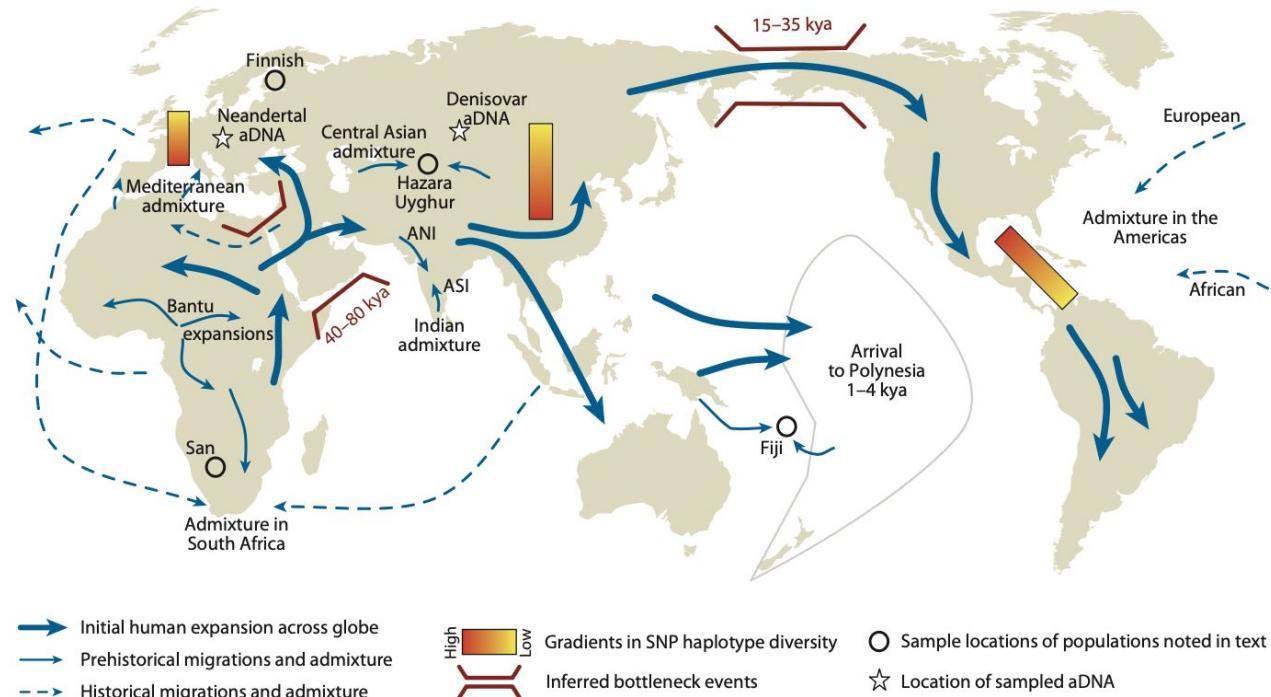


Figure 7

A schematic of human demographic history, highlighting hypotheses investigated by recent single-nucleotide polymorphism (SNP) studies discussed in this review. Numbers indicate the estimated number of years before present at which migrations took place across continents. The distinction between prehistorical and historical is in some cases approximate—the timing and duration of these events is an ongoing area of study. Abbreviations: aDNA, ancestral DNA; ANI, Ancestral North Indian; ASI, Ancestral South Indian.

Conclusions du chapitre

Génomique des populations

- Repose sur le génotypage d'un grand nombre d'individus, soit par profils génétiques basés sur des biopuces, soit par séquençage génomique. Les profils génomiques permettent de **retracer l'origine d'un individu**, sur base de la similarité de son génome avec celui d'autres individus issus de la même région. Cette analyse peut être menée de façon plus précise au niveau de chaque région chromosomique. On observe généralement des **origines diverses chez un même individu**, qui reflètent l'histoire des migrations de ses ancêtres.

Paléogénomique

- On a également séquencé les génomes d'espèces éteintes (Homme de Neanderthal, de Denisova) à partir d'os fossiles. Ces études ont démontré deux introgressions vers Homo sapiens.
 - Neandertal : 1 à 4% du génome d'Homo sapiens d'Européens et d'Asiatiques
 - Denisova : 5% dans les génomes de Mélanésiens, et faible présence dans les génomes d'Asiatiques
 - Ces introgressions ont eu lieu après la deuxième sortie d'Afrique, et on ne trouve donc pas ces gènes dans les génomes d'Africains.
- La génomique des populations et la paléogénomique permettent d'établir de nouveaux scénarios concernant les migrations historiques et préhistoriques des populations humaines, ou de conforter certaines hypothèses qui avaient été formulées sur base de données paléontologiques.

Compléments d'information (pas vus au cours, ne font pas partie de la matière d'examen)

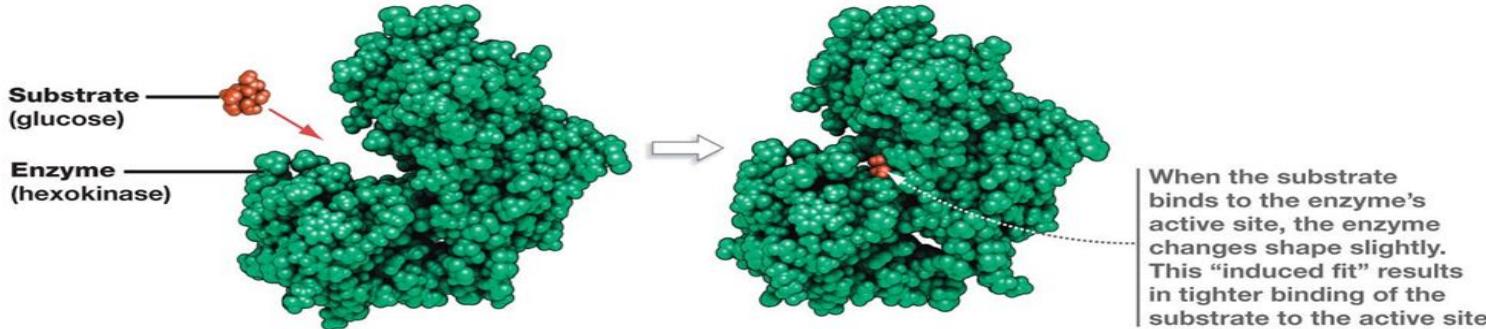
Une classe de protéine: les enzymes

La fonction des protéines résulte de leur structure tridimensionnelle, des propriétés biochimiques de leurs résidus (acides aminés), et de leur capacité à interagir avec d'autres molécules.

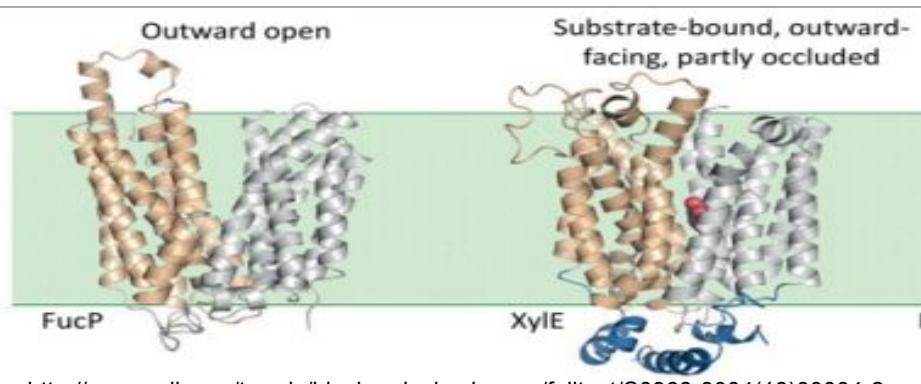
Exemple: les enzymes

- protéines qui peuvent catalyser une réaction spécifique, en interagissant avec des petites molécules (substrats) et en facilitant les échanges atomiques qui produisent d'autres molécules (produits).
- L'interaction avec les substrats repose sur un site spécifique (site catalytique) formé par la conformation tridimensionnelle de la protéine.
- Pendant la réaction, la conformation de la protéine change.

Une enzyme (hexokinase) liée à son substrat (glucose)



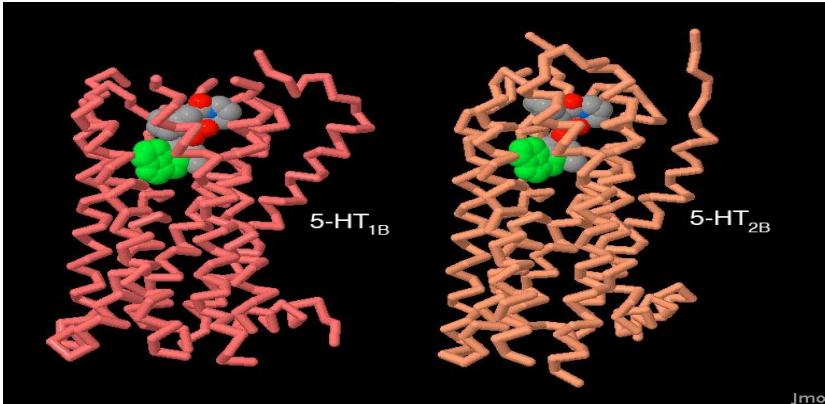
Transporteur



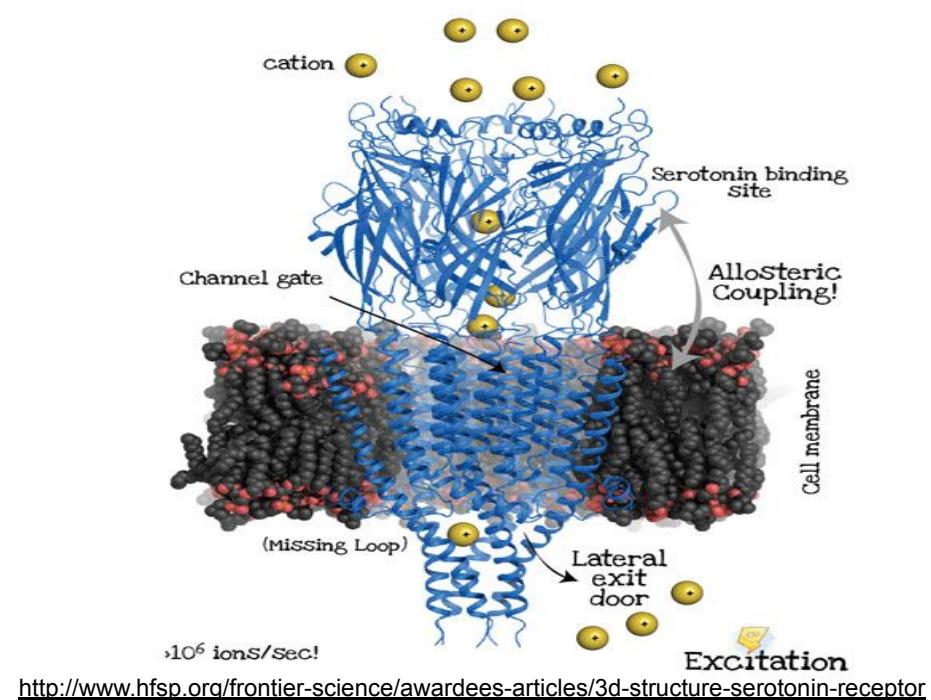
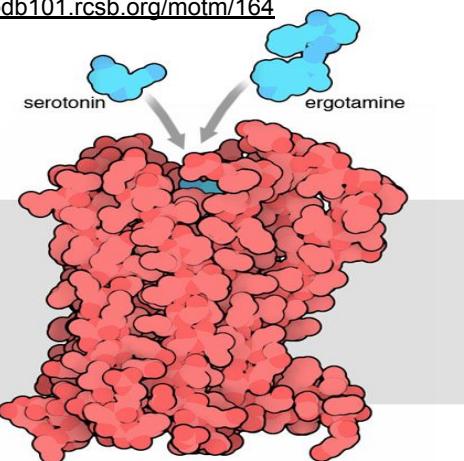
- Les transporteurs sont des protéines qui assurent le transport de petites molécules à travers les membranes cellulaires.
- Ils sont généralement composés de plusieurs hélices hydrophobes qui traversent la membrane.
- Ces hélices transmembranaires forment un canal qui interagit de façon spécifique avec certaines molécules, et les laisse passer d'un côté à l'autre de la membrane.

[http://www.cell.com/trends/biochemical-sciences/fulltext/S0968-0004\(13\)00004-2](http://www.cell.com/trends/biochemical-sciences/fulltext/S0968-0004(13)00004-2)

Récepteur de sérotonine



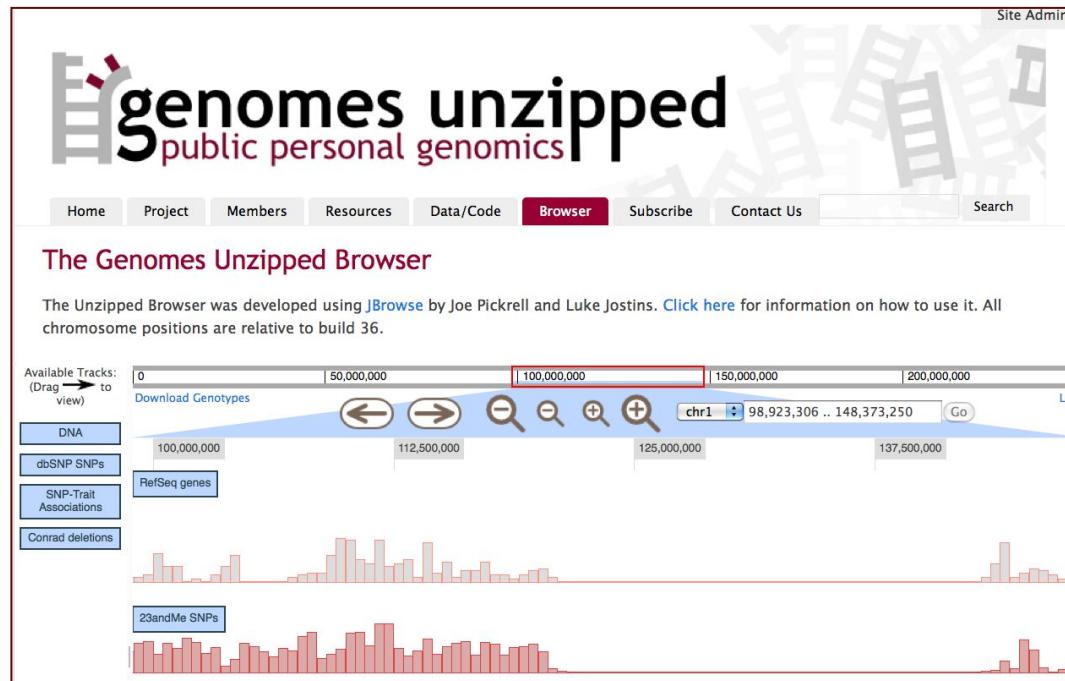
<http://pdb101.rcsb.org/motm/164>



Giving open access to your genome ?

- <http://www.genomesunzipped.org/>

- The goal of the Genomes Unzipped project is to provide genetic testing consumers with independent and informed analysis of developments in the field of genetics and the genetic testing industry. To better illustrate the uses and limitations of genetic information we have chosen to take an unusual approach: sharing our own genetic data publicly.



Exemples de profils génétiques

- L'identification des personnes ne nécessite pas de connaître l'entièreté de leur génome.
- La police scientifique utilise des profils génétiques basés sur une vingtaine de microsatellites. Ce petit nombre de marqueurs suffit à identifier une personne sans équivoque possible.
- Les marqueurs sont choisis pour avoir de bonnes qualités pour l'identification.
 - Ils permettent par exemple d'identifier une personne sur base de traces biologiques laissées sur les lieux d'un délit, à condition de disposer du profil de cette même personne (ou d'un parent proche) dans la base de données.
- Ces marqueurs n'ont pas de valeur prédictive
 - Pour des raisons éthiques, ces microsatellites ont été choisis pour ne pas être porteurs d'information phénotypiques, afin d'éviter d'exposer des données à caractère personnel (par exemple les risques pour un individu de développer une maladie particulière).

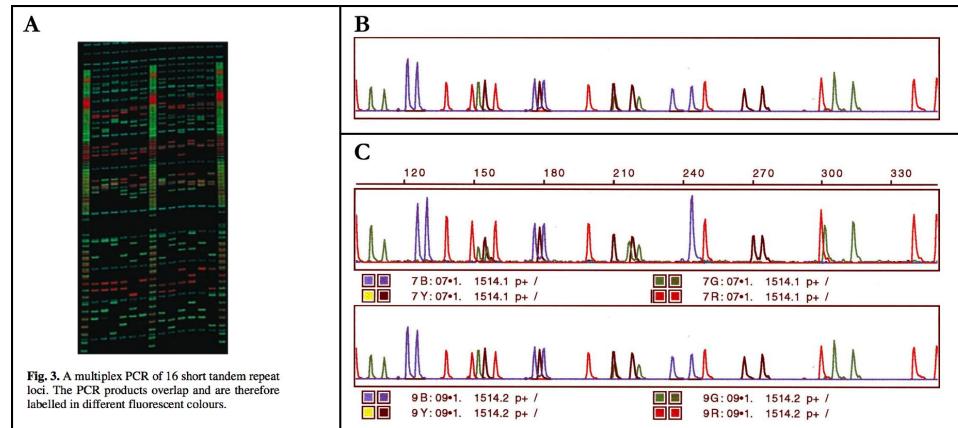


Figure . Exemples de profil génétiques basés sur les microsatellites (source : Pierre Van Renteghem). **A:** multiplex montrant les profils génétiques d'une série d'individus. **B:** profil génétique obtenu à partir d'un échantillon trouvé sur les lieux d'un crime. **C:** profils génétiques de deux suspects.

Bases de données de la police scientifique en Europe

- Les Etats constituent des bases de données nationales de profils génétiques.
- Ces profils sont basés sur quelques microsatellites, utilisés par la police scientifique à des fins d'identification, mais qui n'ont pas de valeur prédictive quant au phénotype d'un individu (traits physiques, risques de maladies, ...).
 - En Europe, chaque pays a sa propre politique en matière de génotypage de la population.
 - En Belgique, on ne conserve les profils génétiques que pour les personnes inculpées pour un crime, ou suspectées dans une enquête en cours.
 - En Angleterre et en France, l'Etat vise à établir les profils génétiques de la population générale, et dispose déjà de profils génotypiques pour des millions d'individus.

Country	Population size	Persons				Remarks
		A	S	CO	T	
Albania	3,600,000					
Armenia	3,000,000					
Austria	8,100,000				197,941	
Belgium	10,400,000	1,207	34,784	35,991		Persons added minus persons removed Including duplicates
Bosnia & Herzegovina	4,400,000					
Bulgaria	7,900,000					
Croatia	4,300,000				31,199	
Cyprus	772,000			414	414	
Czech Republic	10,553,800				171,519	Persons added minus persons removed Including duplicates
Denmark	5,500,000				112,829	Persons added minus persons removed
Estonia	1,311,800				47,618	Persons added minus persons removed
Finland	5,475,866				157,303	Persons added minus persons removed
France	66,030,000	139,727	2,567,112	501,131	3,068,243	Including duplicates
Georgia	4,700,000		1,517		5,819	All person profiles ever added
Germany	81,000,000				849,907	
Greece	10,600,000				8,362	
Hungary	9,982,000		108,600	29,061	137,661	Persons added minus persons removed Including duplicates
Iceland	315,000					
Ireland	4,200,000		7		7	
Italy	58,000,000					
Kosovo	1,800,000					
Latvia	2,000,000		42,446	10,095	52,541	Including duplicates
Liechtenstein	37,000					
Lithuania	2,960,000				76,317	
Luxembourg	570,000		264	2,097	2,361	
Macedonia	2,100,000				17,094	All person profiles ever added
Malta	400,000		30		30	
Montenegro	650,000					
Netherlands	17,000,000				224,669	Persons added minus persons removed No duplicates
Northern Ireland	1,685,000					
Norway	5,000,000		9,755	56,321	66,076	No duplicates
Poland	38,200,000				42,753	No duplicates
Portugal	10,300,000				4,664	All person profiles ever added
Romania	22,000,000		1,839	30,310	32,149	
Russia	143,800,000					
Scotland	5,500,000		174,219	136,888	311,107	
Serbia	7,335,000					
Slovakia	5,500,000				51,826	
Slovenia	2,000,000				31,003	Including duplicates Persons added minus persons removed
Spain	46,700,000				319,837	Persons added minus persons removed
Sweden	9,845,155		13,498	138,433	151,931	Including duplicates
Switzerland	7,779,000				176,758	Persons added minus persons removed
Turkey	66,800,000					No duplicates
UK (England & Wales)	53,700,000				4,691,350	
Ukraine	47,600,000					
Total	801,401,621				11,071,460	

S: suspect; CO: convicted; T: total

Source: DNA DATABASE MANAGEMENT REVIEW AND RECOMMENDATIONS. ENFSI DNA Working Group April 2017.

<http://enfsi.eu/wp-content/uploads/2017/09/DNA-databasemanagement-review-and-recommendations-april-2017.pdf>

Semi-annual ENFSI DNA database overview 2014

S: suspect; CO: convicted; T: total

Country	Population size	Persons				Stains	Matches				Stain-person matches per person		
		Persons		S	CO	T	Person/Stain		Stain/Stain	Total			
		A	S				CO	T					
Austria	8.100.000			173.338	61.115		17.804	7.856	25.660	aug-13	0,11		
Belgium	10.400.000	n/a	26.467	26.467	32.625	346	2.062	2.408	4.144	6.552	jun-13	0,09	
Bulgaria	7.900.000			17.618	1.147		377	122	499	499	jul-09	0,02	
Croatia	4.600.000			29.293	5.180		3.752	1.519	5.271	5.271	feb-10	0,13	
Cyprus	772.000	n/a	305	305	10.628	n/a	78	78	212	290	jun-13	0,26	
Czech Republic	10.515.000	2.436	110.760	113.196	18.727		7.703	2.896	10.599	10.599	jun-13	0,07	
Denmark	5.500.000			92.206	44.740		20.798	4.284	25.082	25.082	jun-13	0,23	
Estonia	1.286.540			34.012	10.475		3.957	932	4.889	4.889	dec-12	0,12	
Finland	5.402.145			136.963	15.633		17.496		17.496	17.496	jun-13	0,13	
France	64.300.000	128.312	1.886.876	427.649	2.314.525	202.427	66.315	19.642	85.957	11.260	97.217	jul-13	0,04
Georgia	4.700.000			793.628	234.205		123.845	32.482	156.327	156.327	jun-13	0,16	
Germany	80.200.000				8.112				520	520	520	jun-13	
Greece	10.600.000												
Hungary	9.982.000	92.614	8.441	101.055	4.454		451		226	677	jul-13	0,00	
Ireland	4.200.000												
Italy	58.000.000												
Kosovo	1.800.000												
Latvia	2.400.000	36.421	9.627	46.048	3.263		1.267	223	1.490	1.490	jul-13	0,03	
Lithuania	2.960.000			66.566	3.786		1.810	378	2.188	2.188	jul-13	0,03	
Luxembourg	500.000		85	1.229	1.314	1.186	197	262	459	2.448	2.907	jul-13	0,35
Former Yugoslavian Republic of Macedonia	2.000.000			7.996	3.145				898	133	1.031	Jun.13	0,11
Malta	400.000												
Montenegro	650.000												
Netherlands	16.100.000			170.788	55.482		38.762	4.868	43.630	43.630	jul-13	0,23	
Northern Ireland	1.665.000												
Norway	5.000.000		9.577	35.043	54.506	9.078		8.967	2.588	11.555	nov-13	0,16	
Poland	38.200.000				32.624	3.600		225	164	389	jun-13	0,01	
Portugal	10.300.000	0	0	1.134	1.134	409	0	4	4	30	34	jul-13	0,00
Romania	22.000.000		930	17.229	18.159	757			3.159	42	3.201	jul-13	0,17
Russia	143.800.000												
Scotland	5.500.000		174.219	136.888	311.107	18.725		31.249	2.556	33.805	apr-13	0,10	
Serbia	7.335.000												
Slovakia	5.500.000				38.559	8.181		4.029	1.391	5.420	jul-13	0,10	
Slovenia	2.000.000				26.548	6.226		3.975	549	4.524	jul-13	0,14	
Spain	44.800.000			244.243	70.380		28.475	34.537	63.012	63.012	jul-13	0,12	
Sweden	9.000.000		13.979	114.039	128.018	26.698	25.674	13.243	36.917	15.617	54.534	jul-13	0,30
Switzerland	7.779.000				152.913	45.796			40.130	9.716	49.846	jul-13	0,26
Turkey	66.800.000												
UK (England & Wales)	53.700.000			4.795.615	414.982		1.905.436	362.252	2.267.688	2.267.688	jul-13	0,40	
Ukraine	47.600.000												
Total	784.266.685				9.928.744	1.321.162		2.392.388	503.945	2.896.333			

L'ADN comme support au portrait-robot

- Une étude d'association à échelle génomique (GWAS) basée sur 10.000 personnes, et 2,5M SNPs.
- L'étude identifie 5 gènes présentant des formes alléliques associées à des caractéristiques morphologiques du visage.
- Elle ouvre la perspective de dresser des portraits-robots à partir de profils génomiques.

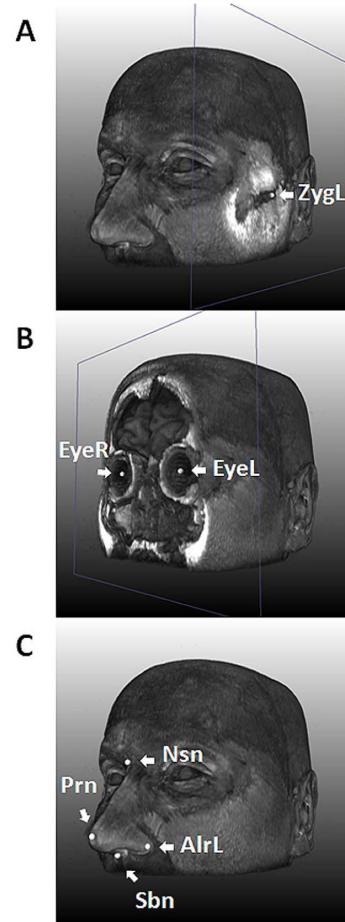


Figure 1. Nine facial landmarks extracted via image registration tools from 3D MRIs. An MRI of one of the authors (MK) is used for illustration. A, with the landmark for left zygion (ZygL) highlighted, where a clipping plane was used to uncover the bone; B, with the landmarks for left (EyeL) and right pupils (EyeR) highlighted, where a clipping plane was used to uncover the vitreous humor; C, with the four nasal landmarks highlighted, including the left alare, nasion (NsN), pronasale (Prn), and subnasale (Sbn). doi:10.1371/journal.pgen.1002932.g001

OPEN ACCESS Freely available online

PLOS GENETICS

A Genome-Wide Association Study Identifies Five Loci Influencing Facial Morphology in Europeans

Fan Liu¹, Fedde van der Lijn^{1,2,3}, Claudia Schurmann⁴, Gu Zhu⁵, M. Mallar Chakravarty^{6,7}, Pirro G. Hysi⁸, Andreas Wollstein¹, Oscar Lao¹, Marleen de Bruijne^{2,3}, M. Arfan Ikram^{3,9}, Aad van der Lugt³, Fernando Rivadeneira^{9,10}, André G. Uitterlinden^{9,10}, Albert Hofman⁹, Wiro J. Niessen^{2,3,11}, Georg Homuth⁴, Greig de Zubiracay¹², Katie L. McMahon¹², Paul M. Thompson¹³, Amro Daboul¹⁴, Ralf Puls¹⁵, Katrin Hegenscheid¹⁵, Liisa Bevan⁹, Zdenka Pausova¹⁶, Sarah E. Medland⁵, Grant W. Montgomery⁵, Margaret J. Wright⁵, Carol Wicking¹⁷, Stefan Boehringer¹⁸, Timothy D. Spector⁸, Tomás Paus^{6,19}, Nicholas G. Martin⁵, Reiner Biffar¹⁴, Manfred Kayser^{1*} for the International Visible Trait Genetics (VisiGen) Consortium