Cahiers de l'Urmis

20 | 2021

Race et biologie

La génomique et la diversité humaine

Bertrand Jordan

Édition électronique

URL : <https://journals.openedition.org/urmis/2387>

DOI : 10.4000/urmis.2387

ISSN : 1773-021X

Traduction(s) :

Genomics and human diversity - URL : <https://journals.openedition.org/urmis/2397> [en]

Éditeur

Urmis

Référence électronique

Bertrand Jordan, « La génomique et la diversité humaine », *Cahiers de l'Urmis* [En ligne], 20 | 2021, mis en ligne le 29 juin 2021, consulté le 15 février 2023. URL : <http://journals.openedition.org/urmis/2387> ; DOI : <https://doi.org/10.4000/urmis.2387>

Ce document a été généré automatiquement le 1 juillet 2021.

Creative Commons - Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International
- CC BY-NC-ND 4.0

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

La génomique et la diversité humaine

Bertrand Jordan

- ¹ Depuis la première lecture complète de l'ADN humain, achevée en 2003 au prix de dix ans d'efforts menés dans des centaines de laboratoires, les techniques d'analyse de l'ADN ont fait d'immenses progrès et ont été appliquées à une échelle de plus en plus large, tant en recherche fondamentale que pour de multiples applications cliniques. L'étude des génomes (la génomique) a ainsi porté sur le patrimoine héréditaire de nombreuses espèces et bien sûr de la nôtre. Elle permet de réenvisager de manière plus précise et plus objective de nombreuses questions de nature anthropologique et notamment celle de l'existence et des caractéristiques de groupes au sein de la population humaine. Le but de cet article est de faire le point sur les résultats actuels de ces recherches et sur leurs implications.

1. Diversité humaine au niveau de l'ADN

- ² Comme chacune le sait aujourd'hui, notre patrimoine héréditaire est inscrit dans notre ADN, une suite de trois milliards de « bases » T, A, G et C contenues dans nos chromosomes et portant les quelque vingt mille séquences (suites de bases) qui constituent nos gènes. C'est cette information qui dirige la synthèse des milliers de protéines assurant la structure et la fonction des deux cents types cellulaires qui constituent un organisme humain, information qui nous vient de nos parents et qui sera transmise à nos enfants. Nous avons tous les mêmes gènes, disposés dans le même ordre sur nos chromosomes, mais ils ne sont pas strictement identiques d'un individu à un autre : maintenant que nous avons les moyens de lire (« séquencer ») l'ensemble de l'ADN d'une personne à un prix accessible (moins de mille euros), on peut comparer les ADN de différents individus et l'on constate qu'entre deux personnes prises au hasard existent environ trois millions de différences ponctuelles dans l'ADN¹. Différences ponctuelles, c'est-à-dire qu'en un certain point d'un certain chromosome on trouvera chez l'un une base donnée (disons un G), chez l'autre une base différente (un A par

exemple). Trois millions sur trois milliards, un pour mille : c'est le célèbre 0,1 % de différence dont on a déduit que « les races humaines n'existent pas » puisque deux personnes prises au hasard ont des ADN identiques à 99,9 %. Nous verrons que la réalité n'est pas si simple, et que cela dépend beaucoup du sens que l'on donne au terme « race »... (Jordan 2008).

- 3 De fait, trois millions de différences, ce n'est *a priori* pas négligeable. Disons tout de suite que la très grande majorité d'entre elles n'a aucune importance, aucun impact sur le phénotype (l'aspect, la physiologie, le comportement...) de la personne à laquelle appartient cet ADN. En fait les gènes proprement dits (les séquences de bases qui codent pour une protéine, les *séquences codantes*) ne représentent qu'environ 1 % de notre ADN : les 99 % restants, parfois appelés *Junk DNA* ou ADN poubelle, semblent dans leur grande majorité ne pas avoir de rôle précis. Du coup, une différence située dans ces régions ne va pas avoir de conséquence fonctionnelle sur la personne qui porte l'ADN correspondant. De plus, même si la différence touche une séquence codante, elle n'est pas nécessairement significative : le « code génétique », qui indique quel acide aminé de la protéine est spécifié par l'ADN, est « dégénéré » (au sens mathématique du terme), c'est-à-dire que plusieurs « codons » (suite de trois bases dans l'ADN) codent pour le même acide aminé. Par exemple l'histidine, un des vingt acides aminés présents dans les protéines, peut être codée dans l'ADN par CAT ou par CAG : du coup, si le T est remplacé par un G dans l'ADN, le même acide aminé sera incorporé dans la protéine, qui ne sera pas différente malgré la différence présente dans l'ADN (c'est ce que l'on appelle une différence (ou une mutation) silencieuse). Au total, seules environ dix mille de ces trois millions de différences sont significatives, c'est-à-dire qu'elles changent la formule d'une protéine. La modification de la protéine (une enzyme digestive, par exemple) peut la rendre moins active, ou plus active, plus ou moins sensible à la température, ou encore ne rien changer : sur les quelques centaines d'acides aminés qui constituent une protéine, tous ne sont pas critiques. Si par contre la différence dans l'ADN fait apparaître un codon Stop (TGA, TAA ou TAG) qui signifie en quelque sorte « fin de message », on aura alors la synthèse d'une protéine tronquée, *a priori* non fonctionnelle : c'est le cas pour une centaine des différences. Cela pourrait sembler catastrophique, mais n'oublions pas que chacun de nos chromosomes est présent en deux exemplaires (l'un provenant du père, l'autre de la mère) : si un gène est inactif, son homologue porté par l'autre chromosome pourra en général le suppléer. On peut aussi avoir des mutations « faux-sens », qui provoquent un changement d'acide aminé et vont parfois (mais pas toujours) altérer ou annuler la fonction de la protéine. Au total, c'est l'ensemble de ces différences génétiques qui entraîne la diversité phénotypique entre individus de notre espèce : caractères visibles comme la couleur de la peau, des yeux, des cheveux, la taille à l'âge adulte, mais aussi différences cachées comme les groupes sanguins, la capacité à digérer le lactose, la tendance à l'obésité, la vulnérabilité à certains cancers... Soulignons que si la séquence de l'ADN détermine de nombreux éléments du phénotype, nous ne sommes actuellement pas capables, sauf exception, de déduire ce phénotype de la lecture de l'ADN. Rappelons aussi que cette diversité génétique est modulée par l'environnement, les conditions de vie et l'histoire personnelle, parfois de manière importante comme pour la taille ou l'obésité.
- 4 L'étude de la diversité humaine au niveau du génome n'a réellement pu se développer que grâce à des avancées scientifiques et technologiques récentes : le séquençage, la lecture d'un premier génome humain, achevée en 2003, puis la mise au point, dans la

décennie qui a suivi, d'outils très performants, les « puces à ADN » ou *microarrays*, dispositifs miniaturisés permettant d'établir rapidement le profil génétique d'un ADN. Plus précisément, ces systèmes permettent d'analyser quelques centaines de milliers de points particulièrement variables au sein d'un ADN (souvent appelés SNP ou Snip pour *Single Nucleotide Polymorphism*) et de déterminer quel allèle (quelle « version ») d'un Snip est présent en chacun de ces points. À l'heure actuelle ces *microarrays* examinent couramment cinq cent mille points dans un ADN humain, pour un coût de quelques dizaines d'euros par échantillon, ce qui permet d'étudier ainsi l'ADN de milliers de personnes (Jordan 2008). Plus récemment, d'incroyables perfectionnements des techniques de lecture (séquençage) de l'ADN ont fait passer le coût de la lecture intégrale d'un génome humain sous la barre des mille euros – alors que le premier séquençage avait mobilisé des centaines de laboratoires durant dix ans et coûté plusieurs milliards de dollars. La lecture intégrale de l'ADN donne évidemment une image plus précise du profil génétique d'une personne (trois milliards de points au lieu de cinq cent mille pour un *microarray*), mais elle n'est pas encore très répandue en raison de son coût et de la complexité de sa mise en œuvre pour des études de population portant sur des milliers de personnes.

- 5 Si l'on a déterminé le profil génétique de plusieurs personnes, par exemple à l'aide de *microarrays*, on va pouvoir définir et calculer une « distance génétique » entre deux personnes. Deux individus génétiquement identiques (deux vrais jumeaux) posséderont le même allèle pour chacun des cinq cent mille Snip analysés, et leur distance génétique sera nulle. En dehors de ce cas, la distance génétique entre deux personnes sera fonction du nombre d'allèles différents repérés par l'analyse, par rapport au nombre total de Snip examinés. Pour deux individus apparentés, la distance sera plus faible, et d'autant plus que la parenté est proche. À ce stade, on imagine bien que l'analyse selon ce principe d'une population va permettre d'examiner s'il existe en son sein des groupes de personnes génétiquement similaires, des *clusters* : c'est de cela que traite la prochaine partie.

2. Analyses de populations et apparition de groupes (« clusters »)

- 6 Nous allons maintenant nous intéresser à l'analyse de populations sur la base des profils génétiques établis pour chaque individu. Il s'agit cette fois d'examiner la distance génétique entre personnes au sein d'une population, et de représenter les résultats d'une manière immédiatement compréhensible. Je prendrai pour illustrer cette approche et ses résultats une étude publiée en 2012 par un consortium international (*The 1000 Genomes Project*) qui avait analysé les ADN de 1092 personnes d'origines diverses – analyse fine puisqu'elle avait été effectuée par lecture complète de chacun de ces ADN, ce qui donne plus de précision à l'étude. On compare donc le profil génétique de chacun de ces 1092 individus à ceux de 1091 autres, et l'on calcule à chaque fois la distance génétique qui les sépare. À l'arrivée, on a donc plus d'un million de valeurs de distances génétiques, et l'on est assez embarrassé pour représenter graphiquement ces résultats puisque les mille points à figurer sont situés dans un espace à mille dimensions... On a alors recours à une technique mathématique devenue classique, la décomposition en « composantes principales »² : la méthode révèle la première composante principale, c'est-à-dire la direction selon laquelle la dispersion

des points est la plus importante, puis la deuxième composante (la direction perpendiculaire donnant la plus grande dispersion), et ainsi de suite... En fait on s'arrête souvent aux deux premières composantes, ce qui permet de représenter les résultats dans un espace à deux dimensions, un graphique sur une feuille de papier. La Figure 1 montre le résultat obtenu :

Figure 1. Répartitions des points indiquant la position de 1092 ADN analysés par rapport aux deux premières composantes principales (PC 1 et PC 2). Chaque point représente la position d'un individu par rapport aux 1091 autres, telle qu'elle apparaît en regardant les données selon les deux premières composantes principales. Extrait partiel et modifié de la figure 4 du supplément de *The 1000 Genomes Project* (2012), licence Creative Commons.

- 7 On voit que les points, dont chacun représente un ADN analysé et donc un individu, ne sont pas répartis de manière homogène, ni regroupés tous ensemble : ils forment une constellation dans laquelle on distingue trois groupes principaux et des nuages moins denses se répartissant entre certains de ces groupes. Soulignons que ce résultat est fondé sur la seule analyse des ADN, intégrant l'ensemble des différences révélées par l'analyse, et que l'origine des personnes impliquées n'a à aucun moment été prise en compte. Il existe donc bien, objectivement, des sous-ensembles de personnes génétiquement proches au sein de l'ensemble de la population étudiée.
- 8 L'étape suivante consiste naturellement à examiner dans quelle mesure les groupes ainsi révélés correspondent à des populations définies d'un point de vue géographique et social. C'est ce que montre la même image dans laquelle chaque population est repérée par la forme et la couleur du symbole.

Figure 2. Comme la Figure 1, mais les origines des différentes populations sont indiquées. On distingue trois groupes compacts, les Européens (GBR, CEU, FIN en bleu), les Africains (YRI, LWK en gris) et les Asiatiques (CHS, CHB, JPT en vert). Les points correspondant aux autres populations (Afro-Américains (ASW) et Mexicains (MXL), notamment) sont plus dispersés, ce qui correspond à leur origine mixte. Abréviations utilisées : GBR, Britanniques ; FIN, Finlandais ; CHS, Chinois Han, Chine du Sud ; PUR, Portoricains ; CLM, Colombiens de Medellin ; IBS, Ibériques en Espagne ; CEU, résidents de l'Utah originaires de l'Europe du Nord ; YRI, Yoruba d'Ibadan au Nigéria ; CHB, Chinois Han de Beijing ; JPT, Japonais de Tokyo ; LWK, Luhya du Kenya ; ASW, Afro-Américains du Sud-Ouest des USA ; MXL, Mexicains de Los Angeles ; TSI, Toscans d'Italie. Extrait partiel de la figure 4 du supplément de *The 1000 Genomes Project* (2012), licence Creative Commons.

- 9 Il apparaît immédiatement qu'en effet les trois zones les plus denses du graphique correspondent à des populations géographiquement définies : Européens pour les points bleus (GBR, FIN, CEU), Asiatiques pour les verts (CHS, CHB, JPT), Africains pour les gris (YRI, LWK). Quant aux points plus dispersés, ils concernent les Afro-Américains des États-Unis (ASW) dont on sait qu'ils ont souvent une part d'ascendance Européenne et qui s'étalent effectivement entre Africains et Européens, les Portoricains (PUR, même remarque), les Mexicains et les Colombiens (MXL, CLM)³. Au total, l'analyse sans hypothèse *a priori* des distances génétiques, pratiquée sur un ensemble de personnes de provenance diverse met donc en évidence de manière objective des groupements, des *clusters*, et ces derniers sont assez bien corrélés avec des groupes « ethniques » socialement définis comme les Européens, les Asiatiques ou les Africains. Ces groupes correspondent en fait à des origines géographiques (Europe, Asie, Afrique) et sont le reflet de l'histoire évolutive de notre espèce. Après son apparition en Afrique il y a deux ou trois cent mille ans, *Homo sapiens* (l'homme moderne) a essaimé, principalement il y a quatre-vingts ou cent mille ans, et a progressivement peuplé toute la Terre. La sélection naturelle agissant dans des environnements très différents, la dérive génétique (perte aléatoire de certains variants en raison du faible effectif des groupes humains) et peut-être la sélection sexuelle (choix du partenaire en fonction de certains caractères visibles) ont poussé à la différenciation de groupes humains,

différenciation limitée car cent mille ans, c'est peu en termes d'évolution, mais différenciation tout de même.

- 10 Un dernier point : la vue du graphique (Figure 2) donne l'impression de populations séparées ayant, du moins pour certaines d'entre elles, des limites assez nettes et pas de recouvrement avec d'autres. Allons-nous donc donner raison à Gobineau qui, dans son « Essai sur l'inégalité des races humaines » (1853), affirmait que « L'humanité est *divisée* en races » ? Non bien sûr : il ne faut pas oublier que l'analyse en composantes principales nous a permis de choisir la représentation qui différencie le plus les points correspondant aux individus. La très grande majorité des aperçus possibles des mêmes données ne différencie pas des nuages de points. Et l'affirmation selon laquelle les différences génétiques au sein d'une population sont plus importantes que la différence moyenne entre deux populations, cette affirmation qui a été considérée comme la preuve scientifique que « les races humaines n'existent pas », reste valable – mais elle s'applique à des caractères considérés isolément. Presque tous les allèles (les variants de Snip) existent dans chaque population, qui est de ce point de vue presque aussi diverse que l'ensemble de l'humanité ; mais les *fréquences* des différents allèles varient entre les groupes et permettent ainsi de les circonscrire. Pour prendre un exemple concret, on peut trouver des individus aux yeux clairs au sein d'une population purement africaine, mais ils sont extrêmement rares, alors qu'ils sont majoritaires parmi les Européens du Nord.

3. Du génotype au phénotype, individus et groupes

Diversité des individus

- 11 L'exemple donné ci-dessus est d'ailleurs impropre, du fait que la couleur des yeux n'est pas déterminée par un seul gène : elle est le résultat de l'expression de plusieurs gènes, au moins une dizaine. C'est le cas pour l'essentiel des caractères visibles (et souvent aussi des autres), et cela nous amène à nous interroger sur la manière dont le *génotype* (ce qui est inscrit dans les gènes) détermine ou influence le *phénotype*, les caractéristiques morphologiques, fonctionnelles, comportementales... d'une personne. Nous n'avons jusqu'ici considéré que la diversité génétique, repérée au niveau de l'ADN, il nous faut maintenant examiner comment elle s'exprime et s'incarne dans les caractères d'une personne. Disons tout de suite qu'élucider ce passage reste, aujourd'hui encore, très malaisé. Cela tient à plusieurs raisons. Tout d'abord, comme je viens de le dire, un caractère est la plupart du temps déterminé par (ou plutôt sous l'influence de) plusieurs gènes. Par exemple, la taille à l'âge adulte, caractère pourtant très héritable, est liée à plus d'un millier de gènes ; la couleur des yeux, une dizaine. Pour la couleur de la peau, c'est plus d'une centaine de gènes qui ont été repérés dans différentes populations. Et le fait d'avoir identifié une série de gènes influençant un caractère ne signifie pas que l'on puisse prédire ce caractère à partir d'une analyse d'ADN : pour cela, il faudrait savoir précisément quel rôle joue chacun de ces gènes et comment il influence le caractère en question, ce qui est rarement le cas. De plus, l'environnement, l'histoire de la personne jouent un rôle parfois important dans l'expression d'un caractère : le potentiel génétique de stature, par exemple, n'est réalisé que si l'enfant puis l'adolescent n'ont pas souffert de carences alimentaires. C'est une meilleure alimentation infantile qui explique que les Hollandais soient passés

d'une taille moyenne de 1m70 (en 1914) à 1m83 en 2014 : leur patrimoine génétique n'a pas significativement changé dans l'intervalle... Le rôle du génotype reste néanmoins important. Prenons par exemple la forme et les traits du visage : il y a à l'évidence détermination génétique stricte, comme en témoigne la ressemblance des vrais jumeaux dont les ADN sont identiques (du moins à la naissance). Pourtant, à l'heure actuelle, et malgré quelques annonces plus publicitaires que scientifiques, il n'est pas possible de dresser un « portrait-robot » précis à partir d'une analyse d'ADN (Jordan 2020). Tout au plus peut-on éventuellement indiquer que la personne doit avoir un aspect plutôt Européen, Asiatique ou Africain : mais cette information sera déduite du fait de son appartenance à un groupe d'ascendance – dont on a vu le caractère approximatif – et non d'une connaissance précise des gènes qui déterminent son aspect. On pourra aussi, grâce cette fois à l'analyse d'une quarantaine de gènes, avoir une idée de la couleur de ses yeux, de ses cheveux et (c'est tout récent) de la teinte de sa peau (Jordan 2020). C'est déjà beaucoup, et cela ne manque pas de soulever quelques problèmes éthiques, mais c'est encore peu pour caractériser un visage dont l'aspect est pourtant très largement déterminé par l'ADN...

- 12 La diversité génétique des individus, combinée avec la diversité de leurs modes de vie, aboutit ainsi à la grande diversité phénotypique que nous observons au sein de l'espèce humaine. Nous sommes « tous parents, tous différents »⁴, et ces différences peuvent être un avantage ou un inconvénient selon les conditions dans lesquelles nous vivons. Pour un chasseur-cueilleur du paléolithique, être myope devait être un lourd handicap ; par contre, porter une mutation qui entraîne une grave dégénérescence neuromusculaire débutant après la quarantaine⁵ n'avait pas grande conséquence, vu la faible espérance de vie humaine à l'époque. Pour un *homo sapiens* contemporain, c'est tout l'inverse... Un individu n'est pas intrinsèquement supérieur ou inférieur à un autre ; mais il peut être plus ou moins bien adapté à un environnement donné.

Diversité des groupes

- 13 Comment se retrouve cette diversité des génotypes et des phénotypes au niveau des groupes humains ? C'est ici que l'on arrive sur un terrain glissant et même scabreux, qui a donné lieu à toute sorte de stéréotypes qui restent encore prégnants : l'Asiatique intelligent et rusé, le Noir paresseux mais doué pour la musique, l'Européen industriel et dominateur... Ces qualificatifs parfois injurieux sont largement liés à un passé colonial d'oppression par les « races supérieures », pour le bien des « races inférieures » comme l'affirmait Jules Ferry⁶. En fait, si l'on veut analyser les phénotypes au niveau de groupes humains, il faut que ces groupes soient définis par leur proximité génétique et non en tant que catégorie sociale. Comme on le voit sur la Figure 2, les Afro-Américains du Sud-Ouest des États-Unis (ASW) ne constituent pas un groupe génétique bien défini (contrairement aux Luhya du Kenya, LWK), pas plus que les Mexicains de Los Angeles (MEX) ou les Portoricains (PUR). C'est un biais fatal dans la plupart des études menées aux États-Unis sur, par exemple, les différences d'incidence de maladies comme le diabète ou le cancer de la prostate entre différents groupes « ethniques ». Comme la « race » fait l'objet de déclaration dans le cadre des recensements périodiques de la population (Figure 3), il est évidemment commode de considérer cette indication comme une approximation de l'appartenance à un groupe. Mais, comme le montrent les exemples cités plus haut, cette approximation est trompeuse, et la délimitation évoquée dans ce cas par la « race » n'est en fait qu'une

catégorie sociale qui ne correspond à rien de précis du point de vue biologique. Pour reprendre le cas des Afro-Américains, on a là une catégorie sociale qui est hétérogène de point de vue génétique, et cela découle de facteurs historiques : la tristement célèbre *one-drop rule*, la règle encore en vigueur jusqu'à la deuxième guerre mondiale et selon laquelle une seule goutte de « sang noir » (un seul ancêtre africain) faisait d'une personne un « Noir », même s'il était génétiquement et phénotypiquement plus « Blanc » que « Noir ».

9. What is Person 1's race?
Mark ☒ one or more boxes AND print origins.

☐ White – Print, for example, German, Irish, English, Italian, Lebanese, Egyptian, etc.

☐ Black or African Am. – Print, for example, African American, Jamaican, Haitian, Nigerian, Ethiopian, Somali, etc.

☐ American Indian or Alaska Native – Print name of enrolled or principal tribe(s), for example, Navajo Nation, Blackfeet Tribe, Mayan, Aztec, Native Village of Barrow, Inupiat Traditional Government, Nome Eskimo Community, etc.

☐ Chinese ☐ Vietnamese ☐ Native Hawaiian

☐ Filipino ☐ Korean ☐ Samoan

☐ Asian Indian ☐ Japanese ☐ Chamorro

☐ Other Asian – Print, for example, Pakistani, Cambodian, Hmong, etc.

☐ Other Pacific Islander – Print, for example, Tongan, Fijian, Marshallese, etc.

☐ Some other race – Print race or origin.

Figure 3. Déclaration de la « race » dans le formulaire de recensement 2020 aux États-Unis (extrait du formulaire). Il s'agit d'une auto déclaration, et il est possible (bien que peu fréquent) de cocher plus d'une case. « Person 1 » est le déclarant. (Formulaire accessible à : <https://www.census.gov/programs-surveys/decennial-census/technical-documentation/questionnaires/2020.html>).

- 14 Que peut-on dire alors des phénotypes liés à un groupe d'ascendance, et de la différence moyenne entre ces groupes ? Pour autant que l'on envisage un groupe bien défini (forte proximité génétique), il y a effectivement une certaine correspondance entre chaque groupe et le phénotype visible : couleur de la peau, des yeux, des cheveux, type de visage et traits... Ce phénotype « moyen » pour chaque groupe résulte de la sélection naturelle (peau claire pour les Européens favorisant la synthèse de vitamine D, par exemple), de la « dérive génétique » (au fil des migrations, certains allèles de certains gènes peuvent être perdus) et sans doute de la sélection sexuelle (choix des partenaires selon certains critères). Pour un caractère phénotypique donné, il reste une grande amplitude de variation au sein du groupe (il existe des Africains d'Afrique dont la peau est assez claire, d'autres très foncés), mais le « type moyen » au sein d'un groupe se distingue clairement de celui d'un autre. C'est une expérience que peut faire chacun en regardant les passants dans la rue, et qui explique le scepticisme provoqué par des affirmations sans nuance comme « les races humaines n'existent pas »... sans autre précision.

- 15 Qu'en est-il pour les phénotypes plus profonds, sensibilité aux maladies, comportements, aptitudes ? On a pu supposer que ces différences étaient moins fortes que celles qui touchent l'aspect, en somme que nous différions au niveau de la « carrosserie » mais que châssis et moteur étaient les mêmes. Cela aurait pu résulter d'un rôle primordial joué par la sélection sexuelle (qui se fonde principalement sur l'aspect des individus), mais l'analyse des génomes a largement démenti cette hypothèse. Nous verrons plus loin la question des maladies ; l'étude des performances physiques et intellectuelles montre à quel point il est difficile de différencier l'apport du génotype et celui de l'environnement, de l'histoire personnelle. Prenons un exemple qui a fait couler beaucoup d'encre, celui des performances intellectuelles. Il a fait l'objet de maintes études, principalement dans le monde anglo-saxon, qui ont porté sur l'héritabilité du quotient intellectuel (QI) et sur sa valeur dans différentes populations. Deux conclusions principales en découlent : d'abord, que l'héritabilité du QI est importante, au moins 0,5 (50 %), et d'autre part que sa valeur moyenne diffère selon les populations. Notons bien qu'une héritabilité de 50 % ne signifie pas que la moitié du niveau de QI soit due aux gènes, l'autre moitié à l'environnement, mais que, dans une population qui présente une certaine variabilité du QI (disons de 90 à 110), la moitié de cette variation (soit dix points de QI) est liée au patrimoine génétique. Quant à la valeur moyenne du QI, elle apparaît plus faible de 5 à 10 points, selon les études, chez les Afro-Américains par rapport aux Européens ou aux Asiatiques. Un caractère fortement héritable, et une différence entre deux groupes « ethniques » : il n'en a pas fallu plus à certains pour conclure que l'« infériorité » intellectuelle des Afro-Américains était due à leurs gènes... Soulignons d'abord que le QI n'est pas une mesure de l'intelligence, mais un paramètre qui lui est (plus ou moins) corrélé, et que la mesure du QI d'un individu repose sur une série de tests (d'association, d'abstraction, de logique...) dont il est bien difficile de s'assurer qu'ils ne sont pas influencés par la culture (la famille, les valeurs, le mode de vie...) à laquelle appartient la personne testée. Admettons que ce test soit réellement objectif, faisons aussi l'impasse sur l'hétérogénéité génétique des Afro-Américains : l'étude prouve-t-elle que la différence entre Européens et Africains est due à leurs gènes ? Non bien sûr : souvenons-nous de ces Hollandais dont la taille moyenne a augmenté de 13 cm en un siècle : ce ne sont pas leurs gènes qui ont changé entre 1914 et 2014, mais bien leurs conditions de vie et leur alimentation. Pourtant la taille à l'âge adulte est un des caractères les plus hératables (héritabilité de 80 %) – mais ceci n'est vrai que « toutes choses égales par ailleurs ». Et il est tristement évident qu'il n'est pas « égal » d'être Européen ou Afro-Américain aux États-Unis...
- 16 Cet exemple montre toute la difficulté qu'il y a à définir la part de l'hérédité et de l'environnement dans les comparaisons entre groupes humains. De la même façon, et malgré la prépondérance des coureurs Kenyans ou Éthiopiens dans les courses de fond, on n'a pas pu identifier des variants de gène(s) capables d'expliquer cette prépondérance, et l'on fait appel à un ensemble de facteurs environnementaux et culturels interagissant avec un génotype globalement favorable pour expliquer ces succès.
- 17 Finalement, on peut conclure que le génotype d'un individu détermine assez largement son phénotype, mais que nous ne savons pas encore, sauf exception, déduire le second du premier ; quant aux groupes, ils présentent bien des différences entre les moyennes des valeurs de telle ou telle composante du phénotype, mais il est en général bien

difficile de déterminer si ces différences sont liées au patrimoine génétique, à l'environnement, ou aux deux.

Gènes, groupes humains et maladies

- 18 Le patrimoine génétique d'une personne joue un rôle dans sa susceptibilité à de nombreuses affections. Ce rôle est prépondérant dans le cas des maladies génétiques « mendéliennes » ou monogénétiques, dans lesquelles l'affection est due à l'altération d'un gène précis, comme c'est le cas pour l'hémophilie, la mucoviscidose, ou la myopathie de Duchenne. L'altération de l'ADN transmise par un parent, ou apparue en raison d'une nouvelle mutation, entraîne alors la production d'une protéine défectueuse qui ne joue pas son rôle dans l'organisme et aboutit à une pathologie bien définie. Ces maladies sont génétiquement déterminées en ce sens que toute personne portant la mutation en cause sera atteinte par l'affection ; elles sont aussi rares car leurs porteurs, lourdement handicapés, sont défavorisés par la sélection naturelle. Mais beaucoup d'affections moins graves et plus répandues, comme le diabète, l'hypertension artérielle ou la maladie de Crohn⁷, présentent une composante génétique significative : le risque de souffrir de diabète, par exemple, dépend fortement des variants de gènes portés par une personne. Cette composante génétique implique en général des dizaines ou des centaines de gènes, chaque variant individuel ayant un effet très faible sur le risque ; celui-ci est également fortement dépendant de l'environnement (régime alimentaire et obésité pour le diabète). Enfin, le risque de cancer est lui aussi influencé par la génétique, comme pour le cancer du sein (les gènes BRCA1 et BRCA2, dont les mutations peuvent fortement élever le risque) mais aussi pour bien d'autres (prostate notamment). Chacun de nous porte donc un assortiment de variants de gènes qui peut, par exemple, lui conférer un risque de diabète très inférieur à la moyenne mais un risque plus élevé de cancer de la prostate.
- 19 On peut alors s'intéresser aux différences d'incidence de diverses maladies selon le groupe humain envisagé. Ces différences sont une réalité en ce qui concerne les affections génétiques mendéliennes : la fréquence de plusieurs de ces maladies diffère entre les groupes, ce qui reflète le passé de ces populations : la mucoviscidose (inactivation du gène codant pour un « canal ionique ») est surtout présente chez les Européens du Nord et leur descendance aux États-Unis ; l'anémie falciforme (une mutation du gène de l'hémoglobine) se retrouve surtout chez les Africains. Dans ce dernier cas, on s'explique cela par le fait qu'à l'état hétérozygote⁸ cette mutation protège contre l'infection par *Plasmodium falciparum*, agent de la malaria qui a sévi essentiellement dans les régions chaudes et humides, elle n'est pathogène qu'à l'état homozygote⁹. On a de même proposé que la mutation qui provoque la mucoviscidose à l'état homozygote protège des diarrhées infantiles (autrefois cause de nombreux décès) à l'état hétérozygote. Pour d'autres maladies génétiques, il semble s'agir d'une mutation relativement récente (quelques milliers d'années) qui est restée confinée au sein d'un groupe humain géographiquement ou culturellement isolé, comme la maladie de Tay-Sachs chez les Juifs ashkénazes. Les exemples sont nombreux et concernent d'ailleurs aussi des mutations adaptatives qui ne sont pas pathogènes, comme la capacité à digérer le lait à l'âge adulte (persistance de l'activité de la lactase), majoritaire chez les Européens et chez quelques groupes Africains de tradition pastorale.

- 20 Pour les maladies plus fréquentes, influencées par de nombreux gènes et par les conditions d'existence, la situation est plus complexe. De nombreuses études menées aux États-Unis (du fait de l'existence de statistiques ethniques) indiquent que l'hypertension, le cancer de la prostate, le cancer du poumon sont notablement plus courants chez les Afro-Américains que chez les « Caucasiens »¹⁰, ces derniers étant au contraire plus souvent affectés par l'ostéoporose ou les cancers cutanés. Mais les Afro-Américains comptent proportionnellement plus d'obèses et plus de fumeurs, ce qui accroît la probabilité d'hypertension et de cancer du poumon ; les Blancs, eux, ont une espérance de vie plus élevée, donc plus de risques d'atteinte par une maladie de la vieillesse comme l'ostéoporose... L'environnement de ces deux groupes diverge donc de manière significative, ce qui rend difficile l'affirmation d'un rôle des gènes – sauf pour les cancers cutanés, liés à l'exposition solaire d'une peau claire dont la couleur est génétiquement déterminée. On a naturellement tenté de tenir compte de ces facteurs socio-économiques, mais c'est très difficile : l'accessibilité aux soins, mais aussi la propension à s'adresser à un médecin, le choix d'un certain type d'approche thérapeutique, sont eux aussi influencés par l'appartenance culturelle, même à niveau économique comparable. En ce sens on peut dire que l'appartenance à une catégorie sociale (ce que recouvre en réalité le terme de « Afro-Américain ») a une certaine valeur prédictive pour différentes affections – mais rien ne prouve que cette prédiction soit liée à la constitution génétique de cette catégorie, qui est d'ailleurs génétiquement hétérogène dans la plupart des cas.
- 21 En résumé, les différences de fréquence de maladies entre groupes humains socialement définis (souvent appelés « groupes ethniques », pour éviter de dire « races ») sont bien réelles, mais, mis à part quelques cas d'affections mendéliennes touchant des populations bien particulières, rien ne prouve que ces différences soient d'origine génétique. C'est même assez peu probable compte tenu de la diversité génétique de certains groupes, comme celui des Afro-Américains (Figure 2).

4. Ambigüité de la « race », réalité des groupes d'ascendance

Un concept bâtard et variable

- 22 L'appartenance à une « race » socialement définie repose largement sur des caractères visibles (couleur de la peau, des yeux, des cheveux, forme du visage) qui sont génétiquement déterminés même si nous ne savons pas encore les déduire tous de l'analyse de l'ADN. Mais cette appartenance est aussi fortement influencée par l'histoire de la société à laquelle appartient l'individu : c'est en vertu de la *one-drop rule*¹¹ que Barack Obama est considéré comme le premier président Noir des États-Unis, alors qu'il est autant Blanc (par sa mère) que Noir (par son père). Elle dépend aussi du « récit » qu'adopte un groupe social qui va se définir lui-même comme un groupe racial alors que, comme pour les soi-disant Aryens du III^e Reich, cette catégorie qui se prétend biologique correspond seulement à un mythe sur son origine. Et bien sûr, ce concept a une histoire dramatique : largement invoqué pour justifier la pratique de l'esclavage et les entreprises de colonisation (le devoir des races supérieures de civiliser les races inférieures, selon Jules Ferry – en accord avec la plupart des penseurs de son époque), consubstantiel plus tard au fascisme de Mussolini et surtout au nazisme, sous-jacent au

génocide Arménien, à celui du Rwanda et bien sûr à la Shoah, sa charge historique résolument négative, liée aux épisodes les plus sombres du passé récent de l'humanité, disqualifie désormais son usage. Il me semble donc légitime de bannir ce terme qui est imprécis, susceptible d'interprétations très diverses, et largement pollué par l'histoire. En ce sens, on peut dire que « les races humaines n'existent pas », si l'on attribue au terme le sens que lui donnait Gobineau et bien d'autres, mais on ne peut pas nier l'existence de groupes humains définis par leur ADN, reflétant l'histoire de notre espèce et présentant des différences significatives (Reich 2018).

« Groupe d'ascendance », une solution ?

- 23 Comment donc désigner ces groupes dont l'existence objective est avérée (Figures 1 et 2), en écartant le terme de « groupe ethnique » qui n'est souvent qu'un euphémisme pour dire « race » et qui d'ailleurs, si on le prend à la lettre, désigne une culture plutôt qu'une entité biologique ? Le terme de groupe d'ascendance n'est pas (encore) courant, mais il a le grand avantage de renvoyer à une définition précise : un ensemble de personnes génétiquement proches en raison de leur origine commune, souvent rattachée à une région de la Terre où ce groupe a été relativement isolé durant quelques milliers d'années. Au fil des siècles, ce groupe a (un peu) évolué grâce aux mécanismes déjà évoqués (sélection, dérive génétique, sélection sexuelle) et a acquis certains caractères communs qui, au moins au niveau de l'apparence, correspondent plus ou moins aux stéréotypes « raciaux ». Mais au-delà de cette correspondance, les groupes d'ascendance ont une réalité objective, et leur origine géographique peut souvent être évaluée grâce à la comparaison de leurs profils génétiques avec des profils de référence établis grâce à l'étude de populations ancestrales. Populations ancestrales : il s'agit là de groupes humains dont l'histoire est suffisamment bien connue pour affirmer qu'ils sont restés statiques et sans mélanges significatifs depuis quelques siècles ou, de plus en plus, de groupes aujourd'hui disparus mais répertoriés grâce à l'ADN extrait de vestiges ou d'ossements, et dont l'analyse détaillée est récemment devenue possible (Reich 2018). La définition des groupes d'ascendance est objective, contrairement à leur rattachement à des populations anciennes qui dépend beaucoup de l'effectif et de la qualité des groupes de référence. Les entreprises comme *23andMe*, qui se sont spécialisées dans l'établissement de tels profils pour une somme modique (99 dollars américains), fournissent à leurs clients des schémas très détaillés (Figure 4) qu'il ne faut pas prendre trop au sérieux lorsqu'ils indiquent, par exemple, 14,6 % d'ascendance amérindienne en oubliant de donner la marge d'incertitude, mais qui sont valables dans leurs grandes lignes : la personne représentée sur la figure 4 a bien comme origines majoritaires (dans l'ordre) l'Afrique subsaharienne, l'Europe et l'Asie (dont proviennent les populations amérindiennes).

Figure 4. Exemple de résultat fourni par l'entreprise 23andMe après analyse de l'ADN d'une personne (« Monica De Armond »). La mappemonde au centre indique à quoi correspondent les couleurs ; la décomposition en grands groupes d'ascendance continentaux est montrée sur le premier cercle concentrique (les blancs correspondent aux composants non identifiés, *unassigned*). Sur le deuxième cercle concentrique sont indiquées les origines plus précises, quand elles ont pu être définies, de même pour le troisième cercle. La précision apparente des valeurs présentées (14,6 % d'origine amérindienne, par exemple) est illusoire car les marges d'erreur sont importantes. Extrait du site de l'entreprise.

- 24 Contrairement à « groupe ethnique », « groupe d'ascendance » n'est pas un substitut politiquement correct au terme de « race humaine ». Ces races, dans leur acception classique, sont des entités homogènes et clairement distinctes : comme l'écrivait Gobineau au milieu de XIX^e siècle (1853), l'humanité est *divisée* en races – et celles-ci sont inégales, la race blanche étant bien sûr supérieure à toutes les autres. Et la définition de la race, bien qu'elle se prétende fondée sur la biologie, est fortement influencée par le comportement, les coutumes, l'habillement... la culture en somme, comme l'a bien démontré Claude Lévi-Strauss (2001). Les groupes d'ascendance, eux, apparaissent de manière objective à partir de la seule analyse de l'ADN, et ne sont rattachés à une origine géographique que secondairement et par des méthodes toujours ouvertes à la critique et très dépendantes de la qualité des groupes de référence utilisés. Et, comme nous l'avons vu, si le phénotype visible est assez largement lié au groupe d'ascendance, si quelques affections, quelques caractères métaboliques peuvent être quasiment spécifiques à certains groupes (où ils sont apparus à la suite de mutations fréquentes), la plupart des éléments importants sont influencés à la fois par la génétique et par l'environnement, ce qui rend leur analyse causale extrêmement difficile. Les techniques d'analyse du génome continuent à progresser, et l'on devient maintenant capable d'évaluer l'effet d'un grand nombre de variants de gènes sur un caractère phénotypique donné, mais il restera toujours très difficile de départager la part de la génétique et celle de l'environnement dans l'évolution d'un individu et plus encore d'un groupe, celui-ci fût-il défini par une analyse objective.

En somme...

- 25 L'information fournie par l'analyse des ADN humains éclaire nombre de questions restées en suspens. Elle révolutionne l'anthropologie en révélant, grâce à l'étude des ADN anciens, la complexité des migrations humaines et des mélanges de population au cours des derniers millénaires de notre histoire (Reich 2018). Comme j'espère l'avoir montré dans les pages qui précèdent, elle bat en brèche la notion de race qui a largement dominé en Occident jusqu'à la moitié du XX^e siècle tout en nuancant l'affirmation un peu simpliste selon laquelle « les races humaines n'existent pas ». Oui, une analyse sans *a priori* des ADN humains montre bel et bien l'existence objective de groupes plus apparentés génétiquement entre eux qu'avec les autres individus ; et oui, les groupes ainsi définis correspondent souvent, bien qu'assez approximativement, aux catégories « raciales » (« ethniques » pour être politiquement correct) qui reposent sur l'aspect mais aussi sur l'appartenance culturelle des personnes concernées. Mais les frontières de ces groupes sont floues, leur diversité interne est élevée : ils n'ont pas le caractère exclusif, fixe et héréditaire de l'ancienne notion de race. C'est le reflet de l'histoire de notre espèce qui, bien que jeune (deux ou trois cent mille ans, à l'échelle de l'évolution, c'est peu de chose), a derrière elle une histoire complexe faite d'évolution différentielle selon les régions du globe et leurs caractéristiques, mais aussi de migrations et de mélanges de populations qui ont rebattu les cartes en associant des caractères initialement apparus dans des groupes distincts. Ce grand métissage, qui se poursuit et s'amplifie aujourd'hui, ne tend pas vers l'uniformité mais au contraire répand les différents allèles de multiples gènes, favorisant ainsi la résilience de notre espèce qui en aura bien besoin face aux défis qui l'attendent...

BIBLIOGRAPHIE

- 1000 Genomes Project Consortium. 2012. « An integrated map of genetic variation from 1,092 human genomes. » *Nature* 49 : 56-65.
- De Gobineau, Joseph-Arthur. 1853. *Essai sur l'inégalité des races humaines*. Paris : Firmin Didot Frères.
- Jordan, Bertrand. 2008. *L'humanité au pluriel, la génétique et la question des races*. Paris : Éd. du Seuil.
- Jordan, Bertrand. 2020. « ADN et portrait-robot : où en est-on ? » *Med Sci (Paris)* 36 : 813-816.
- Lévi-Strauss, Claude. 2001. *Race et Histoire* [1952] ; *Race et Culture* [1971]. Paris : Albin Michel/Unesco.
- Reich, David. 2018. *Who We Are and How We Got Here: Ancient DNA and the New Science of the Human Past*. Oxford : Oxford University Press.

NOTES

1. Pour une bonne discussion chiffrée des différences entre individus, voir : <http://book.bionumbers.org/how-genetically-similar-are-two-random-people/>.
  2. https://fr.wikipedia.org/wiki/Analyse_en_composantes_principales.
  3. Les Amérindiens (non représentés sur cette figure) sont génétiquement assez proches des Asiatiques, qui sont à l'origine du peuplement des Amériques il y a environ quinze mille ans. L'étalement des points correspondant aux Mexicains et Colombiens traduit leur ascendance mixte, amérindienne et européenne.
  4. Selon le titre de l'exposition qui a eu lieu au musée de l'Homme de 1990 à 2015.
  5. Comme la mutation responsable de la maladie de Huntington.
  6. Discours à l'Assemblée nationale, le 28 juillet 1885.
  7. Inflammation chronique qui touche les parois du tube digestif.
  8. Gène muté sur un chromosome, normal sur l'autre.
  9. Mutation sur les deux chromosomes.
  10. Européens, Blancs...
  11. *A white person is one with no trace whatever of blood from another race*, formulaire de recensement de l'État de Virginie en 1924.
-

RÉSUMÉS

Le séquençage du génome humain (2003) a été suivi de progrès techniques qui permettent aujourd'hui la caractérisation détaillée (y compris la séquence intégrale) de l'ADN de milliers d'individus. Cela a fourni une estimation de la diversité génétique humaine, soit environ 3 millions de différences dans notre génome qui contient 3 milliards de bases. Bien que la divergence entre deux individus soit faible (environ 0,1 %), elle est responsable de la majeure partie de la diversité phénotypique observée au sein de la population humaine. Les techniques actuelles permettent de mesurer la distance génétique entre deux individus dont l'ADN a été analysé. En appliquant des méthodes de représentation multidimensionnelle, les résultats obtenus pour un grand ensemble d'individus montrent l'existence de groupements (*clusters*) regroupant des personnes plus similaires entre elles qu'avec des personnes d'autres groupes. Bien que ces groupes aient été définis uniquement par l'analyse moléculaire de l'ADN, ils s'avèrent correspondre approximativement aux groupes dits « ethniques » (par exemple Européens, Africains et Asiatiques). La diversité interne de chaque groupe est importante, mais l'analyse de milliers de marqueurs génétiques permet de les distinguer. Le terme de « Race », tels qu'il est généralement compris, ne correspond pas à la réalité de ces groupes : il suggère une forte homogénéité à l'intérieur de chaque groupe, implique des notions de comportement et de « caractère » sans fondement scientifique, et est marqué par son utilisation ancienne en tant que justificatif d'oppression et même de génocide. « Ascendance » ou « Groupe d'ascendance » sont des termes bien préférables, d'autant plus qu'aujourd'hui beaucoup de personnes présentent un mélange d'ascendances. Les méthodes décrites ici ont été largement appliquées aux populations actuelles et à de l'« ADN ancien » provenant de restes humains ou même de fossiles, et ont permis de grands progrès dans la compréhension de l'histoire de notre espèce.

The sequencing of the human genome (2003) has been followed by a number of technical developments that allow detailed characterization (including complete sequencing) of the DNA of thousands of individuals. This has provided an estimate of human genetic diversity: approx. 3 million base substitutions within our genome that includes 3,000 million bases. Although the divergence between any two individuals is small, it is responsible for much of the phenotypic diversity observed within the human population, and current techniques make it possible to measure genetic distances between any two individuals whose genome has been analysed. When results from a large sample of persons are analysed by sophisticated multi-dimensional representation, fairly distinct clusters appear in the population, indicating groups of individuals that are more similar to each other than to persons from other groups. Although these groups are defined by genome analysis, without any *a priori* information on their origin, it turns out that they largely correspond to so-called “ethnic” categories that reflect genetic ancestry (e.g. Europeans, Africans and Asians). Every group is internally quite diverse, but analysis of thousands of genetic markers does indeed differentiate between them. “Race”, as commonly defined, is a very imperfect proxy for these ancestry groups, as it suggests high homogeneity within each group, implies assumptions about behaviour and “character” that have no scientific basis, and is heavily tainted by past uses in support of oppression and even genocide. Thus “ancestry” or “ancestry group” is a highly preferable term, especially now that more and more people are of mixed ancestry. These methods have been widely applied to current populations and to “ancient DNA” from historical samples, and have made possible great strides in understanding the history of our species.