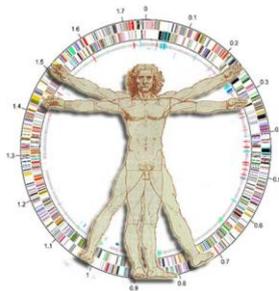


## La Saga du Génome Humain a 20 ans



**Auteurs : Fredj Tekaiia et Abdellatif Boudabous**

### Résumé :

Depuis les publications des premières versions du génome humain, le développement de la génomique humaine a fait des avancées extraordinaires dans la connaissance de l'origine et la caractérisation des maladies génétiques ainsi que dans l'identification des gènes, des mutations et des variations génomiques qui contribuent au développement des maladies complexes (cardiaques, neurologiques, cancers, diabète,...). De grands projets emblématiques de séquençage pour l'étude de ces nombreuses maladies ont permis d'établir des catalogues de variations génétiques et leurs éventuelles conséquences phénotypiques et cliniques à partir des génomes de populations de descendance essentiellement Européenne. Ces études ont néanmoins pointé la nécessité de couvrir toute la diversité génétique humaine, afin d'établir des catalogues universels.

Un autre développement spectaculaire concerne les génomes humains anciens obtenus à partir d'échantillons archéologiques (cheveux, fragment d'os,...) qui ont ouvert une ère nouvelle de la lecture de l'évolution de l'homme et de l'histoire de la migration d'*Homo sapiens* à partir de l'Afrique il y a environ 70000 ans, pour peupler le monde, s'adapter et se mélanger avec les populations locales en particulier avec le Neandertal (en Europe) et le Dénisovien (en Sibérie). Des legs génétiques du Neandertal et du Dénisovien, conséquences de ces mélanges, impactent encore aujourd'hui la biologie de certaines populations par des répercussions favorables ou défavorables sur leur santé et leur mode de vie.

Pour conclure, nous proposons d'intégrer ces avancées technologiques et scientifiques dans l'Enseignement et la Recherche Scientifique en Tunisie. Nous faisons nôtre la devise d'un programme stratégique canadien d'intégration des connaissances génomiques dans la santé de précision :

«Think Big, Start Small, Learn Fast».

Notez que nous avons tenu à inclure un nombre important de références bibliographiques afin de permettre au lecteur intéressé d'approfondir les riches informations que nous rapportons.

Il y a 20 ans (15 Février 2001), les deux premières versions des séquences du génome humain ont été simultanément publiées par les journaux *Nature*, pour le projet public [1] et *Science*, pour le projet privé de Celera Genomics [2], aboutissement d'un programme lancé en 1990. Les deux versions du génome ont été séparément générées à partir de donneurs volontaires anonymes de différentes origines ethniques.

Ces deux journaux consacrent des articles [3,4] à cette épopée pour célébrer les avancées extraordinaires initiées par le séquençage du génome humain.

Dans la version du Consortium international, il était indiqué que la séquence produite correspondait environ à 94% du génome total et incluait de 30000 à 40000 gènes codants pour des protéines. La version Celera correspondait à environ 95% (2.91Gb) du génome total et incluait 39114 gènes codants pour des protéines. Ce nombre de gènes, environ le double de ce qu'on avait trouvé dans le génome du verre de terre *Caenorhabditis elegans* déjà séquencé en 1998, avait alors surpris plus d'un scientifique qui pensait que la complexité de la machinerie de l'homme nécessitait un nombre beaucoup plus grand que celui du petit verre de terre.

Après ces deux premières versions, il était évident qu'il fallait construire un génome humain de référence [5] aussi complet que possible, ce qui fut fait en 2004 mais avec 300 trous (Gaps ou fragments incomplets, non séquencés). Ce génome de référence est depuis continuellement complété et mis à jour avec de nouvelles données. La version actuelle du génome de référence humain GRCh38.p13 est composée de 3.27 Giga bases (Gb) et 19116 gènes codants pour des protéines. Ce qui est remarquable, c'est que le nombre de gènes a vertigineusement chuté par rapport aux premières estimations et les indications des deux premières publications. Le nombre exact des gènes du génome humain n'est d'ailleurs toujours pas connu à ce jour. On sait aujourd'hui que selon l'épissage des exons, plusieurs gènes codent pour plusieurs protéines à la fois et contrôlent plusieurs fonctions, ce qui réduit le nombre final de gènes humains et en plus rend difficile son établissement.

La séquence du génome humain a marqué la recherche scientifique en ouvrant une ère nouvelle de séquençage touchant tous les organismes vivants ou éteints. Ces publications sont le début d'une ère génomique qui a transformé le paysage de la biologie, de la médecine et de l'histoire de l'évolution de l'homme. Ces développements ont été facilités par les avancées technologiques et la chute du coût de séquençage [6] ainsi que par les avancées extraordinaires de la Bioinformatique qui offrent de nouvelles méthodes et des outils permettant d'analyser les données et de répondre aux questions de plus en plus pointues posées par l'accumulation de masses des données génomiques [7] relatives à différents organismes.

Nous nous limitons ici à commenter les avancées concernant l'homme, uniquement sur le plan de la santé et de son évolution depuis le départ d'*Homo sapiens* de l'Afrique.

## **Le Génome Humain et la Santé**

Les variations génétiques sont connues pour être la source de maladies héréditaires et de maladies complexes comme les différents types de cancer, le diabète, les maladies cardiaques ou neurologiques [8]. Par exemple la maladie Charcot-Marie Tooth (maladie qui touche les nerfs périphériques) est due à une

insertion ; le syndrome de William-Beuren (anomalie de développement), la dystrophie musculaire de Duchenne, sont tous deux dus à une délétion ; les Aphasies Primaires Progressives (APP) dans la maladie d'Alzheimer sont dues à une duplication dispersée (dispersed duplication) ; le syndrome de Down est dû à une translocation ; la maladie de Hunter est due à une inversion et la maladie de Huntington est due à une duplication en tandem. Il a été aussi démontré que l'autisme qui est caractérisé essentiellement par un déficit de communication, d'interaction sociale et d'isolement, est d'origine génétique due à l'héritage et au développement de variant *de novo*. Les répétitions en tandem de séquences de 1 à 20 nucléotides répétées consécutivement sont la source majeure des variants *de novo* chez l'homme. Il a été noté qu'un excès de ce type de variant est corrélé avec l'autisme [9].

Les efforts récents pour caractériser les variations génétiques dans le génome humain ont permis le développement de nouvelles approches pour l'identification des gènes et des mutations qui contribuent au développement des maladies complexes. Afin de permettre l'étude de ces maladies, plusieurs projets «1000 projets génomes humains » nationaux et internationaux ont été mis en place (en réalité contrairement à ce que laisse penser la dénomination, beaucoup plus que 1000 individus ont été considérés dans ces programmes). Nous citons dans ce qui suit, certains de ces projets emblématiques. Les variations du génome humain sont aussi bien situées dans les exons (régions codantes 3 à 10%) que dans les régions non codantes (introns 90 à 95%). Il est encore difficile de prévoir toutes les implications des mutations, insertions, délétions et recombinaisons des gènes humains. Le séquençage d'un grand nombre de génomes et l'amélioration des algorithmes d'analyses permettront une meilleure exploitation des génomes.

-En 2007, un consortium «HapMap project » a publié une première carte génétique [10] incluant 3 Millions de SNPs (mutations) découverts dans 270 individus d'origines diverses (Japon, Chine, USA et Nigéria). Cette étude a montré comment notre ADN est hérité par blocs (groupes ou paquets de gènes) et comment ces blocs varient à l'intérieur d'une même population et entre populations.

-Le «Humain 1000 genomes project (<https://www.internationalgenome.org/data/>)» commencé en 2006 et terminé en 2015, a abouti à l'établissement d'un catalogue de la variabilité génétique humaine (rare et commune), en séquençant les génomes de 2504 humains sélectionnés à travers le monde (26 groupes à travers les 5 continents) [11]. Ce catalogue inclut : 88 Millions de variants dont 84.7 Millions de SNPs, 3.6 Millions de courtes insertions/délétions (indels) et 60000 variants structuraux.

-Le « International Cancer Genome Consortium » (<https://icgc.org/>) est constitué de 90 équipes internationales de recherche qui s'occupent de 84 types de tumeurs différentes. Il a pour objectif de séquencer 25000 génomes de cancers non traités (initiative 25K). Un autre consortium « Pan-Cancer Project » s'occupe de l'analyse de ces données dont certaines ont été récemment publiées [13].

-Une base de données « gnomAD » (<https://gnomad.broadinstitute.org>) agrège environ 17000 génomes complets et environ 126000 exomes (les parties codantes des génomes) afin d'étudier les variants génétiques qui inactivent les gènes codants

pour des protéines et qui impliquent donc des conséquences phénotypiques. L'analyse de ces données a permis de constituer un catalogue des différents types de variation, révélant ainsi leur impact fonctionnel et leurs utilités cliniques pour les diagnostics médicaux afin d'identifier les mutations qui causent des maladies et pour guider dans le choix de protocoles thérapeutiques.

- Un grand programme de séquençage de «1 Million Genomes Initiative (1+MG) » (<https://b1mg-project.eu>) est un accord entre 23 pays Européens pour réaliser le séquençage d'un million de génomes complets d'ici 2022. Ce projet fait partie d'un projet plus large « Beyond 1 Million Genomes » pour créer un réseau de données génétiques et cliniques à travers l'Europe.

Il est important de signaler à ce stade que la découverte de variants inconnus à partir d'étude de populations Africaines et Asiatiques (en particulier de l'Océanie) montre la limite de l'utilité des découvertes à partir de génomes centrés sur des individus d'origine Européenne. En effet l'utilité clinique des variants observés dans une population peut s'avérer inadéquate pour d'autres populations.

Plusieurs études récentes montrent que ces catalogues sont loin de couvrir la totalité de la diversité humaine. Par exemple une étude de 300 génomes provenant de 142 populations différentes [12] a montré qu'une proportion de variation (11% KhoeSan (habitants du Désert du Kalahari en Afrique du Sud) et 5% nouveaux Guinéens et d'Australiens, ne se trouve pas dans les catalogues actuels de variants humains.

Un autre exemple est donné par une étude africaine récente qui concerne le séquençage complet de 426 individus Africains sub-sahariens appartenant à 50 groupes ethnolinguistiques différents a permis d'explorer la diversité génétique à l'échelle africaine sub-saharienne et de découvrir 3 Millions de nouveaux variants (inconnus à ce stade) ainsi que 62 nouvelles localisations chromosomiques soumises à une pression de sélection incluant des gènes impliqués dans l'immunité virale, la réparation de l'ADN et le métabolisme [14]. Cette étude n'inclut malheureusement aucun échantillon provenant de l'Afrique du nord (d'Est en Ouest).

Des développements qui visent à construire un génome de référence régional (ou même par pays) sont en cours par exemple :

-Le projet «The GenomeAsia 100K project» (<https://genomeasia100k.org>). Ce projet se concentre sur les génomes asiatiques afin d'augmenter la diversité génétique et de chercher des spécificités à la population asiatique. Une première analyse a été publiée [15]. Elle est relative à 1739 individus appartenant à 219 groupes de populations provenant de 64 pays à travers l'Asie et qui a abouti à l'établissement d'un catalogue des variations génétiques et des associations génotypiques et phénotypiques.

Plusieurs autres grands projets de séquençage à grande échelle de génomes humains sont en cours dont par exemple en Corée

([https://en.wikipedia.org/wiki/Korean\\_Genome\\_Project](https://en.wikipedia.org/wiki/Korean_Genome_Project)) et en Chine avec le génome des Hans (<https://www.hanchinesegenomes.org/HCGD/index>).

-En Tunisie bien qu'elle ne considère pas de génomes complets, une étude récente a permis d'établir une carte de 1083 CNVs (Copy Number Variations) s'étendant sur plus 61 Mb du génome humain, dans une population constituée de 102 Tunisiens en

bonne santé. Ces résultats constituent un apport à la diversité génétique humaine et sont utiles en médecine génomique [16].

Notons que 20 ans après la publication du premier génome humain, la dernière version du génome humain de référence publié par le « Genome Reference Consortium » (GRCh38:[https://www.ncbi.nlm.nih.gov/assembly/GCF\\_000001405.39](https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.39)) ne contient aucun chromosome complètement séquencé de bout en bout. Il manque encore environ 100 trous ou gaps qui totalisent environ 150 Mb de séquences nécessaires pour les combler. Ces séquences manquantes incluent des répétitions dans les centromères et des zones de duplication segmentaire qui peuvent receler des variations inexplorées aux conséquences inconnues. Un consortium (Telomere-to-Telomere : <https://sites.google.com/ucsc.edu/t2tworkinggroup/home>) se propose de combler la séquence et de fournir un génome humain complet.

Rappelons que le génome est à 99.9% commun à tous les humains. Le reste (0.1%) est dû à des différences et des spécificités de populations ayant des ancêtres différents. Le 0.1% de différence correspond à 3 Millions de sites où les génomes individuels peuvent varier. De plus la distribution de ces sites sur le génome n'est pas aléatoire. Certaines de ces variations sont anodines et ne causent aucun problème, d'autres peuvent être dommageables pour leurs porteurs.

Ainsi les résultats cliniques relatifs aux analyses des variants en rapport avec les maladies ne concernent pas forcément toute l'humanité. Une étude a montré en se basant sur environ 6000 individus appartenant à 21 origines ancestrales qu'en moyenne chaque individu a 4 origines ancestrales différentes, démontrant par là même que l'origine ancestrale est corrélée avec la langue et que la race n'est pas une classification génomique objective [17].

Il est donc nécessaire d'avoir plusieurs références incluant toutes les diversités humaines. Un nouveau projet appelé «Human Pangenome Project» se propose de séquencer complètement (Telomere-to-Telomere) les génomes de 300 humains.

-Le « Human Pangenome Reference Consortium » (<https://humanpangenome.org>) a entre autres objectifs, la création d'un génome humain de référence qui inclut la diversité génétique humaine et qui complète et corrige la référence actuelle (Suivre les publications correspondantes : <https://humanpangenome.org/publications/>).

-Dans les publications célébrant le 20<sup>ième</sup> anniversaire du génome humain, un chercheur Africain propose de séquencer 3 Millions d'Africains [18] et d'autres appellent à augmenter la diversité afin d'universaliser le génome humain de référence et les banques de variants génomiques [19]. Moins de 2% des génomes séquencés à ce stade proviennent des populations africaines, et ceci malgré sa grande diversité génétique comparée à d'autres continents [18,19]. De même les populations indigènes en Asie devraient être prises en considération en s'impliquant dans le séquençage de leurs génomes. Les bénéfices cliniques des analyses actuelles des génomes humains risquent de ne pas profiter à tous les humains, tant qu'on n'a pas pris en compte l'entière diversité humaine.

Le génome nous renseigne non seulement sur notre état génétique actuel mais aussi sur notre passé lointain préhistorique et historique. Le séquençage de génomes complets a ouvert la voie au séquençage de l'ADN ancien chez des organismes éteints particulièrement chez l'homme ancien [20].

## Le Génome Humain et l'Évolution de l'Homme

La possibilité de séquencer des génomes à partir de matériels biologiques anciens a fourni une source d'information riche pour la biologie évolutive et a éveillé un intérêt public considérable pour l'évolution de l'homme et les organismes anciens y compris les espèces éteintes à partir de l'analyse de l'ADN ancien (aDNA). La génomique ancienne s'est spectaculairement développée durant cette dernière décennie grâce aux développements technologiques du séquençage et des techniques de laboratoire permettant d'extraire de l'ADN exploitable à partir d'échantillons archéologiques anciens (cheveux, fragments d'os,...) ainsi qu'au fait de pouvoir les authentifier et d'éviter les contaminations par du matériel moderne. Le séquençage des fragments archaïques a permis de reconstituer les génomes d'espèces éteintes ou archaïques et a permis de réévaluer l'histoire de l'évolution de l'homme, de son émigration pour peupler le monde ainsi que la diversité des populations et leurs environnements. Avec le séquençage de l'ADN ancien, on a pu reconstruire l'histoire de l'adaptation de l'homme aux conditions rencontrées dont par exemple, le froid, la domestication d'animaux, la période et le lieu des événements de mélange entre anciens humains : Neandertal, Denisovien et *Homo sapiens* entre autres. Différentes études montrent que Neandertal et Denisovien sont nos cousins les plus proches.

Le séquençage de l'ADN ancien a aussi permis de reconstituer les génomes d'anciennes bactéries et de retrouver les pathogènes anciens responsables de maladies et d'épidémies dévastatrices qui ont eu lieu dans l'histoire.

Les travaux de reconstruction des génomes archaïques et anciens sont parmi les plus intéressants à la fois sur le plan scientifique et intellectuel. Ils ont permis de mener des recherches fascinantes sur la reconstitution de l'évolution de l'homme et de retracer ses parcours pour le peuplement du monde, donnant un degré de précision sans précédent aux datations basées sur les variations génétiques des différents épisodes de l'histoire de l'homme tels que découverts par les Anthropologues. Ces datations sont susceptibles de changer au gré de nouvelles découvertes et des précisions de calcul.

Les données fossiles pointent vers une origine Africaine d'*Homo sapiens* («Homme sage» ou «Homme moderne»): Jebel Irhoud [21] au Maroc (315000 ans ou 315 ka), Florisbad en Afrique du Sud [22] (260 ka), en Éthiopie Kibish [23] (195 ka) et Herto [24] (160 ka).

*Homo sapiens* est apparu pour la première fois en Afrique il y a plus de 200000 ans. Il est sorti d'Afrique il y a 70000 à 50000 ans en prenant probablement différentes directions. Il a commencé sa migration par le Moyen Orient ensuite vers le sud de l'Asie et vers l'Océanie il y a environ 50000 ans. Il a émigré vers l'Europe il y a environ 40000 ans [25]. D'autres directions possibles passent par l'Afrique du Nord.

Le premier génome humain ancien séquencé est celui d'un Paléo-Eskimo qui vivait il y a 4000 ans et appartenait à la culture Saqqaq (développée au village de Saqqaq qui se situe sur la côte Sud-ouest du Groenland sur la période comprise entre 2500 et 800 avant JC). Le séquençage a été possible à partir de cheveux conservés dans le permafrost [26].

*Homo sapiens* s'est mélangé avec l'Homme de Neandertal en Europe et l'Homme de Denisova (Sibérie) en Asie. Des fragments archaïques d'hominidés trouvés dans des

grottes à Denisova (Sibérie), à Vindija (Croatie) et en Altaï (Sibérie) ont permis de reconstruire les génomes de l'homme de Neandertal et de Denisova (en fait une femme).

Le premier génome du Paléo-Eskimo a été suivi par les premiers génomes séquencés du Neandertal [27] à partir de restes trouvés dans la cave de Vindija (Croatie) et datant d'il y a 33-44000 années et celui de Denisova [28]. Plus tard un autre génome d'un hominidé Neandertal de l'Altaï (Sibérie) a été séquencé [29]. Ces génomes constituent les références utilisées pour toutes les comparaisons à l'échelle génomique ancienne ou moderne concernant l'évolution de l'homme.

Le plus ancien génome humain séquencé à ce jour est celui du Neandertal de Sima de Los Huesos (Espagne) et date de 430000 années [30]. Notez que les génomes les plus anciens à avoir été séquencés à ce jour sont ceux de deux mammouths récemment publiés et datent respectivement de 1.65 Millions d'années (Ma) et de 1.34 Ma [31], détrônant le record jusqu'alors détenu par celui d'un cheval ancien qui date de 700000 années [32], tous conservés dans le permafrost sibérien.

Les comparaisons des génomes de l'homme moderne aux génomes anciens montrent que tous les non-Africains modernes possèdent des legs du Neandertal plus développés chez les non-Africains de l'Est ainsi que chez les Australo-Mélanésiens et qu'à un degré moindre les non-Africains de l'Est possèdent des legs du Denisovien [12].

Plus de 2000 génomes archaïques humains et d'individus hominidés sont disponibles [33]. Ainsi, on a pu définir que la zone de mélange du Neandertal est l'Europe alors que celle du Denisovien est l'Asie. Neandertal et Denisovien sont éteints il y a environ 40000 ans [34], mais ils continuent à avoir un impact biologique (protecteur ou nuisible) sur les humains actuels non-Africains, pour lesquels ils ont pu léguer un héritage génétique. On estime que 1-2% du génome actuel Européen est hérité du génome du Neandertal. Une étude récente concernant environ 28000 génomes complets islandais a révélé la nature de l'introggression du génome du Neandertal dans leurs génomes en détaillant la distribution de fragments génomiques archaïques sur chacun des chromosomes actuels [35]. Concernant l'impact de l'héritage du Neandertal, l'analyse des données génétiques et phénotypiques de participants à la banque UK Biobank [36] ayant hérité des legs génomiques du Neandertal montre que ceux-ci portent certaines caractéristiques phénotypiques comme la prédisposition au coup de soleil, au couche-tard, à la dépression ou au sentiment de solitude.

Plus récemment, une étude concernant des malades du Covid-19 a montré l'existence d'un cluster de gènes d'environ 50 kb sur le chromosome 3, qui prédispose ces patients à une défaillance respiratoire sévère lors de leur infection par le SARS-CoV-2. Ce cluster est porté par environ 65% de la population de l'Asie du sud, environ 16% de la population Européenne et il est absent en Asie de l'Est. Il a été démontré que ce segment chromosomique a été hérité du Neandertal [37]. Plus récemment encore, les mêmes auteurs ont montré qu'une autre région située sur le Chromosome 12 nécessite des soins intensifs lors de l'infection par le Covid-19 [38]. Cette région est aussi héritée du Neandertal et contrairement au segment du Chromosome 3, celle-ci est protectrice contre la sévérité de l'infection (réduction d'environ 22% du risque relatif d'être sévèrement affecté).

Les populations d'Asie et d'Océanie (Australie, Mélanésie, Philippines) portent un héritage estimé à environ 5% du Denisovien dans leur génome [12,27].

Parmi les legs bénéfiques du Denisovien, les Tibétains ont hérité un gène (EPAS1) qui leur permet de s'adapter à la vie en altitude extrême [39].

Les populations africaines sub-sahariennes n'ont pas eu de contact avec l'homme de Neandertal, mais elles portent des fragments archaïques dont les origines ne sont pas encore pleinement établies. Une étude récente concernant quatre populations ouest-africaines a montré que 2–19% de leur génome provient d'une introgression archaïque d'une population qui a divergé avant la séparation de Neandertal de l'homme moderne [40].

Il est à noter que peu ou pas de génomes Nord-Africains ont été séquencés. Pourtant les populations Nord-Africaines ont été historiquement soumises à des échanges migratoires venant de l'Est, de l'Ouest et du Nord de la Méditerranée ainsi que du sud du Sahara et donc peuvent potentiellement présenter un intérêt majeur pour les analyses de la diversité génomique. Une étude a été néanmoins entreprise et a montré qu'il existe des signatures de mixité avec le Neandertal [41] dans les populations Nord-Africaines et particulièrement dans la population Tunisienne ayant un héritage ancestral Pré-Néolithique, prouvant que l'héritage du Neandertal date d'avant les migrations en provenance du Proche-Orient ou de l'Europe [42].

Les *Homo sapiens* qui ont quitté l'Afrique il y a environ 70000 ans et se sont dispersés à travers le monde, n'ont emporté avec eux que la diversité génomique humaine qu'ils possédaient. Ce qui implique qu'à ce jour la grande diversité génomique n'est portée que par des Africains. Différentes études ont montré qu'une plus grande diversité existe chez les Africains comparés aux non-Africains [43] ce qui implique que des parties de la diversité humaine ne peuvent être étudiées qu'à partir de populations africaines.

En plus de l'homme de Neandertal et de Denisova, d'autres hominidés ont été découverts en Indonésie [44] (*Homo floresiensis*, découvert en 2003 dans la grotte Lang Boa sur l'île de Flores en Indonésie et qui a vécu de 90000 à 12000 ans), aux Philippines [45] (*Homo Luzonensis* découvert dans une cave dénommée Gallao sur l'île de Luçon datant de 67000 à 50000 ans), dont les génomes ne sont malheureusement pas encore séquencés.

Ces études laissent penser qu'il y a 50000 ans, au moins 5 espèces d'hommes étaient contemporaines mais ne vivaient pas dans les mêmes lieux ont pu néanmoins se rencontrer : Neandertal, *Homo sapiens* (en Afrique), Denisovien, *floresiensis* et *Luzonensis*.

## Conclusions et Perspectives

La saga du génome humain a permis d'accumuler des connaissances extraordinaires sur le patrimoine génétique de l'homme, particulièrement en rapport avec les maladies génétiques et les maladies complexes ou rares ainsi que sur l'histoire de plus en plus précise de l'évolution de l'homme moderne depuis son départ de l'Afrique pour le peuplement du monde il y a 70000 ans.

Ce développement a été encouragé par la vision historique des scientifiques de l'ère des premiers génomes séquencés, qui ont imposé le partage des données en les rendant publiquement accessibles sans restriction. L'Internet a permis aussi

l'échange des programmes d'analyses et l'accessibilité rapide aux données, favorisant ainsi les collaborations scientifiques.

Les milliers de séquences de génomes d'humains, de microbes, de plantes et d'animaux ont transformé les sciences biologiques et ont permis le développement extraordinaire de la Bioinformatique. Le fait que ces données soient publiques a permis à des scientifiques de différents domaines, ayant des compétences et des perspectives différentes, de travailler sur ces données permettant ainsi l'émergence de nouvelles idées en rehaussant l'innovation et la découverte scientifiques. Les collaborations multidisciplinaires entre chercheurs se sont naturellement imposées par les projets de séquençage.

Des développements mathématiques, statistiques et informatiques ont permis la mise au point de méthodes et d'outils pour résoudre des problèmes posés par la génomique. Certains de ces développements n'auraient pas pu voir le jour sans les données génomiques.

Sur le plan de la santé humaine, la génomique a bouleversé la recherche biomédicale et a introduit la médecine génomique personnalisée qui a pour objectif d'utiliser le profil génomique personnel afin de guider le suivi, le traitement et la prévention des maladies. Il y a un effort international pour appliquer les découvertes génomiques à l'amélioration de la santé humaine. L'édition des génomes, un nouveau domaine en plein développement, va aider à introduire de nouvelles thérapies pour les maladies héréditaires et les maladies complexes [46].

Le Microbiome humain (l'ensemble des génomes des microorganismes qui vivent sur et dans le corps humain) qui n'a pas été abordé ici, a aussi un impact fondamental sur la santé de l'homme.

La génomique humaine joue un rôle central dans les recherches fondamentales et cliniques pour améliorer la santé humaine. En effet on peut imaginer que dans un proche avenir tout nouveau-né va disposer de son génome à la naissance ce qui va permettre de savoir s'il est porteur d'une des sources des maladies complexes et donc avoir la possibilité de le suivre adéquatement sur le plan médical.

La génomique a eu un impact sur l'éducation des jeunes générations et sur la recherche scientifique. L'enseignement a été adapté aux nouvelles connaissances. Des nouveaux programmes ont été établis et les profils des jeunes chercheurs actuels sont significativement différents de ceux d'il y a 30 ans.

Les publications scientifiques relatives à la génomique sont nombreuses et de grande qualité. Des travaux qui apportent les nouvelles connaissances acquises à partir des analyses génomiques sont régulièrement publiées. Le suivi régulier de cette littérature scientifique, aujourd'hui favorisé par les publications électroniques ouvertes, est fondamental pour les jeunes générations de chercheurs. Il est impératif de suivre régulièrement l'évolution des connaissances dans ce domaine [47] et de favoriser le travail multidisciplinaire en équipe pour former des compétences locales dans les domaines que nous venons de présenter.

Malgré ces développements et ces ouvertures, la génomique reste largement euro centrique : les études sont biaisées directement ou indirectement vers des études spécifiques des descendants européens. Il n'est pas sûr que les retombées cliniques

de la génomique puissent être profitables à toute l'humanité. Il reste beaucoup à faire pour que la génomique tienne compte de la diversité des hommes. Nous l'avons vu, moins de 2% des génomes séquencés à ce stade proviennent des populations africaines, malgré sa grande diversité génétique comparée à d'autres continents. Probablement aucun génome Nord-Africain n'a été à ce stade complètement séquencé malgré sa position géographique, historiquement lieu d'échanges intenses avec les populations Européennes et Africaines Subsahariennes.

Les pays en voie de développement n'ont pas encore mis en place les changements nécessaires pour s'adapter aux nouvelles connaissances surtout au niveau de l'Enseignement et de la Recherche Scientifique. A l'ère de la science ouverte où l'accès est libre et sans restriction aussi bien aux données qu'aux publications scientifiques, chacun doit prendre conscience de l'importance de la génomique et de la recherche scientifique qu'elle implique. Il est important d'améliorer l'Enseignement, d'encourager les collaborations et de former des compétences locales.

Concernant l'Enseignement Supérieur et la Recherche Scientifique en Tunisie, il est important de mettre en place l'infrastructure (Laboratoires, Séquenceurs, Ordinateurs) et les moyens adéquats (Budgets) pour favoriser l'émergence des compétences (Formation) afin de maîtriser et de participer aux découvertes scientifiques qu'offre la génomique humaine. Plusieurs initiatives sont engagées et sont en cours de réalisation, dans la formation des jeunes chercheurs (Masters et Thésards) au sein de plusieurs 3èmes cycles et laboratoires de génétique humaine, Immunologie et maladies infectieuses. Il serait urgent de renforcer et d'améliorer les formations existantes et de s'inspirer des programmes stratégiques en génomique humaine des pays développés.

Comme indiqué plus haut [6], le coût du séquençage d'un génome humain est en dessous de 1000 US\$, monter un projet de séquençage des génomes d'une centaine de Tunisiens n'aurait pas un coût excessif. Disposer de ces génomes permettrait à plusieurs jeunes de participer à des projets de recherches de hauts niveaux qui pourraient aboutir à un génome de référence du Tunisien et à rehausser l'image de la recherche tunisienne à l'échelle internationale. Le génome de référence du Tunisien n'est pas uniquement un symbole, c'est utile comme nous l'avons vu plus haut pour le suivi et la détection des maladies complexes et héréditaires spécifiques. Aussi avoir le génome de référence du Tunisien ouvrirait de vastes possibilités de collaborations internationales pour l'étude des populations autour de la Méditerranée et Subsahariennes.

L'actualité ne semble pas favoriser ce genre de réflexions, mais il nous appartient d'éclairer les décideurs d'aujourd'hui de ne pas hypothéquer l'avenir des jeunes chercheurs de demain. Nous adhérons à la devise d'un programme stratégique canadien d'intégration des connaissances génomiques à la santé:

« Think Big, Start Small, Learn Fast » !

## Références :

Note: des définitions de termes utilisés peuvent être trouvées dans ces 2 sites :

<http://www.ensembl.org/info/website/glossary.html>

<http://www.ensembl.org/info/genome/variation/prediction/classification.html>

1. International Human Genome Sequencing Consortium, *et al.* (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921. <https://doi.org/10.1038/35057062>

2. Venter JC. *et al.* (2001). The Sequence of the Human Genome. *Science*; 291 (5507), 1304-1351. <https://doi.org/10.1126/science.1058040>

3. The Human Genome at 20: <https://www.nature.com/collections/febighfbg>

4. Complicated legacies: The human genome at 20: <https://science.sciencemag.org/content/371/6529/564>

5. International Human Genome Sequencing Consortium. (2004). Finishing the euchromatic sequence of the human genome. *Nature* 431, 931–945.

6. Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP) Available at: [www.genome.gov/sequencingcostsdata](http://www.genome.gov/sequencingcostsdata). Accessed [February, 2021].

7. Genome OnLine Database : <https://gold.jgi.doe.gov>

8. Weischenfeldt, J., Symmons, O., Spitz, F. *et al.* (2013). Phenotypic impact of genomic structural variation: insights from and for human disease. *Nat Rev Genet* 14, 125–138. <https://doi.org/10.1038/nrg3373>

9. Mitra, I., Huang, B., Mousavi, N. *et al.* (2021). Patterns of de novo tandem repeat mutations and their role in autism. *Nature* 589, 246–250. <https://doi.org/10.1038/s41586-020-03078-7>

10. The International HapMap Consortium. (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449, 851–861. <https://doi.org/10.1038/nature06258>

11. The 1000 Genomes Project Consortium. (2015). A global reference for human genetic variation. *Nature* 526, 68–74. <https://doi.org/10.1038/nature15393>

12. Mallick, S., Li, H., Lipson, M. *et al.* The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* 538, 201–206 (2016). <https://doi.org/10.1038/nature18964>

13. The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium., Campbell, P.J., Getz, G. *et al.* (2020). Pan-cancer analysis of whole genomes. *Nature* 578, 82–93. <https://doi.org/10.1038/s41586-020-1969-6>

14. Choudhury, A. *et al.* (2020). High-depth African genomes inform human migration and health. *Nature* 586, 741–748. <https://doi.org/10.1038/s41586-020-2859-7>

15. GenomeAsia100K Consortium., Wall, J.D., Stawiski, E.W. *et al.* (2019). The GenomeAsia 100K Project enables genetic discoveries across Asia. *Nature* 576, 106–111. <https://doi.org/10.1038/s41586-019-1793-z>

16. Romdhane, L., Mezzi, N., Dallali, H. *et al.* (2021). A map of copy number variations in the Tunisian population: a valuable tool for medical genomics in North Africa. *npj Genom. Med.* 6, 3. <https://doi.org/10.1038/s41525-020-00166-5>

17. Baker, J.L., Rotimi, C.N. & Shriner, D. Human ancestry correlates with language and reveals that race is not an objective genomic classifier. *Sci Rep* 7, 1572 (2017). <https://doi.org/10.1038/s41598-017-01837-7>

18. Wonkam A. (2021). Sequence three million genomes across Africa. *Nature* 590, 209-211. <https://doi.org/10.1038/d41586-021-00313-7>
19. Rotimi CN, Adeyemo AA. (2021). From one human genome to a complex tapestry of ancestry. <https://doi.org/10.1038/d41586-021-00237-2>.
20. Veeramah KR, Hammer MF. (2014). The impact of whole-genome sequencing on the reconstruction of human population history. *Nat Rev Genet.*15(3):149-62. <https://doi.org/10.1038/nrg3625>.
21. Hublin, JJ., Ben-Ncer, A., Bailey, S. *et al.* (2017). New fossils from Jebel Irhoud, Morocco and the pan-African origin of *Homo sapiens*. *Nature* 546, 289–292. <https://doi.org/10.1038/nature22336>
22. Grün, R., Brink, J., Spooner, N. *et al.* (1996). Direct dating of Florisbad hominid. *Nature* 382, 500–501. <https://doi.org/10.1038/382500a0>
23. McDougall, I., Brown, F. H. & Fleagle, J. G. (2005). Stratigraphic placement and age of modern humans from Kibish, Ethiopia. *Nature* 433, 733–736. <https://doi.org/10.1038/nature03258>
24. White, T., Asfaw, B., DeGusta, D. *et al.* (2003). Pleistocene *Homo sapiens* from Middle Awash, Ethiopia. *Nature* 423, 742–747. <https://doi.org/10.1038/nature01669>
25. Delson E. (2019). An early dispersal of modern humans from Africa to Greece. *Nature* **571**, 487-488. <https://doi.org/10.1038/d41586-019-02075-9>
26. Rasmussen, M., *et al.* (2010). Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* 463, 757–762. <https://doi.org/10.1038/nature08835>
27. Green RE, *et al.* (2010). A draft sequence of the Neandertal genome. *Science*. 328(5979):710-722. <https://doi.org/10.1126/science.1188021>
28. Reich, D., *et al.* (2010). Genetic history of an archaic hominin group from Denisova cave in Siberia. *Nature* 468, 1053–1060. <https://doi.org/10.1038/nature09710>
29. Prüfer, K., *et al.* (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505, 43–49. <https://doi.org/10.1038/nature12886>
30. Meyer M, *et al.* *Nature*. 531, 504–507. (2016). Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. <https://doi.org/10.1038/nature17405>
31. van der Valk, T., Pečnerová, P., Díez-del-Molino, D. *et al.* (2021). Million-year-old DNA sheds light on the genomic history of mammoths. *Nature*. <https://doi.org/10.1038/s41586-021-03224-9>
32. Orlando, L., Ginolhac, A., Zhang, G. *et al.* (2013). Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* 499, 74–78 (2013). <https://doi.org/10.1038/nature12323>
33. Marciniak, S., Perry, G. (2017). Harnessing ancient genomes to study the history of human adaptation. *Nat Rev Genet* 18, 659–674. <https://doi.org/10.1038/nrg.2017.65>
34. Higham, T., Douka, K., Wood, R. *et al.* (2014). The timing and spatiotemporal patterning of Neanderthal disappearance. *Nature* 512, 306–309. <https://doi.org/10.1038/nature13621>
35. Skov, L., *et al.* (2020). The nature of Neanderthal introgression revealed by 27,566 Icelandic genomes. *Nature* 582, 78–83 (2020). <https://doi.org/10.1038/s41586-020-2225-9>
36. Gibbons A. (2019). Spotting Evolution Among us. *Science* 363, 6422, 21-23. <https://doi.org/10.1126/science.363.6422.21>

37. Zeberg, H. & Pääbo, S. (2020). The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. *Nature* 587, 610–612. <https://doi.org/10.1038/s41586-020-2818-3>.
38. Zeberg H, Pääbo S. (2021). A genomic region associated with protection against severe COVID-19 is inherited from Neandertals. *Proc Natl Acad Sci U S A*. 118(9):e2026309118. <https://doi.org/10.1073/pnas.2026309118>.
39. Hu H, et al. (2017). Evolutionary history of Tibetans inferred from whole-genome sequencing. *PLoS Genet* 13(4): e1006675. <https://doi.org/10.1371/journal.pgen.1006675>
40. Durvasula and Sankararaman (2020). Recovering signals of ghost archaic introgression in African populations. *Sci. Adv.* 6: eaax5097. <https://doi.org/10.1126/sciadv.aax5097>
41. Sánchez-Quinto F, et al. (2012). North African Populations Carry the Signature of Admixture with Neandertals. *PLoS ONE* 7(10): e47765. <https://doi.org/10.1371/journal.pone.0047765>
42. Henn BM, Botigué LR, Gravel S, Wang W, Brisbin A, Byrnes JK, et al. (2012). Genomic Ancestry of North Africans Supports Back-to-Africa Migrations. *PLoS Genet* 8(1): e1002397. <https://doi.org/10.1371/journal.pgen.1002397>
43. Campbell MC. and Tishkoff SA. (2008). African Genetic Diversity: Implications for Human Demographic History, Modern Human Origins, and Complex Disease Mapping. *Annu Rev Genomics Hum Genet.* 9: 403–433. <https://doi.org/10.1146/annurev.genom.9.081307.164258>
44. Argue d, Groves CP, Lee MLS, Jungers WL. (2017). The affinities of *Homo floresiensis* based on phylogenetic analyses of cranial, dental, and postcranial characters. *Journal of Human Evolution.* 107, 107e133. <http://dx.doi.org/10.1016/j.jhevol.2017.02.006>
45. Détroit, F., Mijares, A.S., Corny, J. *et al.* (2019). A new species of *Homo* from the Late Pleistocene of the Philippines. *Nature* 568, 181–186. <https://doi.org/10.1038/s41586-019-1067-9>
46. Chari, R., Church, G. Beyond editing to writing large genomes. (2017). *Nat Rev Genet* 18, 749–760 <https://doi.org/10.1038/nrg.2017.59>
47. Boudabous A, Tekai F. (2020). Enhancing Bioinformatics and Genomics Courses: Building Capacity and Skills via Lab Meeting Activities. *Bioessays.* 42(10):e2000134. <https://doi.org/10.1002/bies.202000134>.

## Auteurs

### Fredj Tekai

Chercheur Invité

Institut Pasteur

28, rue du Dr Roux

75724 Paris cedex 15 France

Page web: "Bioinformatics and Genome Analyses":

[https://webext.pasteur.fr/tekaia/BGA\\_courses.html](https://webext.pasteur.fr/tekaia/BGA_courses.html)

### Abdellatif Boudabous

Professeur

Université Tunis El Manar, Faculté des Sciences de Tunis,

Laboratoire Microorganismes et Biomolécules Actives, Campus

Universitaire Farhat Hached, El Manar, Tunis, Tunisie

e-mail : [abdellatif.chihi@mes.rnu.tn](mailto:abdellatif.chihi@mes.rnu.tn)