

► hybrides issus de croisements plus anciens. Mais ces unions seraient quasiment indétectables sur la base de preuves fossiles, le registre étant très lacunaire. Des indices les trahissant ne seraient laissés que dans l'ADN de certains individus, et, quand bien même, ils seraient plus ténus encore que les fragments d'ADN laissés par les Néandertaliens et les Denisoviens dans notre génome. Des modèles statistiques aident les chercheurs à repérer ces populations sans données fossiles. Ainsi, en 2013, la même équipe de Leipzig a identifié des motifs de variation génétique chez des humains modernes et anciens pointant vers une population humaine inconnue s'étant reproduite avec les Denisoviens (ou leurs ancêtres). Toutefois, selon plusieurs experts, ces méthodes négligent inévitablement beaucoup d'épisodes de cette nature.

Qui d'autre a contribué aux génomes humains d'aujourd'hui ? À quoi ressemblaient ces « populations fantômes » ? Où vivaient-elles et à quelle fréquence interagissaient-elles et se reproduisaient-elles avec d'autres espèces humaines ? En 2019, Oscar Lao, de l'institut des sciences et des techniques de Barcelone, en Espagne, et ses collègues ont montré le potentiel de certaines techniques d'apprentissage profond (*deep learning*) pour apporter des éléments de réponse. Et même, révéler des événements que personne n'aurait pu imaginer. De fait, ils ont mis en évidence une autre population fantôme : un ancêtre humain inconnu en Eurasie, probablement un hybride Néandertalien-Denisovien ou un proche de la lignée denisovienne.

Ces résultats confirment l'utilité de l'intelligence artificielle en paléontologie, non seulement pour identifier des populations fantômes jusqu'alors inconnues, mais aussi pour mettre au jour les étapes évolutives qui nous ont façonnés en tant qu'espèce.

LA CHASSE AUX SIGNES SUBTILS

Les méthodes statistiques actuelles consistent à comparer simultanément quatre génomes. C'est un test de similitude, mais pas nécessairement d'ascendance, car il existe de nombreuses façons d'interpréter les petits mélanges génétiques qu'il met en lumière. Par exemple, de telles analyses peuvent suggérer qu'un Européen moderne, à l'inverse d'un Africain actuel, partage certains caractères génétiques avec les Néandertaliens. Cela ne signifie pas pour autant que ces gènes sont issus d'un métissage entre les Néandertaliens et les ancêtres des Européens. Ces derniers pourraient s'être hybridés avec une autre population, de parenté proche de l'homme de Néandertal, mais pas avec les Néandertaliens eux-mêmes.

Nous n'avons pour l'heure aucune certitude, car en l'absence de preuves physiques

L'INTELLIGENCE ARTIFICIELLE A CONCLU QU'UN GROUPE HUMAIN JUSQU'ALORS INCONNU AVAIT CONTRIBUÉ À L'ÉVOLUTION DES PEUPLES ASIATIQUES

indiquant quand, où et comment ces anciennes sources hypothétiques de variation génétique pourraient avoir vécu, il est difficile de déterminer laquelle de ces nombreuses ascendances supposées est la plus probable. Cette technique « est puissante par sa simplicité, mais elle laisse beaucoup de flou dans notre compréhension de l'évolution », explique John Hawks, paléo-anthropologue de l'université du Wisconsin, à Madison, aux États-Unis.

La nouvelle méthode d'apprentissage profond vise à faire mieux, en cherchant à expliquer des flux génétiques trop faibles pour les approches statistiques classiques. Par entraînement, le réseau de neurones artificiels peut apprendre à classer différents modèles de données génomiques selon leurs histoires démographiques, sans qu'il soit nécessaire de lui inculquer comment établir ces liens.

Cet usage de l'apprentissage profond peut mettre au jour des fantômes dont nous ne suspicions pas même l'existence. Il n'y a aucune raison de penser que les Néandertaliens, Denisoviens et hommes modernes étaient les trois seules populations en jeu. Selon John Hawks, il pourrait bien y en avoir eu des dizaines.

Jason Lewis, anthropologue à l'université de Stony Brook, à New York, est du même avis. « Notre imagination a été limitée par l'attention que nous portons aux individus vivants et aux fossiles que nous avons trouvés en Europe, en Afrique et en Asie, estime-t-il. Avec les techniques d'apprentissage profond, nos recherches ne sont plus restreintes par notre imagination. »

L'apprentissage profond semble au premier abord inapproprié pour résoudre les problèmes des paléontologues, car il nécessite d'ordinaire d'énormes quantités de données d'entraînement. Prenons l'une des applications les plus connues, comme un classificateur d'images. Lorsque les experts élaborent un système dédié à l'identification d'images de chats, ils l'entraînent avec des



Quanta
magazine

Cet article a d'abord été publié en anglais par *Quanta Magazine*, une publication en ligne indépendante soutenue par la Simons Foundation afin de favoriser la diffusion des sciences : <http://bit.ly/QM-IA-Ghost>

NOS ANCÊTRES INCONNUS

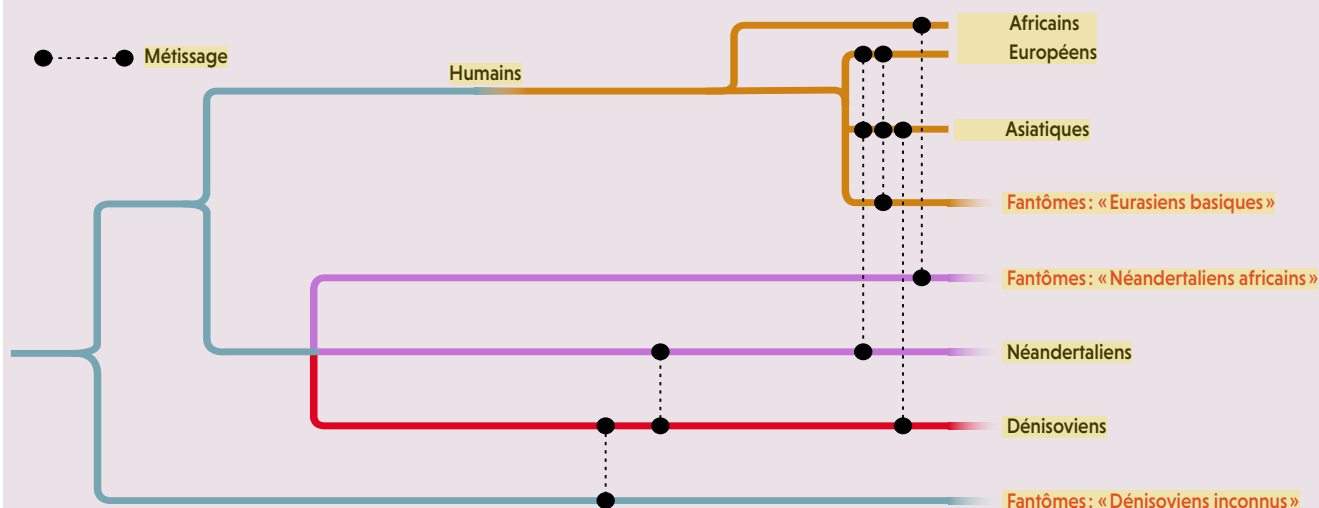
Depuis l'avènement de la paléogénétique, on a mis en évidence les traces de nombreux mélanges entre espèces du genre *Homo*. La plupart sont connues (Néandertal, Denisova...), mais quelques fois, on découvre des espèces avec qui nos ancêtres se sont reproduits et dont on n'a jamais entendu parler par ailleurs, aucune trace fossile : ce sont des fantômes. Le premier a été repéré par David Reich, de l'université Harvard, aux États-Unis. L'analyse du génome de 44 individus ayant vécu au Moyen-Orient entre 3400 et 14000 ans a révélé les traces d'un ancien groupe d'humains installés dans la région depuis plus de 45000 ans. On ne sait rien de ces «Eurasiens basiques» ! L'hypothèse la plus probable est qu'au sortir de l'Afrique, un groupe d'*Homo sapiens* s'est rapidement retrouvé isolé, tandis que les autres ont rencontré les Néandertaliens avec qui ils se sont reproduits. Un autre fantôme a été découvert en Afrique par Joshua Akey, de l'université de Princeton, et Sarah Tishkoff, de l'université de Pennsylvanie. Ils ont étudié le génome d'Africains modernes appartenant à deux groupes ancestraux, d'une part, les Bakas (voir la photo ci-dessus) du Cameroun et, d'autre part, les Hadzas et les Sandawes de Tanzanie. Tous sont des chasseurs-cueilleurs. Les analyses ont révélé des fragments d'ADN qui semblent provenir d'une autre espèce d'hominidés inconnue. Cet ADN étant absent du génome des Eurasiens, on suppose que le métissage a eu lieu après la sortie d'Afrique d'*Homo sapiens*, probablement au cours des 30000 dernières années. Ainsi,



au moins une autre espèce d'hominidés vivait à nos côtés en Afrique ! Et peut-être même plus selon Joshua Akey. Ce dernier interprète ses résultats ainsi : « Il y a 700000 ans, un groupe s'est séparé de la lignée qui allait donner l'homme moderne pour devenir, après migration hors d'Afrique les Néandertaliens que l'on connaît. Mais parmi les populations restées sur le continent, une autre séparation a eu lieu, conduisant à l'apparition d'une sorte de "Néandertalien africain" ». Qui était-il ? Les candidats sont nombreux : *Homo naledi* ? *Homo heidelbergensis* ? Une sous-population d'*Homo sapiens* isolée suffisamment longtemps pour se démarquer génétiquement ? La réponse ne pourra venir que du séquençage de l'ADN de fossiles

africains. Joshua Akey a montré en 2018 que les Denisoviens auraient également une espèce fantôme en leur sein. Quoi qu'il en soit, toutes ces études, même si leurs résultats doivent être précisés, confirment une chose. Notre arbre généalogique n'est pas une simple suite de bifurcations. Au contraire, de nombreuses passerelles relient les branches (voir ci-dessous). La séparation en groupes, en sous-populations, qui évoluent indépendamment les unes des autres, puis se retrouvent, parfois plusieurs centaines de milliers d'années après, et se mélangent semble avoir été la règle par le passé. Nous sommes le fruit d'un important brassage incluant des amours avec des fantômes !

LOÏC MANGIN



À en croire les analyses génétiques, notre arbre généalogique héberge au moins trois espèces fantômes (en rouge).

➤ milliers d'images. Et l'efficacité est tout de suite vérifiable, car tout le monde sait à quoi ressemble un chat.

Toutefois, le manque de données anthropologiques et paléontologiques pertinentes oblige les chercheurs à faire preuve d'inventivité et à créer leurs propres données. «Nous avons un peu triché, admet Oscar Lao. Nous pouvions utiliser une quantité infinie de données pour entraîner le système d'apprentissage, parce que nous utilisions des simulations.»

Les chercheurs ont imaginé des dizaines de milliers d'histoires évolutives, reposant sur des combinaisons différentes de détails démographiques: le nombre de populations d'humains ancestraux, leur taille, les dates de divergence, leur taux de métissage... À partir de ces histoires simulées, ils ont généré les nombreux génomes des individus modernes qui pourraient en résulter, ces génomes ayant nourri leur algorithme d'apprentissage. De la sorte, le programme a appris quels modèles évolutifs sont les plus susceptibles de produire des génomes virtuels.

L'équipe a ensuite laissé le champ libre à l'intelligence artificielle pour qu'elle identifie les histoires correspondant le plus aux véritables données génomiques. Finalement, le système a conclu qu'un groupe humain jusqu'alors inconnu avait contribué à l'évolution des peuples asiatiques. Ces humains constituaient soit une population distincte résultant du métissage des Denisoviens et des Néandertaliens il y a 300 000 ans, soit un groupe descendant de la lignée dénisovienne peu de temps après cette époque.

Ce n'est pas la première fois que l'apprentissage profond est utilisé pour élucider des questions liées à l'évolution. En 2016, un groupe de recherche, dirigé par Andrew Kern de l'université de l'Oregon, aux États-Unis, a utilisé une approche basée sur des simulations et des techniques d'apprentissage profond pour différencier plusieurs modèles d'évolution des espèces, comme les humains. Ils ont découvert que la plupart des adaptations favorisées par l'évolution ne reposent pas sur l'émergence de nouvelles mutations bénéfiques, mais plutôt sur l'expansion de variants génétiques préexistants.

UN NOUVEL OUTIL ENTRE ESPOIR ET ENGOUEMENT

Bien entendu, cette méthode appelle à la prudence. En effet, si l'histoire évolutive humaine ne ressemble pas aux modèles simulés par lesquels ces systèmes d'apprentissage profond sont entraînés, les résultats produits seront incorrects. Ce problème, Andrew Kern et d'autres tentent de le surmonter.

«Je pense que les applications de l'intelligence artificielle en génomiques sont survenues, concède Joshua Akey, chercheur en

écologie et biologie évolutive à l'université de Princeton, aux États-Unis. L'apprentissage profond est un nouvel outil formidable, mais il s'agit simplement d'une autre méthode, et elle ne va pas résoudre tous les problèmes ni nous révéler tout ce que nous souhaitons savoir sur l'évolution humaine.» D'ailleurs, le même Joshua Akey a utilisé d'autres méthodes pour repérer les fantômes de notre passé (voir l'encadré page précédente).

L'apprentissage profond va donner une impulsion nouvelle à la génétique des populations

Certains experts sont encore plus sceptiques. Toutefois, d'autres paléontologues et généticiens y voient un progrès, une technique utile pour des prédictions quant aux futures découvertes de fossiles et aux variations génétiques qui devraient avoir existé au sein des populations humaines il y a des milliers d'années. «Je pense que l'apprentissage profond va donner une impulsion nouvelle au domaine de la génétique des populations», estime Oscar Lao.

Il pourrait en être de même pour d'autres domaines pour lesquels nous avons des données, mais pas d'idées sur les processus qui les ont produits. En 2016, des physiciens ont eux aussi développé un système d'intelligence artificielle pour filtrer les quantités astronomiques de données produites par le LHC. La recherche en géologie et les méthodes de prédiction des séismes commencent également à bénéficier de ces approches d'apprentissage profond.

Pour Nick Patterson, de l'institut de technologie du Massachusetts, et à l'université Harvard, à Cambridge, aux États-Unis, «nous ne savons pas vraiment où cela nous mènera, nous verrons bien. Mais il est toujours bon de voir naître de nouvelles méthodes. Nous utiliserons toutes les approches disponibles si elles semblent pertinentes.» ■

BIBLIOGRAPHIE

M. MONDAL ET AL., Approximate Bayesian computation with deep learning supports a third archaic introgression in Asia and Oceania, *Nature Communications*, vol. 10, art. 246, 2019.

S. BROWNING ET AL., Analysis of human sequence data reveals two pulses of archaic Denisovan admixture, *Cell*, vol. 173, pp. 53-61, 2018.

V. SLON ET AL., The genome of the offspring of a Neanderthal mother and a Denisovan father, *Nature*, vol. 561, pp. 113-116, 2018.

D. SCHRIDER ET A. KERN, S/HIC: Robust identification of soft and hard sweeps using machine learning, *Plos Genet.*, vol. 12(3), art. e1005928, 2016.

J. LACHANCE ET AL., Evolutionary history and adaptation from high-coverage whole-genome sequences of diverse African hunter-gatherers, *Cell*, vol. 150, pp. 457-469, 2012.