

La généalogie à l'ère de la génomique

S. MANRUBIA, B. DERRIDA et D. ZANETTE

Les ancêtres que vous partagez avec votre voisin de palier ou avec le Président de la république sont moins éloignés dans le passé que nous ne le pensons.

Est-il filiation directe plus claire que celle-là ? Ses prétentions à la couronne, Henri les tire de Jean de Gand, le quatrième fils ; York tire les siennes du troisième fils [Lionel]. Tant que Lionel aura descendance, l'autre branche ne doit point régner.

Shakespeare, *Henri VI*.

ien ne fut simple chez les Plantagenêts, la dynastie qui régna sur l'Angleterre de 1154 à 1485. La famille faillit s'éteindre, à force de luttes fratricides, au cours de la tristement célèbre guerre des Deux-Roses qui opposa au XV^e siècle les maisons de Clarence (les descendants de Lionel, troisième fils d'Édouard III), de Lancastre (fondée par Jean de Gand, le quatrième fils) et d'York (celle du cinquième fils, Edmond). Seule la défaite du dernier roi Plantagenêt, Richard III, face à Henri VII Tudor, descendant de Jean de Gand apaisa les troubles. Le vainqueur consolida son pouvoir grâce à un mariage intrafamilial avec Élisabeth d'York. Leur fils, Henri VIII, descendait ainsi d'Édouard III par quatre lignées différentes, marquant chacune une alliance clef et un tournant dans l'histoire d'Angleterre.

L'histoire des familles royales d'Angleterre illustre comment la généalogie peut présider au destin des nations, mais aussi combien le phénomène de coalescence, la fusion des branches dans un arbre généalogique, traduisant le mariage de deux individus ayant un ancêtre commun, est courant au sein de toute population fermée.

Ces dernières années, la génomique a bouleversé notre perception de la proximité généalogique des êtres humains. L'étude de l'ADN mitochondrial (les mitochondries sont des petits organites cellulaires qui sont transmis sans changement, hormis les mutations, de la mère à la fille) et de certains gènes du chromosome Y (transmis de père en fils) a permis aux généticiens de situer l'époque de l'« Ève mitochondriale » et de l'« Adam du chromosome Y » dans un passé étonnamment proche. Les mitochondries portées par toutes les cellules humaines seraient l'héritage global d'une seule femme. En 1987, l'équipe de Rebecca Cann, Mark Stoneking et Allan Wilson, de l'Université de Berkeley, ont estimé que cette femme vivait il y a entre 140 000 et 290 000 années.

Ces analyses ne révèlent qu'une partie de l'histoire, car elles sont fondées sur un héritage monoparental. Or la majeure partie de notre génome est héritée à la fois de la mère et du père, et leurs gènes

sont brassés par des recombinaisons de l'ADN (des chromosomes échangeant du matériel). Nous verrons que notre ADN raconte toute la richesse de notre histoire, pour autant que nous puissions la lire. Chacun d'entre nous découvrira certainement une ascendance aussi embrouillée que celle d'Henri VIII. Nous décrirons aussi le pendant masculin de la transmission de l'ADN mitochondrial, uniquement féminine, il s'agit du nom de famille, transmis de façon exclusivement paternelle.

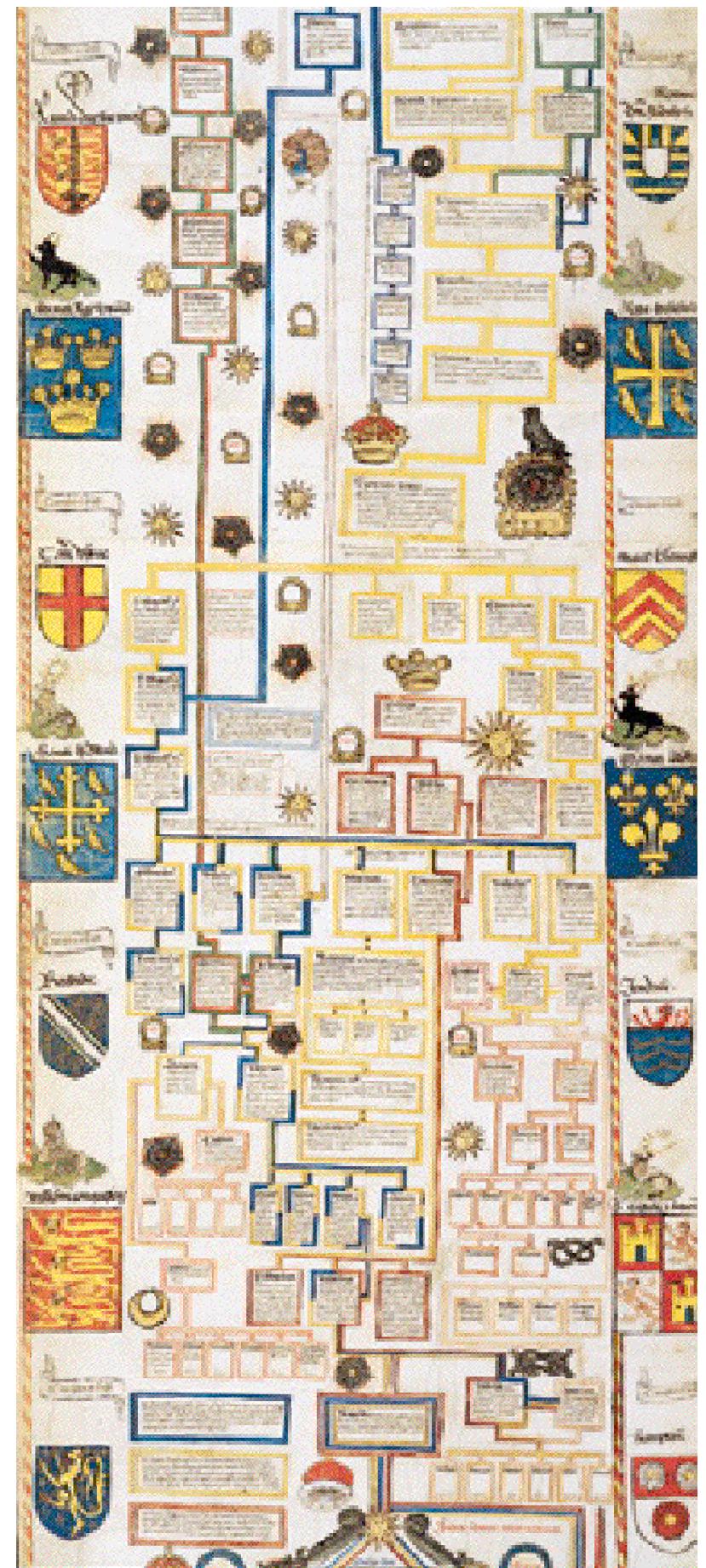
Famille en voie de disparition

L'ADN mitochondrial est un outil puissant, car il élague notre arbre généalogique pour ne faire apparaître qu'un seul rameau, mais pour cette raison, il simplifie notre passé. Pour comprendre comment les gènes et les lignées évoluent sur des dizaines ou des centaines de générations, nous devons utiliser des modèles mathématiques et des simulations informatiques, car nous ne possédons pas de données généalogiques remontant aussi loin dans le passé. Avec ces modèles biparentaux, nos ancêtres communs sont beaucoup moins éloignés dans le temps que ceux déterminés par les modèles monoparentaux, fondés, par exemple, sur l'ADN mitochondrial ou le nom de famille.

La première tentative sérieuse de résolution mathématique d'un problème généalogique fut la conséquence d'une controverse mettant en scène l'un des plus célèbres scientifiques de l'ère victorienne, Sir Francis Galton. Avec un certain sens de l'à-propos, Galton, cousin germain de Charles Darwin, avait écrit un livre intitulé *Le génie héréditaire*, où il tentait d'expliquer le déclin des grandes familles, phénomène illustré par la raréfaction, voire la disparition de patronymes autrefois répandus. Selon lui, l'explication qui prévalait (l'augmentation du confort matériel et de la capacité intellectuelle s'accompagne nécessairement d'une diminution de la fertilité) était inexacte et il en proposa une autre qui, rien de surprenant à l'époque, en imputait la responsabilité aux femmes : « les hommes qui ont récemment vu leur statut s'élever tendraient à consolider leur position en épousant des hérétiques, qui sont par définition des femmes issues de familles sans fils. » Ces femmes, pensait-il, avaient elles-mêmes tendance à donner naissance à moins de fils.

Cependant, le botaniste suisse Alphonse de Candolle fit remarquer qu'une autre explication était possible : le hasard. Pour départager les hypothèses, il convenait donc de connaître la probabilité d'extinction d'un patronyme par des processus aléatoires. Pour ce faire, en 1874, Galton enrôle un mathématicien, le révérend Henry William Watson. L'approche de Watson était ingénieuse. Parce qu'il voulait évaluer le rôle du hasard, Watson supposa que tous les mâles avaient la même fertilité, de sorte

1. CE MANUSCRIT ENLUMINÉ, daté de 1461, illustre la complexité des généalogies royales. Ce rouleau de vélin prétend retracer l'ascendance du roi Édouard IV d'Angleterre jusqu'à Noé, établissant ainsi sa légitimité en tant que monarque. Les soleils et les roses sont des symboles de la maison d'York. La partie du rouleau présentée ici (*environ un sixième de l'ensemble*) couvre la période allant d'Édouard III à Richard, Duc d'York (*Ricardus, en bas au centre*). L'encadrement multicolore de la case de Richard indique ses liens avec les autres familles régnantes d'Europe. À l'inverse, les encadrements jaunes simples des rois de Lancastre, Henri IV, V et VI, dénotent leurs connexions royales relativement modestes, et donc l'infériorité de leurs prétentions au trône.



que les différences de nombre d'enfants étaient aléatoires. Ainsi, chaque mâle avait une probabilité p_0 de n'avoir aucun fils ; une probabilité p_1 d'avoir un seul fils ; une probabilité p_2 d'avoir deux fils, etc. La probabilité de l'extinction d'un patronyme après une génération, nommons-la q_1 , est donc égale à p_0 .

Les choses se compliquent un peu aux générations suivantes. Voyons pour l'extinction d'une lignée après deux générations. Un homme peut avoir un fils (probabilité p_1) qui lui-même n'aura aucun fils (probabilité p_0) : la probabilité que sa lignée s'éteigne de cette façon est alors le produit des deux probabilités, soit $p_1 p_0$. Ou bien, l'homme peut avoir deux fils (proba-

bilité p_2), qui n'auront tous deux aucun fils (probabilité p_0^2) : la probabilité de cet événement est $p_2 p_0^2$. L'addition des probabilités de chacun de ces événements est la probabilité que la lignée s'éteigne après deux générations, soit $q_2 = p_0 + p_1 p_0 + p_2 p_0^2 + p_3 p_0^3 + \dots$

Watson eut l'intuition de génie que le terme de droite, une fonction génératrice, contenait l'information sur la probabilité d'extinction dans les générations suivantes. Voyons comment. Le calcul de la probabilité d'extinction consistait donc à appliquer la fonction génératrice encore et encore, dans un processus itératif. Mathématiquement, il a défini la fonction génératrice $f(x)$ en remplaçant dans cette expression chaque p_0 (sauf le premier) par une variable x : $f(x) = p_0 + p_1 x + p_2 x^2 + p_3 x^3 + \dots$

Il a ensuite montré que la probabilité d'extinction pour chaque génération est obtenue en réintroduisant la probabilité d'extinction de la génération précédente dans cette fonction : $q_1 = f(0)$, $q_2 = f(q_1)$, $q_3 = f(q_2)$...

On retrouve les expressions de q_1 et q_2 précédentes. Pour q_3 , on obtient $f(q_2) = p_0 + p_1(q_2) + p_2(q_2)^2 + p_3(q_2)^3 + \dots$ Chaque terme représente la probabilité qu'un homme n'ait aucun fils (p_0) ; un fils (p_1) dont la lignée s'éteint après deux générations (q_2) ; deux fils (p_2) dont les lignées respectives s'éteignent après deux générations ($(q_2)^2$) et ainsi de suite.

Quelle serait la probabilité d'extinction q_∞ après un nombre infini de générations ? Ce serait tout simplement une récursion sur elle-même ! En d'autres termes, on cherche la solution de l'équation $f(q_\infty) = q_\infty$, c'est-à-dire le point fixe de la fonction génératrice.

Cette équation fournit la probabilité qu'une lignée quelconque s'éteigne à terme, que l'extinction intervienne après une génération ou 150. À ce stade, Watson, tout près d'une élégante solution, a commis une grossière erreur. Sans données démographiques sur les probabilités d'avoir un fils, deux fils... ou aucun, il a simplement hasardé l'estimation suivante : $f(x) = (3+x)^{5/4}$. Cette estimation n'était pas mauvaise en soi, mais Watson a alors fait une erreur mathématique en passant à côté d'une solution de son équation. Il pensait que la seule solution était $f(1) = 1$, c'est-à-dire $q_\infty = 1$: toute lignée est condamnée à s'éteindre !

L'analyse de Watson était correcte pour les populations en régression ou à taille constante. En revanche, pour une population de taille croissante, il existe une autre solution pour q_∞ . Ainsi, pour une population qui augmente de huit pour cent environ par génération, $f(0,05) = 0,55$: il y a 55 pour cent de chance qu'une lignée donnée s'éteigne, et une probabilité de 45 pour cent qu'elle se perpétue à jamais. Hélas, on pensait que les calculs de Watson avaient clos le débat et personne ne s'aperçut de l'erreur avant 50 ans.

Les extinctions de patronymes

Dans les années 1920, une nouvelle génération de biologistes et de mathématiciens pose les bases de la génétique des populations, et à tort fait de découvrir l'erreur de Watson. Dans une population croissante, toute lignée a une probabilité non nulle de survivre indefiniment. En 1939, Alfred Lotka a utilisé les données du recensement des États-Unis de 1920 pour estimer p_0 , p_1 ... et a calculé que $q_\infty = 0,82$. Dans les États-Unis de cette époque, la probabilité de la survie définitive d'un patronyme était de 18 pour cent.

Ces déclarations chiffrées semblent être des vérités absolues, mais elles dépendent d'hypothèses mathématiques particulières qui peuvent être infirmées. Dans le modèle de Watson, nommé processus de Galton-Watson, certaines des hypothèses sont discutables. Par exemple, tous les mâles ont-ils vraiment la même fertilité innée ? La fertilité de chaque mâle est-elle vraiment indépendante de celle des autres mâles, et invariante dans le temps ? Que se passe-t-il lorsque nous prenons en compte la « mutation » des patronymes, s'opérant par le biais de l'immigration ou de l'évolution de l'orthographe ?

La mutabilité des patronymes varie selon les cultures. En Chine, les patronymes ont été bien conservés pendant plusieurs millénaires. Une enquête, effectuée pour le compte de l'empereur Tang Taizhong en 627 de notre ère, a comptabilisé 593 patronymes différents. En 960, le livre *Patronymes de cent familles* enregistrait 438 noms

de famille. Aujourd'hui, environ 40 pour cent de la population chinoise porte l'un des dix noms les plus répandus, et 70 pour cent, l'un des 45 noms les plus répandus. Cette constance est peut-être inhérente au système d'écriture chinois, qui représente chaque nom de famille au moyen d'un seul caractère.

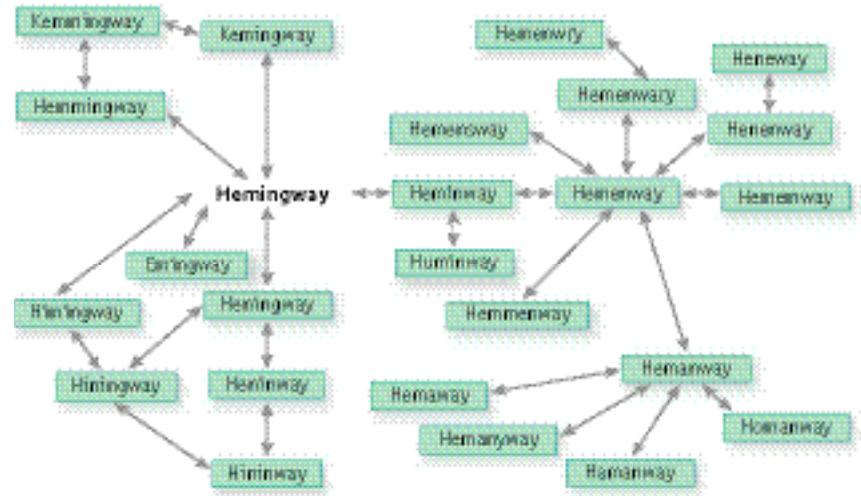
À l'inverse, les États-Unis et le Canada présentent la plus grande diversité de patronymes au monde, conséquence de leur construction par immigration. La grande mutabilité de l'orthographe anglaise a également contribué à augmenter la variété des patronymes (voir la figure 3).

Nous avons étudié la distribution des patronymes à l'aide d'un modèle intégrant une petite probabilité de mutation à tout instant, et incluant également un taux de mortalité flexible qui peut être rendu égal, supérieur ou inférieur au taux de natalité. Avec ce modèle, nous avons trouvé des différences spectaculaires entre les populations en accroissement et les populations statiques, où les taux de natalité et de mortalité sont égaux.

Dans une population en accroissement, la diversité des noms augmente toujours avec le temps. Après un temps suffisant, le nombre de noms appartenant à y personnes exactement, ou $n(y)$, devient proportionnel à $1/y^2$ pour y suffisamment grand. Ainsi, il devrait y avoir 100 fois plus de noms appartenant à 20 personnes que de noms appartenant à 200 personnes.

Dans une population statique le taux de mutation orthographique est déterminant. Quand le taux de mutation est trop faible, la diversité décroît jusqu'à ce qu'il n'y ait plus qu'un patronyme dominant. À l'inverse, lorsque le taux de mutation est élevé, la fonction de fréquence $n(y)$ approche un état stationnaire, mais où les ensembles de noms de petite taille sont plus répandus que dans une population croissante.

Ces distributions d'états stationnaires n'apparaissent qu'après de nombreuses générations. Les déviations par rapport à l'état stationnaire attendu peuvent refléter des événements historiques récents. Par exemple, les patronymes japonais modernes ne sont apparus qu'il y a



3. L'ORTHOGRAPHIE DES PATRONYMES ANGLAIS est flexible. On voit ici un petit ensemble de variations sur le nom Hemingway, non exhaustif (les flèches indiquent les mutations apparues dans les différents recensements des États-Unis depuis 1800). Les auteurs ont étudié un modèle d'héritage du patronyme qui prévoit une probabilité petite, mais non nulle, qu'un nom quelconque mute quand il est transmis à la génération suivante. Ce modèle prédit que les mutations assurent une augmentation constante de la diversité des patronymes dans toute population dont l'effectif croît, et une diminution quand cet effectif est constant.

120 ans. On peut donc s'attendre à ce que la distribution des tailles de famille porte encore la trace de l'« état initial » d'il y a un siècle.

Du patronyme aux mitochondries

Une comparaison entre des données réelles provenant de trois sources (l'annuaire téléphonique de l'Argentine pour 1996, les noms commençant par A dans l'annuaire de 1996 de Berlin, et la liste des patronymes de cinq villes japonaises autour de 2000) confirment ces conclusions (voir la figure 2).

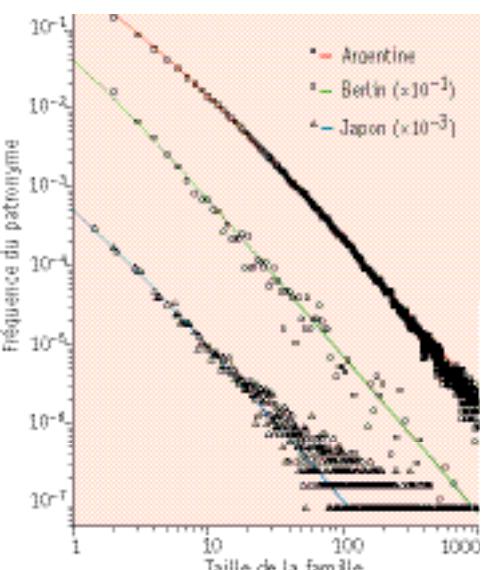
Dans une population statique le taux de mutation orthographique est déterminant. Quand le taux de mutation est trop faible, la diversité décroît jusqu'à ce qu'il n'y ait plus qu'un patronyme dominant. À l'inverse, lorsque le taux de mutation est élevé, la fonction de fréquence $n(y)$ approche un état stationnaire, mais où les ensembles de noms de petite taille sont plus répandus que dans une population croissante.

En revanche, les données japonaises s'écartent notablement de la distribution d'état stationnaire, avec un excès significatif de grandes familles. Dans un ou deux siècles, nous trouverions probablement une distribution plus serrée autour de la

droite. Cependant, cette préiction ne serait pas validée si la population du Japon devait connaître une période de stagnation prolongée.

Les scientifiques modernes attachent sans doute moins d'importance que les victoriens aux patronymes ou à la mort des « grandes familles », mais ils se soucient du devenir de l'ADN mitochondrial, qui se transmet selon le même mode que les patronymes. L'ADN mitochondrial d'une mère est transmis à tous ses enfants, mais seules les filles peuvent propager cet ADN jusqu'à la génération suivante. L'ADN mitochondrial est dotée d'une région particulière, la région de contrôle, de 500 paires de nucléotides, qui évoluerait de façon neutre : ce segment ne semble pas avoir de fonction, et ses mutations n'offrent pas d'avantage en termes de survie. Ainsi, la lente dérive génétique aléatoire de l'ADN mitochondrial est une excellente horloge génétique qui indique si deux personnes, ou deux groupes de personnes, ont un ancêtre commun et de quand il date.

La découverte de cette horloge a clos un certain nombre de débats historiques. Par exemple, l'ADN mitochondrial d'une femme que certains pensaient être la princesse Anastasia, la fille du dernier tsar de Russie, n'avait aucun lien avec celui des Romanov encore vivants. Les mitochondries des



2. LE MODÈLE DE PATRONYMES prédit le nombre de noms (exprimé en fréquence) qui auront une « taille de famille » donnée (ici, une famille est l'ensemble de tous les individus qui ont le même nom de famille). La fréquence prédictive pour chaque taille de famille dans le modèle (lignes colorées) est comparée aux données des annuaires de l'Argentine, de Berlin et du Japon. Les données de Berlin et du Japon ont subi une réduction d'échelle pour les besoins de la présentation. La pente des droites colorées est égale à environ -2, les échelles étant logarithmiques, on en déduit que le modèle est en $1/y^2$: en d'autres termes, le nombre de noms appartenant à y personnes exactement est proportionnel à $1/y^2$ quand y est suffisamment grand et que la population a atteint son état stationnaire. Les distributions de taille pour l'Argentine et pour Berlin s'accordent avec le modèle, mais les données japonaises s'en écartent notablement. Cet écart reflète probablement l'introduction relativement récente des patronymes au Japon : la distribution de taille des familles n'est pas encore stabilisée !

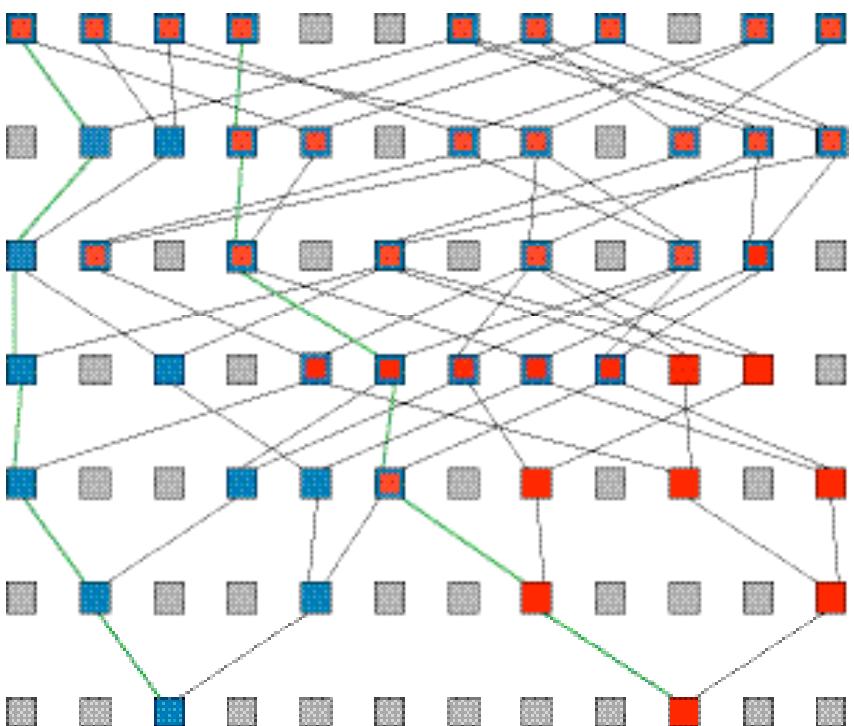
habitants des îles du Pacifique prouvent que leurs ancêtres venaient d'Asie, et non pas des Amériques. Enfin, les analyses d'ADN mitochondrial ont établi que les hommes de Néandertal se sont séparés de la lignée qui a produit l'homme moderne il y a environ 500 000 ans.

L'ADN mitochondrial a apporté un éclairage nouveau sur l'histoire humaine, mais il ne donne qu'une vision incomplète du tableau. Nos gènes renferment, potentiellement, autant de contributions qu'il y a d'ancêtres dans notre arbre généalogique. L'*« Ève mitochondriale »* et l'*« Adam du chromosome Y »* ne sont pas nécessairement contemporains, pas plus qu'ils n'habitaient nécessairement la même région, et ne sont pas forcément non plus les contributeurs les plus importants à notre patrimoine génétique. Si nous avions un ancêtre commun à un moment particulier de notre histoire, nous en avons probablement un grand nombre. L'*Ève*

mitochondriale est simplement la mère de la mère de la mère... de notre mère. L'analyse mitochondriale ne peut pas nous dire qui était le père de la mère du père de la mère de notre père... Certains de ces ancêtres communs indétectables ont peut-être vécu beaucoup plus récemment que l'*Ève mitochondriale*.

Un parent, ou deux ?

Par ailleurs, les ancêtres communs ne contribuent pas de façon égale à notre génome. Nos parents nous donnent chacun 50 pour cent de notre matériel génétique, mais la contribution de nos grands-parents n'est pas forcément de 25 pour cent, des changements ayant pu se produire. En reculant dans le passé, certains ancêtres ont pu voir leur contribution renforcée par coalescence généalogique : dans n'importe de branches nous rattachant à eux, leurs chances de nous avoir transmis de leur ADN augmentent.



4. LA COALESCENCE dans un modèle à deux parents survient beaucoup plus rapidement que dans un modèle à un parent. Chaque rangée représente une génération dans une population de taille constante de 12 individus. On choisit deux personnes dans la génération actuelle (en bas), et l'on suit leurs ancêtres sur six générations. L'ancêtre commun le plus récent (le carré rouge et bleu) apparaît dès la génération des grands-parents. Dans les générations précédentes, le nombre d'ancêtres partagés augmente jusqu'à la première génération où les descendants bleus et rouges se superposent. En revanche, quand le mode d'héritage ne fait intervenir qu'un seul parent, comme pour l'ADN mitochondrial, on ne détecte pas d'ancêtre commun en remontant six générations (lignes vertes).

Deux études récentes, l'une effectuée par notre équipe et l'autre par Joseph Chang, de l'Université Yale, ont mis en relief la différence entre les approches génétique et généalogique de la coalescence. Nos modèles mathématiques de généalogie et ceux que Chang sont semblables et peuvent être étendus à des populations de tailles diverses. Ces modèles remontent du présent vers le passé. Nous avons supposé que chaque individu « choisit » au hasard ses deux parents dans la génération précédente.

Dès lors, on peut étudier une variété de questions. Par exemple, si vous prenez deux personnes au hasard dans le présent, à combien de générations faut-il remonter pour trouver un ancêtre commun ? Pour que tous les ancêtres soient communs ?

Le nombre de génération pour retrouver le premier ancêtre commun, dans une population constante de n individus, est le logarithme à base 2 de n . Selon Chang, le nombre de générations, G , jusqu'à ce que deux individus aient le même ensemble d'ancêtres, est égal à 1,77 fois le logarithme à base 2 de n . On pourrait appeler cette valeur le « temps de coalescence » de la population. Pour une population de douze individus, on obtient environ 6,3 générations (*voir la figure 4*).

Nous avons choisi une approche différente de celle de Chang, en comparant le nombre de fois qu'un ancêtre donné apparaît dans deux arbres généalogiques distincts. Environ $\log n$ générations sont nécessaires pour qu'un ancêtre donné apparaisse dans tout arbre, avec une transition abrupte à quatorze générations environ (indépendante de la taille de la population) où la similarité (la probabilité qu'un individu ait comme ancêtres tous les individus présents à la génération n dans le passé qui ont des descendants à l'heure actuelle) passe brusquement de 1 à 99 pour cent (*voir la figure 5*).

Enfin, nous avons découvert qu'il n'y a pas seulement un ancêtre commun universel, mais plutôt une population ancestrale universelle. À l'époque de la coalescence, une dichotomie complète émerge, où chaque individu est soit un ancêtre de toutes les personnes de la génération actuelle, soit l'ancêtre d'aucune d'entre elles.

Ces modèles remettent donc en question les idées reçues en termes de généalogie et d'*« Èves mitochondrielles »*. Il est important de passer au crible les hypothèses que nous avons faites, pour vérifier ce qui est raisonnable et ce qui ne l'est pas.

Dans le monde réel, la sélection des parents (ou dans un modèle descendant le choix des conjoints) n'est pas aléatoire et dépend de la géographie, la race, la religion et la classe sociale. Cependant, dans des sous-populations où le mariage aléatoire peut être une réalité, un ensemble d'ancêtres communs émerge avec une surprenante rapidité, en quelques centaines d'années, et non pas quelques centaines de milliers d'années.

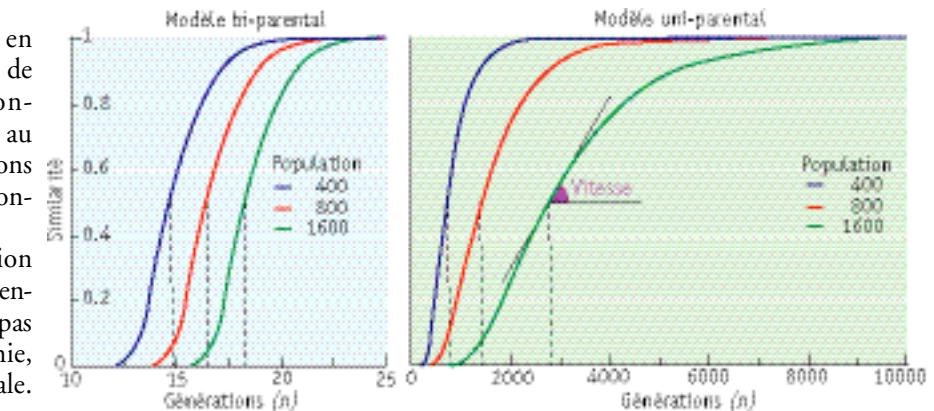
À l'inverse, l'homogénéité génétique au sein d'une population met beaucoup plus longtemps à émerger. Bien qu'un arbre généalogique ait la propriété de doubler le nombre d'ancêtres à chaque génération, ce n'est pas vrai pour les gènes individuels, qui sont nécessairement hérités le long de branches uniques et se conforment donc au modèle monoparental.

Ève de nouveau chassée

Ainsi, on pourrait définir un temps de coalescence génétique qui serait le nombre de générations requises pour atteindre un ancêtre commun pour tout allèle non recombinant donné (c'est un problème similaire à celui de l'ADN mitochondrial où chaque individu n'est relié qu'à un parent). Sir John Kingman, à l'Université de Cambridge, a montré que cette coalescence requiert un nombre de générations égal à la taille de la population où la similarité (la probabilité qu'un individu ait comme ancêtres tous les individus présents à la génération n dans le passé qui ont des descendants à l'heure actuelle) passe brusquement de 1 à 99 pour cent.

Enfin, nous avons découvert qu'il n'y a pas seulement un ancêtre commun universel, mais plutôt une population ancestrale universelle. À l'époque de la coalescence, une dichotomie complète émerge, où chaque individu est soit un ancêtre de toutes les personnes de la génération actuelle, soit l'ancêtre d'aucune d'entre elles.

L'analyse de l'ADN mitochondrial a été riche d'enseignements sur



5. LES MODÈLES À DEUX PARENTS montrent que dans les populations fermées où les couples se forment de façon aléatoire, une coalescence totale des descendants survient très vite. Ici, deux mesures différentes de la « similarité » génétologique sont utilisées. Dans un modèle à deux parents (a), la « similarité » représente la probabilité qu'un individu ait comme ancêtres tous les individus présents à la génération n dans le passé qui ont des descendants à l'heure actuelle. La vitesse de la transition (la pente des parties rectilignes des courbes) de l'absence universelle de parenté à une parenté universelle ne dépend pas de la taille de la population. Ce n'est pas le cas du moment (les étoiles) où se produit cette transition. Dans un modèle à un parent (b), la « similarité » représente la probabilité que tous les individus de la population actuelle partagent l'ADN mitochondrial de la même personne il y a n générations. Pour des populations de quelques centaines de personnes, il faut plus de 1 000 générations avant d'avoir une forte probabilité que toutes les personnes aient le même ancêtre mitochondrial. Dans ce cas, la vitesse de la transition dépend de la taille de la population.

l'évolution humaine. Bien qu'essentiel, l'ADN mitochondrial représente une petite partie de notre génome. Sa pertinence en termes d'origines et de relations entre les différents groupes humains tient à son mode particulier de transmission par la lignée maternelle, analogue à celle des patronymes par la lignée paternelle. Cependant, notre descendance génétique est beaucoup plus vaste, car nous savons qu'une large fraction de toute population vivant il y a de nombreuses générations est incluse dans notre arbre généalogique. Notre patronyme, comme notre ADN mitochondrial, n'apporte qu'une petite quantité d'information sur nos origines.

Une partie importante de l'information contenue dans les gènes mitochondriaux concerne la production d'énergie. Toutefois, l'essentiel de l'information qui nous caractérise en tant qu'individu réside dans nos gènes nucléaires, qui constituent 99,99 pour cent du génome humain. Ces gènes se mélangent à chaque fois qu'un couple se reproduit, par le processus de recombinaison. Si nous pouvions suivre toutes les branches par où nous ont été transmis nos gènes, nous trouverions probablement que toutes ces personnes figurant sur notre arbre généalogique ont contribué à notre

héritage génétique. En même temps que l'*Ève mitochondriale*, la plupart de ses contemporains ont probablement laissé leur empreinte silencieuse dans le génome collectif.

La prochaine fois que vous entendrez quelqu'un se vanter d'une ascendance royale, ne vous laissez pas intimider : il y a de grandes chances pour que vous aussi ayez de nobles ancêtres. L'entre croisement rapide des branches généalogiques, en quelques dizaines de générations seulement, rendrait le contraire surprenant. La seule question est celle de la quantité de « sang royal » dont votre ami peut se targuer. Généalogie ne signifie pas gènes, et la question de notre proximité génétique reste posée.

Nous remercions la revue *American Scientist* de nous avoir accordé la permission de reproduire cet article.

Susanna MANRUBIA travaille au Centre d'astrobiologie de Madrid. Bernard DERRIDA est professeur de physique à l'Université Pierre et Marie Curie. Damián ZANETTE est professeur à l'Institut Balseiro, en Argentine.

S. MANRUBIA et D. ZANETTE, *At the boundary between biological and cultural evolution: The origin of surname distributions*, in *Journal of Theoretical Biology*, vol. 216, pp. 461-477, 2002.

