médecine/sciences 2024 ; 40 : 965-70

Des nouvelles du passé Histoires de paléogénétique (2)

La paléo-épigénétique : quand les dommages de l'ADN deviennent des traces épigénétiques

Céline Bon

L'immense majorité des travaux sur l'ADN ancien portent sur l'analyse de la séquence nucléotidique du génome, qui permet d'identifier ses mutations et de reconstituer son histoire évolutive. Pourtant, d'autres informations portées par la molécule d'ADN permettent de comprendre comment s'expriment les gènes. C'est l'objet de l'épigénétique, ou, quand elle est appliquée à l'ADN ancien, de la paléo-épigénétique.

Contrairement à la paléogénétique, pour laquelle les dommages taphonomiques[1] subis par l'ADN au cours du temps constituent un problème, la paléo-épigénétique s'appuie sur l'identification de certains de ces dommages et des traces qu'ils laissent dans la molécule d'ADN pour en extraire des informations sur l'activité fonctionnelle des génomes. Deux types de modifications, en particulier, sont analysées : les cassures de la molécule d'ADN et la désamination des cytosines.

Les dommages de l'ADN détectables dans l'ADN ancien

L'épigénétique étudie l'ensemble des processus transmissibles lors de la division cellulaire, réversibles et adaptatifs, qui modifient l'expression des gènes sans affecter la séquence nucléotidique de l'ADN. Il s'agit, dans la plupart des cas, de méthylation, acétylation, ou phosphorylation des histones, ou de la méthylation des cytosines en 5-méthylcytosines au sein des dinucléotides CpG. En utilisant les traces taphonomiques sur l'ADN ancien, des études paléogénétiques tentent également aujourd'hui d'en extrapoler les marques épigénétiques [1].

UMR7206 Éco-Anthropologie (EA), CNRS, Muséum national d'Histoire naturelle, Université Paris-Cité, Paris, France.
celine.bon@mnhn.fr

Des cassures de l'ADN qui ne sont pas aléatoires

Chaque nucléotide de l'ADN est formé de l'association entre un sucre (désoxyribose) et une base azotée (adénine, thymine, guanine, ou cytosine), unis par une liaison N-glycosylique. Dans des conditions physiologiques, la rupture de cette liaison se produit à un rythme de $3{\cdot}10^{-11}$ évènement par seconde si la base azotée est une purine (adénine ou guanine), et beaucoup plus lentement, de l'ordre de $1{,}5{\cdot}10^{-12}$ évènement par seconde, si la base azotée est une pyrimidine (cytosine ou thymine) [2]. Dans un génome humain composé de $3{\cdot}10^{9}$ nucléotides, une dépurination se produit donc approximativement toutes les 100 secondes, et 10 000 sites « abasiques » (principalement apuriniques) sont produits, en moyenne, chaque jour.

Dans une cellule vivante, la réparation de ces dommages est effectuée par le système de réparation d'excision de base (*base-excision repair*). En revanche, dans le cas de l'ADN ancien, une réaction chimique se produit entre le site abasique de l'ADN et une molécule d'acétaldéhyde, ce qui conduit à l'élimination du résidu phosphate et à une cassure du brin d'ADN (*Figure 1A*). Dans des conditions physiologiques, cette réaction s'effectue avec une demi-vie de 200 heures, mais elle peut être accélérée par un environnement basique (pH > 7) ou en présence d'autres composés organiques, comme la putrescine (1,4 diaminobutane) présente dans les cadavres.

Vignette.(© Céline Bon).

[1] La taphonomie est la discipline qui étudie tous les processus intervenant après la mort d'un organisme dans les sites archéologiques et paléontologiques (ndlr).

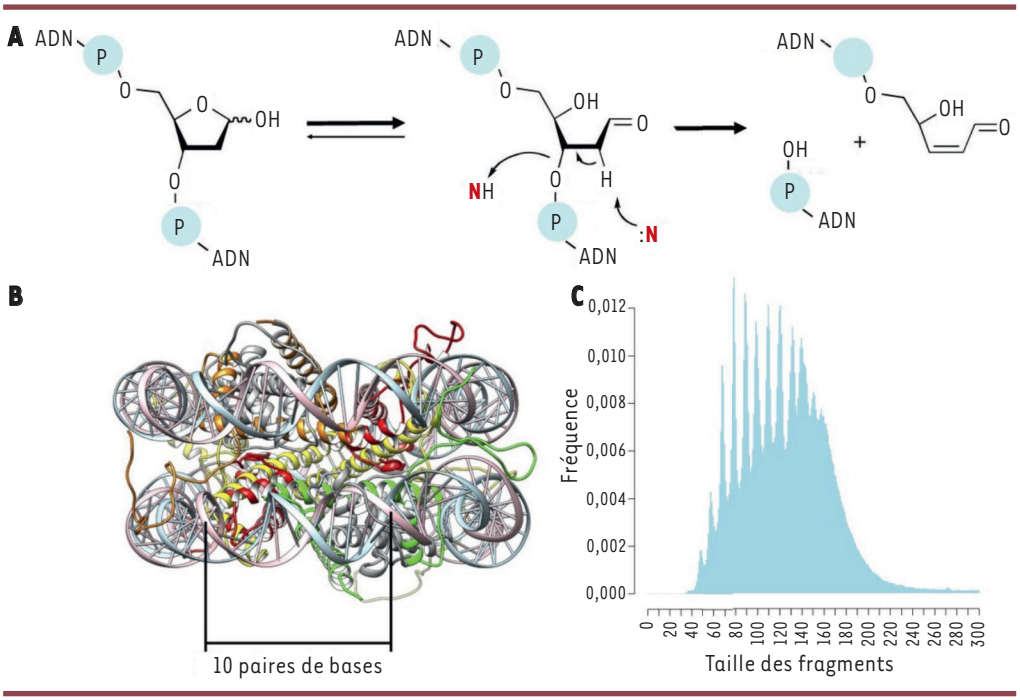


Figure 1. A. Réaction de coupure du brin d'ADN pouvant survenir dans l'ADN ancien. Un équilibre se forme entre un nucléotide abasique et un acétaldéhyde ouvert. En présence d'une molécule nucléophile (N, en rouge), cela peut conduire à l'élimination du résidu phosphate (disque bleu) en 3', et à une rupture du brin d'ADN. ***B. Les brins d'ADN*** *(en bleu/rose)* ***s'enroulent autour d'un complexe de huit histones*** *(en vert/orange/jaune/rouge)* ***pour former l'unité de base de la chromatine : le nucléosome.*** La taille du fragment de la molécule d'ADN constitutif d'un nucléosome est d'environ 146 paires de bases (pb). Sur un même brin d'ADN, un nucléotide est exposé vers l'extérieur toutes les 10 bases. ***C. Profil des tailles des fragments d'ADN dans un échantillon d'ADN ancien conservé dans un climat sec (désertique).*** La taille de la plupart des fragments varie de 10 en 10 pb, et le rythme se perd après 150 pb, taille du fragment d'ADN constitutif d'un nucléosome.

Aux échelles de temps de l'ADN ancien, le clivage du brin d'ADN après une réaction de dépurination est donc considéré comme presque immédiat. En raison de la différence entre les vitesses des réactions de dépurination et de dépyrimidination, les ruptures de la molécule d'ADN sont situées, dans leur immense majorité, après des purines [3]. En utilisant des collections d'ADN dans lesquelles les deux brins de chaque fragment d'ADN ont été séquencés indépendamment, et en comparant les positions initiales et finales de chacune des deux lectures, Bokelmann *et al.* [4] ont pu montrer que la rupture sur l'un des brins induit généralement une rupture sur l'autre brin à une distance de moins de 5 paires de bases (pb). Les fragments d'ADN ancien possèdent donc des extrémités simple brin de quelques nucléotides.

L'analyse de 187 génomes anciens [5] a montré que, dans le cas des échantillons conservés dans un climat sec (froid ou désertique), la taille des fragments suit une périodicité de 10 pb, soit le nombre de bases séparant deux nucléotides d'un même brin exposé sur la face externe du nucléosome (*Figure 1B, C*). Cette rupture non aléatoire de la double hélice serait due à une relative préservation des protéines histones, et donc à la conservation dans le temps de l'enroulement de l'ADN autour du nucléosome.

La désamination des cytosines

Un autre dommage très fréquent de l'ADN est la désamination des cytosines, qui résulte d'une réaction d'hydrolyse conduisant à la perte d'un groupement amine et à la transformation de la cytosine en uracile. Si la cytosine a été précédemment méthylée (comme souvent dans les zones riches en dinucléotides CpG), elle est transformée en thymine (*Figure 2A*). Dans les organismes vivants, des mécanismes de réparation permettent de reconnaître l'uracile et de l'exciser, grâce à l'uracile-ADN glycosylase, avant de réparer le brin.

Cette désamination reste relativement rare lorsque l'ADN est sous sa forme native de double brin (sa demi-vie est de 30 000 à 85 000 ans en conditions physiologiques) [2], mais la réaction est beaucoup plus rapide lorsque l'ADN est sous forme de simple brin (demi-vie de 200 ans). Ce type de dommage de l'ADN demeure donc rare, sauf, dans le cas de l'ADN ancien, dans des portions d'ADN simple brin qui sont présentes aux extrémités des fragments d'ADN dégradés (*Figure 2B*). L'étude des profils de désamination des cytosines dans des prélèvements archéologiques [5] montre que l'intensité de ce type de dommage de l'ADN, mesurée par la fréquence des transformations de cytosines en uracile ou en thymine par rapport à une séquence de référence, est corrélée non seulement à l'ancienneté des restes, mais également à la température à laquelle les échantillons ont été préservés.

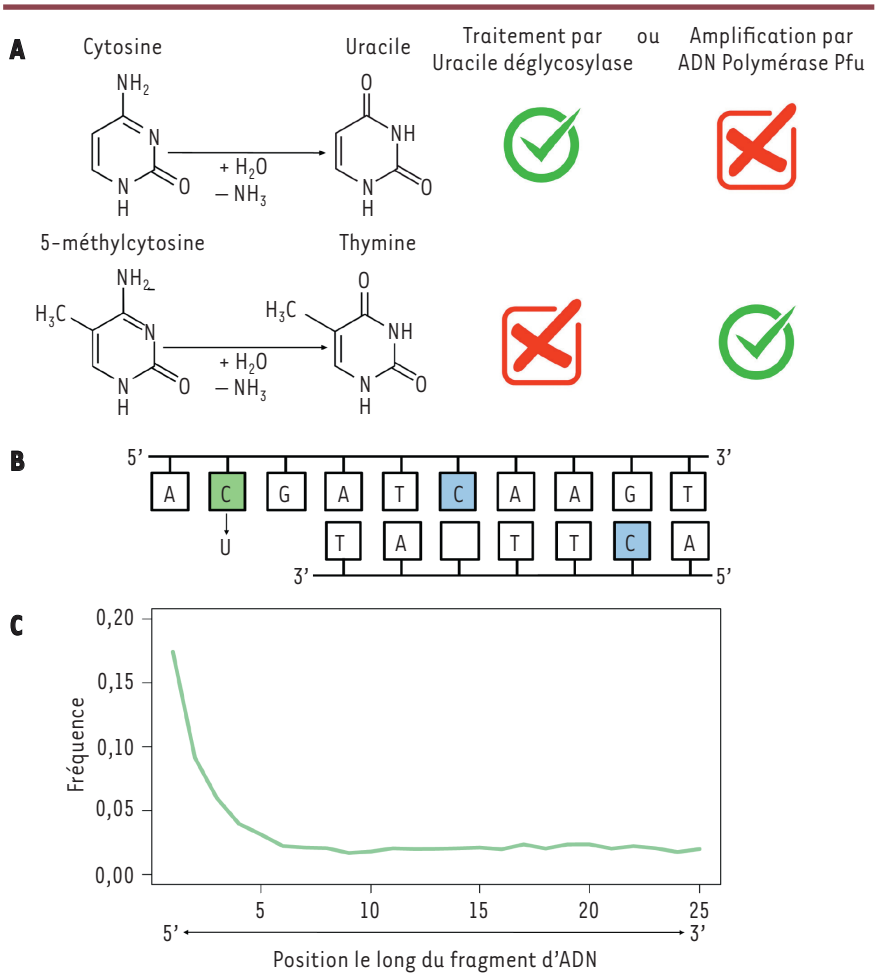


Figure 2. Les endommagements des bases azotées donnent accès au paléoépigénome

A. La désamination des cytosines non méthylées conduit à leur transformation en uracile. Ces bases uraciles sont éliminées par un traitement avec l'uracile-ADN déglycosylase (UDG). Ces uraciles ne peuvent également pas être amplifiés par la technique de PCR (*polymerase chain reaction*) avec une ADN polymérase thermostable telle que l'ADN polymérase Pfu. En revanche, la désamination des cytosines méthylées conduit à leur transformation en thymines, insensibles au traitement par l'enzyme UDG, et les fragments qui les contiennent peuvent donc être amplifiés par l'ADN polymérase Pfu. ***B.*** Ces désaminations de cytosines se produisent particulièrement dans les zones où l'ADN a une structure en simple brin, à l'extrémité d'un fragment (cytosine verte), et sont beaucoup plus rares dans les zones où l'ADN a une structure en double brin (cytosines bleues). ***C.*** Cela conduit à un excès de transformations des cytosines méthylées en thymine (C⟶T) aux extrémités 5' des fragments d'ADN ancien (observés ici avec le logiciel mapDamage) par rapport à une séquence d'ADN de référence.

De la détection des dommages de l'ADN ancien aux analyses paléo-épigénétiques

Quelles traces laissent ces modifications après plusieurs milliers d'années ?

Puisque la préservation des protéines histones et de leur liaison avec la molécule d'ADN améliore sa conservation, il devient possible, en analysant la profondeur de séquençage du génome, de reconstituer les régions génomiques riches en histones (les mieux couvertes) et les plus pauvres en histones (moins bien couvertes), et donc d'en déduire une carte de compaction de l'ADN ancien. Cependant, pour des raisons techniques, cette approche n'est que rarement utilisée en pratique.

D'un autre côté, les méthylations des cytosines dans les dinucléotides CpG sont des modifications stables qui perdurent au cours du temps, et qui peuvent être mises en évidence grâce à la désamination des cytosines. En effet, alors qu'un traitement par l'uracile-ADN glycosylase permet de dégrader l'uracile issu d'une cytosine désaminée, la 5-méthylcytosine désaminée forme une thymine, qui résistera à ce traitement. En absence de traitement par l'enzyme, la différence d'amplification des séquences contenant ces bases modifiées, en fonction du type d'ADN polymérase utilisée, permettra aussi de les différencier (*Figure 2A*). Ces méthodes permettent donc de distinguer, dans l'ADN, les zones fortement méthylées de celles qui ne le sont pas.

Cette approche théorique fonctionne dans la pratique. L'analyse de l'ADN extrait de cheveux d'un inuit de culture Saqqaq[2] [6] et datant de plus de 4 000 ans montre, après amplification par une ADN polymérase Pfu[3], une multiplication par 5 du taux de conversion des cytosines en thymine dans les dinucléotides CpG par comparaison au reste du génome : dans ces régions, les cytosines sont donc fortement méthylées et sont transformées en thymine après désamination. L'analyse fonctionnelle de ces régions hyperméthylées montre qu'elles sont associées à une forte expression des gènes codant les kératines 71 et 85 (*KRT71*, *KRT85*), et à une sous-expression du gène codant la kératine 79 (*KRT79*), un gène non exprimé dans les cheveux. Ce profil d'expression génique, déduit de l'analyse de l'ADN extrait de ces

[2] Culture préhistorique du Groenland.

[3] ADN polymérase thermostable provenant de l'archée hyperthermophile *Pyrococcus furiosus*.

cheveux datant de 4 000 ans, est donc similaire à celui de cheveux actuels, et est différent de celui d'autres tissus. Même si ce constat n'est pas étonnant en soi, il confirme qu'un signal épigénétique est bien conservé dans certains échantillons archéologiques !

Malheureusement, les tissus mous sont rarement retrouvés en archéologie, et la quasi-totalité des données obtenues proviennent de tissus osseux ou dentaires, et sont donc uniquement informatives sur la régulation des gènes exprimés dans les cellules du squelette. Une approche paléo-épigénétique peut en revanche être déployée pour comprendre les mécanismes biochimiques à l'origine des différences morphologiques entre Homme de Néandertal, Homme de Denisova, et *Homo sapiens* [6]. Une analyse comparative, entre ces trois espèces humaines, des zones hyperméthylées ou hypométhylées dans l'ADN extrait du tissu osseux a révélé des différences dans la régulation de 1 100 régions génomiques. Parmi les régions concernées par ces différences de méthylation, on trouve, par exemple, une séquence régulatrice du groupe de gènes du complexe *HOXD* (*homeobox D cluster*), qui est impliqué dans le développement des membres : alors que cette région est hypométhylée chez *Homo sapiens*, elle est hyperméthylée dans les génomes des espèces archaïques Denisova et Neandertal. Or, la morphologie des membres de l'Homme de Néandertal est effectivement différente de celle de *Homo sapiens* : les membres de Néandertal sont, en particulier, plus robustes que ceux de Sapiens. Les résultats d'une étude chez la souris ont montré que l'introduction expérimentale de modifications pour moduler de façon similaire l'expression de ce groupe de gènes conduit à des articulations plus larges, des radius et fémurs courbés, et des membres plus courts : précisément ce qui est observé chez Néandertal [7].

Il devient donc tentant d'appliquer cette approche aux individus ayant vécu la plus grande transformation économique et sociale qu'ait connue l'humanité, c'est-à-dire le passage d'un mode de vie de chasseur-cueilleur à celui d'éleveur-agriculteur au Néolithique. La comparaison de profils de méthylation entre six éleveurs-agriculteurs et chasseurs-cueilleurs européens [8] montre que les profils de méthylation des individus post-Néolithique sont plus proches entre eux qu'ils ne le sont de ceux des chasseurs-cueilleurs. En particulier, les gènes des éleveurs-agriculteurs sont plus souvent hyperméthylés que ceux des chasseurs-cueilleurs.

Cependant, la même analyse montre également que des biais importants peuvent être introduits lors de l'extraction de l'ADN et de son analyse. Cette question a été étudiée de façon plus systématique par les chercheurs d'un autre laboratoire, utilisant le *pipeline*[4] PALEOMIX [9] pour déterminer les traces épigénétiques laissées par le Néolithique. Trente-quatre génomes séquencés et annotés par six laboratoires différents, et couvrant la fin du Paléolithique et le début du Néolithique à travers l'Eurasie, ont été analysés [10], conduisant à la conclusion que le principal facteur explicatif de la variabilité demeure le laboratoire d'origine de l'échantillon et non le mode de vie de l'individu dont l'échantillon est issu, sa population d'origine ou son sexe génétique.

Est-ce à dire que toute analyse épigénétique d'ADN ancien est vouée à l'échec ? Pas inévitablement. En revanche, il est nécessaire d'homogénéiser les pratiques de laboratoire, d'augmenter la profondeur d'analyse des échantillons testés, voire de développer de nouveaux protocoles mieux adaptés aux études paléo-épigénétiques.

Une autre piste d'application se dessine : la détermination de l'âge au décès. Les études archéo-anthropologiques sur le squelette permettent de déterminer assez précisément (à deux ans près) l'âge au décès des enfants et des jeunes adultes, en se fondant sur les calendriers d'éruption dentaire et de disparition des différents cartilages métaphyso-épiphysaires (cartilages de croissance). Cependant, au-delà de 25 ans, le squelette n'évolue plus, laissant un flou sur l'âge au décès des individus étudiés.

Or, un changement dans la signature de méthylation de centaines de sites dans l'ADN se produit au cours du temps, ce qui permet de définir un âge épigénétique [11]. En analysant le profil de méthylation de génomes anciens à ces emplacements, il serait alors possible de déterminer l'âge au décès des individus étudiés. Cette analyse a, par exemple, été conduite sur l'ADN de chevaux anciens, et a permis de définir leur âge au décès avec une marge d'erreur de moins d'un an [12]. Une approche similaire serait donc envisageable pour les êtres humains, même si quelques limites persistent. La principale est la création d'un référentiel fondé sur l'ADN issu de tissu osseux, car la plupart des études menées sur des êtres humains modernes ne se fondent que sur de l'ADN extrait de cellules sanguines. Afin de prendre en compte les transformations taphonomiques de l'ADN, ce référentiel devrait être lui aussi ancien, nécessite l'accès à des collections de restes humains pour lesquels la détermination de l'âge au décès est précise (par exemple, des restes provenant d'hôpitaux), et devrait avoir été établi en suivant un protocole similaire à celui utilisé pour les échantillons étudiés, ce qui restreint *de facto* son utilisation par d'autres laboratoires. Ces approches seraient donc très coûteuses, et ne seraient pas forcément adaptées à une utilisation « routinière » par les archéo-anthropologues. Par ailleurs, la précision de cet âge épigénétique est également discutable : en effet, cette signature est très sensible aux conditions environnementales qu'a connues l'individu (en particulier au stress et aux maladies), voire à son origine populationnelle.

[4] Processus informatique de stockage des données brutes, qui sont ingérées à partir de diverses sources, transformées, puis transférées vers un entrepôt de données à des fins d'analyse (ndlr).

De la structure des chromosomes à leur conformation tridimensionnelle dans le noyau

C'est parfois la conformation de l'ADN au sein du noyau qui peut être conservée. Des chercheurs ont pu reconstituer la conformation tridimensionnelle d'un génome de mammouth à partir de cellules de peau d'un individu femelle en utilisant la technique Hi-C [13]. Cette technique permet de reconstituer la conformation de la chromatine native en reliant artificiellement les fragments de chromatine spatialement proches, puis, après digestion de l'ADN par des enzymes de restriction, en séquençant ces fragments préalablement rattachés l'un à l'autre par ligation, et enfin, en les localisant, par bio-informatique, dans la molécule d'ADN [14] (➜).

(➜) Voir le Dossier technique de G. Umlauf, *m/s* n° 3, mars 2015, page 304

En modifiant cette technique pour l'adapter à de l'ADN ancien, Sandoval-Velasco *et al.* ont pu démontrer que la conformation tridimensionnelle du génome était conservée dans le cas du mammouth. Vingt-huit régions chromosomiques ont même pu être identifiées et localisées, fournissant ainsi un résultat identique à celui obtenu chez les autres espèces d'éléphantidés. Mieux encore, la technique a permis de montrer un nombre réduit d'interactions spatiales entre les fragments de chromatine issus du chromosome X et ceux issus des autres chromosomes dans les cellules de ce mammouth femelle, une preuve indirecte de l'existence de corpuscules de Barr[5] chez les individus femelles.

L'approche Hi-C permet, en particulier, de distinguer les régions composées de la chromatine active (euchromatine), contenant les gènes pouvant être transcrits, de celles composées de chromatine inactive (hétérochromatine), contenant des gènes dont l'expression est réprimée. De cette façon, il a été possible d'identifier les régions qui étaient les plus exprimées dans le génome de mammouth : la signature correspond à celle des cellules de la peau des éléphants, plus qu'à celle des cellules de tout autre tissu.

Ainsi, il semble que, dans ce cas particulier, la structure tridimensionnelle du noyau, révélatrice des régions exprimées ou réprimées de la chromatine, a été conservée. Les résultats obtenus ont montré que 820 gènes sont exprimés différemment dans la peau des mammouths par comparaison aux gènes du même tissu des autres éléphantidés, environ la moitié de ces gènes étant surexprimés, et l'autre moitié sous-exprimés. Parmi ces gènes, on retrouve le gène *Edaradd*, interagissant avec le gène *Edar* (un gène connu pour être sous pression de sélection chez les Inuits adaptés au froid), ou le gène *Egfr*, dont la mutation chez les êtres humains est associée à une croissance pileuse excessive. Cela pourrait expliquer l'aspect « laineux » du mammouth : les gènes associés à la pilosité sont ceux dont l'expression est la plus différente entre ce mammouth et les autres éléphantidés étudiés.

Évidemment, ces résultats n'ont pu être obtenus que grâce à un concours de circonstances extraordinaire : après son décès, le corps de ce mammouth a été rapidement congelé et dessiqué, atteignant, selon les auteurs, un état proche de la vitrification. Cet état aurait conservé la structure du noyau cellulaire, malgré le fractionnement de l'ADN. Ce phénomène pourrait être fréquent dans le pergélisol : en effet, des structures tridimensionnelles du génome ont pu être déterminées chez un autre mammouth conservé dans des conditions similaires. Il est donc probable que l'application de cette technique à la paléontologie humaine reste limitée à un cas très particulier : celui des momies conservées dans le sol gelé.

Après ce rapide tour d'horizon, que pouvons-nous dire des approches paléo-épigénétiques ? L'expérience a montré qu'il était possible de reconstituer des cartes épigénétiques des génomes anciens, et d'identifier les régions actives et inactives de la chromatine. Les applications pourraient être multiples : de la détermination des âges au décès, et des facteurs de stress dans des populations anciennes, jusqu'à la compréhension des mécanismes évolutifs à l'œuvre dans la spéciation et l'extinction d'espèces, en passant par une meilleure compréhension des mécanismes épigénétiques d'adaptation des populations humaines aux changements environnementaux et culturels. Mais pour ce faire, il est important de démontrer qu'il est possible de s'extraire d'une approche individuelle, et que les résultats obtenus par l'analyse de quelques échantillons sont généralisables, si ce n'est à tous les échantillons, au moins à une part significative d'entre eux. ◊

Paleoepigenetics: when DNA damages become epigenetic traces

LIENS D'INTÉRÊT

L'auteure déclare n'avoir aucun lien d'intérêt concernant les données publiées dans cet article.

[5] Les femelles des espèces de mammifères possèdent deux chromosomes X, tandis que les mâles n'en ont qu'un seul. Chez les individus femelles, afin d'éviter un surdosage de l'expression des gènes portés par le chromosome X, l'un des deux chromosomes X est inactivé et étroitement compacté dans chacune de leurs cellules, formant le « corpuscule de Barr ».

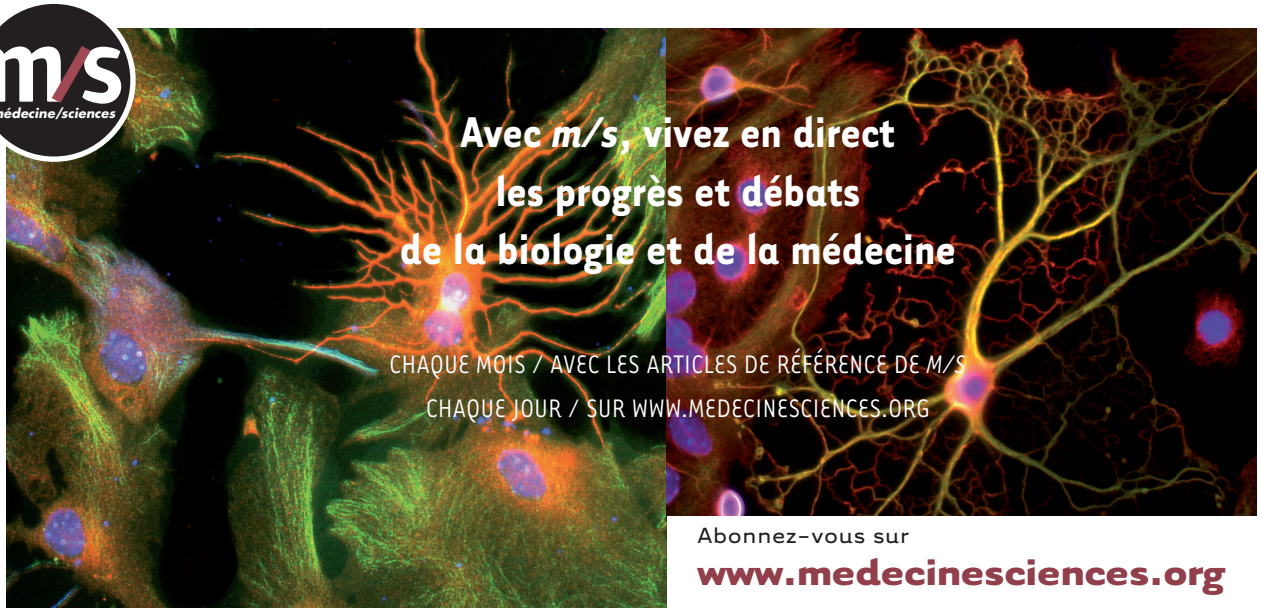
RÉFÉRENCES

1. Orlando L, Gilbert MTP, Willerslev E. Reconstructing ancient genomes and epigenomes. *Nat Rev Genet* 2015 ; 16 : 395-408.
2. Gates KS. An Overview of Chemical Processes That Damage Cellular DNA: Spontaneous Hydrolysis, Alkylation, and Reactions with Radicals. *Chem Res Toxicol* 2009 ; 22 : 1747-60.
3. Briggs AW, Stenzel U, Johnson PLF, *et al.* Patterns of damage in genomic DNA sequences from a Neandertal. *Proc Natl Acad Sci U S A* 2007 ; 104 : 14616-21.
4. Bokelmann L, Glocke I, Meyer M. Reconstructing double-stranded DNA fragments on a single-molecule level reveals patterns of degradation in ancient samples. *Genome Res* 2020 ; 30 : 1449-57.
5. Kistler L, Ware R, Smith O, *et al.* A new model for ancient DNA decay based on paleogenomic meta-analysis. *Nucleic Acids Res* 2017 ; 45 : 6310-20.
6. Pedersen JS, Valen E, Velazquez AMV, *et al.* Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. *Genome Res* 2014 ; 24 : 454-66.

RÉFÉRENCES

7. Zakany J, Duboule D. The role of Hox genes during vertebrate limb development. *Curr Opin Genet Dev* 2007 ; 17 : 359-66.
8. Seguin-Orlando A, Donat R, Sarkissian C Der, *et al.* Heterogeneous Hunter-Gatherer and Steppe-Related Ancestries in Late Neolithic and Bell Beaker Genomes from Present-Day France. *Curr Biol* 2021 ; 31 : 1072-83.e10.
9. Schubert M, Ermini L, Sarkissian C Der, *et al.* Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nat Protoc* 2014 ; 9 : 1056-82.
10. Çokoğlu SS, Koptekin D, Fidan FR, *et al.* Investigating food production-associated DNA methylation changes in paleogenomes: Lack of consistent signals beyond technical noise. *Evol Appl* 2024 ; 17.
11. Simpson DJ, Chandra T. Epigenetic age prediction. *Aging Cell* 2021 ; 20.
12. Liu X, Seguin-Orlando A, Chauvey L, *et al.* DNA methylation-based profiling of horse archaeological remains for age-at-death and castration. *iScience* 2023 ; 26 : 106144.
13. Sandoval-Velasco M, Dudchenko O, Rodríguez JA, *et al.* Three-dimensional genome architecture persists in a 52,000-year-old woolly mammoth skin sample. *Cell* 2024 ; 187 : 3541-62.e51.
14. Umlauf D. Le génome intime et en trois dimensions. *Med Sci (Paris)* 2015 ; 31 : 304-11.

TIRÉS À PART
C. Bon



Abonnez-vous sur
www.medecinesciences.org