Chapitre 29

Les stratégies et enjeux de l'analyse de l'ADN ancien

Eva-Maria Geigl et Thierry Grange
Institut Jacques Monod, Equipe « Epigénome et Paléogénome », CNRS, UMR 7592, Paris, France

29.1 Introduction

Le patrimoine biologique ancien peut être, grâce aux approches paléogénétiques et paléogénomiques, l'objet d'analyses génétiques qui permettent d'accéder à des informations phénotypiques et populationnelles sur les organismes passés. Ceci nous permet d'éclairer d'un jour nouveau les études de la biodiversité du passé et de son évolution sous l'influence du climat et des êtres humains. En effet, l'analyse de l'acide désoxyribonucléique (ADN) ancien récupéré à partir d'os et de dents nous procure l'information sur la biologie des plantes, des animaux et des humains du passé. On peut ainsi déterminer leurs taxons et suivre la dynamique des populations (migrations, extinctions, expansions, remplacements, etc.). On peut déceler certains caractères phénotypiques d'organismes anciens et, récemment, déchiffrer l'information génétique complète, le génome. L'étude génétique de cet ADN ancien correspond à la discipline appelée paléogénétique, que l'on appelle paléogénomique quand elle s'intéresse à l'analyse simultanée de nombreux marqueurs génétiques, voire du génome entier. Cette discipline est une branche de la génétique et de la génomique qui met un accent particulier sur l'analyse évolutive des êtres vivants. Elle déborde largement de ce cadre puisqu'elle peut apporter des résultats importants pour d'autres disciplines comme l'anthropologie, l'histoire, l'archéologie et ses sous-disciplines que sont l'archéozoologie et l'archéobotanique. Si les connaissances qu'elle peut apporter à ces autres disciplines peuvent donner à penser à certains que la paléogénétique est une technique au service des disciplines historiques, cette vue réductrice reflète surtout la méconnaissance de ce domaine et l'adoption d'un point de vue limité à la discipline que l'on pratique. Les innovations futures dans le domaine impliqueront les scientifiques qui sauront faire interagir fructueusement les diverses disciplines sans en instrumentaliser certaines au service d'autres, et restreindre ainsi leurs dynamiques évolutives.

Dans le matériel biologique frais, l'ADN intact est présent sous forme de très grandes molécules en double hélice de plusieurs dizaines, voire centaines, de millions de nucléotides, ses unités de base. Par contre, l'ADN est très dégradé dans les tissus anciens qui sont, pour les animaux et les êtres humains, surtout des os et dents, ces tissus étant généralement les seuls préservés. En effet, plus de 99% de l'ADN d'origine est perdu et il n'en reste, dans le meilleur des cas, que des petites molécules de moins de 100 nucléotides en quantité infime. Pour cela, l'analyse de l'ADN ancien exige l'utilisation des méthodes expérimentales de biologie moléculaire les plus avancées. Leur adaptation à ce substrat particulier qu'est l'ADN ancien requiert de nombreuses optimisations méthodologiques, ainsi que des laboratoires de haut-confinement et des procédures d'analyse extrêmement strictes. Le problème majeur rencontré lors de l'analyse de l'ADN ancien est la production de résultats artéfactuels résultant de faux négatifs et de faux positifs. Si les méthodes ciblent des fragments d'ADN trop longs, ou ne sont pas optimisées pour permettre la détection de molécules très peu abondantes présentes à des nombres unitaires de molécules, l'ADN endogène peut échapper à la détection (faux négatifs). Lorsque des méthodes ultrasensibles sont utilisées, le risque est de détecter de l'ADN moderne contaminant, lui aussi présent à l'état de traces et de le confondre avec de l'ADN ancien (faux positifs). Il est donc nécessaire d'utiliser des conditions expérimentales extrêmement rigoureuses pour prévenir la contamination par de l'ADN moderne. Les sources de contamination par de l'ADN moderne étant extrêmement diverses, il est indispensable d'utiliser des procédures complexes attaquant toutes les sources potentielles. Plus les méthodes de détection sont sensibles, plus le risque de contamination est élevé. Il faut donc développer en parallèle les approches d'optimisation de détection et de prévention de la contamination pour pouvoir analyser des spécimens contenant peu d'ADN. La rigueur expérimentale mise en œuvre, l'expérience et un équipement sophistiqué sont des conditions qui ont un impact déterminant pour pouvoir poser certaines questions nécessitant l'obtention d'un grand nombre de résultats et/ou s'attaquant à des questions concernant des spécimens anciens provenant de régions connaissant des températures élevées.

Le sort normal d'un organisme vivant et de ses parties squelettaires après sa mort étant sa totale disparition, voire son recyclage dans la nature, il est rare de trouver des ossements anciens et encore plus rare d'en trouver qui contiennent de l'ADN endogène autre que l'ADN des microorganismes qui les colonisent dès la mort et qui constituent la plus grande partie de l'ADN présent dans un os ou une dent archéologique. Ce sont les conditions climatiques et physico-chimiques qui décident quels types de microorganismes se reproduisent le mieux dans le sédiment d'enfouissement et dans l'os ou la dent. C'est le métabolisme des microorganismes ainsi que les conditions physico-chimiques agissant directement sur les molécules d'ADN qui décident du sort de l'ADN. Plus la température lors de l'enfouissement est élevée, moins l'ADN va être préservé. Cet effet sera encore accentué si l'organisme est enfoui dans un environnement chaud et humide. Par contre, l'ADN dans les organismes enfouis dans les régions géographiques très froides, et surtout dans le permafrost, est le mieux conservé. Les os et dents enfouis dans la glace du permafrost se comportent lors de l'analyse paléogénétique presque comme des spécimens biologiques frais. Ce

n'est pas le cas pour les os et dents enfouis dans des régions chaudes, comme les déserts, ou chaudes et humides, comme les régions tropicales, mais aussi dans des régions où le sol est caractérisé par une acidité importante. Pour ces échantillons, le taux de réussite des analyses paléogénétiques peut être très faible et il faut analyser beaucoup d'ossements pour en trouver qui contiennent suffisamment de molécules d'ADN préservé pour qu'une analyse paléogénétique soit possible, fiable et couronnée de succès. Ce sont aussi ces échantillons où le danger de la production de résultats faux-positifs est le plus élevé.

29.2 Analyse de l'ADN ancien

29.2.1 Stratégies d'analyse

Lorsque l'on cherche à analyser les traces d'ADN préservé dans des ossements ou des dents archéologiques, on est confronté à des choix expérimentaux stratégiques. Selon les questions auxquelles on cherche à répondre, on choisira soit l'analyse d'une cible unique sur l'ADN, comme c'est le cas pour l'ADN mitochondrial, qui est plutôt utilisé pour faire des analyses phylogénétiques de populations, ou l'analyse d'un ou de plusieurs gènes nucléaires pour chercher des variations génétiques associées à quelques caractères phénotypiques particuliers, comme la couleur du pelage par exemple (voir article de Pruvost et Geigl dans ce volume), ou encore des génomes entiers, si l'on veut analyser de façon systématique de multiples aspects de l'évolution des spécimens.

La mitochondrie est un organite présent dans le cytoplasme et qui permet aux cellules de respirer. Elle contient un petit génome d'environ 16000 nucléotides, ce qui est 2 millions de fois plus petit que le génome présent dans le noyau d'une cellule de mammifères. L'avantage de l'ADN mitochondrial est sa présence en multiples copies, plusieurs centaines, voire milliers par cellule, ce qui augmente la probabilité de trouver des traces de ces molécules. C'est la raison pour laquelle ce fut pendant longtemps la molécule de prédilection des paléogénéticiens. Un autre avantage de cette molécule, qui se comporte comme un locus unique, est l'absence de recombinaison, c'est-à-dire la production de séquences chimériques à partir des séquences paternelles et maternelles, comme cela est le cas pour les gènes portés par les chromosomes dans le noyau des cellules. Ceci résulte du fait que l'ADN mitochondrial est transmis par voie maternelle, les mitochondries des spermatozoïdes étant perdues au moment de la fécondation. Ceci permet de suivre les lignées maternelles au cours de l'évolution dans le temps comme dans l'espace. Par contre, cette molécule livre peu d'information phénotypique. Elle contient un nombre limité de gènes dont la plupart jouent un rôle essentiel dans la respiration des cellules. Les mutations délétères sont efficacement purgées par la sélection et il n'est pas clair qu'il y ait des mutations donnant lieu à des propriétés phénotypiques adaptatives sous sélection positive. Une petite fraction est particulièrement étudiée pour les analyses phylogénétiques. Il s'agit d'une région appelée « région de contrôle » ou « région hypervariable ». Cette région ne code pas pour une protéine mais a une fonction de contrôle, ce qui explique son nom. Elle ne subit pas une pression de sélection forte et évolue donc plus rapidement et est beaucoup plus variable au sein d'une population que le reste

du génome mitochondrial ou les gènes nucléaires. Cette propriété permet de suivre la diversification génétique d'une population sur des distances évolutives courtes, ce qui est une troisième raison de l'intérêt que de nombreux chercheurs lui ont portée pour effectuer des analyses phylogénétiques.

Pour analyser des traits phénotypiques, comme la couleur de la robe des animaux, il faut connaître les gènes nucléaires qui codent pour le caractère en question ainsi que les mutations qui distinguent les allèles codant pour les différents traits phénotypiques. Généralement, cela implique qu'une étude génétique approfondie des populations modernes ait été effectuée pour que les mutations responsables des traits phénotypiques aient été identifiées, comme c'était les cas chez les chevaux (voir Pruvost et Geigl, ce volume). Deux versions de chacun de ces gènes sont présentes dans les noyaux, une copie maternelle et une copie paternelle. La probabilité que l'on trouve encore de ces molécules après des milliers d'années est beaucoup plus faible que pour les multiples copies de l'ADN mitochondrial mais cette approche a été couronnée de succès avec des échantillons provenant de régions froides.

C'est avec l'optimisation des méthodes du séquençage massivement parallèle que l'analyse de l'ADN ancien a vraiment pris son essor puisqu'il est désormais possible d'analyser des génomes entiers, mitochondriaux et même nucléaires, et des fragments d'ADN encore plus petits d'environ 30-70 nucléotides qui sont préservés dans plus d'ossements que les molécules plus longues. Ce séquençage ne cible pas un locus génétique mais toutes les molécules d'ADN présentes dans l'extrait.

29.2.2 Extraction de l'ADN

Afin de pouvoir extraire l'ADN ancien des os et dents des vertébrés, il faut le « libérer » de la matrice osseuse constituée d'hydroxyapatite. Pour cela, on utilise des produits chimiques appelés chélateurs qui ont la propriété de piéger les cations divalents, comme le Ca^{2+} et le Mg^{2+}, constituants essentiels de la matrice osseuse, qui forment des cristaux avec les groupes phosphates et hydroxyles. Ainsi l'os est déminéralisé et l'ADN qui était adsorbé à cette matrice solubilisé. Il faut ensuite purifier cet ADN ce qui peut être fait par différentes méthodes. La méthode qui semble la plus adaptée actuellement est basée sur l'adsorption des molécules d'ADN sur une matrice d'oxyde de silice, suivi par un lavage et une élution dans des conditions de pH et de concentration saline bien définies. Cette méthode de purification enlève une grande part des autres molécules qui auraient pu interférer avec l'analyse de l'ADN.

29.2.3 Multiplication de l'ADN par PCR

La très faible quantité des molécules d'ADN ancien présentes dans les vestiges biologiques du passé et en particulier dans les os et dents des vertébrés, nécessite la multiplication du matériel afin d'en obtenir une quantité suffisante pour pouvoir l'analyser avec les méthodes de biologie moléculaire. La multiplication de l'ADN est possible grâce à des réactions de polymérisation en chaîne, appelées PCR, à l'aide d'une enzyme, une ADN polymérase thermorésistante, et d'amorces d'ADN, c'est-à-dire, de petites molécules d'ADN simple brin complémentaires aux séquences d'ADN que l'on veut analyser [1]. Grâce à ces amorces complémentaires aux extrémi-

tés de la région d'ADN ciblée (Figure 1A), il sera possible de doubler à chaque cycle la quantité d'ADN que l'on cible, ce qui permet de produire à partir d'une molécule d'ADN initiale des milliers de milliards de copies au bout d'une trentaine de cycles. La PCR est donc un outil génétique extrêmement puissant qui est au cœur des analyses génétiques depuis de nombreuses années. Toutefois, les réactions en jeu étant complexes, le potentiel qu'offre la procédure d'amplifier une seule molécule à partir d'un mélange d'ADN très hétérogène requiert une optimisation expérimentale significative pour éviter que des réactions parasites réduisent l'efficacité du phénomène. L'ADN ancien étant très dégradé, plus la séquence d'ADN ciblée est courte, plus la probabilité de trouver une molécule de cette taille devient importante. Par contre, plus la séquence est courte, moins l'information génétique obtenue sera élevée, d'autant plus que l'information génétique correspondant aux séquences complémentaires aux amorces utilisées est perdue. Il faut donc trouver le meilleur compromis entre quantité d'information obtenue par échantillon dans une expérience et proportion d'échantillons qui pourront produire des résultats paléogénétiques.

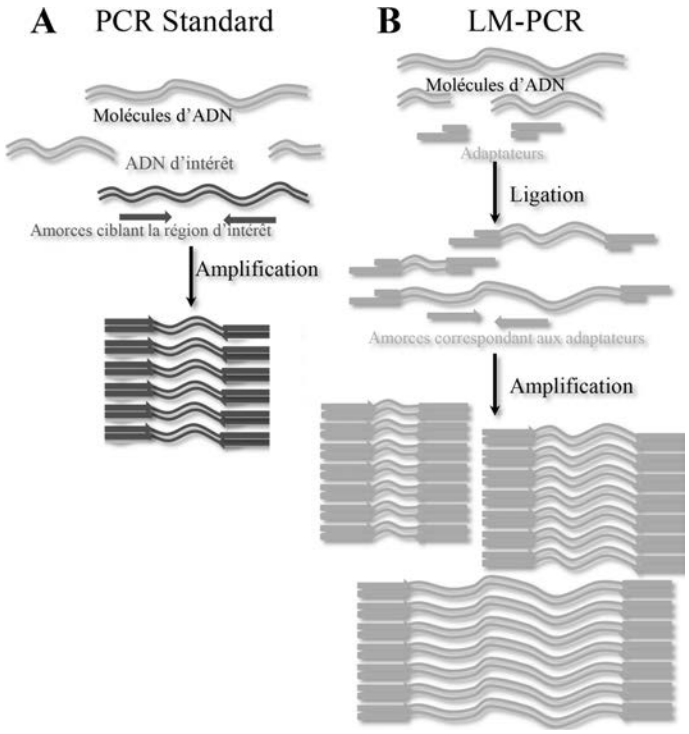


Figure 1 – Comparaison du principe des méthodes d'amplification par PCR qui soit cible une région génomique précise (A : PCR standard), soit permet d'amplifier toutes les molécules après ligation d'un adaptateur (B : ligation-mediated PCR).

D'autre part, les extraits préparés à partir d'échantillons fossiles peuvent contenir malgré leur purification d'autres molécules qui interfèrent avec ou qui inhibe la PCR. Les procédures de purification doivent être optimisées pour trouver le meilleur compromis entre rendement de récupération de l'ADN contenu dans le fossile et

élimination des molécules inhibitrices. Ce compromis peut être insatisfaisant lorsque les fossiles contiennent un grand nombre de molécules inibitrices dont les propriétés physicochimiques ressemblent suffisamment à celles de l'ADN pour qu'elles ne puissent être séparées d'elles lors de la purification. Il faut donc pouvoir quantifier l'effet inihibiteur de chaque extrait indépendamment de la quantification des molécules d'ADN cibles qu'il contient pour utiliser les conditions optimales d'amplification par PCR.

Pour réaliser toutes ces quantifications nécessaires aux optimisations qui peuvent augmenter considérablement la capacité d'analyse d'une grande proportion d'échantillons difficiles, nous utilisons la PCR quantitative en temps réel pour toutes les analyses que nous effectuons. Cette méthode permettant l'analyse de la cinétique de la réaction d'amplification est très performante pour optimiser la PCR, pour mesurer la puissance de l'effet inhibiteur de l'extrait osseux tout en facilitant l'identification de produits parasites de la PCR qui peuvent fausser les résultats [2]. Elle est particulièrement adaptée à l'analyse efficace d'un grand nombre de PCRs.

Une fois que l'ADN est amplifié, son analyse implique généralement la détermination de sa séquence nucléotidique. Pour cela, les produits de PCR sont ensuite purifiés et séquencés par des méthodes chimiques, comme la méthode de Maxam et Gilbert, basée sur la dégradation chimique sélective des molécules d'ADN [3], ou enzymatiques basées sur le principe de séquençage par synthèse *in vitro* de copies d'ADN complémentaire à une matrice par les polymérases, méthode inventé par Sanger [4]. En utilisant des nucléotides fluorescents et l'électrophorèse capillaire, cette dernière méthode a été automatisée. Une autre approche, le pyroséquençage, ajoute les nucléotides non pas tous en même temps, comme le fait la méthode selon Sanger, mais l'un après l'autre. Lorsqu'un nucléotide complémentaire au nucléotide situé dans le brin d'ADN cible est incorporé, une suite de réactions enzymatiques permet la production d'un signal lumineux qui est enregistré par une caméra haute définition [5]. Alternativement, les produits de PCR peuvent être clonés dans des bactéries avant séquençage, une stratégie qui a été largement répandue au début, mais dont l'utilité, au-delà de l'identification des modifications chimiques (diagénétiques) des nucléotides dues aux attaques chimiques au cours du temps, n'est pas évidente et nous paraît superflue (voir discussion dans [6 , 7]).

29.2.4 La contamination : problème et solutions

Par contre, cet outil fantastique qu'est la PCR est une épée à double tranchant : d'un côté, il permet la multiplication par milliards d'un très faible nombre de molécules d'ADN ancien, de l'autre côté, il peut être trompeur puisque des molécules d'ADN moderne environnemental sont aussi multipliées, et cela même mieux puisqu'elles ne contiennent pas de lésions. Il est donc très facile de produire des résultats artéfactuels et il faut utiliser des protocoles extrêmement stricts pour éviter et éliminer la contamination avec l'ADN moderne, source majeure de faux positifs et de conclusions erronées dans le domaine. La contamination pouvant être introduite à n'importe quelle étape du processus d'analyse, il faut la prévenir ou s'en débarrasser à toutes les étapes. Une seule faille peut compromettre la fiabilité de toute l'analyse. De plus, il faut considérer qu'aucune procédure n'est fiable à 100%, au mieux

leur efficacité varie entre 95 et 99,9%. Il est aussi important d'être conscient que la procédure de décontamination utilisée peut être rendue inefficace par un autre composant ou réactif utilisé et/ou peut diminuer considérablement la sensibilité de la méthode de détection utilisée. Il est donc nécessaire de mesurer l'efficacité de chaque procédure utilisée dans les conditions d'utilisation pour éviter les comportements « mystiques » inefficaces ou délétères basés sur des rumeurs ou des a priori non fondés (voir [8]). D'autre part, il est recommandé de combiner des approches redondantes s'attaquant à la même source de contamination de façon indépendante. Ainsi, si l'on combine deux méthodes différentes qui préviennent ou éliminent chacune 99% des contaminants, on peut atteindre une efficacité de 99,99%. L'expérience que nous avons accumulée dans la matière montre aussi que le combat est incessant : lorsque la source majoritaire de contamination a été traitée, on voit apparaître la contamination provenant d'une autre source moins abondante dont l'effet était masqué auparavant. Il faut donc à chaque fois identifier la source de contamination et la traiter, ce qui nécessite souvent d'utiliser de nouvelles stratégies adaptées à la source de contamination. Les précautions doivent commencer dès le prélèvement des échantillons qui, surtout dans le cas d'échantillons humains, doivent être effectués en évitant tout contact avec la peau, les cheveux et la sueur des expérimentateurs. Pour cela, il faut porter des gants et utiliser des instruments décontaminés avec l'eau de Javel, produit chimique dégradant efficacement l'ADN et donc l'ADN contaminant les surfaces de travail, les outils et les gants [8]. Au laboratoire de paléogénétique à l'institut Jacques Monod (`http://www.univ-paris-diderot.fr/Mediatheque/spip.php?rubrique76`), le traitement des échantillons anciens est ensuite effectué dans un laboratoire de haut confinement où aucun échantillon d'ADN moderne n'a jamais été introduit, où tout matériel entrant est décontaminé avec l'eau de Javel et où les expérimentateurs portent des vêtements et gants javellisés, des combinaisons, des masques. De plus, les différentes étapes expérimentales sont séparées les unes des autres en les effectuant dans des enceintes nettoyées à l'eau de Javel, aux rayons UV et à l'ozone. Les étapes expérimentales qui suivent l'extraction, la purification de l'ADN ancien et la mise en place de la réaction de PCR, sont effectuées dans d'autres laboratoires. En effet, la multiplication enzymatique de l'ADN ancien via la PCR ainsi que la purification des produits de PCR pour le séquençage sont séparés physiquement des étapes précédentes pour minimiser le transfert d'ADN entre différents laboratoires, stations de travail, équipements, échantillons, et expérimentateurs. Cette minimisation et ce confinement physique sont nécessaires pour ce travail, mais ils ne sont malheureusement pas suffisants. Une des sources de contamination avec l'ADN moderne est l'étape expérimentale de la multiplication par PCR au cours de laquelle sont produites des milliards de copies de la molécule ciblée par l'analyse. Les aérosols formés lors de l'ouverture des tubes contenant ces produits de PCR contiennent des milliers de molécules de cibles qui contaminent immanquablement les mains gantés de l'utilisateur et tout le matériel manipulés à la suite. Cette contamination dispersée dans un environnement plus ou moins large selon la rigueur des procédures et l'attention portée aux gestes pratiqués peut ensuite contaminer l'expérimentateur, ou tout autre expérimentateur pratiquant dans les mêmes lieus, plusieurs jours, semaines, voire mois ou années, plus tard. Même si une seule molécule est transférée dans un tube dans lequel l'ADN extrait de l'échantillon ancien est

amplifié, on court le risque que l'analyse porte sur les molécules d'ADN moderne et non pas sur les anciennes. Aucune des procédures et confinements physiques utilisés par les laboratoires de paléogénétique ne peut être considérées suffisamment strictes pour garantir que cet événement ne se produise pas compte tenu du nombre de molécules initiales concernées et de l'efficacité relative des procédures de prévention de la contamination suivie. C'est pour cela que l'équipe de paléogénétique à l'institut Jacques Monod effectue toutes les PCRs en présence d'un nucléotide qui naturellement ne se trouve pas dans l'ADN et qui « marque » donc toutes les molécules synthétisées lors de la PCR. Si, malgré la rigueur de la procédure expérimentale, des molécules provenant d'étapes de PCR précédentes « atterrissent » dans les mélanges réactionnels constitués pour amplifier l'ADN extrait d'échantillons anciens (ce que l'on appelle contamination « carry-over»), la majeure partie va être détruite avant l'amplification de l'ADN ancien grâce à un traitement enzymatique qui précède chaque amplification [9]. Ainsi en combinant des procédures visant à limiter la diffusion des molécules produites lors de l'analyse des produits de PCR, en effectuant un contrôle physique strict pour prévenir l'entrée de ces molécules dans le laboratoire de préparation de l'ADN à partir des fossiles et en éliminant enzymatiquement les éventuelles molécules qui auraient échappées aux procédures précédentes, on peut garantir un niveau de sécurité qui permet d'avoir la certitude de l'authenticité de tout résultat dupliqué.

Une autre source importante de contamination qui est souvent sous-estimée est celle provenant des réactifs utilisés pour l'analyse de l'ADN qui sont produits par les sociétés de biotechnologie et qui contiennent des traces d'ADN originaires d'animaux domestiqués, comme le bœuf et le cochon [8, 10]. En effet, les enzymes utilisées pour effectuer la PCR sont stabilisées avec des protéines extraites à partir de tissus de ces animaux. La purification de ces protéines n'est pas suffisante pour que quelques molécules d'ADN des animaux en question ne soient copurifiées. Les réactifs sont de plus significativement contaminés par de l'ADN humain et bactérien car les procédures utilisés par les employés des sociétés de biotechnologie lors de la production des réactifs sont loin d'être aussi rigoureuses que celles mises en œuvre dans un laboratoire de paléogénétique exigeant comme le nôtre. Ainsi on peut trouver de tels contaminants jusque dans les amorces nécessaires pour la PCR. La décontamination de cet ADN contaminant sans affecter l'ADN synthétique de ces amorces pose un défi important car il s'agit de distinguer des molécules ayant les mêmes propriétés physicochimiques. Nous avons dû développer des stratégies spécifiques pour traiter cette source de contamination. Ainsi, nous utilisons des combinaisons de traitements enzymatiques ainsi que physiques, c'est-à-dire à l'aide de rayons UV ou gamma, pour détruire des molécules contaminant ces différents réactifs [8].

Finalement, il est important d'avoir bien à l'esprit qu'il ne suffit pas de faire un petit nombre de contrôles expérimentaux pour conclure que l'on n'est pas victime de la contamination. Lorsque la contamination est présente à l'état de traces, et que l'on analyse des échantillons contenant de l'ADN à l'état de traces, la confusion est aisée. De simples considérations statistiques permettent de réaliser l'ampleur de la rigueur nécessaire et de voir qu'il faut réaliser un grand nombre de contrôles expérimentaux

et utiliser des tests statistiques rigoureux pour s'assurer de la fiabilité des analyses [8].

L'application de l'ensemble de ces procédures constitue un filet de sécurité avec des mailles extrêmement fines qui permettent de réduire considérablement l'impact des molécules modernes contaminantes. Ceci est particulièrement critique pour analyser des ossements difficiles contenant peu d'ADN assez dégradé. Ainsi, nous avons pu analyser beaucoup d'ossements néolithiques difficiles, soit parce qu'ils étaient préservés dans un sol acide comme celui du Lœss, ou dans une aire géographique d'un climat chaud, conditions qui sont peu favorables à la conservation de l'ADN. Les méthodes de prévention de la contamination, particulièrement celles traitant aussi les réactifs, se sont avérées essentielles pour prévenir les faux-positifs et ainsi analyser avec fiabilité la domestication des bovidés en Asie-du Sud-Ouest (Pruvost et Geigl, ce volume). Grâce à la rigueur de nos procédures, nous avons pu identifier de nombreuses sources de production de faux-positifs et sommes en mesure d'apprécier que le taux d'erreur peut être important lorsque ces procédures ne sont pas utilisées, ce qui nous amène à faire preuve d'un certain scepticisme quant au taux de fiabilité de résultats préalablement publiés.

29.2.5 Dommages à l'ADN ancien et conséquences

Au cours du vieillissement, l'ADN qui a échappé à la dégradation biologique initiale est transformé par des réactions chimiques qui dégradent lentement les brins ou modifient les bases. Particulièrement, les bases puriques (Adénine et Guanine) sont séparées des sucres qui les lient au polymère, ce qui favorise ensuite la rupture du polymère au niveau du sucre abasique. D'autre part, l'ADN ancien contient des nucléotides qui sont modifiés chimiquement au cours de la diagénèse de l'os, majoritairement des cytosines désaminées que l'on trouve enrichies aux extrémités des molécules d'ADN, probablement à cause de la nature simple brin des extrémités produites par le processus de rupture chimique de ces brins [11]. Ces cytosines désaminés seront interprétées comme des thymines par la polymérase synthétisant des nouvelles molécules d'ADN lors de la PCR et l'enzyme introduira de manière erronée une adénine à la place de la guanine qu'elle aurait dû insérer face à la cytosine dans le brin matrice. Ces erreurs de lecture conduisent à l'incorporation de mauvais nucléotides qui changent l'information génétique de la molécule d'ADN. Ceci peut devenir un problème lorsque la PCR a démarré à partir d'une seule molécule d'ADN ancien car ces « lésions diagénétiques » seront interprétées comme étant des vraies substitutions de nucléotides. C'est pour cela qu'il faut toujours reproduire au moins deux fois indépendantes chaque amplification d'ADN ancien car il est hautement improbable que ces lésions diagénétiques surviennent au même nucléotide dans deux molécules d'ADN ancien différentes. La reproductibilité des résultats est donc une exigence incontournable de toute recherche en ADN ancien.

Dans le système de décontamination des produits « carry-over » de la PCR, les mêmes cytosines désaminées sont détruites. En effet, les cytosines désaminés sont éliminées par l'enzyme que nous utilisons pour détruire la contamination « carry-over » (uracil-N-glycosylase). Ce traitement présente donc l'avantage de réduire le problème de la contamination tout en éliminant une autre source d'erreur de séquence due à la

transformation diagénétique principale de l'ADN. Nous avons pu démontrer l'intérêt de cette approche dans la fiabilisation des données paléogénétiques [6]. Certains chercheurs du domaine qui n'utilisent pas cette stratégie malgré le fait qu'elle soit reconnue comme une procédure précieuse pour les applications biomédicales et judiciaires de la PCR le justifient en argumentant avec une crainte hypothétique que cela pourrait diminuer le taux de succès des analyses en éliminant des molécules qui auraient pu être amplifiées si elles gardaient ces erreurs. On peut s'interroger sur la pertinence de la préférence donnée à la production de résultats erronés plutôt qu'à l'absence de production de résultats. Au-delà de cet aspect, cette crainte n'est pas corroborée par les données expérimentales. D'une part, nous n'avons pas observé de différence significative du taux de succès des amplifications avec et sans traitement préalable par l'UNG d'un grand nombre d'échantillons fossiles. D'autre part, la majorité de ces cytosines désaminés se trouve aux extrémités des brins d'ADN à des positions où leur élimination à peu de chance d'affecter l'hybridation des amorces utilisées pour la PCR [11], donnant ainsi une base théorique solide à notre observation que le traitement ne diminue pas significativement le taux de succès des amplifications. Le risque majeur de l'utilisation de l'UNG est donc la diminution du taux d'erreurs et de faux-positifs.

29.2.6 Séquençage massivement parallèle de nouvelle génération

Depuis 2005, on dispose de méthodes de séquençage, dites de « nouvelle génération » permettant le séquençage à très haut débit grâce aux approches massivement parallèles. Le principe de base des techniques actuelles de séquençage à haut débit est que l'étape de clonage de l'ADN dans des vecteurs procaryotes a été éliminée et que les molécules d'ADN sont soit amplifiées *in vitro* de manière clonale (on parle de « polonies », contraction de PCR et colonie) et séquencée directement, soit séquencée à partir de molécules uniques non amplifiées. Ceci permet de séquencer simultanément plusieurs centaines de milliers à plusieurs dizaines de millions de polonies ou de molécules. Les différentes technologies sur le marché se distinguent par le nombre total de lectures simultanées, la longueur des lectures, le coût par base, lecture ou réaction, et le type et la fréquence d'erreur de séquençage. Elles profitent d'une immobilisation de toutes les molécules d'ADN d'une « banque » sur un support, billes ou lames de verre, suivie le plus souvent par une amplification clonale par PCR pour générer des polonies sur ce support, suivies d'une lecture de la séquence par des ADN polymérase ou des ADN ligases en analysant simultanément l'ensemble des polonies. La lecture de la séquence repose soit sur une lecture de lumière en utilisant le pyroséquençage mentionné auparavant (technologie 454 ou Roche GSFlex), soit en incorporant des nucléotides fluorescents avec une ADN polymérase (technologie Solexa ou Illumina), soit des oligonucléotides fluorescents avec une ADN ligase (technologie Solid), soit en mesurant grâce à des puces électroniques les microchangements de pH résultant de l'incorporation de nucléotides naturels par une ADN polymérase (technologie Ion Torrent). Les technologies analysant des molécules uniques reposent elle aussi sur l'incorporation de nucléotides fluorescents (technologies Helicos et Pacific Biosciences), ou de microchangements de courant électrique (technologie Oxford Nanopore). Ces méthodes de molécules uniques n'ont pas encore réussi à offrir des débits, fiabilité de séquençage et rap-

port qualité-prix suffisants pour être généralisées et occuper une part de marché économiquement viable.

Ces nouvelles méthodes de séquençage massivement parallèle ont révolutionné la paléogénétique et l'on fait rentrer dans l'ère de la paléogénomique. Ainsi ont été séquencé les génomes anciens d'un mammouth [12], d'un paléoesquimo [13], des Néanderthaliens [14], d'une lignée humaine ancienne inconnue auparavant et appelée Dénisovienne [15], de l'agent pathogène de la peste noire médiévale [16], ainsi que des multiples génomes mitochondriaux d'organismes anciens. Il est même devenu possible de séquencer des molécules uniques d'ADN, ce qui a déjà donné lieu au séquençage d'un génome mitochondrial de 700 000 ans d'un cheval préservé dans le permafrost [17]. Ces techniques offrent maintenant la possibilité d'une part d'analyser de manière exhaustive l'information génétique contenue dans un petit nombre d'échantillons, mais aussi d'analyser moins d'information génétique mais sur un grand nombre simultané d'échantillons à moindre coût et beaucoup plus grande efficacité qu'avec les techniques disponibles auparavant.

29.2.7 Construction des banques d'ADN pour le séquençage

Pour construire les banques d'ADN pour le séquençage massivement parallèle en utilisant des approches qui reposeront sur la production de polonies, les molécules d'ADN vont être préparées pour permettre une amplification par PCR de chaque molécule individuellement sans cibler de séquences particulières. Cette amplification reposera sur la ligation d'adaptateurs d'ADN synthétique aux extrémités de toutes les molécules d'ADN. Ces adaptateurs serviront de cible pour une PCR utilisant des amorces leur correspondant (Figure 1B).

Généralement, la grande majorité des molécules d'ADN présentes dans un fossile est originaire de l'environnement dans lequel les os et les dents sont enfouis. Lorsque l'ensemble des molécules d'ADN est séquencé, on n'obtient ainsi qu'une faible proportion de séquences que l'on peut attribuer à l'organisme à l'origine du fossile, la majeure partie sera soit attribuable à des microorganismes (bactéries et champignons), des plantes dans une moindre proportion, ou ne sera pas attribuable à des organismes connus, la diversité des microorganismes environnementaux n'ayant pas encore été correctement référencée dans les banques de séquences. Ainsi, lors du premier séquençage du génome néandertalien, seulement 3.4 % des séquences produites ont pu être attribuées au Néandertalien [18]. Exceptionnellement, il arrive que l'on trouve des restes qui contiennent une proportion beaucoup plus importante de séquences endogènes, comme dans le cas des restes d'organismes préservés dans le permafrost, par exemple les mammouths (e.g., [12]), ou des cas rarissimes de préservation exceptionnelles, comme la phalange de la fille de Dénisova [15], dont l'exceptionnelle préservation semble plus relever du miracle dont il ne faut pas attendre qu'il se renouvelle souvent.

Puisque c'est un très grand gaspillage de ressources de ne produire que quelques pourcents de séquences utiles, plusieurs stratégies d'enrichissement en information utile ont été développées, bien que celles actuellement publiées offrent un taux de succès modérés et causent la perte d'une partie du matériel génétique ancien que

l'on souhaite analyser [14, 19]. La capture par hybridation à des séquences complémentaires fixées sur un support solide permet d'enrichir en séquences d'intérêt [20]. D'autres approches reposant sur la capture appliquées à l'ADN ancien sont actuellement en cours de développement dans plusieurs laboratoires dont le nôtre.

L'impact de la contamination comme source de faux positifs lors de la production de données paléogénomiques est moins bien exploré. Puisque la méthodologie permet d'amplifier des fragments d'ADN de petite taille qui sont les plus abondants dans un extrait de fossile, un certain nombre de chercheurs considèrent que le risque d'avoir des résultats faussés par la contamination est moindre que lorsque l'on analyse des fragments d'ADN plus grands. Ceci est basé sur le principe que l'ADN ancien serait plus dégradé que l'ADN contaminant, et donc contiendrait une beaucoup plus grande proportion de fragments courts que de fragments longs, tandis que l'ADN moderne étant moins dégradé, montrerait une taille moyenne des fragments plus importantes, avec moins de fragments courts et plus de fragments longs. Lorsque l'on analyse des fragments courts, il faudrait un degré de contamination beaucoup plus important pour que les fragments courts modernes soient plus abondants que les fragments courts anciens, alors que lorsque l'on analyse les fragments longs, un degré de contamination beaucoup moins important suffirait pour que les contaminants modernes soient majoritaires. Ce raisonnement est étayé par l'analyse d'un des premiers jeux de données paléogénomiques sur des ossements Néandertaliens qui a montré que les fragments longs séquencés correspondaient surtout à de l'ADN humain moderne, alors que les plus courts contenaient plus d'ADN néandertalien [21]. Bien qu'il soit raisonnable de considérer que l'ADN contaminant moderne ait une taille moyenne de fragments plus élevée que l'ADN ancien, la différence n'est pas nécessairement aussi importante que cela. D'une part, selon l'origine de l'ADN contaminant, la taille moyenne des fragments peut être petite. En effet, si l'ADN provient de cellules qui sont mortes par mort cellulaire programmée (apoptose), l'ADN est fragmenté naturellement en fragments d'environ 150 paires de base dès qu'il est libéré dans l'environnement. Sa taille peut ensuite être réduite encore plus à la suite de l'attaque par les microorganismes de l'environnement. Si la contamination est un peu ancienne, elle peut aussi avoir été affectée par les processus chimiques d'hydrolyse, et voir ainsi sa taille moyenne encore réduite. Par contre, les fragments d'ADN que l'on trouve dans les fossiles peuvent parfois avoir des tailles moyennes assez élevées, car cet ADN peut provenir de cellules qui n'ont pas subies cette mort programmée et qui étaient de plus vraisemblablement préservées dans un environnement favorable pour la conservation à long terme. L'argument que la taille moyenne des fragments devrait permettre de discriminer à tous coups l'ADN ancien de l'ADN moderne et rendre caduque le problème de la contamination ne repose donc pas sur une base théorique solide et n'est pas clairement étayé par les données actuelles. Il faut garder à l'esprit que la situation peut être différente d'un fossile à l'autre, et d'une source de contamination à l'autre. Il est important de toujours bien effectuer les contrôles pour quantifier la teneur maximale en contaminant dans les conditions d'obtention des données utilisées (voir par exemple, les estimations de la teneur maximale en séquences contaminées effectuées par le groupe de S. Pääbo dans le cas du génome des Néandertaliens [22]).

Certains chercheurs proposent d'utiliser en plus de la taille, un autre critère d'authenticité : la présence de lésions diagénétiques identifiables aux extrémités des fragments d'ADN, particulièrement la présence de cytosine désaminées, identifiable dans les données de séquence par des transitions C->T et G->A [11 , 23]. Ce critère ne peut être considéré suffisant à lui seul car si la contamination est un peu ancienne, de telles transformations peuvent aussi avoir eu lieu [23]. D'autre part, il semble préférable d'éliminer les cytosines désaminées avant de construire les banques pour pouvoir obtenir des données de séquence contenant moins d'erreurs, comme cela a été fait systématiquement par le groupe de S. Pääbo pour produire la meilleure séquence génomique actuellement disponible, celle du Dénisovien [15]. Hormis pour des ossements très anciens et précieux qui ont été contaminés lors du prélèvement et que l'on souhaite néanmoins analyser, il ne semble donc pas des plus pertinents de sacrifier la qualité des séquences pour obtenir un critère d'authenticité pas très fiable et qui peut être substitué par la réalisation de contrôles expérimentaux rigoureux.

Une fois que les adaptateurs ont été ajoutés aux extrémités des molécules d'ADN, le risque de contamination par de l'ADN moderne décroit très significativement. En effet, puisque seules les molécules qui portent les adaptateurs seront amplifiées, la plupart des contaminants que l'on serait susceptible d'introduire après la ligation ne seraient pas amplifiables étant dépourvues d'adaptateurs. Il y a toutefois une source de contamination qui reste alors dangereuse, ce sont les produits d'amplification des banques précédemment réalisées, qui eux portent de tels adaptateurs. Le problème est encore plus important dans le cas des projets de capture d'une fraction du génome (par exemple, le génome mitochondrial complet, [24]) car les séquences amplifiées après capture sont très enrichies dans un petit nombre de séquences différentes dont l'abondance relative devient alors très importante. Une solution pour identifier de tels contaminants est d'utiliser des étiquettes « code-barres »spécifiques de chaque projet. Ces étiquettes correspondent à une petite séquence nucléotidique de 6 à 12 paires de bases que l'on ajoute aux adaptateurs et dont la séquence sera lue en même temps que les fragments d'ADN seront séquencés. Cette approche est recommandée, mais il faut être conscient toutefois que le nombre de « code-barres » disponibles n'est pas infini et qu'une plate-forme de séquençage qui fonctionne normalement aura vite utilisée les combinaisons disponibles au moins une fois. Il faut donc utiliser en paléogénomique toutes les précautions d'usage utilisées en paléogénétique : confinement physique extrême en amont de la préparation des banques, séparation physique des laboratoires ADN ancien et ADN moderne, et des laboratoires pré et post-PCR, procédures strictes d'élimination des contaminants éventuels à l'entrée et à la sortie des pièces ainsi que dans les réactifs, etc. . .

29.3 Préservation de l'ADN

29.3.1 Préservation de l'ADN lors de l'enfouissement

Dans des conditions « normales », l'ADN disparaît rapidement après la mort d'un organisme par des mécanismes biologiques (autolyse des cellules puis dégradation de la matière organique par les insectes et les microorganismes), chimiques (hydrolyse, oxydation, condensation, alkylation) et physiques (rayons UV et ionisantes, tempé-

rature élevée). Pour que l'ADN puisse « survivre » pendant des milliers d'années, il faut que ces processus de dégradation soient ralentis, voire que certains n'aient pas lieu. Hormis la température, peu de facteurs agissant sur la vitesse de dégradation de l'ADN ont été rigoureusement identifiés. Plusieurs études visant à estimer la limite temporelle absolue de préservation de l'ADN ont été réalisées, mais soit elles se basaient sur les propriétés de l'ADN en solution aqueuse [25], ce qui ne reflète pas la situation dans les os et dents au cours de fossilisation, soit elles généralisaient abusivement à partir d'observations sur un nombre trop limité d'espèces, de type d'enfouissements, de régions géographiques (par exemple [26]). De plus, un certain nombre d'études étant réalisées dans des conditions insatisfaisantes de prévention de la contamination, il paraît incertain que l'ensemble des données utilisées soient authentiquement anciennes, ce qui fragilise les conclusions tirées. A l'heure actuelle, aucune étude ne permet de conclure de manière convaincante quant à la limite absolue dans le temps de la préservation de l'ADN dans les os et les dents, si tant est qu'une telle limite existe.

La nature et/ou la vitesse des processus de dégradation peuvent varier d'un os à l'autre à l'intérieur d'un même site d'enfouissement. Par conséquent, le taux de réussite de l'analyse paléogénétique d'ossements d'un site archéologique n'est jamais de 100%, mais toujours inférieur, entre 0 et ca. 80% selon les conditions taphonomiques. De plus, l'os et la dent ne sont pas des tissus homogènes. Il n'est donc pas étonnant que nous avons observés que l'ADN n'est pas préservé de manière homogène dans les os et dents archéologiques [27]. En effet, plusieurs études indiquent que l'ADN peut « survivre » particulièrement longtemps dans des niches moléculaires de préservation [28]. Un type de niche serait constitué de cristaux d'hydroxyapatite interconnectés [29], mais d'autres types sont susceptibles de servir comme niches moléculaires de préservation, comme les colloïdes d'acides humiques [30]. Malheureusement, il y a peu d'études qui s'attaquent vraiment à l'élucidation des mécanismes de la préservation de l'ADN en utilisant des méthodes strictes qui ne confondent pas contamination avec l'ADN exogène et préservation de l'ADN endogène. Ainsi on ne peut être sûr que d'un facteur qui joue un rôle clé dans la préservation de l'ADN, la température [31].

En analysant 20 parties anatomiques différentes dans 239 ossements et dents d'herbivores (bœufs et équidés) de différents sites archéologiques néolithiques et de l'Âge du Bronze en Europe et au Proche Orient, nous avons pu constater que l'ADN n'est pas conservé de la même manière dans toutes les différentes parties du squelette. En effet, l'analyse d'un grand nombre d'échantillons nous a permis d'utiliser des tests statistiques pour évaluer la significativité de nos résultats. La seule différence qui a satisfait les exigences du test Chi2 était celle qui a distingué les os de l'intérieur du corps des animaux associés aux gros muscles, les « os à viande » (tibia, fémur, humérus, ulna, côtes, scapula, coxal), et les autres (vertèbres, crâne, pattes) : Nous avons eu un taux de succès de 31% pour les premiers et de 14% pour les autres [6]. Nous avons attribué cette différence à une différence dans les pratiques de boucherie et de cuisine après la mort. Les parties de la carcasse qui étaient destinées à la consommation, auraient été très vite décarnées, cuites ou grillées ce qui aurait arrêté la prolifération bactérienne, un facteur important de la dégradation de l'ADN. Les autres parties des

carcasses auraient soit été jetées dans des fosses où leur matière organique aurait été digérée par des bactéries, soit soumis à des traitements de tannage (dans les cas des pattes qui souvent restaient associés aux peaux) [6]. Ce seraient donc les premières étapes de la diagénèse des os qui étaient concernées par ces processus, surtout le déroulement de la putréfaction, pour lesquelles nos données indiquent une influence décisive sur la « survie » de l'ADN.

29.3.2 Préservation de l'ADN après prélèvement

Notre équipe a pu démontrer que l'ADN peut se dégrader vite après prélèvement des os hors de l'environnement qui leur avaient permis jusque-là d'être préservés. Notre analyse d'un grand nombre d'os et de dents originaires de nombreux sites archéologiques européens et proche-orientaux a montré que le taux de succès des amplifications par PCRq de l'ADN ancien était plus élevé à partir d'ossements analysés directement après prélèvement sans être lavés qu'à partir d'ossements prélevés, lavés et stockés selon les méthodes standards en archéologie (46% de succès contre seulement 18% [32]. Cette différence était statistiquement significative. De plus, la quantité de molécules d'ADN initiales et leur longueur moyenne étaient également plus élevées dans les ossements fraîchement prélevés et non-lavés. Finalement, nous avons pu confirmer de manière remarquable ces corrélations globales grâce à l'analyse d'ossements provenant du squelette d'un même aurochs prélevés d'une part en 1947 en utilisant des méthodes standards et d'autre part en 2004 sans laver les os et en les gardant au congélateur jusque leur analyse paléogénétique. L'amplification d'un fragment d'ADN de 152 nucléotides a été réalisé pour 100% des tentatives à partir des os prélevés fraîchement aseptiquement et sans lavage, alors que cette efficacité a chuté à 0% à partir des os prélevés 57 ans auparavant. Nous avons modélisé sur la base d'une cinétique de dégradation du premier ordre que l'ADN se serait dégradé 70 fois plus rapidement dans ces os après prélèvement que pendant les 3200 ans de leur enfouissement [32]. Plusieurs facteurs peuvent expliquer ce résultat, comme l'augmentation de la température dans les conditions de stockage par rapport au site d'enfouissement. Toutefois, selon toute probabilité, c'est le lavage des os après prélèvement qui est le facteur le plus néfaste pour la préservation de l'ADN. Le lavage pourrait à la fois éliminer la fraction la plus soluble de l'ADN des ossements, modifier les conditions physicochimiques de l'environnement dans lequel cet ADN se trouve, et augmenter la probabilité de contamination de l'os avec de l'ADN exogène. Evidemment, les conséquences délétères de traitements post-fouille inadaptés se feront sentir avec d'autant plus d'acuité que la quantité d'ADN préservé dans l'os était faible, et que les os sont phylogénétiquement proche des contaminants introduits lors du lavage. Il est opportun de souligner que nos données ne signifient pas que ces conséquences délétères se manifesteront systématiquement. C'est pour cela qu'il était indispensable de réaliser cette étude sur un grand nombre d'ossements correspondant à une grande diversité de situations et d'utiliser des approches quantitatives et statistiques rigoureuses pour pouvoir mettre en évidence suffisamment de situations pour lesquelles le phénomène est observable.

29.4 Analyse des séquences d'ADN ancien

Une fois que les séquences d'ADN sont produites, il faut les analyser en utilisant des modèles et calculs de génétique de populations et/ou les méthodes de l'assemblage de séquences et la comparaison de génomes. Les méthodes de génétique de populations ont bien progressé au cours de ces dernières années et permettent maintenant de modéliser des migrations, goulots d'étranglements, croissances de populations en se basant sur des séquences d'ADN ancien (par exemple, [33 , 34]. Un grand corpus de données est toutefois nécessaire pour effectuer des modélisations plausibles. De plus, il est important de garder à l'esprit que la modélisation implique un grand nombre d'hypothèses simplificatrices qui sont assez éloignées de la réalité des phénomènes biologiques. Elle fournit donc un guide pour interpréter les données mais son pouvoir démonstrateur réel est faible avec les données paléogénétiques. Comme il est rarement possible de concevoir ou mettre en œuvre les expériences qui permettront de tester les conclusions des modèles, ces conclusions ont des durées de vie élevées qui reflètent rarement leur robustesse.

En ce qui concerne l'analyse des génomes entiers, leur assemblage à partir des petits fragments d'ADN de 30 à 100 nucléotides constitue encore un énorme défi. Actuellement, il n'est guère envisageable de le faire sans disposer du génome d'une espèce actuelle très proche contre lequel les séquences obtenues seront alignées. Ceci limite donc l'analyse des espèces pour lesquelles on ne dispose pas d'un génome moderne proche déjà séquencé. L'importance de ce problème va décroitre avec le temps du fait de l'ambition affichée par les génomiciens de séquencer une grande proportion de la diversité biologique. La comparaison entre un génome ancien et des génomes modernes permet une meilleure reconstruction phylogénétique de l'évolution des branches du vivant considérée (par exemple, [13 , 35 , 36]). Il est aussi possible d'utiliser les génomes anciens pour identifier les gènes qui se trouvent sous sélection positive dans les lignées modernes depuis la divergence avec l'ancêtre commun à la lignée ancienne et ainsi de mettre en évidence des bases génétiques pour des adaptations physiologiques, morphologiques ou comportementales [14 , 15]. Ces analyses sont susceptibles de déceler des processus d'une grande importance pour l'archéologie, comme cela est le cas pour les changements génétiques survenus au cours de l'évolution de *l'Homo sapiens* ou au cours de la domestication des animaux et plantes.

29.5 Conclusion

La paléogénétique et plus encore, son nouvel essor la paléogénomique, sont des disciplines susceptibles d'éclairer de plus en plus les événements biologiques du passé et ainsi contribuer énormément à l'archéologie, la paléontologie, la paléoanthropologie, la paléomicrobiologie. Avec l'avènement de la paléogénomique, nous assistons à un vrai changement de paradigme. En effet, le séquençage à haut débit permet de multiples applications et modifie profondément des pans entiers de la recherche phylogénétique et archéologique qui s'intéresse à l'évolution des animaux, végétaux et humains ainsi que leurs relations.

Cependant, ces avancés de la recherche génétique seront utiles à l'archéologie et la paléontologie sous condition d'une collaboration étroite avec les généticiens permet-

tant de connaître les limites méthodologiques de chacune des disciplines impliquées dans un projet archéologique. Ces limites sont importantes dans toutes les disciplines cherchant à interpréter le passé à partir de données fragmentaires, morcelées et incomplètes. Il est donc important que chacun ait une bonne conscience des limites expérimentales et conceptuelles des autres disciplines afin que les attentes soient réalistes et que la collaboration permette de distinguer clairement les faits de leur interprétation et de réduire efficacement la distance séparant l'interprétation des données de la réalité.

Bibliographie

[1] R. K. Saiki, Enzymatic amplification of b-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science. 230, 1350-1354, 1985.

[2] M. Pruvost, E.-M. Geigl, Real-time quantitative PCR to assess the authenticity of ancient DNA amplification. Journal of Archaeological Science. 31, 1191-1197, 2004.

[3] A. M. Maxam, W. Gilbert, A new method for sequencing DNA. Proceedings of the National Acadademy of Science USA. 74, 560-564, 1977.

[4] F. Sanger, S. Nicklen, A. R. Coulson, DNA sequencing with chain-terminating inhibitors. Proceedings of the National Acadademy of Science USA. 74, 5463-5467, 1977.

[5] M. Ronaghi, M. Nygren, J. Lundeberg, P. Nyren, Analyses of secondary structures in DNA by pyrosequencing. Analytical Biochemistry. 267, 65-71, 1999.

[6] M. Pruvost, R. Schwarz, V. Bessa Correia, S. Champlot, T. Grange, E.-M. Geigl, DNA diagenesis and palaeogenetic analysis : critical assessment and methodological progress. Palaeogeography, Palaeoclimatology, Palaeoecology. 266, 211-219, 2008.

[7] E.-M. Geigl, Palaeogenetics of cattle domestication : methodological challenge for the study of fossil bones preserved in the domestication centre in Southwest Asia. Comptes Rendus Palévol. 7, 99-112, 2008.

[8] S. Champlot, C. Berthelot, M. Pruvost, E. A. Bennett, T. Grange, E. M. Geigl, An efficient multistrategy DNA decontamination procedure of PCR reagents for hypersensitive PCR applications. PLoS One. 5, e13042, 2010.

[9] M. Pruvost, T. Grange, E.-M. Geigl, Minimizing DNA contamination by using UNG-coupled quantitative real-time PCR on degraded DNA samples : application to ancient DNA studies. BioTechniques. 38, 569-575, 2005.

[10] J. A. Leonard, O. Shanks, M. Hofreiter, E. Kreuz, L. Hodges, W. Ream, R. K. Wayne, R. C. Fleischer, Animal DNA in PCR reagents plagues ancient DNA research. Journal of Archaeological Science. 34, 1361-1366, 2007.

[11] A. W. Briggs, U. Stenzel, P. L. F. Johnson, R. E. Green, J. Kelso, K. Prüfer, M. Meyer, J. Krause, M. T. Ronan, M. Lachmann, S. Pääbo, Patterns of damage in genomic DNA sequences from a Neandertal. Proceedings of the National Acadademy of Science USA. 104, 14616-14621, 2007.

[12] H. N. Poinar, C. Schwarz, J. Qi, B. Shapiro, R. D. E. MacPhee, B. Buigues, A. Tikhonov, D. H. Huson, L. P. Tomsho, A. Auch, M. Rampp, W. Miller, S. C. Schuster, Metagenomics to Paleogenomics : Large-Scale Sequencing of Mammoth DNA. Science. 311, 392-394, 2006.

[13] M. Rasmussen, Y. Li, S. Lindgreen, J. S. Pedersen, A. Albrechtsen, I. Moltke, M. Metspalu, E. Metspalu, T. Kivisild, R. Gupta, M. Bertalan, K. Nielsen, M. T. Gilbert, Y. Wang, M. Raghavan, P. F. Campos, H. M. Kamp, A. S. Wilson, A. Gledhill, S. Tridico, M. Bunce, E. D. Lorenzen, J. Binladen, X. Guo, J. Zhao, X. Zhang, H. Zhang, Z. Li, M. Chen, L. Orlando, K. Kristiansen, M. Bak, N. Tommerup, C. Bendixen, T. L. Pierre, B. Gronnow, M. Meldgaard, C. Andreasen, S. A. Fedorova, L. P. Osipova, T. F. Higham, C. B. Ramsey, T. V. Hansen, F. C. Nielsen, M. H. Crawford, S. Brunak, T. Sicheritz-Ponten, R. Villems, R. Nielsen, A. Krogh, J. Wang, E. Willerslev, Ancient human genome sequence of an extinct Palaeo-Eskimo. Nature. 463, 757-762, 2010.

[14] R. E. Green, J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, H. Li, W. Zhai, M. H. Fritz, N. F. Hansen, E. Y. Durand, A. S. Malaspinas, J. D. Jensen, T. Marques-Bonet, C. Alkan, K. Prufer, M. Meyer, H. A. Burbano, J. M. Good, R. Schultz, A. Aximu-Petri, A. Butthof, B. Hober, B. Hoffner, M. Siegemund, A. Weihmann, C. Nusbaum, E. S. Lander, C. Russ, N. Novod, J. Affourtit, M. Egholm, C. Verna, P. Rudan, D. Brajkovic, Z. Kucan, I. Gusic, V. B. Doronichev, L. V. Golovanova, C. Lalueza-Fox, M. de la Rasilla, J. Fortea, A. Rosas, R. W. Schmitz, P. L. Johnson, E. E. Eichler, D. Falush, E. Birney, J. C. Mullikin, M. Slatkin, R. Nielsen, J. Kelso, M. Lachmann, D. Reich, S. Paabo, A draft sequence of the Neandertal genome. Science. 328, 710-722, 2010.

[15] M. Meyer, M. Kircher, M. T. Gansauge, H. Li, F. Racimo, S. Mallick, J. G. Schraiber, F. Jay, K. Prufer, C. de Filippo, P. H. Sudmant, C. Alkan, Q. Fu, R. Do, N. Rohland, A. Tandon, M. Siebauer, R. E. Green, K. Bryc, A. W. Briggs, U. Stenzel, J. Dabney, J. Shendure, J. Kitzman, M. F. Hammer, M. V. Shunkov, A. P. Derevianko, N. Patterson, A. M. Andres, E. E.

Eichler, M. Slatkin, D. Reich, J. Kelso, S. Paabo, A high-coverage genome sequence from an archaic Denisovan individual. Science. 338, 222-226, 2012.

[16] K. I. Bos, V. J. Schuenemann, G. B. Golding, H. A. Burbano, N. Waglechner, B. K. Coombes, J. B. McPhee, S. N. DeWitte, M. Meyer, S. Schmedes, J. Wood, D. J. Earn, D. A. Herring, P. Bauer, H. N. Poinar, J. Krause, A draft genome of Yersinia pestis from victims of the Black Death. Nature. 478, 506-510, 2011.

[17] L. Orlando, A. Ginolhac, M. Raghavan, J. Vilstrup, M. Rasmussen, K. Magnussen, K. E. Steinmann, P. Kapranov, J. F. Thompson, G. Zazula, D. Froese, I. Moltke, B. Shapiro, M. Hofreiter, K. A. Al-Rasheid, M. T. Gilbert, E. Willerslev, True single-molecule DNA sequencing of a pleistocene horse bone. Genome Research. 21, 1705-1719, 2011.

[18] K. Prufer, U. Stenzel, M. Hofreiter, S. Paabo, J. Kelso, R. E. Green, Computational challenges in the analysis of ancient DNA. Genome Biology. 11, R47, 2010.

[19] A. W. Briggs, J. M. Good, R. E. Green, J. Krause, T. Maricic, U. Stenzel, S. Paabo, Primer extension capture : targeted sequence retrieval from heavily degraded DNA sources. Journal of Visualized Experiments. 1573, 2009.

[20] H. A. Burbano, E. Hodges, R. E. Green, A. W. Briggs, J. Krause, M. Meyer, J. M. Good, T. Maricic, P. L. Johnson, Z. Xuan, M. Rooks, A. Bhattacharjee, L. Brizuela, F. W. Albert, M. de la Rasilla, J. Fortea, A. Rosas, M. Lachmann, G. J. Hannon, S. Paabo, Targeted investigation of the Neandertal genome by array-based sequence capture. Science. 328, 723-725, 2010.

[21] J. D. Wall, S. K. Kim, Inconsistencies in Neanderthal genomic DNA sequences. PLoS Genet. 3, 1862-1866, 2007.

[22] R. E. Green, A. W. Briggs, J. Krause, K. Prüfer, H. A. Burbano, M. Siebauer, M. Lachmann, S. Pääbo, The Neandertal genome and ancient DNA authenticity. The EMBO Journal. 28, 2494-2502, 2009.

[23] S. Sawyer, J. Krause, K. Guschanski, V. Savolainen, S. Paabo, Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. PLoS one. 7, e34131, 2012.

[24] T. Maricic, M. Whitten, S. Paabo, Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. PLoS one. 5, e14004, 2010.

[25] T. Lindahl, Instability and decay of the primary structure of DNA. Nature. 709-715, 1993.

[26] M. E. Allentoft, M. Collins, D. Harker, J. Haile, C. L. Oskam, M. L. Hale, P. F. Campos, J. A. Samaniego, M. T. Gilbert, E. Willerslev, G. Zhang, R. P. Scofield, R. N. Holdaway, M. Bunce, The half-life of DNA in bone : measuring decay kinetics in 158 dated fossils. Proceedings of the Royal Society B : Biological Sciences. 279, 4724-4733, 2012.

[27] E.-M. Geigl, A personal analysis of the high failure rate of ancient DNA research. Geoarchaeol. Bioarchaeol. Studies. 3, 463-466, 2005.

[28] E.-M. Geigl, On the circumstances surrounding the preservation of very old DNA. Archaeometry. 44, 337-342, 2002.

[29] M. Salamon, N. Turoos, B. Arensburg, S. Weiner, Relatively well preserved DNA is present in the crystal aggregates of fossil bones. Proceedings of the National Acadademy of Science USA. 102, 13783-13788, 2005.

[30] E.-M. Geigl, " Why ancient DNA research needs taphonomy " dans Biosphere to Lithosphere - New studies in vertebrate taphonomy, Proceedings of the 9th Conference of the International Council of Archaeozoology, Durham, August 2002, ed. T. O'Connor, Oxbow Books, Oxford, p. 78-85, 2005.

[31] C. I. Smith, A. T. Chamberlain, M. S. Riley, A. Cooper, C. B. Stringer, M. J. Collins, Neanderthal DNA. Not just old but old and cold ? Nature. 410, 771-772, 2001.

[32] M. Pruvost, R. Schwarz, V. Bessa Correia, S. Champlot, S. Braguier, N. Morel, Y. Fernandez-Jalvo, T. Grange, E.-M. Geigl, Freshly excavated fossil bones are best for amplification of ancient DNA. Proceedings of the National Acadademy of Science USA. 104, 739-744, 2007.

[33] M. S. Gill, P. Lemey, N. R. Faria, A. Rambaut, B. Shapiro, M. A. Suchard, Improving Bayesian Population Dynamics Inference : A Coalescent-Based Model for Multiple Loci. Molecular Biology and Evolution. 2012.

[34] I. Alves, A. Sramkova Hanulova, M. Foll, L. Excoffier, Genomic data reveal a complex making of humans. PLoS Genetics. 8, e1002837, 2012.

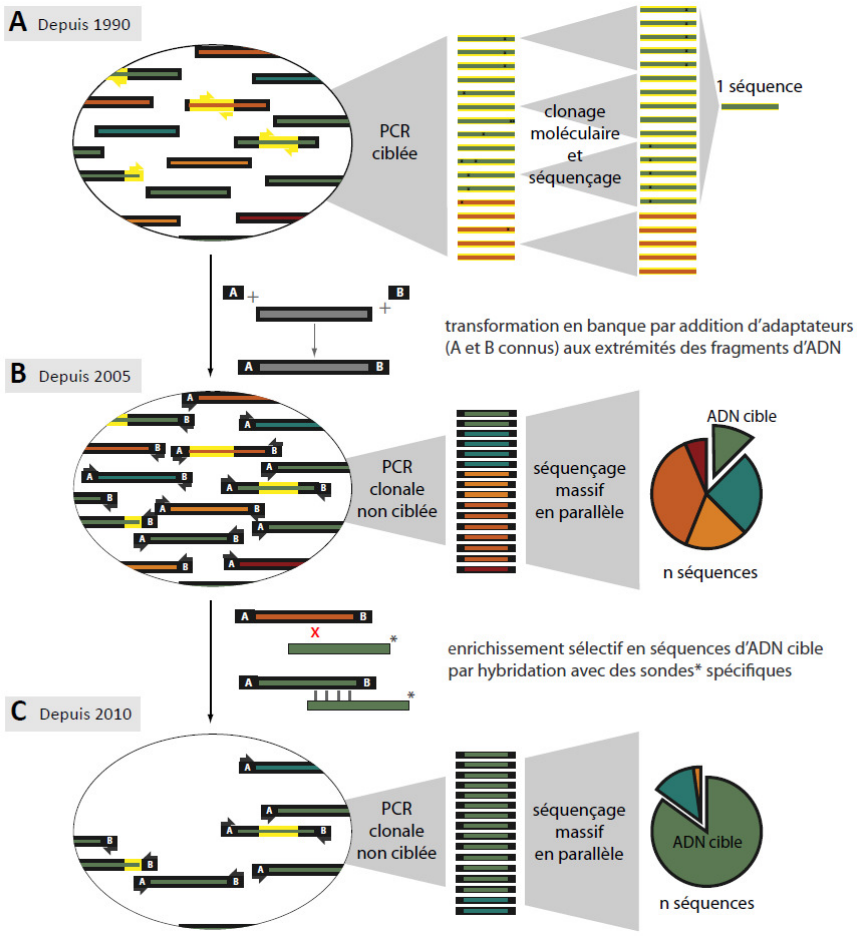
[35] P. F. Campos, E. Willerslev, A. Sher, L. Orlando, E. Axelsson, A. Tikhonov, K. Aaris-Sorensen, A. D. Greenwood, R. D. Kahlke, P. Kosintsev, T. Krakhmalnaya, T. Kuznetsova, P. Lemey, R. MacPhee, C. A. Norris, K. Shepherd, M. A. Suchard, G. D. Zazula, B. Shapiro, M. T. Gilbert, Ancient DNA analyses exclude humans as the driving force behind late Pleistocene musk ox (Ovibos moschatus) population dynamics. Proceedings of the National Acadademy of Science USA. 107, 5675-5680, 2010.

[36] E. D. Lorenzen, D. Nogues-Bravo, L. Orlando, J. Weinstock, J. Binladen, K. A. Marske, A. Ugan, M. K. Borregaard, M. T. Gilbert, R. Nielsen, S. Y. Ho, T. Goebel, K. E. Graf, D. Byers, J. T. Stenderup, M. Rasmussen, P. F. Campos, J. A. Leonard, K. P. Koepfli, D. Froese, G. Zazula, T. W. Stafford, Jr., K. Aaris-Sorensen, P. Batra, A. M. Haywood, J. S. Singarayer, P. J. Valdes, G. Boeskorov, J. A. Burns, S. P. Davydov, J. Haile, D. L. Jenkins, P. Kosintsev, T. Kuznetsova, X. Lai, L. D.

Martin, H. G. McDonald, D. Mol, M. Meldgaard, K. Munch, E. Stephan, M. Sablin, R. S. Sommer, T. Sipko, E. Scott, M. A. Suchard, A. Tikhonov, R. Willerslev, R. K. Wayne, A. Cooper, M. Hofreiter, A. Sher, B. Shapiro, C. Rahbek, E. Willerslev, Species-specific responses of Late Quaternary megafauna to climate and humans. Nature. 479, 359-364, 2011.

Encart 16.- Aperçu des méthodes d'amplification et de séquençage d'ADN ancien

Régis Debruyne
UMS2700 CNRS/MNHN, Paris, France

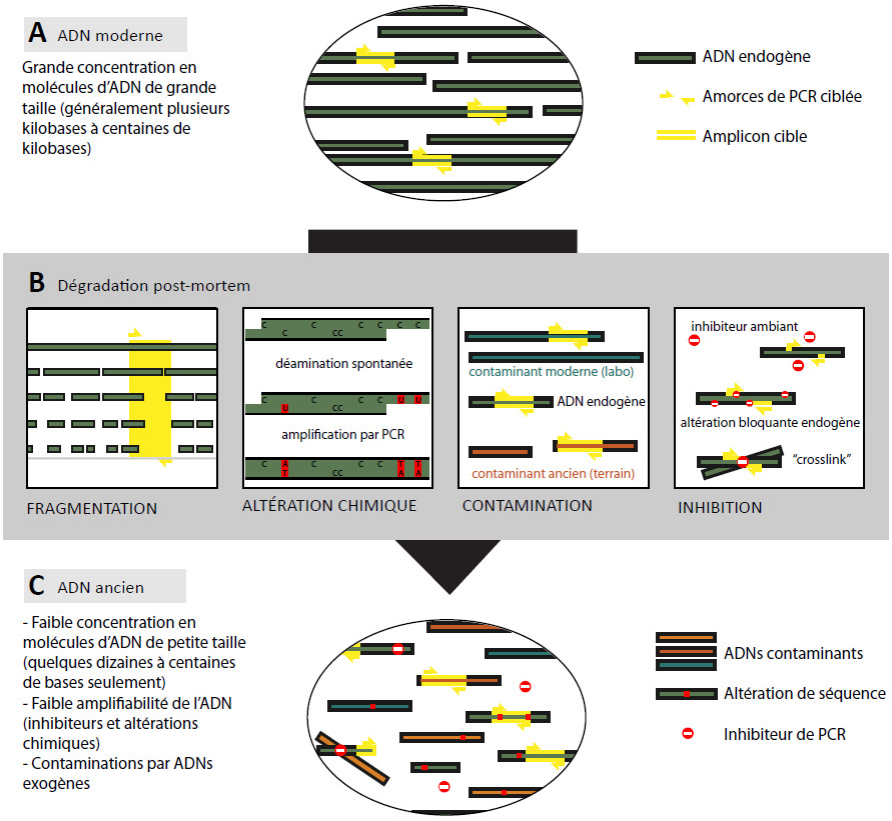


Deux méthodes principales d'amplification d'ADN peuvent être distinguées. L'amplification ciblée (A) vise à multiplier le nombre de copies d'un fragment génomique particulier, choisi comme référence pour des comparaisons entre individus. Son efficacité dépend du nombre de copies du fragment d'intérêt présentes dans l'extrait. On lui adjoint souvent une étape de clonage moléculaire. Le séquençage des clones permet d'identifier les altérations de séquence et/ou contaminations (*cf.* encart "Dé-

gradation post-mortem de l'ADN dans les restes archéologiques"). La méthode d'amplification non ciblée (B, C) repose sur la transformation préalable de l'extrait en une banque où les fragments d'ADN se voient ajouter à leurs extrémités des séquences flanquantes connues qui serviront d'amorces d'amplification et de séquençage. Cette approche, indépendante de la séquence, amplifie tous les ADNs présents en fonction de leur quantité relative dans l'extrait. Elle est suivie d'un séquençage en parallèle qui permet d'obtenir des millions de fragments issus aléatoirement de la banque. Cette technique présente le désavantage de séquencer tous les contaminants contenus dans l'extrait ancien et de produire proportionnellement peu de séquences du génome d'intérêt (B). Elle peut être affinée par l'ajout d'une étape d'enrichissement qui repose sur l'hybridation de fragments du génome d'intérêt avec des sondes spécifiques. Dans ce cas, la représentation du génome d'intérêt est largement améliorée par rapport aux éventuels contaminants (C).

Encart 17.- Dégradation post-mortem de l'ADN dans les restes archéologiques

Régis Debruyne
UMS2700 CNRS/MNHN, Paris, France



Plusieurs phénomènes naturels interviennent lors de la dégradation post-mortem de l'ADN, aggravés par les conditions de manipulation et d'analyse des restes archéologiques sur le terrain et en laboratoire. Ces phénomènes concourent aux difficultés d'analyse du matériel génétique ancien (C), comparativement à l'ADN extrait de tissus frais (A) et sont responsables de la faible amplifiabilité des ADNs anciens par approche de PCR ciblée (*cf.* encart "Aperçu des méthodes d'amplification et de séquençage d'ADN ancien"). On peut les classifier en quatre grands groupes (B). Premièrement, la dégradation post-mortem conduit à la fragmentation progressive des molécules d'acides nucléiques et à la dilution de l'ADN endogène. Deuxièmement, certaines altérations chimiques des bases ne produisent pas de rupture du squelette des molécules d'ADN, mais induisent une modification des acides nu-

cléiques. Le type le mieux documenté en est la déamination spontanée des cytosines qui produit des uracils (interprétées comme des thymines par les polymérases) : elles conduisent à l'enrichissement artificielle des séquences en bases T (et donc A) par rapport aux C (et G). Troisièmement, la faible concentration des ADNs anciens les rend très sensibles aux contaminations par des ADNs exogènes, qu'elles soient le fait de contaminants précoces (e.g. champignons et bactéries venus dégrader la matière organique) ou récents (e.g. ADN humain moderne des manipulateurs). Quatrièmement, les ADNs anciens réagissent souvent assez faiblement aux protocoles d'analyse classique en laboratoire du fait de leur potentiel inhérent à inhiber ces réactions. Ces phénomènes d'inhibition peuvent dériver de substances inhibitrices ambiantes ou d'altérations des séquences d'ADN bloquant les enzymes comme les polymérases. Les phénomènes de condensation d'ADN ("crosslink") sont également responsables d'inhibition, et sont particulièrement bien documentés dans les paléofecès.