

ADN fossile : quand la PCR remonte le temps...

Historique de la discipline

L'histoire de l'ADN ancien (ADNa) commence au milieu des années quatre-vingt lorsque deux équipes reportent consécutivement l'extraction d'ADN de fossiles. L'un était un animal disparu de la surface du globe depuis plus d'un siècle, le [quagga](#), l'autre une momie égyptienne, datant du V^{ème} siècle avant Jésus-Christ. Si la preuve était faite que l'ADN pouvait résister aux ravages du temps, on pouvait encore douter de l'utilité et de la portée de ces études. Dans les deux cas en effet, les fossiles étudiés avaient connu des traitements particuliers juste après leur mort pour minimiser leur dégradation.



Le quagga avait été taxidermisé pour être exposé sous son plus beau jour dans les collections des musées, et les rites funéraires égyptiens servaient à rendre les corps des morts... éternels. Or, la très grande majorité des fossiles répandus dans les gisements Quaternaires n'ont pas bénéficié de telles conditions de conservation : bien souvent l'ADN, s'il était présent, restait à l'état de traces. Ainsi, le [clonage](#) par transformation bactérienne, seule technique d'amplification du matériel génétique alors disponible, et très exigeante quant au matériel source, ne permettait pas d'extraire et d'étudier l'ADN de fossile. L'enthousiasme alors naissant pour l'étude du matériel génétique du passé allait-il avorter *in utero* ? Non, puisque dans le même temps était mise au point la technique d'amplification de l'ADN par PCR (*Polymerase Chain Reaction*) ; la possibilité d'amplifier en plusieurs centaines de milliers de copies une seule molécule d'ADN de départ semblait venir à point. L'enthousiasme naissant pour les séquences du passé avait trouvé son outil de prédilection et les études de paléogénétique allaient pouvoir se multiplier.

Limites techniques

De l'ADN ancien, d'accord mais pas trop Les premières études se sont cantonnées à des fossiles d'âge relativement récent, allant de quelques centaines à quelques dizaines de milliers d'années. Mais évidemment, très vite, on a voulu défier la limite du temps et remonter jusqu'à des périodes très reculées. D'ailleurs, la découverte que des peptides pouvaient être conservés au-delà de quelques centaines de milliers d'années allait alimenter la course à la recherche des séquences les plus vieilles. On atteignait 17 millions d'années avec des feuilles de *Magnolia* sédimentées en milieu lacustre anoxique, puis 30-35 millions d'années avec divers insectes englués dans l'ambre ; bientôt un os de dinosaure, vieux de 65 millions d'années défrayait la chronique. La progression semblait irrésistible ; parti de la fin du quaternaire, une anecdote à l'échelle des temps géologiques, on arrivait à la limite entre l'ère Tertiaire et l'ère Secondaire. Et on n'allait pas s'arrêter en chemin : le record absolu était décerné à des charançons préservés dans l'ambre, vieux de 125 millions d'années. Aucune époque géologique ne semblait hors de portée de la paléogénétique. Un examen plus attentif des séquences amplifiées allait cependant stopper net la machine à remonter le temps. Le prétendu ADN amplifié à partir du charançon n'était rien d'autre que celui d'un champignon tout à fait actuel ! L'histoire allait inmanquablement se répéter avec chacune de ces séquences : l'ADN soi-disant de dinosaure n'était qu'un contaminant... humain. La très sensible technique d'amplification par PCR qui, pensait-on, avait permis d'amplifier les infimes vestiges d'ADN contenus dans ces fossiles vieux de plusieurs millions d'années était en fait tout aussi sensible pour amplifier des traces d'ADN moderne, un bien meilleur substrat. Il a donc fallu se rendre à l'évidence : l'ADN n'est pas éternel dans les fossiles. D'ailleurs, sur la base théorique de la demi-vie des molécules d'ADN en solution aqueuse, on estime que seules des conditions exceptionnelles permettent de préserver une information génétique au-delà de 50 à 100 mille ans. Seul le Quaternaire récent peut donc être soumis à l'enquête paléogénétique. Pourquoi ? parce que *post-mortem*, les acides-nucléiques sont soumis à toute une diversité de processus de dégradation. Par exemple, les tissus d'Ötzi (la fameuse momie de cinq mille ans retrouvée dans un glacier alpin) ne contiennent plus que l'équivalent de dix à vingt génomes nucléaires par gramme de tissu, soit un million de fois moins que chez un organisme vivant. Bien qu'elles restent à l'état double-brin, les traces d'ADN fossile sont excessivement fragmentées (segments de 100-200pb) ; des bases nucléotidiques sont perdues par hydrolyse de liaisons N-glycosidiques, et la nature chimique des bases restantes a pu être modifiée au cours du temps : produits d'oxydation (8-oxo Guanine), ou d'irradiation (photoproduits, 5-hydroxy 5-méthyl hydantoïne).

La PCR, une incontournable difficulté L'ADN des fossiles est donc fragmenté, dégradé et chimiquement modifié. Cette matrice n'est donc pas un substrat optimal pour la *Taq polymerase*. Par exemple, l'existence de *nicks* sur les fragments anciens bloque l'élongation catalysée par la *Taq* polymérase. L'encombrement stérique lié à la présence de bases chimiques atypiques (dimères de pyrimidines par exemple) a le même effet. Mais puisque l'étape d'amplification des traces fossiles est indispensable compte tenu de la quantité d'ADN présent, il a fallu contourner un certain nombre de contraintes liées à des réactions de PCR réfractaires.

- L'inhibition de la Taq... Une première difficulté peut venir de la présence d'inhibiteurs de la *Taq* polymérase, coextraits avec l'ADN. Ces inhibiteurs agissent en *trans* puisque la simple dilution des produits d'extraction suffisent généralement à abolir leur effet et que certains protocoles d'extraction réduisent leur copurification. Relativement peu de choses est connue sur la nature de ces inhibiteurs. Certains seraient dérivés des constituants même des tissus de l'organisme fossile (produits de Maillard, résidus porphyriques, collagène de type I), d'autres seraient des composants même du sol (acides humiques, fulviques, tannins). Généralement, l'ajout de fortes concentrations en BSA (*Bovine Serum Albumine*) dans les réactifs de PCR (jusqu'à 1mg/ml) suffit à contenir ces effets inhibiteurs.
- Des recombinaisons aléatoires lors de la PCR... Une seconde difficulté provient de l'état particulièrement fragmenté de l'ADN fossile. Il arrive, lorsqu'on part d'un substrat dégradé, que le fragment amplifié soit d'une taille plus grande qu'attendue (*jumping PCR*). Il s'agit, dans les cas où ce ne sont pas des contaminations aspécifiques, d'amplicons chimères produits lors des premiers cycles d'amplification au cours desquels l'information présente en différents *loci* génétiques est recombinaisonnée aléatoirement. Par nature ininterprétables, de tels produits ont conduit, aux balbutiements de la discipline, à des conclusions erronées. Les séquences reportées pour les termites piégées dans l'ambre de Saint Domingue (vieilles de 30 MA) n'étaient en fait que des molécules chimères de plusieurs espèces, dont des champignons et la drosophile .
- Les contaminations... Si les prétendues séquences datant de plusieurs millions d'années ont servi à dévoiler le phénomène jusqu'alors ignoré de *jumping PCR*, elles ont également révélé le problème majeur de toute étude d'ADN ancien, les contaminations. L'ADN des fossiles n'étant présent qu'à l'état de traces dégradées, tout contaminant d'ADN moderne, intègre, est susceptible d'être préférentiellement amplifié lors de l'amplification par PCR. Généralement, le choix d'amorces très spécifiques de l'espèce et du locus génétique à étudier permet de s'affranchir des contaminants. Cependant, cette stratégie est mise à défaut dans au-moins deux situations fréquentes : lorsque le marqueur génétique est particulièrement conservé entre différents groupes, il n'est guère possible de recourir à des amorces d'amplification totalement spécifiques ; et, lorsqu'il s'agit d'hommes fossiles, il n'est point d'amorces qui s'affranchissent de la source principale de contamination, l'ADN des gens qui ont été au contact de l'échantillon. Dans ce cas, on établit alors le profil génétique de toute personne ayant été au contact avec le fossile, depuis les fouilles jusqu'au laboratoire, afin de remonter à tout événement de contamination éventuel.
- Les erreurs d'élongation Puisque la *Taq* polymérase ne possède aucun système de correction des erreurs d'élongation (*DNA proofreading*), tout mésappariement est conservé au cours de l'amplification. Si de telles erreurs sont peu fréquentes sur des substrats d'ADN moderne, elles sont tout au contraire particulièrement générées lors de l'amplification des substrats anciens. Par exemple, parmi les principales lésions formées à la suite de l'exposition à des conditions oxydantes, la 8OH G s'apparie préférentiellement à la base A plutôt qu'à la base C. A chaque cycle d'élongation, la *Taq* polymérase est donc susceptible d'engendrer des transversions C A, brouillant considérablement l'information présente dans la séquence primaire du fragment ancien. Les études de paléogénétique portant nécessairement sur des séquences courtes (100 à 200 pb), on comprend combien ces artefacts peuvent fausser le peu d'information recueillie. Plusieurs

stratégies ont été développées pour contourner ces problèmes et authentifier les mutations observées. La première consiste à comparer les sites variables sur l'ADN à ceux d'espèces plus ou moins apparentées. Si un changement est observé au niveau d'un site très conservé, il est peu vraisemblable que la mutation amplifiée ait une réalité biologique. Ce raisonnement simple a permis de repérer que les deux mutations observées entre des segments de la cytochrome oxydase du quagga et le zèbre des plaines étaient des artefacts puisque chez les Vertébrés les codons " mutés " du quagga étaient invariants du Xénope à l'Homme. Ces mutations faux-sens sont donc vraisemblablement apparues *post-mortem* du fait des modifications chimiques de l'ADN. Cependant, tous les marqueurs génétiques étudiés ne permettent pas d'aligner aussi fidèlement les séquences. Une seconde méthode consiste alors à cloner systématiquement les produits d'amplification pour étudier la distribution des mutations dues aux erreurs de la *Taq* polymérase. La séquence majoritaire est ensuite retenue, en faisant le raisonnement que les mutations liées à l'activité de la polymérase sont aléatoires et qu'il est donc peu vraisemblable qu'elles affectent systématiquement les mêmes sites.

Caractéristiques de l'ADN ancien

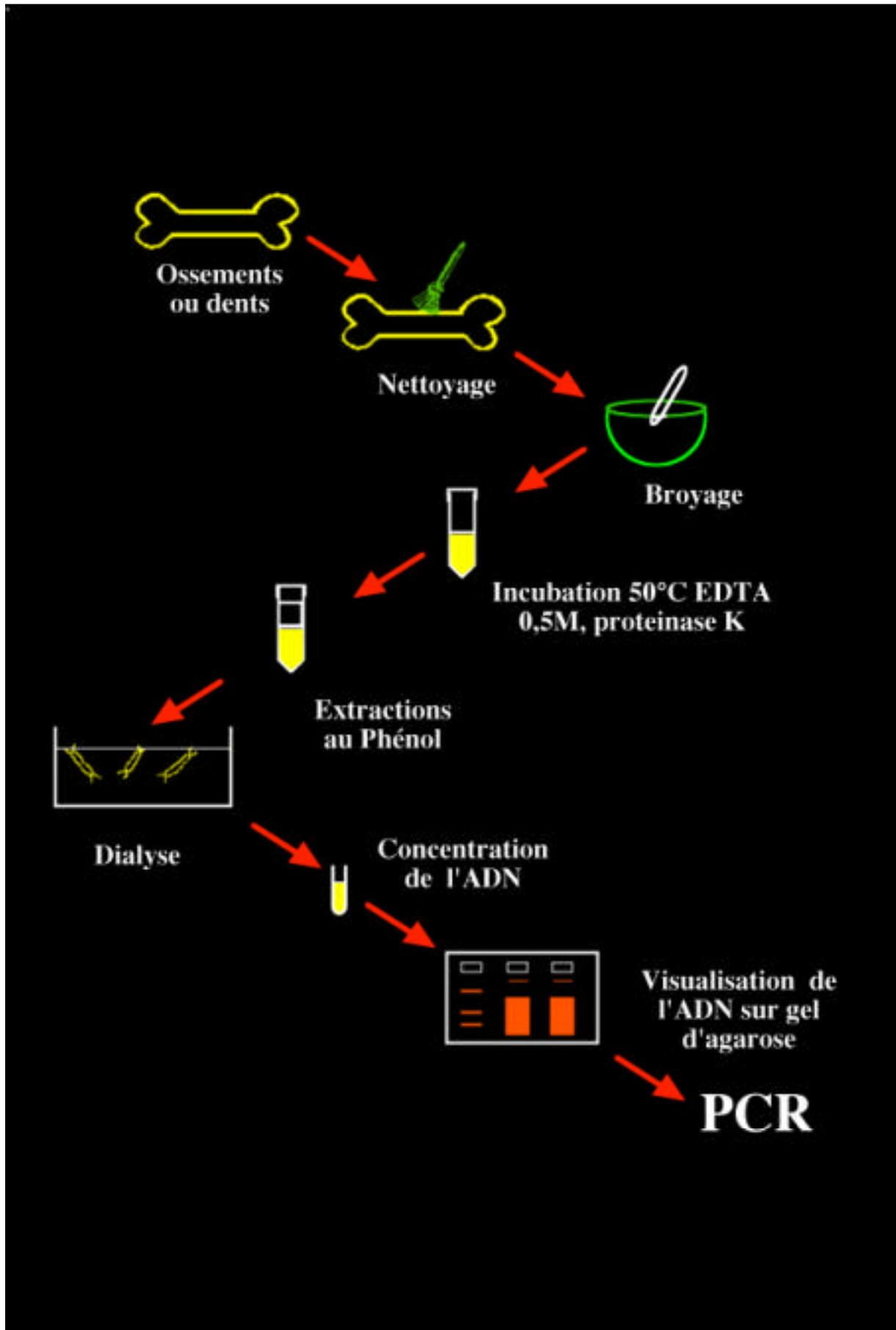
- Très petites quantités
- Dégradé
- Chimiquement modifié
- Présence d'inhibiteurs
- Risques de contamination par de l'ADN exogène

→ Compétences spécifiques

Méthodes (extraction, PCR)
Analyse des résultats (authenticité)



Caractéristiques de l'ADN ancien et conséquences



Principe de l'extraction puis de l'analyse d'un ADN ancien

Les champs d'étude

Les paléogénéticiens se sont attaqués à des questions tout azimut. A quels animaux aujourd'hui encore vivants s'apparentent les espèces disparues de la surface du globe, comme le tigre à dents de sabre, le quagga, le mammouth ou l'ours des cavernes ? Quels hominidés ont participé à la diversité de notre pool génétique ? L'homme de Néanderthal par exemple est-il notre proche parent, ou bien nos génomes ne se sont-ils jamais jamais mélangés ? Quelles routes ont suivi les premiers hommes préhistoriques pour peupler la planète ? Quelles étaient les conditions d'hygiène des populations du passé ? Comment sont apparues les premières grandes épidémies ? Enfin, le passage du statut de chasseur/cueilleur à celui d'agriculteur/éleveur lors du grand épisode de domestication Néolithique, a-t'il eu un impact sur la variabilité génétique des espèces ? Bref, autant de questions visant à préciser le mystère de nos origines et à déchiffrer le monde tel qu'il était il y a encore quelques milliers d'années.

□ Position phylogénétique d'espèces éteintes

Ours des cavernes (*Ursus spelaeus*)

Mammouth

Tigre à dents de sabre

□ Préservation d'espèces en voie de disparition et réimplantation d'espèces dans leurs biotopes d'origine

Canard de l'île de Laysan

□ Régimes alimentaires d'espèces aujourd'hui disparues

Paresseux

□ Populations animales et paléogéographie

Ours brun (*Ursus arctos*)

□ Impact génétique de la domestication des espèces

Blé

Maïs

Auroch

□ Origine du pool génique de l'homme moderne

Homme de Néanderthal

□ Vagues de migration et colonisations humaines

Colonisation du nouveau-monde

Colonisation de la Polynésie

Colonisation du Japon

□ Analyse génétique des nécropoles

Tumulus et disposition sociale versus familiale des défunts

□ Emergence des paléopathologies

Tuberculose (*Mycobacterium tuberculosis*)

Peste (*Yersinia pestis*)
Grippe espagnole (virus de la grippe, Influenza)
Virus HTLV-1

Perspectives

Toutes les techniques mises au point par les paléogénéticiens n'ont qu'un but : amplifier de l'ADN particulièrement abimé, dégradé, et présent à l'état de traces. Les enquêteurs de criminologie se trouvent confrontés aux mêmes contraintes : l'assassin ne laisse souvent traîner qu'un minimum d'indices. C'est donc grâce à la paléogénétique qu'ont pu être appliquées des techniques de génétique moléculaire en criminologie. Dans le monde de l'industrie agro-alimentaire, les produits fabriqués sont pour des raisons d'hygiène et de conservation, pasteurisés (processus d'*appertisation*). Là encore, l'ADN initialement présent dans les aliments est dégradé et fragmenté du fait de conditions de pression et de température délétères. Mais avec les techniques de paléogénétique, toute trace encore présente est susceptible d'être décelée. Avis aux fraudeurs : il est désormais possible de détecter les espèces animales et végétales présentes dans les produits de grande distribution. Toute indéclicatesse d'étiquetage pourra désormais être trahie par une analyse de l'ADN...

Ludovic ORLANDO.

Ludovic Orlando est un ancien élève de l'Ecole Normale Supérieure de Lyon. Agrégé des Sciences de la Vie et de la Terre, il exerce la fonction de Moniteur à la préparation à l'Agrégation SVT de l'ENS Lyon. Son travail de thèse s'effectue dans le laboratoire de Catherine Hänni.

Laboratoire de Catherine HÄNNI, chargé de recherche au CNRS (CR1)
Evolution Moléculaire et ADN fossile – CNRS UMR 5534
Centre de Génétique Moléculaire et Cellulaire
Université Claude Bernard, Lyon1 – Bâtiment 741
69624 Villeurbanne cedex
France

04 72 43 29 50

orlando@maccgmc.univ-lyon1.fr

Références

Orlando L, Hänni C. Du nouveau pour l'ADN ancien. *Médecine/Sciences* 2000. 8-9 ; I-XVI.

Orlando L, Hänni C. De l'ADN fossile à l'ARN ancien : le cas du virus de la « grippe espagnole ». *Médecine/Sciences* 2000. 11 ; *in press*.

Hänni C, Laudet V, Stéhélin D, Taberlet P. Tracking the origins of the cave bear (*Ursus spelaeus*) by mitochondrial DNA sequencing. *Proc Nat Acad Sci U S A*. 1994; 91 : 12336-40.

Higuchi R, Bowman B, Freiburger B, Ryder OA, Wilson AC. DNA sequence from the quagga, an extinct member of the horse family. *Nature* 1984; 312 : 282-4.

Pääbo S. Molecular cloning of Ancient Egyptian mummy DNA. *Nature* 1985; 314 : 644-5.

Pour en savoir plus...

Bailey JF, Richards MB, Macauley VA, Colson IB, James IT, Bradley DG, Hedges RE, Sykes B. Ancient DNA suggests a recent expansion of European cattle from a diverse wild progenitor species. *Proc R Soc Lond B Biol Sci* 1996; 263 : 1467-73.

Cooper A, Rhymer J, James HF, Olson SL, McIntosh CE, Sorenson MD, Fleischer RC. Ancient DNA and island endemics. *Nature* 1996; 381 : 484.

Drancourt M, Aboudharam G, Signoli M, Dutour O, Raoult D.. Detection of 400-year-old *Yersinia pestis* DNA in human dental pulp: an approach to the diagnosis of ancient septicemia. *Proc Natl Acad Sci U S A* 1998; 95 : 12637-40.

Eglinton G, Logan GA. Molecular preservation. *Philos Trans R Soc Lond B Biol Sci* 1991; 333 : 315-27; discussion 327-8.

Golenberg EM, Bickel A, Weihs P. Effect of highly fragmented DNA on PCR. *Nucleic Acids Res* 1996; 24 : 5026-33.

Goloubinoff P, Pääbo S, Wilson AC. Evolution of maize inferred from sequence diversity of an *Adh2* gene segment from archaeological specimens. *Proc Natl Acad Sci USA* 1993; 90 : 1997-2001.

Greenwood AD, Capelli C, Possnert G, Pääbo S. Nuclear DNA sequences from late Pleistocene megafauna. *Mol Biol Evol* 1999; 16 : 1466-73.

Guttiérrez G, Narin A. The most ancient DNA recovered from an amber-preserved specimen may not be as ancient as it seems. *Mol Biol Evol* 1998; 15 : 926-929.

Handt O, Richards M, Trommsdorff M, Kilger C, Simanainen J, Georgiev O, Bauer K, Utermann G, Pääbo S. Molecular genetic analyses of the Tyrolean Ice Man. *Science* 1994; 264:1775-8.

Höss M. Neanderthal population genetics. *Nature* 2000; 404 : 453-4.

Leonard JA, Wayne RK, Cooper A. Population genetics of ice age brown bears. *Proc Natl Acad Sci USA*. 2000; 97 : 1651-4.

Lindahl T. Instability and decay of the primary structure of DNA. *Nature*. 1993; 362 : 709-15.

Pääbo S. Ancient DNA : extraction, characterization, molecular cloning, and enzymatic amplification. *Proc Nat Acad Sci USA* 1989; 86 : 1939-1943.

Poinar HN, Hofreiter M, Spaulding WG, Martin PS, Stankiewicz A, Band H, Evershed RP, Possnert G, Pääbo S. Molecular coproscopy: dung and diet of the extinct ground sloth *Nothrotheriops shastensis*. *Science* 1998; 281 : 402-6.

Pusch CM, Giddings I, Scholz M. Repair of degraded duplex DNA from prehistoric samples using *Escherichia coli* DNA polymerase I and T4 DNA ligase. *Nucl Acids Res* 1998; 26 : 857-859.

Richards B, BC Sykes. The origins of the Polynesians: an interpretation from mitochondrial lineage analysis. *Am J Hum Genet* 1995; 57 : 1463-75.

Rogan P, Salvo JJ. Study of nucleic acids isolated from ancient remains. *Yearbook of Physical anthropology* 1990; 33 : 195-214.

Roy MS, Girman DJ, Taylor AC, Wayne RK. The use of museum specimens to reconstruct the genetic variability and relationships of extinct populations. *Experientia* 1994; 50 : 551-7.

Scholz M, Giddings I, Pusch CM. A polymerase chain reaction inhibitor of ancient hard and soft tissue DNA extracts is determined as human collagen type I. *Anal Biochem* 1998; 259 : 283-6.

Taubenberger JK, Reid AH, Krafft AE, Bijwaard KE, Fanning TG. Initial genetic characterization of the 1918 “ Spanish ” Influenza virus. *Science* 1997; 275 : 1793-1796