



HAL
open science

ÉTUDE PALÉOGÉNOMIQUE DES OSSEMENTS DU DOLMEN DE SAINT-EUGÈNE

Samantha Brunel, E Andrew Bennett, Thierry Grange, Eva-María Geigl,
Mélanie Pruvost

► **To cite this version:**

Samantha Brunel, E Andrew Bennett, Thierry Grange, Eva-María Geigl, Mélanie Pruvost. ÉTUDE PALÉOGÉNOMIQUE DES OSSEMENTS DU DOLMEN DE SAINT-EUGÈNE. Jean Guilaine. Le dolmen de Saint-Eugène: autopsie d'une sépulture collective néolithique, 1 vol., archives d'écologie préhistorique, pp.339-345, 2020, 9782358420266. halshs-03099714

HAL Id: halshs-03099714

<https://shs.hal.science/halshs-03099714v1>

Submitted on 6 Jan 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ÉTUDE PALÉOGÉNOMIQUE DES OSSEMENTS DU DOLMEN DE SAINT-EUGÈNE

Samantha Brunel, E. Andrew Bennett, Thierry Grange,
Eva-Maria Geigl et Mélanie Pruvost

CONTEXTE DE L'ÉTUDE

Dans le cadre du projet ANCESTRA (ANR JCJC 15-CE27-0001), l'équipe « Epigénome et paléogénome » de l'Institut Jacques Monod (CNRS, UMR7592, Paris) a réalisé l'analyse paléogénomique de 11 sujets inhumés dans le dolmen de Saint-Eugène et l'analyse du microbiome du tartre provenant de 5 dents isolées. Le but du projet Ancestra est d'étudier l'impact des mouvements de populations depuis le Néolithique jusqu'au Haut-Moyen Âge sur la structure des populations au niveau régional grâce à l'ADN ancien. L'équipe s'intéresse aux marqueurs génétiques liés à l'origine géographique, des caractères phénotypiques (couleur des yeux, de la peau, des cheveux...) ainsi que ceux liés à des résistances ou prédisposition à certaines maladies. Au total, ce sont plus de 180 individus néolithiques du territoire français actuel qui ont été analysés en 2016 et 2017 dont près de 65 pour la moitié sud de la France, dans des contextes chronologiques s'échelonnant du Néolithique ancien au début de l'Âge du Bronze. L'analyse des sujets de Saint-Eugène s'inscrit donc dans un projet beaucoup plus vaste. Néanmoins ces résultats apportent un éclairage nouveau sur le mode d'utilisation de ce monument funéraire.

LES ÉCHANTILLONS

Au total une trentaine d'os temporaux et dents ont été isolés par Jean Zammit et Jean Guilaine parmi les fragments osseux issus de la vidange de la crypte réalisée lors des fouilles menées par Germain Sicard en 1924-1928 et par Jean Guilaine en 1992-1993. Afin de ne pas risquer d'analyser deux fois le même sujet, seuls les os temporaux issus du même côté du crâne ont été sélectionnés. Parmi ces échantillons, onze d'entre eux ont été analysés par Samantha Brunel et Mélanie Pruvost au laboratoire Epigénome et Paléogénome de l'Institut Jacques Monod. Cinq dents présentant des dépôts de tartre ont également été sélectionnées, trois provenaient du couloir fouillé en 1992-1993 (D1, D2 et D3) et deux autres du tamisage des déblais des fouilles de

Germain Sicard (D5 et D6). Le tartre a été prélevé et étudié par Andrew Bennett à l'Institut Jacques Monod. L'échantillon D6 présentait des concrétions foncées pouvant correspondre à un dépôt de tartre ou de sédiments.

Les modifications chimiques et enzymatiques subies par l'ADN juste après la mort de l'individu sont généralement responsables de sa dégradation complète. Cependant, dans des conditions taphonomiques particulières, l'ADN peut parfois être préservé dans les restes fossilisés. L'ADN retrouvé dans ces conditions est fortement dégradé, fragmenté et présent en faibles quantités, ce qui accroît les risques d'analyser des molécules d'ADN moderne contaminant. Toutes les étapes d'extraction, de purification et de préparation des mélanges réactionnels sont réalisées dans un laboratoire de haut-confinement au sein de l'Institut Jacques Monod en employant une procédure de prévention des contaminations d'une extrême rigueur (Champlot *et al.*, 2010). Avant de réaliser les prélèvements, la surface externe des ossements qui est la plus susceptible d'avoir été contaminée lors des manipulations est retirée à l'aide d'un scalpel stérile. Les prélèvements sont réalisés avec une scie au niveau de la partie pétreuse de l'os temporal. Environ 200 mg de poudre d'os ont été utilisés pour réaliser chaque extraction d'ADN.



Fig. 1. Photos avant prélèvement des os temporaux P1, P8 et P11 sélectionnés pour les analyses paléogénomiques.

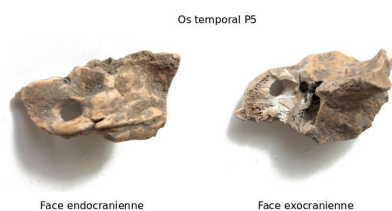


Fig. 2. Prélèvement de la partie pétreuse de l'os temporal : Le prélèvement est effectué dans le laboratoire de haut-confinement dans des conditions de haute stérilité (port de blouse, gants et masque et travail sous enceinte). Vue de l'os temporal P5 après prélèvement.



Fig. 3. Dépôts de tartre présents sur 3 dents isolées D1, D3 et D5. La dent D1 présente également une cavité de carie.

En ce qui concerne le tartre, la surface des dents a été nettoyée à l'eau de javel pour éliminer tout ADN moderne pouvant provenir de l'environnement puis le tartre a été prélevé par grattage pour chaque dent sélectionnée à l'aide d'un scalpel stérile.

Quatre des six échantillons analysés ont fait l'objet d'une datation ^{14}C par AMS1. Trois se rapportent au Néolithique final avec deux dates comprises entre 2950 et 2880 av. n.è. (95% ; P 3 et P 8) et une troisième légèrement plus récente, entre 2875 et 2620 av. n.è. (P 5). La dernière se rapporte au Bronze ancien, entre 2030 et 1830 av. n.è. (P 11). Aucune dent n'a été datée.

MÉTHODES D'ANALYSE ET RÉSULTATS PRÉLIMINAIRES

Préservation et authenticité de l'ADN

Dans un premier temps, nous avons caractérisé le contenu en ADN total des échantillons, qui comprend l'ADN endogène humain ainsi que l'ADN d'origine environnementale (microorganismes du sol...), grâce à un fluorimètre QubitTM. Nous avons ensuite amplifié par PCR quantitative en temps réel un fragment de 156 paires de bases (pb) de la région hypervariable mitochondriale HVR1-2 humaine. Grâce à cette expérience, nous avons pu déterminer la quantité d'ADN initiale ayant servi à l'amplification. Ces résultats sont présentés dans le tableau ci-dessous. Si la quantité d'ADN total extrait est relativement homogène entre les échantillons (autour de 4 ng/ μL), les quantités d'ADN humains ayant une taille compatible avec l'amplification du fragment ciblé varient énormément d'un échantillon à l'autre avec au maximum un facteur 100 entre deux échantillons. Il y a donc de grandes différences de préservation entre les échantillons et seuls six d'entre eux ont été sélectionnés pour la suite des expériences.

Tableau 1 :

Echantillon	Ossement	Dates	Quantité d'ADN total (ng/ μL)	Concentration ADN mt HVR12 (156bp) (pg/ μL)	Nbre seq total	ADN humain	% ADN endogène	Ratio $X_{\text{obs}}/X_{\text{calc}}$	Sexe
P1	Os temporal		4,67	0,2	1030014	42818	4,2%	0,422564	M
P2	Os temporal		4,07	3,3	7454	30	0,4%	1,494165	n.c.
P3	Os temporal	2942-2877 cal BC	4,04	0,4	n.c.	n.c.	n.c.	n.c.	n.c.
P4	Os temporal		4,96	0,2	n.c.	n.c.	n.c.	n.c.	n.c.
P5	Os temporal	2874-2621 cal BC	3,45	27	n.c.	n.c.	n.c.	n.c.	n.c.
P7	Os temporal		4,39	0	n.c.	n.c.	n.c.	n.c.	n.c.
P8	Os temporal	2925-2871 cal BC	3,45	1,2	924556	133445	14,4%	0,937112	F
P9	Os temporal		4,39	0,1	n.c.	n.c.	n.c.	n.c.	n.c.
P10	Os temporal		2,87	0,5	901897	24875	2,8%	0,914677	F
P11	Os temporal	2028-1874 cal BC	3,1	5	1020000	155243	15,2%	0,471002	M
P12	Os temporal		2,87	3	1073693	137410	12,8%	0,95033	F

Tableau 1. Résultats des analyses paléogénétiques. Le tableau représente la liste des échantillons étudiés avec les principaux résultats obtenus pour chacun.

Afin de mieux évaluer l'état de préservation de l'ADN dans ces échantillons, des banques d'ADN destinées au séquençage haut-débit ont été réalisées sur les extraits. Celles-ci ont ensuite été séquencées sur un séquenceur de deuxième génération MiSeq (Illumina), sur la plateforme de génomique de l'Institut Jacques Monod. La proportion de séquences alignées sur le génome humain de référence par rapport au nombre total de séquences obtenues pour chaque individu donne une estimation du taux d'ADN endogène préservé dans ces échantillons. Chez les individus du Dolmen de Saint-Eugène, le taux d'ADN endogène varie entre 0,4% et 15,2%.

L'authenticité de l'ADN endogène ainsi identifié est établie selon deux critères liés aux processus de dégradation post-mortem : la fragmentation et la modification de certains nucléotides. La distribution

de taille des fragments d'ADN présents dans les échantillons (Fig. 4 A) révèle une longueur moyenne d'environ 50 paires de bases pour tous les échantillons, cohérente avec une dégradation importante de l'ADN. Parmi les modifications nucléotidiques que l'on peut attribuer aux processus taphonomiques, la plus importante est la désamination qui transforme la cytosine en uracile et qui se traduit par la substitution d'une cytosine C par une thymine T ou d'une guanine G par une Adénine A. Cette signature caractéristique de l'ADN ancien est d'autant plus visible que l'on est proche des extrémités des fragments d'ADN. Chez les échantillons du Dolmen de Saint-Eugène, la fréquence de cette substitution à la dernière position des fragments d'ADN atteint 30% en moyenne (Fig. 4 B), ce qui indique que l'ADN humain présent dans ces échantillons est principalement ancien et authentique.

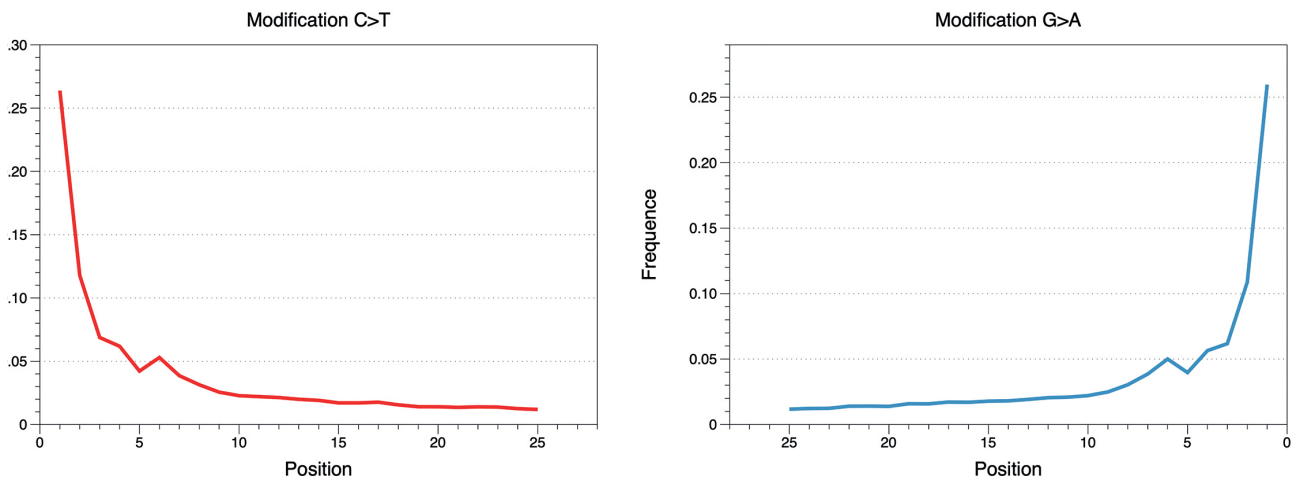


Fig. 4B. Fréquence des substitutions caractéristiques de l'ADN ancien observées en fonction de la position sur la séquence – Exemple de l'échantillon P11.

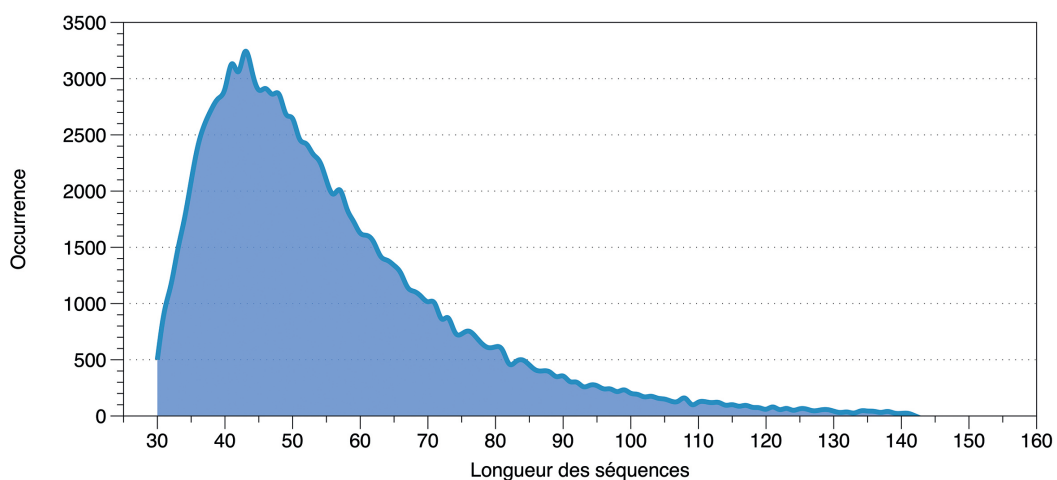


Fig. 4A. Distribution de la taille des fragments d'ADN – Exemple de l'échantillon P11.

Détermination du sexe biologique

La détermination du sexe biologique des échantillons repose sur deux critères : la présence de séquences d'ADN spécifiques du chromosome X ou, éventuellement Y, et la quantité de séquences d'ADN s'alignant au chromosome X, par rapport à celle s'alignant sur les autosomes (chromosomes non-sexuels). En effet, l'être humain possède 22 paires d'autosomes ainsi qu'une paire de chromosomes sexuels. Chez la femme, le chromosome X est présent en deux exemplaires, tandis que l'homme possède un seul chromosome X et un chromosome Y. Ainsi, la proportion attendue de séquences d'ADN correspondant au chromosome X chez un individu masculin est deux fois moindre que celle de tous les autres chromosomes, en tenant compte de leur taille respective. Pour déterminer le sexe, on préfère utiliser principalement le chromosome X car il procure des données plus fiables que le chromosome Y, ce dernier étant constitué en grande partie de séquences répétées et possédant une région très proche de celle du chromosome Y. Le graphe ci-dessous (Fig. 5) représente le rapport du nombre de séquences alignées par rapport à celui attendu pour les 22 autosomes (cercle) et pour le chromosome X (triangle). Pour les chromosomes 1 à 22, la valeur attendue de ce rapport est ainsi de 1, quel que soit le sexe de l'individu. Pour le chromosome X, ce rapport peut prendre les valeurs de 1 ou de 0,5, cette dernière valeur indiquant que l'individu

considéré est un homme. La sensibilité et la fiabilité de cette méthode dépendent du nombre total de séquences prises en compte : plus ce nombre est important, moins le calcul du rapport sera sensible aux artefacts expérimentaux pouvant engendrer des valeurs ambiguës (comprises entre 0,6 et 0,8). Ainsi, si des fluctuations importantes de la représentativité des autosomes sont observées, la détermination du sexe ne sera pas fiable.

Détermination des lignées mitochondriales

L'ADN mitochondrial est historiquement un marqueur privilégié pour l'étude de l'ADN ancien, et permet d'obtenir des informations sur la lignée maternelle des individus. Bien qu'il soit présent en multiples copies dans chaque cellule, il ne représente qu'une fraction de l'ADN humain total contenu dans un extrait en raison de sa petite taille comparative au génome nucléaire (16 kilobases contre 3,2 milliards de paires de bases). Une méthode reposant sur l'enrichissement par capture, mise au point dans l'équipe Epigénome et Paléogénome de l'Institut Jacques Monod (Massilani *et al.*, 2016), a ainsi été employée afin d'obtenir l'intégralité du génome mitochondrial à très haute couverture à l'issue du séquençage. L'échantillon P2 présentant un très faible taux d'ADN endogène a été retiré des expériences et remplacé par l'échantillon P5.

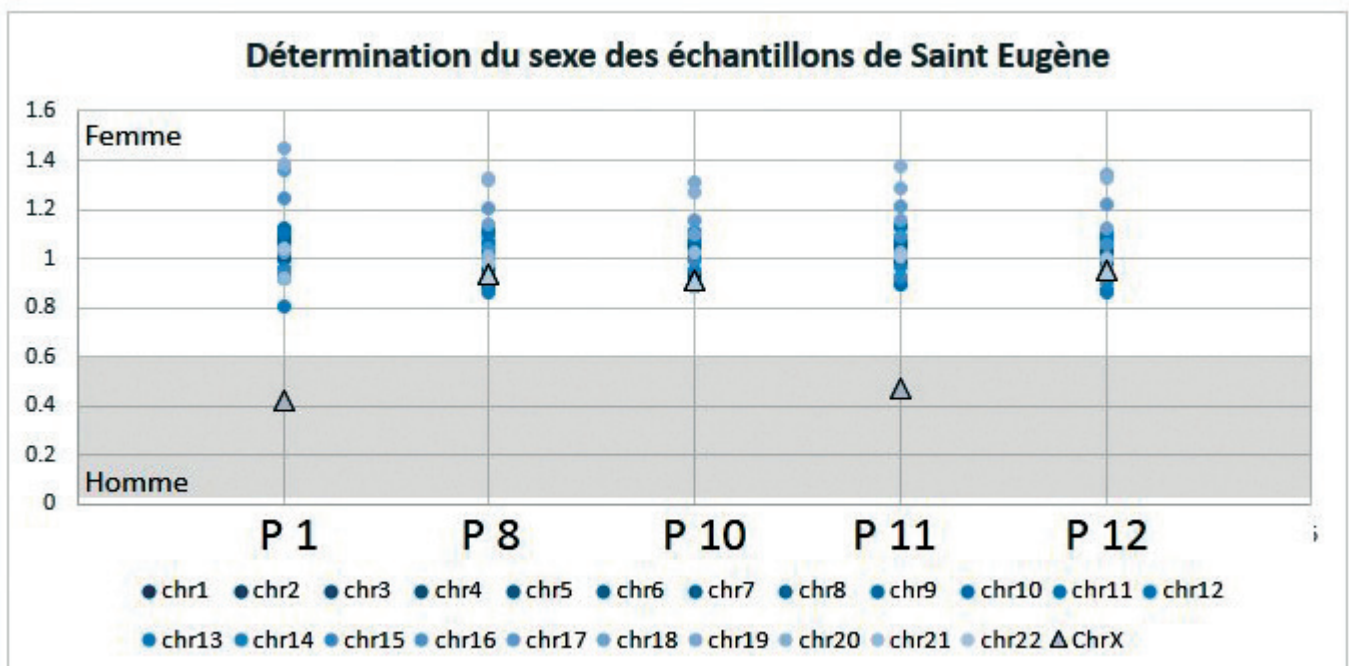


Fig. 5. Détermination du sexe des échantillons du Dolmen de Saint-Eugène. Représentation du rapport entre le nombre de séquences observées et celui calculé pour chaque chromosome au nombre de séquences totales après normalisation par la taille relative de chaque chromosome. Les valeurs pour le chromosome X sont représentées par un triangle.

Grâce à cet enrichissement, nous avons pu obtenir pour les six échantillons analysés le génome mitochondrial complet et déterminer ainsi leurs haplogroupes mitochondriaux. Aucun des sujets étudiés ne partage le même haplotype mitochondrial, ils ne sont donc pas apparentés au niveau maternel. Les six individus appartiennent à des lignées mitochondriales très différentes (X2, H1, T2, K1 et V) mais que l'on retrouve fréquemment au Néolithique, dans le sud de la France et dans le reste de l'Europe. Replacés dans le contexte du Néolithique final languedocien, avec notamment les échantillons du Dolmen des Fades, les individus de Saint-Eugène s'inscrivent dans la continuité de sites antérieurs comme le Crès (Béziers, Hérault) et le Pirou (Valros, Hérault) (Néolithique moyen), ce qui indique une relative homogénéité des lignées maternelles de la population du sud de la France durant le Néolithique (Fig. 6).

Le microbiome buccal des sujets du Dolmen de St Eugène

Des banques d'ADN ont été réalisées sur les extraits d'ADN provenant du tartre puis séquencées à une profondeur allant de 0,1 à 1 million de séquences, à l'exception de D3 qui a produit plus

de 7 millions de séquences. Les analyses métagénomiques ont été réalisées en utilisant DIAMOND, un programme qui recherche les séquences d'ADN correspondant à une base de données de protéines non redondantes, et MEGAN6 qui facilite l'analyse et la détermination taxonomique des séquences alignées. Les séquences ont également été alignées directement sur une base de données comprenant plus de 5000 génomes complets bactériens.

Les premières analyses réalisées sur D6 montrent un profil métagénomique majoritairement constitué de bactéries du sol avec une très faible proportion de bactéries de la flore buccale, ce qui indique que le dépôt observé sur D6 était constitué majoritairement de sédiments. Cet échantillon a donc été retiré des analyses.

Les quatre échantillons restants montrent une grande diversité dans leur profil métagénomique aussi bien au niveau du type de bactéries présentes que de leur abondance relative. Le tartre de la dent D1 (seule dent avec une carie) est constitué majoritairement de bactéries pathogènes (34%) associées avec des périodonties chroniques et des infections orales y compris les membres du RED complexe *Tannerella forsythia* et *Treponoma denticola*. Le tartre des dents D3 et D5 contient à l'inverse une large proportion d'espèces commensales trouvées

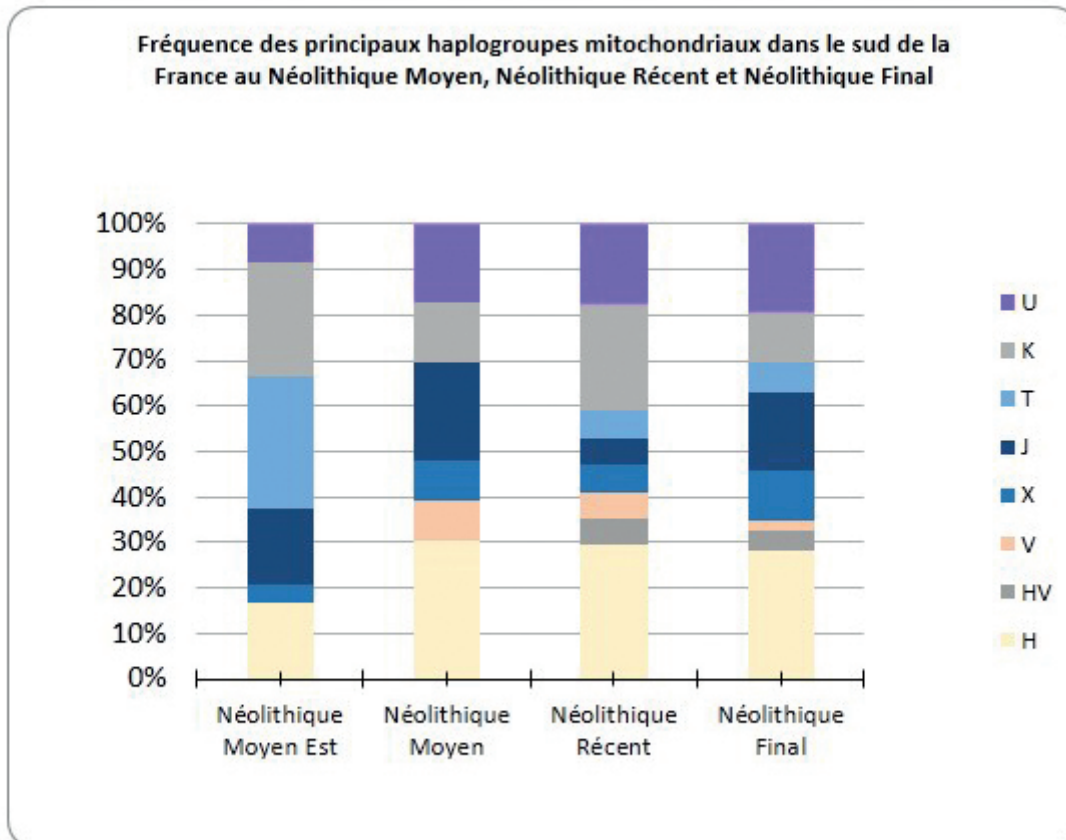


Fig. 6. Évolution de la fréquence des principaux haplogroupes mitochondriaux dans le sud de la France au cours du Néolithique. Les résultats représentés correspondent aux échantillons analysés au cours du projet Ancestra (Brunel et al., en cours). Le Néolithique moyen à l'Est est représenté par des sites d'Alsace et permet de mettre en évidence les haplogroupes caractéristiques du Sud de la France. Les échantillons de Saint-Eugène appartiennent au groupe du Néolithique final.

chez des individus sains tels que *Streptococcus sanguinis* et *Streptococcus oligofermentans* dont la présence contrecarrerait l'établissement de colonies de *Streptococcus mutans*, un autre pathogène associé à la formation des caries. De plus, D3 a un faible taux de bactéries associées à des abcès et des gingivites, bactéries qui étaient complètement absentes de D5. Finalement D2 présente des taux élevés de pathogènes responsables des périodonties et gingivites mais aussi d'autres bactéries absentes ou rares chez les autres défunts de Saint-Eugène.

La profondeur de séquençage pour l'échantillon D3 nous a permis d'obtenir le génome complet couvert à 3-4 x pour plusieurs populations de bactéries buccales. Ces génomes anciens de populations de pathogènes vont pouvoir être comparés aux génomes modernes pour mieux comprendre l'évolution du microbiome buccal.

CONCLUSIONS ET PERSPECTIVES

Issus de fouille parfois très anciennes, les échantillons du Dolmen de Saint-Eugène ne présentent pas des conditions de préservation exceptionnelle au niveau de l'ADN. Cependant grâce aux qualités de la partie pétreuse de l'os temporal très dense et aux avancées techniques en paléogénomique, il a été possible d'extraire et d'authentifier l'ADN pour 6 individus sur onze soit un taux de réussite de 54%.

Dans ce contexte d'inhumations multiples et de remaniements s'étalant sur plusieurs millénaires, les résultats génétiques portant sur seulement cinq individus ne peuvent faire l'objet que d'interprétations préliminaires et devront s'inscrire dans un contexte plus large. La proportion des sexes- au sein des inhumés est de 3 femmes pour 2 hommes et aucun individu n'est apparenté au niveau de sa lignée maternelle. Cette étude s'inscrit dans un contexte plus large au sein du projet Ancestra et les résultats du site du Dolmen de St Eugène, associés à d'autres sites de la fin du Néolithique en Languedoc-Roussillon (Dolmen des Fades), permettent de caractériser le profil génétique des populations languedociennes. Ces résultats feront donc partie d'une publication portant sur le Néolithique en France (Brunel *et al.*, en cours de rédaction).

Les résultats préliminaires obtenus sur l'analyse du microbiome buccal de 5 défunts de Saint-Eugène montre une grande variabilité entre individus. Deux individus ont une bonne hygiène buccale, un individu a une flore microbienne cohérente avec la présence de caries observée sur la dent étudiée et finalement le dernier échantillon pourrait provenir d'un individu présentant une infection qui expliquerait les changements observés au niveau de son profil métagenomique. Ces résultats vont être repris dans le cadre du projet Ancestra et comparés à plus de deux cent autres profils obtenus à partir d'échantillons datés du Néolithique à nos jours sur toute la France.

Pour conclure ce chapitre sur les analyses paléogénétiques des défunts de Saint-Eugène, il est intéressant d'évoquer le cas particulier de l'individu P11. Cet échantillon a été sélectionné pour une analyse plus profonde de son génome avant que sa datation ^{14}C ne nous soit connue. Plus de 37 millions de molécules d'ADN ont ainsi été séquencées sur une plateforme illumina NextSeq (plateforme de séquençage de l'IRBA Division Défense NRBC), permettant l'obtention de son génome partiel avec une couverture de 0.14x. Il a ainsi été possible de comparer cet individu à d'autres individus issus de populations anciennes et modernes dont les génomes complets étaient publiés. Ces résultats feront également l'objet d'une publication dans une revue internationale et ne seront donc pas développés en détail ici. Cependant nos différentes analyses génomiques, telles que l'analyse en composantes principales présentée ci-dessous (Fig. 7), ont révélé que cet échantillon est plus proche génétiquement d'individus datant de l'âge du Bronze que des autres échantillons néolithiques. Ceci nous a amené à faire effectuer la datation au ^{14}C de cet échantillon qui a montré que cet individu correspond à une phase plus récente d'occupation du Dolmen de Saint-Eugène, confirmant ainsi les résultats génomiques.

BIBLIOGRAPHIE

DANIEL H. HUSON, SINA BEIER, ISABELL FLADE, ANNA GÓRSKA, MOHAMED EL-HADIDI, SUPARNAMITRA, HANS-JOACHIM RUSCHEWEYH, REWATITAPPU. MEGAN COMMUNITY EDITION - Interactive exploration and 2 analysis of large-scale microbiome sequencing data, 2016 PLoS Computational Biology 12(6): e1004957. doi:10.1371/journal.pcbi.1004957

BUCHFINK B., XI^F. C. AND HUSON, D.H. Fast and sensitive protein alignment using DIAMOND, 2014 Nature Methods doi:10.1038/nmeth.3176.

JENS KRETH, JUSTIN MERRITT, AND FENGXIA QI. Bacterial and Host Interactions of Oral Streptococci DNA Cell Biol. 2009 Aug; 28(8): 397-403. doi: 10.1089/dna.2009.0868

MASSILANI D, GUIMARAES S, BRUGAL JP, BENNETT EA, TOKARSKA M, ARBOGAST RM, BARYSHNIKOV G, BOESKOROV G, CASTEL JC, DAVYDOV S, MADELAINE S, PUTELAT O, SPASSKAYA NN, UERPMANN HP, GRANGE T, GEIGL EM. Past climate changes, population dynamics and the origin of Bison in Europe. BMC Biol. 2016 Oct 21; 14(1):93.

CHAMPLLOT S, BERTHELOT C, PRUVOST M, BENNETT EA, GRANGE T, GEIGL EM. An efficient multistrategy DNA decontamination procedure of PCR reagents for hypersensitive PCR applications. PLoS One. 2010 Sep 28;5(9). pii: e13042. doi: 10.1371/journal.pone.0013042.

REMERCIEMENTS

Cette étude a été réalisée grâce au financement de l'ANR ANCESTRAL (ANR JCJC 15-CE27-0001). La plateforme de paléogénomique de l'Institut Jacques Monod a été soutenue par le programme "Actions de recherches structurantes" de l'Université Paris Diderot. La plateforme de séquençage de l'Institut

Jacques Monod a été financée par l'Université Paris Diderot, la Fondation pour la Recherche Médicale (DGE20111123014), et la Région Ile-de-France (11015901). Nous aimerions remercier l'équipe de la plateforme de séquençage de l'IRBA (Division

Défense NRBC, Département des Services, Unité de Biologie Moléculaire) pour le séquençage avec la technologie Nextseq d'Illumina. Nous aimerions également remercier l'ensemble des archéologues ayant collaboré au projet.

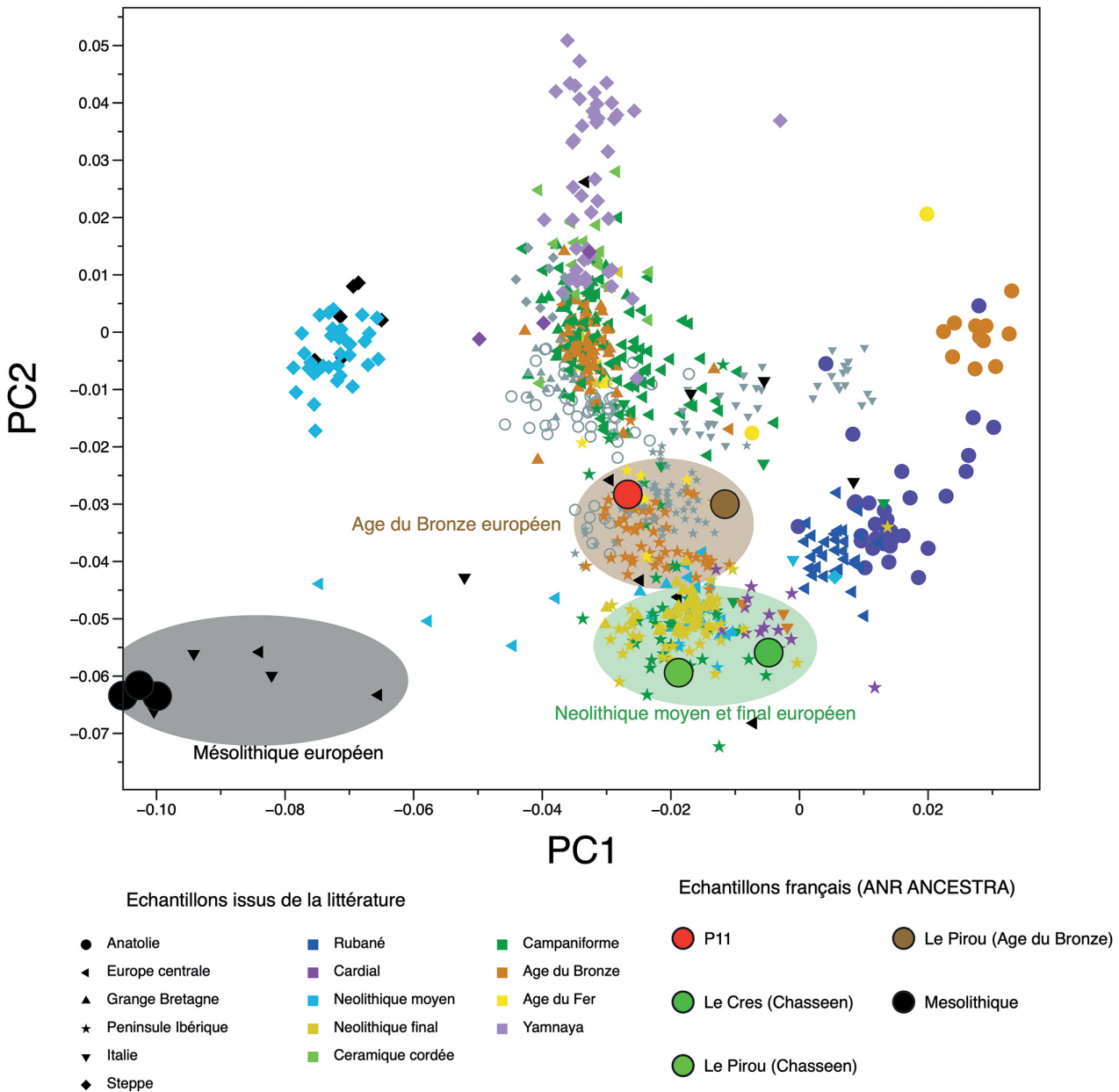


Fig. 7. Analyse en composante principale (ACP) des données paléogénomiques : Les échantillons modernes d'Eurasie provenant d'une étude sur la diversité génomique humaine moderne (Human Genome Diversity Project et Centre d'Etudes du Polymorphisme Humain) ont été utilisés pour calculer l'ACP sur plus de 600 000 marqueurs. Les échantillons anciens issus de différentes publications ont été projetés sur les deux premières composantes. Les échantillons analysés dans le cadre du projet Ancestra sont représentés par des cercles colorés : vert pour 2 individus du chasséen des sites du Crès et du Pirou, marron clair pour un individu du site du Pirou de l'âge du Bronze, noir pour des individus d'un site mésolithique français et rouge pour l'individu P11 de Saint-Eugène.

